THE UNIVERSITY *of* EDINBURGH

Unveiling the Impact of Neuromotor Disorders on Speech: a Structured Approach Combining Biomechanical Fundamentals and Statistical Machine Learning

Andrés Gómez Rodellar

Thesis in fulfilment of the requirements for the degree of

Doctor of Philosophy

Supervisors:

Prof. Athanasios Tsanas (UoE), Prof. Saturnino Luz (UoE), Prof. Victor Nieto-Lluis (UPM), Prof. Per Svenningsson (KTH)

January 2024

To my family with love,

To my friends with gratitude,

To science with devotion,

And to those who left with remembrance.

Unveiling the Impact of Neuromotor Disorders on Speech: a Structured Approach Combining Biomechanical Fundamentals and Statistical Machine Learning

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Summary

Speech has been shown to convey clinically useful information in the study of Neurodegenerative Disorders (NDs), such as Parkinson's Disease (PD). Traditionally the use of speech as an exploratory tool in People with Parkinson's (PwP) has focused on the estimation of acoustic characteristics and their study at face value, analysing the physio-acoustical markers and using them as features for the differentiation between Healthy Controls (HC) and PwP. The present work takes a step further, given the intricate interoperation between neuromotor activity, responsible for both planning and driving the system, and the production of the acoustic speech signal; by the study of speech, this relationship may be properly exploited and analysed, providing a non-invasive method for the diagnosis, analysis, and observation of NDs. This work aims to introduce a working model that is capable of linking both domains and serves as a projection tool to provide insights about a speaker's neuromotor state. This is based on a review of the neurophysiological background of the structure and function of the nervous system, and a review of the main nervous system dysfunctions involved in PD and other related neuromotor disorders. The role of the respiratory, phonatory, and articulatory systems is reviewed in the production of voice and speech under normal and pathological circumstances. This setting might allow for speech to be considered a useful trait within the precision medicine framework, as it provides a personal biometric marker that is innate and easy to elicit, can be recorded remotely with inexpensive equipment, is non-invasive, cost-effective, and easy to process.

The problem can be divided into two main categories: firstly, a binary detection task distinguishing between healthy controls and individuals with NDs based on the projection model and phonatory estimates; secondly, a progression and tracking task providing a set of quantitative indices that enable clinically interpretable scores. This study aims to define a set of features and models that help to characterise hypokinetic dysarthria (HD). These incorporate the neuroscientific knowhow semantically and quantitatively to be used in clinical decision support tools that provide mechanistic insight on the processes involved in speech production, incorporating into the algorithmic element neuromotor considerations that add to better interpretability, consequently leading to improved clinical decisions and diagnosis.

An overview of the acoustic signal processing algorithms for use in speech articulation and phonation system inversion regarding neuromotor disorder assessment is provided. An algorithmic methodology for model inversion and exploration has been proposed for the functional characterization and system inversion of each subsystem involved under the neuro-biomechanical foundations exposed before.

A description of the vocal fold biomechanics using the glottal source, and formant dynamics provides the base for specific mapping to articulation kinematics. The statistical methods used in performance evaluation are based on three-way comparisons and transversal and longitudinal assessment by classical hypothesis testing.

Three related experimental studies are shown to empirically illustrate the potential of phonation and articulation analysis: the characterization of PD from glottal biomechanics based on the amplitude distributions of the glottal flow and on the vocal fold body stiffness in assessing the efficiency of transcranial magnetic stimulation, and the description of PD dysarthria through an articulation projection model.

The results from the biomechanical analysis of phonation showed that the behaviour of glottal source amplitude distributions from PD and healthy controls using three-way comparisons and hierarchical clustering were essentially distinguishable from those from normative young participants with the best accuracy scores produced by SVM classifiers of 94.8% (males) and 92.2% (females). Nevertheless, PD participants were barely separable from age-matched controls, possibly pointing to confounding factors due to age. The outcomes from using vocal fold stiffness in assessing the efficiency of transcranial magnetic stimulation showed mixed results, as some PD participants reflected clear improvements in phonation stability after stimulation, whereas some others did not. Some cases of sham controls experienced also minor improvements of unknown origin, possibly expressing a placebo effect. The overall results on the efficiency of stimulation showed an accuracy global score of 67% over the 18 cases studied. The results from articulation projection modelling showed the possibility of formulating personalised models for PD and control participants to transform acoustic formant dynamics into articulation kinematics. This might open the possibility of characterising PD dysarthria based on speech audio records.

The most remarkable findings of the study include the determination of the glottal source amplitude distribution behaviour of normative and PD participants; the impact of age effects in phonation as a confounding factor in neuromotor disorder characterization; the importance of ensuring that the classification of speech dysarthria is based on principles that can be explained and interpreted; the need of taking into account the effects of medication when framing new classification experiments; the potential of using EEG-band decomposition to analyse vocal fold stiffness correlates, as well as the possibility of using these descriptions in longitudinal monitoring of treatment efficiency; the feasibility of establishing a relationship between acoustic and kinematic variables by projection model inversion; and the potential of these descriptions for estimating neuromotor activities in midbrain related to phonation and articulation activity.

The most important outcome to be brought forth from the thesis is that the methodology used throughout the project uses a bottom-up approach based on speech model inversion at the acoustical, biomechanical, and neuromotor levels allowing to estimate glottal signals, biomechanical correlates, and neuromotor activity from speech alone, establishing a common neuromechanical characterisation framework on its own.

Acknowledgements

The last few years after embarking on this Ph.D. journey have been quite turbulent, both at a personal and professional level; the COVID-19 pandemic, the lockdowns, my grandmother's passing, war in Ukraine and the untimely passing of both my childhood dog and my mother within a week apart. All of these things combined with the demanding activity of a PhD, the pressure to publish, and the uncertainty of pursuing research with unclear outcomes, would have made the task unbearable were it not for those who have accompanied me throughout the journey. The support and care of those around me have been invaluable, I am beyond thankful.

When beginning the challenging journey of pursuing a PhD, there are certain aspects that one often takes for granted, but they may not necessarily be so. During conversations with fellow PhD students and friends, I have come to appreciate how incredibly fortunate I have been regarding my PhD supervisor. I want to express my gratitude to Thanasis, who has been an exceptional mentor, caring, supportive, and going above and beyond. Sadly, due to the pandemic contact with Thanasis has been somewhat hindered by distance and I know I would have benefited far more from his knowledge, guidance, and good nature had I been able to meet in person with him regularly. Nonetheless he has provided me with insightful, caring, and supportive guidance, help, and counselling. I am deeply thankful for our time together. I know I am not the only one who feels this way as I am aware that other of his Ph.D. students share the same opinion. As part of my supervisory team, I also want to thank, Saturnino Luz Filho, Per Svenningsson, and Victor Nieto Lluis. I would also like to thank my yearly examiners Matt Bourmane and Martyn Pickersgill for their dedication and their feedback during the yearly reviews.

I also would like to thank the Usher Institute and the University of Edinburgh for providing me with this outstanding opportunity, not just academically, but also through the tight-knit support network that the institutions have in place, both professional and personal, ranging from the IAD courses to the student support services. I want to also thank and acknowledge the UK Medical Research Council for providing the funding to carry out this research and financially supporting me throughout these years. Although the University of Edinburgh has hosted me there are other institutions and associations that I have come into contact with during the completion of this Ph.D. that require mentioning; substantial collaboration has been undertaken with the NeuSpeLab at the Centre for Biomedical Technology of the Universidad Politécnica de Madrid which Prof. Victor Nieto Lluis was part of until his retirement, for their support and kindness in providing access to their recording equipment and knowhow. To the group in the Czech Republic at the Technical University of Brno and Masaryk University both institutions located in the city of Brno, specifically to Prof. MUDr. Irena Rektorová and doc. Ing. Jiří Mekyska for not only allowing me access to their resources but for extending their friendship beyond academic collaboration. I would like to thank two associations involved with Parkinson disease. Firstly, I would like to thank the association APARKAM, for their support and the coordination of the recording sessions that took place in Madrid of the biometric signals of individuals with Parkinson's disease. I want to personally thank Azucena Balandín and Zoraida Romero Martínez for their kindness and their wilfulness and all of the volunteers who took part in the recording sessions with contagious enthusiasm. Finally, I would like to thank the Edinburgh Parkinson's Research Interest Group for providing formative sessions and allowing me the chance to present my work to both technical professionals and people with Parkinson's, and for recognizing my work with the best presentation of the session and a monetary price.

Finally, I would like to thank the administrative staff of the Precision Medicine program for their support and diligence in handling all my queries, requests and quick replies.

Regarding life outside academia, I would first like to thank my family, for their invaluable support, love, and warmth. Firstly, I want to start by thanking my mother, sadly she was not able to see me defend my work, read this document, or share my life after my PhD, she passed away on a rainy day on the 26th of May 2023 after a complicated cancer diagnosis. I have good reason to believe that my mother was an exceptional woman, not only through my experiences with her but also through the love she inspired among close friends, colleagues, coworkers, and in some instances complete strangers. I want to thank her for her warmth and the love that only a loving mother can provide, and for that wondrous sixth sense she had for social situations, pulling win-win solutions to apparent impasses and charming people with her good spirit and honesty. A person who gave her absolute best with the causes she took and the people she became involved with. She taught me the value of patience, honesty, hard work and finding a way to work with people. She is and always will be missed. Next, I want to thank my father for his dedication, industriousness, and kindness. Throughout all my life he has been a teacher, friend, and accomplice, a selfless man who has taught my brother and me the value of hard work and perpetual learning. A walking encyclopaedia with the precise anecdote or factoid, ranging from botanic science to history, awing anyone who is willing to listen. Finally, I want to thank my brother, for his level-headedness, sense of humour, and integrity. Carlos has the gift of rearranging problems to observe them from a new perspective, calmly splitting them to show that it was not such a big deal after all, I want to thank him for picking my brain and making me try to do better on many occasions. We have been accomplices in crime on many occasions and I hope for many more to come.

I also want to thank Aiste, for her sweetness, kindness, and quirky sense of humour, your occasional cat transformation is something to witness.

I want to thank all my friends for the respite they have provided throughout this journey. I firstly want to thank Valeriano, a force of nature like no other, since we met at the age of three and I bit his hand at the daycare we became inseparable. I want to thank him for his sense of humour and general good nature, you are a truly valued friend. I want to thank Victor who truly is a good soul, reflective, kind, and always willing to help, thanks for sharing some "orejita rica" with me and for occasionally turning into a fun gremlin. I want to thank Irene for her playful and optimistic nature, posing sometimes as a happy little goat who can't stay still in one place. Torralba you had for breakfast a bowl of dice; you are, a gust of wind, a ball of randomness a wild card, you name it, it is always fun to have you around, but you can be serious when the situation demands it and I truly appreciate your friendship. I want to thank Inés for keeping him in check, making him behave, and keeping our collective sanity somewhat intact. I want to thank Alain and Romi, thanks Alain for the absolute nonsense you bring to the table, it is freeing and a pleasure to be an idiot alongside you, but beyond your sense of humour I value your reflective nature and your wise counselling, and for that wondrous skill you possess in splitting dinner checks, thanks Romi for always taking me to new places and showing me great places to eat and for putting up with our general stupidity. Thanks, Patricia and Vesselin for your warmth and kindness during hard times, and for being the most awesome nerd power couple I know. Thanks, Evan for your laid-back and optimistic nature, for kicking off the paddle fever and for those times nerding out at your living room table, I bleep-ing hate when your stupid monkey steals my *threeferi*.

Thanks to Arturo and Enrique, thanks Arturo for being someone whom you can talk to from politics to philosophy, with a little monkeying around mixed in there, thanks Enrique for your loyalty, your honesty, and your evil sense of humour. I want to thank my friend Jorge for his reflectiveness and his playful mischievousness, I hope you can return to Madrid soon. I must thank Juanvi for our stimulating talks, time passes extremely quickly with you my friend, even when we disagree the conversation never ceases to be entretainig and profound. I want to thank Antonio a friend in the distance, I hope you can return to Spain soon and we can recuperate our long-lost sessions and enjoy some nice nerdy-time. I want to thank Alessandra for dragging me to the cinema (with varying degrees of success), and for your fierce idealism, we never seem to end the topics we can argue about, sometimes even civilly. Lastly, I want to thank Layla, thank you for extending a lovely friendship beyond what would simply be regular violin lessons, I am delighted to have crossed paths with someone as lovely as you, I truly enjoy our lessons and our little beer after class during summer, or morning coffee in the winter.

Thank you all for providing a short while, a trip, a session, a dinner, a lunch, or just a simple walk in the park. You have helped me in more ways than you can imagine.

Thank you all for everything, this work has been possible thanks to you all.

Abbreviations

AAW	Average Acoustic Wave
ACC	Accuracy
AD	Alzheimer's disease
AKV	Absolute Kinematic Velocity
AKM	Articulation Kinematic Model
ALS	Amyotrophic Lateral Sclerosis
ANS	Autonomic Nervous System
APM	Articulatory Projection Model
AUC	Area Under the Curve
BA4	Brodmann area 4
BG	Basal Ganglia
BIA	Biomechanical Instability of Articulation
BiCr	Bisector Criterion
BIP	Biomechanical Instability of Phonation
CAG	Cytosine-Adenine-Guanine segment
CB	Cerebellum
CNS	Central Nervous System
CPP	Cepstral Peak Prominence
CVA	Cerebrovascular Accident
CTM	Cricothyroid Muscle
DBS	Deep Brain Stimulation
DsP	Dorsolateral Peak
DsP	Dorsolateral Peak
EEG	Electroencephalography

FCP	Final Common Pathways
FIR	Finite Impulse Response Filter
fMRI	Functional Magnetic Resonance Imaging
GF	Glottal Flow
GFAD	Glottal Flow Amplitude Distribution
GPe	*Globus Pallidus Pars Externa*
GPI	*Globus Pallidus Pars Interna*
GS	Glottal Source
HC	Healthy Controls
HD	Hypokinetic Dysarthria
HiCl	Hierarchical Clustering
HNR	Harmonic To Noise Ratio
HRF	Harmonic Richness Factor
HTT	Hutingtin Gene
HVPS	Human Voice Production System
IAIF	Iterative Adaptive Inverse Filtering
IT	Information Theory
JSD	Jensen-Shannon Divergence
LBD	Lewy Body Dementia
LBMS	Laryngeal Biomechanical System
L1	First Lumbar Vertebra
LC	Laryngeal Cortex
L-F	Liljencrant-Fant Profile
LMN	Lower Motor Neuron
LTAS	Long-Term Average Spectrum
LTI	Linear Time-Invariant
MB	Midbrain

MEG Magnetoencephalography

MFDR Maximum Flow Declination Rate

MI Mutual Information

MS Multiple sclerosis

MU Motor Unit

MWC Mucosal Wave Correlate

NA Nucleus Ambiguus

NAQ Normalized Amplitude Quotient

NDs Neurodegenerative Disorders

NMA Neuromotor Activity

NMDs Neuromotor Disorders

ONPT Oro-Naso-Pharyngeal Tract

PARCOR Partial Correlation

PC Premotor Cortex

PELF Prediction-Error Lattice Filter

PET Positron Emission Tomography

PCVA Primary Motor Cortex Vocalization Area

PD Parkinson's Disease

PDD Parkinson's Disease Dementia

PMC Primary Motor Cortex

PNS Peripheral Nervous System

PPC Posterior Parietal Cortex

PwP People With Parkinson's Disease

QOQ Quasi-Open Quotient

RAS Reticular Activation System

RBF Radial Basis Function

RLN Recurrent Laryngeal Nerve

rTMS	Repetitive Transcranial Magnetic Stimulation
sEMG	Surface Electromyography
SLN	Superior Laryngeal Nerve
SMA	Supplementary Motor Area
SNC	*Substantia Nigra Pars Compacta*
SNS	Somatic Nervous System
SPC	Specificity
SPECT	Single Photon Emission Computerized Tomography
STN	Subthalamic Nucleus
STV	Sensitivity
SVM	Support Vector Machine
T1	First Thoracic Vertebra
TAM	Thyroarytenoid Muscle
TMS	Transcranial Magnetic Stimulation
TBI	Traumatic Brain Injury
TH	Thalamus
UMN	Upper Motor Neuron
UPDRS	Unified Parkinson's Disease Rating Scale
V3	Third Subdivision Of Cranial Nerve V
VFBS	Vocal Fold Body Stiffness
VmCSP	Ventromedial Central Sulcus Peak
VNG	Extracranial Vagus Nerve Ganglia
WHO	World Health Organization

Frequently used notation

The following mathematical notational conventions are used throughout this thesis:

Vectors are written in bold lowercase letters, for example, **x**; matrices are written in bold capital letters, for example, **X**. $\{\cdot\}_{i,j}$ denotes the i^{th} row, j^{th} column matrix entry. The subscript n in the form x_n indicates the n^{th} element of a vector. The expectation and the conditional expectation operators used are: $E[\cdot]$ and $E[(\cdot\,|condition)]$. Also, $(\cdot)^{\mathrm{T}}$ denotes the transpose of a matrix, and $\frac{df}{dx}$ represents the differentiation of a function f concerning x. The covariance of two random variables X, Y is defined by $Cov(X,Y) = E[X,Y] - E[X] \cdot E[Y]$. The convolution operator is denoted by $\otimes$, and the distance metric is represented by $\|\cdot\|$. Unless otherwise specified, the Euclidean distance is used. In the context of this thesis we work with (a) real numbers (represented with $\mathbb{R}$), (b) natural numbers (represented with $\mathbb{N}$), and (c) integer numbers (represented with $\mathbb{Z}$).

List of Figures

List of Appendix Figures

List of Tables

List of Appendix Tables

Contents

CHAPTER 1

1 Introduction

The relationship between the increase in life expectancy and the manifestation of neurodegenerative disorders (NDs) is well established, their impact is expected to rise with the overall increase of the global life expectancy (Hou et al., 2019). This is due to the steady progression of material wealth, improvement of living conditions, and medical advancements that have reduced the impact on factors affecting historical mortality records, such as food or water poisoning, violent death, infectious diseases, starvation, infant death, and exposure (Griffin, 2008). The world average life expectancy has dramatically risen over the past two centuries, with an average lifespan of around 30 years[1] throughout most of human history to today's world average of 72.6-73.2 years (World Health Organization, 2021). Consequently, throughout most of human history, certain types of diseases did not have enough time to develop and exhibit observable symptoms as individuals would expire before from other causes. In rare and uncommon cases such diseases would be documented as oddities and unexplained anomalies. Thus, as individuals' lives progress the nature of the diseases that are likely to affect them changes as well. Infectious diseases, for instance, are more prevalent in children than cancer or diabetes.

[1] The low average life expectancy is largely influenced by high infant mortality rates. However, once individuals reached adulthood, the average life expectancy worldwide was approximately 75 years although this figure varies between studies and time periods.

There is a technological aspect to take into account as well, as technology develops new manufacturing techniques that allow for the development of more complex machinery, tools, and sensors causing the emergence of new devices that solve key unsurmountable problems. Given that diseases with more acute onset and severity historically caused the highest number of fatalities, medical research, and practice have prioritised addressing them, generating a cyclical pattern of identification and the development of effective treatments. Medical advances such as the discovery of penicillin, pacemaker technology, and the evolution of organ transplants changed medical research priorities and have thus shifted focus between medical fields. Over the last century, advancements in cancer and cardio-respiratory diseases have been among the most remarkable medical accomplishments. Probably because of longer life expectancy, greater numbers of people with NDs are being diagnosed these days, driving interest from the medical community, and promoting research to shift gradually toward NDs, to provide a better understanding of this specific field.

1.1 Neurodegenerative disorders

NDs are a group of conditions that affect the neurons in the human nervous system (Hardiman et al., 2011). These pathologies are characterised by a progressive loss of susceptible neuron populations due to toxic or metabolic disorders, leading to impairment or loss of certain brain functions and severely worsening the behaviour of individuals affected by them. The categorization of NDs can be classified according to primary clinical features (e.g., dementia, Parkinsonism, or motor neuron disease), anatomic distribution of neurodegeneration (e.g., frontotemporal degenerations, extrapyramidal disorders, or spinocerebellar degenerations), or principal molecular abnormality (Dugger and Dixon, 2017).

The main group of interest for the present work is Neuromotor Disorders (NMDs); these produce predominantly motor affections that manifest in any task involving the movement of biomechanical body systems. NMDs seem to be a consequence of neurons degenerating past a series of structures known as the Basal Ganglia (BG), which are directly involved in motor control, cognition, and emotion regulation. These structures are the striatum, the *globus pallidus*, the subthalamic nucleus, and the *substantia nigra* (see Figure 1.1).

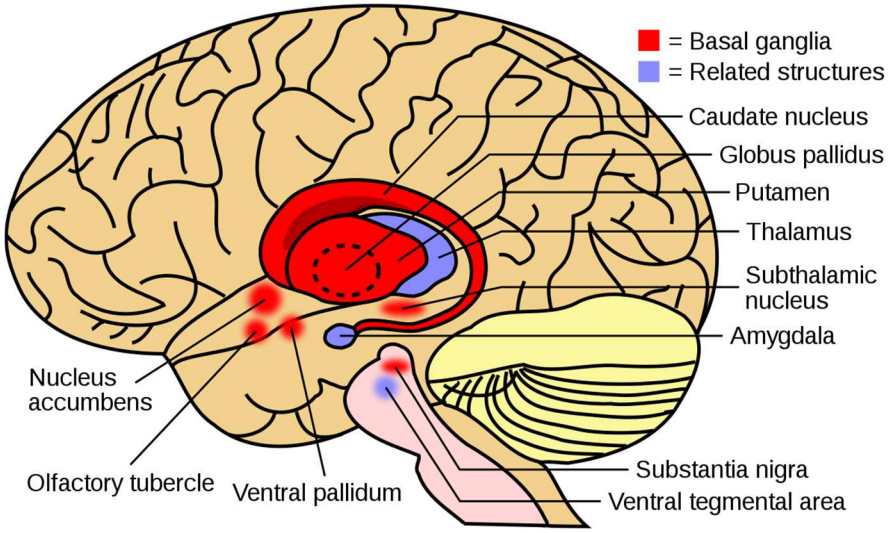

Figure 1.1 BG and associated structures.
Description: BG (red) and associated structures[2] (blue).
https://creativecommons.org/licenses/by-sa/4.0/

The BG serve as a pivotal link between higher cortical functions involved in cognitive planning and the neuromotor pathways that transmit signals to motor neurons in the medulla oblongata and the spinal cord.

[2] Source: https://commons.wikimedia.org/wiki/File:Basal_ganglia_and_related_structures_(2).svg, Licence: CC BY-SA 4.0 (Authorizing free share and remix). No changes made with respect to the original file. Retrieved 2023/10/02.

This network of structures is primarily responsible for control functions and the precise modulation of responses. By integrating information from cognitive processes with motor outputs, the BG contribute to the refinement and coordination of motor actions, ensuring accurate execution and appropriate adjustments in response to environmental demands.

The study of NDs is as long as humans have been recording history (Papavramidou, 2018). The understanding and categorization of these disorders have evolved, gradually unravelling their intricate nature. The earliest observations of these disorders date back to ancient civilizations. For instance, descriptions of dementia-like symptoms can be found in ancient Egyptian, Greek, and Roman texts (Boller and Forbes, 1998). Alzheimer's disease, the most prevalent form of dementia, was named after Dr Alois Alzheimer, who in 1906 identified unique brain abnormalities in a patient exhibiting severe memory loss and cognitive decline (Auguste Detter). The latter half of the 20th century witnessed significant breakthroughs in molecular biology and genetics, which revolutionised the study of NDs and their characterization. The identification of specific genes associated with inherited forms of NDs, such as Huntington's disease, allowed for deeper insights into the underlying mechanisms.

Despite extensive research efforts, effective treatments to halt or reverse the progression of NDs remain elusive. Although various therapeutic approaches, including medication, physical therapy, and supportive care, have been shown to alleviate symptoms and enhance the quality of life for individuals affected by these disorders (Blom, Emmelot-Vonk, and Koek, 2013). Therefore, their study poses a major challenge to the development of effective treatments and interventions. So far exploring potential biomarkers, developing novel therapeutic targets, and investigating strategies are crucial objectives in the pursuit of finding more effective treatments and ultimately discovering a cure (Emamzadeh and Surguchov, 2018).

A brief description of the most prevalent NDs is given in Table 1.1.

Table 1.1 Brief description of the most prevalent NDs[3]

Disorder	Description
Alzheimer's disease	Alzheimer's disease is a neurodegenerative disorder that leads to a progressive loss of memory and cognitive abilities. The exact causes are unclear, but it involves a combination of genetic factors, environmental influences, and lifestyle choices. A key feature of Alzheimer's is the buildup of beta-amyloid plaques and tau tangles in the brain, which are abnormal protein deposits that disrupt brain function.
Amyotrophic lateral sclerosis	Amyotrophic lateral sclerosis (ALS) is marked by the breakdown of motor neurons, which manage voluntary muscles. While most ALS cases occur randomly without a known cause, a minority are linked to certain genetic mutations.
Creutzfeldt-Jakob disease	Creutzfeldt-Jakob's disease is a rare degenerative brain disorder caused by the abnormal folding of prion proteins. It can occur spontaneously, be inherited, or be transmitted through exposure to infected tissues.
Frontotemporal dementia	Frontotemporal dementia is a group of disorders characterised by the degeneration of nerve cells in the frontal and temporal lobes of the brain. It can be caused by genetic mutations or occur sporadically without a known cause.
Huntington's disease	Huntington's disease is an inherited neurodegenerative disorder caused by a mutation in the huntingtin gene. The mutation leads to the production of a toxic protein that damages nerve cells in the brain, particularly in the BG.
Lewy Body Dementia (LBD)	LBD is associated with the buildup of abnormal protein deposits called Lewy bodies in the brain. The exact cause is unknown, but it is believed to involve a combination of genetic and environmental factors.
Multiple sclerosis (MS)	MS is an autoimmune disease where the immune system mistakenly attacks the protective covering of nerve fibres in the central nervous system. The exact cause of MS is unknown, but it is thought to involve a combination of genetic and environmental factors.
Parkinson's disease	Its a neurodegenerative disorder characterized by the loss of dopamine-producing cells in the substantia nigra part of the brain. The causes of PD are not completely known, but it's thought to be due to a mix of genetic and environmental factors, along with the buildup of abnormal alpha-synuclein protein.

[3] https://www.ninds.nih.gov/health-information/disorders

1.1.1 Parkinson's Disease

The neurodegenerative disorder first described by Dr. James Parkinson as shaking palsy, known since then as Parkinson's disease (PD), is the second most prevalent among NDs (Hardiman et al., 2011). At the time that Dr. Parkinson described six cases of this disorder, it was considered an unclassified disabling disease, of novel and rare character. Since then, morbidity and mortality rates due to PD are increasing faster than concerning any other neurological disorder (Dorsey et al., 2018a, b). The prevalence of PD is expected to double every 20 years (Simon, Tanner, and Brundin, 2020). Global estimates in 2019 showed over 8.5 million individuals with PD, an increase of 81% since 2000, 329,000 deaths being attributed to this disorder, an increase of over 100% since 2000 (World Health Organisation, 2023). When considering the causes of PD, three factors seem to be the most relevant: ageing, genetics, and environment (De Lau and Breteler, 2006). Social, political, and economic reasons are mentioned as possible causes for its geographical expansion, linked to processed food and drinks, alcohol, tobacco, environmental factors, and social habits (limited physical activity among them).

PD is the most prevalent NMD, quantifying its incidence in 15 cases per 100,000, with a prevalence ranging from 100 to 200 cases per 100,000 (Tysnes and Storstein, 2017). The way PD generally manifests itself is through unintended or uncontrollable movements, such as shaking, rigidity, and difficulty with balance and coordination, among others. Moreover, there are some behavioural changes such as sleep problems, depression, memory difficulties, and fatigue. The highest risk factor associated with the development of PD is unclear but one of the strongest factors is age, although there are early cases that can manifest before the age of 50 (approximately 5-10%). A comprehensive description of the characteristics of the disease among other relevant aspects can be found in Sapir (2014).

There is increasing evidence that PD is an umbrella term that refers to the range of heterogeneous symptoms, which as it affects most motor functions is also reflected in voice (Tsanas and Arora 2022). The heterogeneity of its potential causes makes PD a prime target for a precision medicine approach from the pharmacological, neurosurgical, and rehabilitative points of view. Although the single most predictive factor for PD is advanced age (typically beyond age 60), almost 25% of affected individuals are younger than 65 years (Bloem, Okun, and Klein, 2021). On average, its incidence is lower, and its onset is higher among women than among men (De Lau and Breteler, 2006). PD manifests itself as a progressive deterioration of motor capabilities developing over many years before clear manifestations become evident (Bloem, Okun, and Klein, 2021). Among the earliest symptoms for prodromal PD diagnosis rapid eye movement disorders during sleep, and speech dysarthria phenomena, seem to be the earliest signs of deterioration. The general symptoms associated with PD are bradykinesia, rigidity, freezing of gait, frozen facial mask (hypomimia), postural sway, and distal limb resting tremor, among others (Jankovic, 2008; Dauer and Przedborski, 2003; Bhat et al., 2018).

According to Tsanas and Arora (2022), “speech as an item within comprehensive PD clinical scales has been previously shown to be very strongly associated with overall PD symptom severity as assessed using standardised clinical metrics”. In this sense, it is well known that HD is one of the characteristic motor symptoms of PD (Duffy, 2013), that affects respiration, phonation, articulation, prosody, and fluency (Titze, 1994a; Tsanas, 2012; Mekyska et al., 2015; Brabenec et al., 2017).

Phonation symptoms, such as *musculus vocalis* hypotonia, vocal fold imbalance, and tremor in voice (altered neuromotor feedback) are some manifestations of PD-related neurodegeneration on speech (Liotti et al. 2003; Sapir et al., 2010; Belalcazar-Bolaños et al., 2015; Hanratty et al., 2016). Precise knowledge of the neural processes and models regulating the respiratory, phonation, and articulation systems in the body is essential to explain the effects of PD on speech (Davis et al., 1996; Schulz et al., 2005; Jürgens, 2009), especially in terms of neurological circuit modelling (Hirschauer, Adeli and Buford, 2015; Yuvaraj et al., 2016; Caiola and Holmes, 2019). There has been a lot of interest in developing signal-processing approaches to mine speech data, extract dysarthria and dysphonia features, and employ statistical machine-learning algorithms for biomedical speech applications (Tsanas, 2013; Brabenec et al., 2017; Arora and Tsanas, 2021; Arora et al., 2021; Tsanas, Little, and Ramig, 2021). However, to a large extent, these studies do not provide the same level of insights that mechanistic models can provide, i.e., models that build on the physical principles of voice production to characterise the underlying vocal production mechanisms related to PD-associated disorders (Duffy, 2013). Exploring a mechanistic model can provide new insights into the underlying physiological processes, which in turn might inspire further signal processing algorithms for the characterisation of speech signals.

1.1.2 Etiology of PD

The BG are the nervous terminations that are responsible for fine-tuning voluntary movements by processing and adjusting the impulses received from the cerebral cortex promoting precise neuromotor actions. They are strongly connected to sensorial areas and integrate information from feedback channels into movement, incorporating information returning from the brain to the planning of new motor orders. This information is then conveyed to the thalamus which relays the impulses back to the cerebral cortex.

Ultimately the fine-tuned processed instructions are relayed to the motor units through the tracts of the pyramidal motor system.

The BG are responsible for mediating some other higher cortical functions, such as planning and modulation of movement, memory, eye movements, reward processing, and motivation. The most associated signs and symptoms of PD occur when the neurons in the BG become impaired or they die. These neurons are responsible for producing dopamine, a neurotransmitter in charge of sending messages between nerve cells closely associated with the ability to think and plan, movement promotion, motivation, mood, memory, and some more tasks. As they degrade dopamine production is reduced which in turn causes a wide display of problems including movement alterations.

There exists a clear relationship between α-synuclein and PD, the presence of Lewy Bodies is a primary component of the neuron degeneration. Depending on the area where neuron decay takes place, the disease is either branded as PD of LBD (Henderson et. al. 2019). Alpha-synuclein is a misfolded protein due to mutations in the gene encoding it. Increasing evidence suggests that this protein can spread throughout a neuron population. *In vitro* models have shown that neurons grafted with α-synuclein proteins from Lewy Bodies form Lewy Body structures of their own, and animal models suggest that rat nigral neurons grafted with human α-synuclein spread α-synuclein to embryonic ventral mesencephalic neurons, showing areas of human α-synuclein surrounded by a larger ring of rat α-synuclein, suggesting a seeding mechanism (Paweł et. al. 2015). As the disease progresses α-synuclein pathology seems to spread throughout the different brain structures, in the early stages affecting primarily motor function regions and progressively moving to cortical structures responsible for higher cognitive processing.

This process seems to be the consequence of a physical transmission of the protein between brain areas, where the protein is misfolded by an originator neuron and then transmitted to proximal vulnerable neurons.

Braak's staging hypothesis (Braak et. al. 2003) proposes that Parkinson's disease is the result of exposure to an exogenous agent that gains entry into the body via the nasal or gastrointestinal pathways. This agent is believed to infiltrate deeper into the nervous system, ultimately reaching the brain either through the pituitary gland or the vagus nerve. The existence of Lewy bodies in the enteric and peripheral nervous systems lends credence to this proposition. The pathological process navigates through the tissues, primarily affecting thin and predominantly unmyelinated neurons. The Braak staging system categorizes the progression of the disease into six distinct stages, each one associated with abnormal pathology in specific neurological structures. In relation to symptomatology, the nature and severity of symptoms correspond to the progression through Braak stages. The initial stages are marked by non-motor symptoms, such as olfactory dysfunction or constipation. Motor symptoms typically manifest around the mid-stage, while cognitive symptoms emerge as the disease progresses to the later Braak stages.

An important distinction between PD and parkinsonism lies in their underlying causes. PD primarily results from neuronal cell death, whereas parkinsonism can have various alternative etiologies. Common alternative causes observed in clinical practice include other neurodegenerative diseases and drug-induced effects. Additionally, less frequent causes include structural brain disorders, head injuries, Wilson's disease, and exposure to toxins (Rajput et al., 2024). One specific condition associated with parkinsonism is Lewy-body dementia. In this disorder, abnormal clumps of α-synuclein (known as Lewy bodies) accumulate inside neurons.

To differentiate between PD and parkinsonism, clinicians often administer levodopa to patients. If symptoms improve in response to the drug, PD is more likely to be the underlying cause.

Some instances of PD seem to show hereditary origin, and some specific cases can be related to specific genetic mutations. While genetics is expected to play a role in PD, in most cases family history does not seem to play a crucial role, which leads researchers to think that the disease is caused by a combination of genetic factors as well as exposure to environmental factors such as toxins.

1.1.3 Symptoms and manifestation

PD is characterized by a range of core symptoms primarily affecting movement. These symptoms can manifest in several ways; one prominent manifestation is the characteristic resting tremor, observed in the hands, arms, legs, jaw, or head. Additionally, individuals with PD may experience muscular stiffness, resulting in prolonged muscle contractions that make movement challenging or even impossible. Another common symptom is hypokinesia, which consists of slowed movements (Duffy, 2013). Furthermore, PD can lead to a loss of balance and coordination, further impacting mobility. In addition to these core symptoms, individuals with PD often display a secondary range of symptoms. These can include psychological changes such as depression, as well as difficulties in controlling secondary motor functions like swallowing, chewing, and speaking. Additionally, PD may give rise to complications such as constipation and urinary problems. The presence of these secondary symptoms further contributes to the overall burden experienced by individuals living with PD.

In the early stages of PD, symptoms seem to manifest with mild motor complications while performing day-to-day tasks, such as tremors or body movement, along with rapid eye movement sleep behaviour disorder, hyposmia and depression (Turcano et. al. 2019, Postuma et. al. 2012). Even though speech is somewhat of an afterthought, most PD cases will develop some form of speech disruptions, such as HD and phonation asthenia. As the disease progresses an important and dangerous manifestation of the disease is the affectation of gait, forcing a forward-leaning posture, quick small steps, and arm stabilising swings. Individuals have trouble initiating or continuing movement, such as the characteristic and potentially dangerous freezing of gait, where the stride becomes locked in place, presenting a fall risk. Another axis that PD manifests is handwriting, lettering becoming smaller and more irregular as the disease settles in.

In most cases, the disease is perceived externally by close ones as the alterations manifest increasingly and more evidently. Symptoms seem too often begin on one side of the body, but as disease progresses, they eventually spread to both sides. However, it is not uncommon for symptoms to be more severe on one side than the other. Many PwPs note that they had other problems before stiffness and tremors manifested, such as sleep disorders, constipation, loss of smell, and restless legs. Some of these symptoms might be relatable to ageing because as stated before PD is more is more prevalent with age.

1.1.4 Effects of PD on cognitive functions

As it was introduced in subsection 1.2.1 dopamine plays a very important role in many cognitive functions such as memory, attention, and the ability to plan and accomplish tasks among others. Some PwPs may experience changes in many of these areas leading to stress and depression. Individuals with PD may experience impairments in various cognitive domains.

One common cognitive effect is executive dysfunction, which can result in difficulties with planning, problem-solving, decision-making, and multitasking. PD can also impact attention and concentration, leading to decreased focus and distractibility. Memory deficits, particularly in working memory and episodic memory, are frequently observed in PD. Additionally, individuals with PD may experience difficulties in language and communication, such as word-finding capability, and reduced verbal fluency. Visuospatial abilities, including spatial perception and navigation, can also be affected. Furthermore, PD can lead to changes in mood and behaviour, including depression, anxiety, apathy, and impulsivity. These cognitive effects of PD can significantly impact daily functioning and quality of life for individuals living with the disease.

As PD progresses, commonly some cases develop a form of dementia which is diagnosed as Parkinson's Disease Dementia (PDD). A consequence of this condition is that people suffering from it may have severe memory and cognitive problems that affect daily life. PDD is characterized by not only prominent cognitive symptoms but also by additional alterations, such as visual hallucinations and motor impairments. The cognitive effects of PDD are diverse and can include deficits in attention and alertness, executive function, and visuospatial abilities. Individuals with PDD may struggle with sustaining attention, and maintaining focus, and easily become disoriented. Executive dysfunction can manifest as difficulties with planning, organizing, problem-solving, and shifting between tasks. Visuospatial impairments can result in challenges to depth perception, spatial navigation, and object recognition. Memory deficits, particularly in episodic memory, may also be present in PDD, although they are generally less severe compared to other forms of dementia such as Alzheimer's disease.

Language and communication can be affected, leading to difficulties in finding words and expressing themselves. Furthermore, PDD is associated with fluctuations in cognitive abilities and attention, with periods of lucidity alternating with periods of confusion or disorientation.

1.1.5 Diagnosis of PD

The presence of α-synuclein can be used to identify the cause of PD even before there are any noticeable symptoms through the test α-synuclein seed amplification assay, where spinal fluid is extracted through a lumbar puncture and then examined for α-synuclein clumps. This procedure is used in clinical research during clinical trials and it is not available in a health care professional's office due to its invasiveness and risks involved, but there is hope that it will be used for the diagnosis of Parkinson's disease in the future. Experts also hope the test could one day be done using blood samples rather than spinal fluid (Mayo Clinic 2023). As a result, the diagnosis primarily relies on the patient's clinical history, symptoms, and physical and neurological examination. Diagnosis is typically undertaken through a combination of exploring a person's medical history and conducting an in-clinic examination. Typically, a specialist, such as a neurologist or a geriatrician, conducts the inspection. The specialist requests the individual to take on a sequence of physical exercises aimed at detecting any potential symptoms or movement-related issues. At this stage diagnostic uncertainty is resolved if the individual presents two of the three following symptoms compatible with a high chance for PD to be present: resting tremor or shaking, bradykinesia or slowed movement, and rigidity or muscle stiffness.

If after the exposure of the patient to levodpa the symptoms subside, the likelihood of PD being the cause of the symptoms dramatically increases. Several disorders that can cause symptoms similar to those of PD such as multiple system atrophy and LBD are labelled with the broader term of Parkinsonism; they share similar features with PD but require different treatments. While these disorders might be initially labelled as PD, certain tests or pharmacological treatments provide a better understanding after the initial diagnosis.

Once the disease has been diagnosed, its progression is commonly assessed according to the Unified Parkinson's Disease Rating Scale (UPDRS, Goetz et al., 2004, 2007 and 2008) which is a widely commonly used metric ranking disease progression according to a clinical opinion and self-assessment questionnaire. The scale is composed of five parts; each covering a different aspect affected by the progression of PD:

- Part I: evaluation of mentation, behaviour, and mood.
- Part II: self-evaluation of daily life activities, including speech, swallowing, handwriting, dressing, hygiene, falling, salivation, turning in bed, walking, and cutting food.
- Part III: clinician-scored monitored motor evaluation.
- Part IV: complications of therapy.
- Part V: Hoehn and Yahr staging of severity of Parkinson's disease.
- Part VI: Schwab and England activities of daily life scale assessing the capacities of people with impaired mobility.

While the medical history and neurological examination play a crucial role in diagnosing non-genetic PD, it is important to note that the diagnosis is primarily clinical and relies on the expertise of the healthcare provider, and this is precisely the key inspiration for this work.

Providing an objective assessment using UPDRS as the only tool might be complicated even for very experienced professionals, given the different subjective factors affecting the final scoring.

1.1.6 Pharmacological treatments

Although there is no cure for PD, however, various treatments and therapies are available to manage the symptoms and improve the quality of life of individuals with PD, such as pharmacological treatment, brain stimulation interventions, and other rehabilitative therapies. The symptoms of PD can be mitigated or reduced by addressing the effect on the different predominant biochemical aspects that cause the distortions, affecting the dopamine supply in the brain, affecting neurotransmitters in the brain, or controlling non-movement-related symptoms. The main drug used in combating PD symptoms is levodopa, a chemical precursor used in the metabolisation of dopamine, as it is absorbed by the brain cells, the medication increases dopamine production to compensate for the declining natural supply. However, like many medications, it can cause adverse effects including nausea, vomiting, low blood pressure, and dyskinesias. To combat the side effects of levodopa, patients tend to take additional medication such as carbidopa. It prevents the breakdown of levodopa before it reaches the brain, this has the added effect of reducing the required levodopa amount to improve the symptoms. There are complications to suddenly stopping the levodopa treatment, as it could cause dopamine levels to abruptly drop leading to the incapacity of movement or difficulty in breathing.

There are several other drugs used to increase the effectiveness of levodopa delivery as listed in Table 1.2.

Table 1.2 Brief description of most commonly used levodopa deliveries for PD.

Drug	Description
Dopamine agonists	These work by stimulating dopamine receptors in the brain, mimicking the effects of dopamine. These medications help increase dopamine activity, compensating for the reduced natural supply in conditions such as Parkinson's disease. However, like other drugs, dopamine agonists may have associated side effects.
Enzyme inhibitors	Such as MAO-B inhibitors and COMT inhibitors, work by blocking specific enzymes in the brain that break down dopamine. By inhibiting these enzymes, more dopamine is available in the brain, which helps alleviate the motor symptoms of PD.
Amantadine	The main use of amantadine is to treat dyskinesias, such as stiffness, tremors, shaking, and uncontrolled muscle movements. It may be used alone or in combination with other drugs like levodopa. Amantadine is a weak antagonist of the NMDA-type glutamate receptor. This action may help reduce abnormal brain activity that contributes to the symptoms of Parkinson's disease. It increases dopamine release and blocks dopamine reuptake in the brain. Although amantadine has dopaminergic-like side effects, it is not clear how it works to treat dyskinesias or "off" episodes in people with Parkinson's disease.
Anticholinergic drugs	These are a class of medications that block the action of acetylcholine on certain receptors, a neurotransmitter in the CNS and PNS. In the context of PD, anticholinergic drugs are sometimes prescribed to help manage specific motor symptoms. They can help reduce tremors and stiffness by balancing the levels of acetylcholine and dopamine in the brain. They are generally not the first-line treatment and are often reserved for patients who do not respond well to other medications or who experience intolerable side effects from other treatments.

1.1.7 Brain stimulation therapies

Brain stimulation therapies involve activating or inhibiting brain areas directly by inducing localised electrical currents, these therapies have shown promise in the management of PD symptoms. When aiming to stimulate the brain, there are certain complications that need to be addressed; the first is the encasement of the brain within the skull, which poses a challenge in accessing the target area. The second complication involves the delivery of the stimulating currents effectively to the brain tissue.

Methodologies can be broadly classified depending on their invasiveness, one way to apply the currents to the desired tissue is to directly apply the currents through electrodes; either noninvasively by placing them on the scalp or invasively by placing them directly into the brain through perforations in the skull. Non-invasive placements are generally safer and easier to administer, however, the electrical currents delivered through the scalp may have limited penetration and specificity, which can hinder their effectiveness. In contrast, invasive methods, like Deep Brain Stimulation (DBS), allow for more precise targeting of specific brain regions. By placing electrodes directly into the brain, the electrical stimulation can be delivered with greater accuracy and depth. This approach is often reserved for conditions that require highly targeted stimulation, such as PD or essential tremor. However, invasive methods carry additional risks associated with surgery, such as infection or damage to surrounding brain structures.

Alternatively, the electrical currents can be induced by applying electromagnetic fields. There is a wide range of techniques that work on inducing currents by magnetic stimulation on the desired areas, changing frequency and field intensity, the most common of them being repetitive Transcraneal Magnetic Stimulation (rTMS). Although this technique was initially conceived as a research tool, there has been great interest regarding its potential clinical role. Presently, it is unclear whether rTMS will have some role as an alternative treatment for PD symptom management. This method is a non-invasive procedure that uses a magnetic coil to generate localized electrical currents in specific areas of the brain. The magnetic coil is placed on the scalp, and the magnetic field it produces penetrates the skull and induces electrical currents in the underlying brain tissue.

It is a relatively safe and well-tolerated procedure compared to surgical interventions like DBS. However, the effects of rTMS are generally temporary and require repeated sessions for sustained benefits; the exact reason explaining why this method is effective is still an issue of ongoing study.

Other therapies that may help manage Parkinson's symptoms are shown in Table 1.3.

Table 1.3 Supporting therapies for PD.

Therapy	Description
Cognitive Therapy	Cognitive changes can occur in PD, and cognitive therapy can help individuals manage and cope with these changes. It involves exercises and strategies to improve memory, attention, problem-solving, and other cognitive functions.
Exercise Programs	Regular exercise has been shown to have numerous benefits for individuals with PD. It can improve motor symptoms, balance, strength, and flexibility, as well as promote overall well-being. Exercise programs may include activities such as walking, cycling, dancing, tai chi, etc.
Occupational Therapy	Occupational therapy aims to enable individuals to perform daily activities more independently. It focuses on enhancing skills related to self-care, work, and leisure, and may involve strategies for adapting the environment to make tasks easier to accomplish.
Physical	Focuses on improving mobility, balance, strength, and flexibility. It involves exercises, stretching, and other techniques to help manage motor symptoms and maintain physical function.
Speech Therapy	Speech therapy helps address the communication difficulties that can arise in PD, such as speech and swallowing problems. Techniques may include exercises to strengthen speech-related muscles, improve articulation, and develop strategies for clearer communication.
Support Groups	Joining a support group can provide a sense of community, a platform for sharing experiences, and emotional support. Support groups can be beneficial for individuals with PD as well as their caregivers and family members.
Supportive Counselling	Counselling or psychotherapy can provide emotional support, help individuals and their families cope with the impact of PD, and address any psychological or emotional challenges that may arise.

1.1.8 Speech alterations as a consequence of NDs

The influence of neurological and cognitive processes on speech is a well-established and recognised fact (Skodda et al., 2013; Sapir, 2014; Rusz et al., 2013). Many studies in the last decade have explored diverse signals such as Electroencephalography (EEG), Magneto Encephalography (MEG), Functional Magnetic Resonance Imaging (fMRI), and other non-invasive methods to provide new insights into the speech production process (Rusz et al., 2013; Oh et al., 2018, Stam, 2010). This is of particular interest when investigating NDs (cognitive and neuromotor) such as Alzheimer's, PD, Amyotrophic Lateral Sclerosis (ALS), Huntington's Chorea, and others related (Skaper, Zusso, and Giusti, 2018).

As introduced previously speech allows contactless remote recording on smart terminals, such as phones, tablets, or laptop computers, it offers the added benefit of mapping acoustic estimates to neuromuscular activity, providing an advantage in the detection and monitoring of disorders dependent on remote neuromotor transmission (Arora, Baghai-Ravary, and Tsanas, 2019). A comprehensive study on the effects of PD on speech (Yunusova, Weismer, and Lindstrom, 2011; Skodda, Visser, and Schlegel, 2011) could provide insights into the underlying physiology, associating speech characteristics to the physical manifestations of the disorder. This can be achieved through the study of phonation, articulation, prosody, and fluency (Mekyska et al., 2015) which would offer valuable information on the activity of specific brain areas involved in speech production, such as motor planning, premotor and motor, and working memory. There is an unmet need to establish a robust and reliable methodology to map estimates extracted from the speech acoustics to motor actions in certain muscles involved in speech articulation and production.

One such example is the masseter muscle, responsible for raising the lower mandible. Such a projection is proposed in this work to transform speech formant dynamics to articulatory kinematics (Dromey, Jang, and Hollis, 2013; Whitfield and Goberman, 2014). First proposals of an inverse model (relating formant dynamics and articulation) are presented; as a result, several indicators were developed to encompass articulatory movements from speech alone (e.g., Absolute Kinematic Velocity, AKV) (Gómez-Vilda et al., 2017a, b). The problem with these first attempts was the lack of a robust model parameter estimation. This led to further exploratory work, where the relationships between masetter surface electromyography (sEMG), accelerometry, and speech were investigated (Gómez-Rodellar et al., 2018). After an in-depth study of the influence of PD on these biometric signals (Gómez-Rodellar et al., 2019b), the conclusions were applied to the characterisation of PD hypokinetic dysarthria (HD), (Gómez-Rodellar et al., 2019a; Gómez-Vilda et al., 2019a).

Speech production is a dynamic neuromechanical activity that involves cognitive and neuromotor resources of extreme complexity, and which is not yet well understood (Duffy, 2013). The natural way in which it is acquired and used shades the sophisticated processes that are placed into work during its normal expression. Speech is instantiated in the linguistic neuromotor cortex (Demonet, Thierry, and Cardebat, 2005), and its execution demands the concourse of cognitive, neuromotor, neuromuscular, and musculoskeletal processes (Duffy, 2013). Through speech, thoughts and emotions are projected to the knowledge of others by cognitive-linguistic messages. These are programmed for their neuromotor expression by the activation, time-alignment and sensorimotor projection, extension, and strength of a large number of diverse muscles, and associated biomechanical structures.

The neuromotor areas from the Central Nervous System (CNS), where planning, programming, and control are provided are responsible for activating the respiratory, phonatory, and articulatory muscular structures innervated by the Peripheral Nervous System (PNS), (Kandel et al, 2013). The resulting speech is a sequence of acoustic interactions between the glottal source signal and the vocal tract cavities, both activated and modulated by neuromotor impulses, imprinting the cognitive-linguistic message. The alteration or dysfunction of any key vocal production mechanisms will result in a deficient production of speech known as a speech disorder. Among them, motor speech disorders are the result of dysfunctional neurological structures involved in the planning, sequencing, activating, and monitoring of the neuromuscular structures responsible for speech sound production, modulation, and projection. One of the most active neuromuscular structures involved in speech production is the masseter-jaw-tongue complex, including part of the lower facial muscles and tissues attached to the mandible (Duffy, 2013). This system is responsible for the production of open or closed, and elongation or retraction of the vocal tract, producing phonations perceptible as vowels and vowel-related sounds (Greenberg et al., 2004). Depending on the quasi-stationarity of this system (for more than 30-50 ms) the outcome is vowel-like phonations, whereas rapid movements are responsible for the acoustical representation of many consonant-like sounds. NDs affect the functional operation of this structure, and its central role in speech articulation suggests it could likely express key pathological changes reflecting the neuromotor behaviour.

PD, also known as shaking palsy has an unclear aetiology in most cases, but evidence suggests that it may be due to different dysfunctions affecting the fine control of muscular actions of cerebral subsystems responsible for musculoskeletal control, such as the hypothalamus, the cerebellum, the primary and secondary neuromotor control areas, and the frontal lobes (Brown et al., 2009). A compelling and comprehensive overall view is given in Duffy (2013): "*The motor system is present at all of the major anatomic levels of the nervous system and is directly responsible for all motor activity involving … to the planning, control, and execution of voluntary movement, including speech.*".

It is a well-established fact that PD causes considerable alterations in speech and phonation (Ricciardi et al., 2016, Brabenec et al., 2017). Broadly speaking, speech alterations may be classified as dysphonia (voice production), dysarthria (speech articulation), dysprosody (the fundamental frequency), and dysfluency (rhythm and sequence of inter-syllabic and intersegmental blocks). These alterations are jointly referred to as hypokinetic dysarthria. Harel, Cannizzaro, and Snyder (2004a) give a summary of the symptoms associated with HD

"*Hypokinetic dysarthria, a speech disorder characterised by indistinctness of articulation, weakness of voice, lack of inflection, burst of speech, and hesitations and stoppages, is an integral part of the motoric changes in PD*". In this same sense, there is "*compelling evidence to suggest that speech can help quantify not only motor symptoms … but generalised diverse symptoms in PD*" (Tsanas, 2012). Godino-Llorente et al. (2017) stress the fact that "*The low levels of dopamine that appear in patients with PD lead to dysfunctions of the basal ganglia… These deficits negatively affect the three main anatomic subsystems involved in the speech production: respiration, phonation and articulation*".

A good description of the neuromotor systems involved in speech production, and how they may be affected by NDs is to be found in Duffy (2013). Therefore, the search for neuromotor degenerative biomarkers in speech is to be concentrated on phonation (glottal signals in terms of distortion and biomechanics), speech articulation (study of acoustic and biomechanical clues as formants and jaw-tongue kinematics), the prosodic flow (concentrated in the time evolution of the fundamental frequency and speech energy stability) and on fluency (syllabic and intersyllabic intervals, duration, stability and fluctuation of the speaking rate). This is well documented in the work of Mekyska et al. (2015). Moreover, speech can be used to investigate the nature and extent of vocal impairment in individuals who are at risk of developing PD and can provide a crucial opportunity to intervene in the prodromal stages. Arora et al. (2021) have reported very compelling findings when comparing speech signals from people diagnosed with sleep behaviour disorder (which is one of the strongest known predictors of PD risk) with a control group and a diagnosed PD group. Furthermore, it has been demonstrated that PD symptom severity can be accurately monitored using speech signals collected over the standard telephone network, thus alleviating the need for frequent physical patient visits to the clinic (Tsanas. Little, and Ramig, 2021). The main compelling facts favouring speech-based PD biomarkers are the low cost of the required equipment with the increase of computational power of smartphones and tablets and their reduction in cost, and the contactless factor, which is particularly useful to facilitate remote studies. Summarising, the acoustic markers induced by HD in PD speech allow us to conclude that speech analysis might become a non-invasive and cost-effective tool to characterise and monitor PD.

The role of speech as a possible biomarker in PD detection is well established in the state-of-the-art research literature, with many studies discussing speech-based PD features sensitive to HD. In the present study, the focus is placed on the study of acoustic and biomechanical clues, such as formants, and jaw-tongue kinematics.

1.2 Motivations

As discussed in the previous section NMDs are a group of NDs that involve the neuromuscular system, severely altering movement activity. PD is among the most prevalent NMDs, which affects 8.5 million individuals globally according to the WHO, with continuously rising prevalence rates. The study and understanding of these diseases have two distinct approaches; the first one is at a biochemical level, by the study of the material processes that lead to neuron deaths. This approach is invasive in its nature, and it has substantial constraints both in time and the environment of the test. The second approach is the study of the affectation of the output tasks, such as movement, speech, or writing. Although there are different degrees of invasive and cumbersome extraction devices generally these approaches tend to be less stress-inducing than any traditional medical tests. The idea behind the use of these systems is that they are minimally invasive and can provide insights into the state of the underlying brain-motor systems. Speech is one of the most key distorted functions, as it is affected both by motor and cognitive impairment, carrying with it significant semantic information about the state of the underlying systems. The processing and study of speech provide a non-invasive setting, with a signal that is easy to extract and record, that can be stored, transported, and processed easily. The study and processing of speech is not exclusively secluded to a neuroacoustic field of research, but it also has immediate applications in the fields of speech therapy and linguistics.

The standard metric for assessing PD is the UPDRS, based on a series of self-assessed metrics as well as an estimation of the symptoms by a trained clinician. This scale is the gold standard upon which the progression of PD is compared and categorised. The inherent problem of such metrics is the lack of objective quantitative tests, even though the standard is quite thorough it still suffers from a high degree of variability due to the individual interpretation of the patient and the clinical staff.

Characterizing PD poses several challenges across various domains, including neuromotor assessment, neuropsychological assessment, diagnosis, pharmacological treatment, and rehabilitation. Traditional methods of PD characterization often rely on subjective assessments, which are prone to variability and subjectivity. There is an increasing need for objective assessment methods to overcome these limitations and provide more accurate and reliable measurements of disease progression and treatment outcomes. Moreover, there is also an increasing need to apply some form of remote diagnosis that is non-invasive and easy to use. As PwP are more affected by the motor symptoms mobility becomes an issue and there is a point that a visit to the health centres or clinics becomes increasingly difficult.

In the realm of neuromotor assessment, objective measures can help quantify motor symptoms such as bradykinesia, tremors, and rigidity. Objective tools, such as motion sensors and wearable devices, can provide quantitative data on motor performance, allowing for more precise monitoring and tracking of symptom severity and progression over time. These assessments can provide valuable insights into the effectiveness of pharmacological treatments and rehabilitation interventions. Neuropsychological assessment is another crucial aspect of PD characterization, as cognitive and behavioural changes often accompany motor symptoms.

Objective assessment methods, such as computerized cognitive tests and neuroimaging techniques like functional MRI (fMRI), can provide objective measures of cognitive functioning, identifying deficits and tracking changes over time. Another dimension of analysis is the use of biomarkers or disease-related markers that can be observed by imaging techniques (e.g., DAT-SPECT, MRI), cerebrospinal fluid analysis, and genetic testing, these provide analysis of the biochemistry of the underlying brain-motor system.

There is another need of paramount importance when dealing with PD and that is the assessment of pharmacological treatment. It is a complex balance of drugs and individual responses to medications can vary. Objective assessment methods can help monitor treatment response objectively by measuring changes in motor symptoms, such as medication-induced dyskinesia or motor fluctuations. Currently, real-time tracking of the effects of medication is exceedingly difficult, as the strain on public healthcare services means that the time per patient a clinician has available is very limited, as a consequence of incremental volume and constricted resources. This combined with the long separation between visits leads to ineffective and in the worst-case scenario detrimental drug dosage and administration. Analogously rehabilitation is a crucial component of PD management, aiming to improve functional abilities, enhance mobility, and maintain quality of life. Objective assessment methods in rehabilitation, such as motion analysis systems, force sensors, and virtual reality-based training programs, provide quantitative data on motor performance and functional outcomes.

These previous methodologies suffer from differing degrees of invasiveness along with the costly and complex infrastructure that confines the collection of data and samples to specialized centres.

They require costly specialized equipment, and a degree of expertise to place, operate, and interpret results heavily restricting the application of therapies and healthcare delivery to PwP. Here is where speech has the space to serve as an invaluable diagnostics tool, as it is non-invasive, can be stored off-site, requires little storage space, and is easy to process and record. Speech has the potential to provide objective measurements and serve as a monitoring and diagnostics vehicle for clinicians, rehabilitator personnel, and caregivers as it could provide real-time information about the underlying state of the speaker.

1.3 Speech as a biomarker of PD

Some of the biomarkers that speech can capture are shown in Table 1.4.

Table 1.4 Speech biomarkers for PD description and characterization.

Biomarker	Description
Articulation and phonation	PD has a clear effect on the articulatory and phonatory aspects of speech production. Individuals with PD exhibit differing degrees of imprecise articulation, reduced speech rate, and variations in pitch and intonation. These changes can be attributed to disruptions in the coordination of the muscles involved in speech production, including the lips, tongue, and vocal folds.
Prosody	In PD, alterations in prosody can occur, leading to decreased variation in pitch, reduced stress patterns, and a monotonous speech pattern.
Speech rate and fluency	PD can affect speech rate and fluency. Individuals with PD may exhibit a slower rate of speech, pauses, and hesitations. These disruptions in speech fluency, known as dysarthria, can be attributed to the bradykinesia and rigidity commonly associated with PD.
Variability and micro-prosody	PD can lead to reduced variability in speech, resulting in a more robotic or monotonic speech pattern. Additionally, alterations in micro-prosody, which refers to the subtle timing variations and expressive cues in speech, may be present in individuals with PD.
Vocal quality	Has an ageing effect on vocal quality, leading to changes such as reduced loudness, monotony, hoarseness, and breathiness. These vocal changes, often referred to as hypophonia, can be indicative of underlying motor impairments affecting the laryngeal muscles.
Voice tremor	PD can manifest as a tremor, including in the muscles involved in voice production. Voice tremors can result in a quivering or shaky quality of the voice.

1.4 Statement of the problem

From a clinical standpoint, managing PD poses several challenges. Currently, there is no definitive cure for PD, and while research endeavours persist in seeking a solution, the primary goal of medical support systems remains focused on symptom management and improving the quality of life for PwP. Achieving this objective necessitates systematic and ongoing patient monitoring to assess and address disease symptoms. However, existing protocols and healthcare structures often fall short in terms of monitoring frequency, hindering comprehensive tracking of disease manifestation, progression, treatment adherence, and medication optimization. This limitation arises from various contributing factors:

- Due to the motor degeneration characteristic of PD, accessing clinical facilities becomes cumbersome and logistically complex for PwP. This situation demands significant effort from both patients affected by the disease and their caregivers or family members. Consequently, PwP tend to visit clinician's offices less frequently than would be ideal.
- Given the current incidence rates of NDs healthcare systems are experiencing strain, and this effect is expected to increase leading to additional demand of dedicated resources. Consequently, regular attendance for check-ups becomes problematic, leading to undesired delays and problems with patient management.
- The progressive nature of PD necessitates adaptive responses in therapy and medication. However, the infrequency of clinical visits results in challenges related to accurate assessment, pharmacological treatment, and dosage management, which can be overly generalized and unresponsive.

- The UPDRS rating system relies on a combination of self-assessed metrics and symptom estimation by trained clinicians. This scale serves as the gold standard for assessing and categorizing PD progression. However, a notable limitation of these metrics lies in their lack of objective quantitative tests, leading to significant variability due to individual patient interpretation and clinical staff assessment.

With this in mind speech may offer support to two unmet clinical needs:

- Serve as a surrogate marker for PD, the alterations produced on speech are observable and measurable. Speech can be regularly recorded at a low cost, processed offsite and produces quasi-instantaneous interpretable data.

- Speech can be used as a telemonitoring tool for PwP. Speech can be recorded on site, but processed elsewhere and transformed into a series of clinical indicators to assess disease progression and patient management.

The aim of this project is to utilize speech as a diagnostic tool to model neuromotor dysfunction. This can be implemented by analysing the affected speech components and tracing the progression of the compromised biomechanical system back to its neuromotor origin. This approach enables the establishment of a methodology to map estimates extracted from speech to motor actions, facilitating the assessment of specific affected areas within a minimally invasive context.

1.5 Thesis Framework

The advances manifested in speech quality evaluation for clinical assessment of different pathologies have a rich history that spans several centuries, reflecting the ongoing interest in understanding and assessing the characteristics of human vocalizations.

While a comprehensive historical review is beyond the scope of this thesis, a brief overview of key developments in voice quality evaluation is due to fix the framework fo the study, which may be summarized as follows:

- Early Observations: Ancient civilizations recognized the importance of voice quality and its impact on communication, and in the assessment of emotions and health. Ancient Greek scholars, such as Aristotle and Galen, made observations about voice characteristics, including pitch, loudness, and clarity.

- Subjective evaluation: In the 19th century, speech evaluation primarily relied on subjective descriptions by clinicians and voice teachers. Experts would assess voice quality based on their own perceptions and qualitative judgments. For instance, James Parkinson informed of voice alterations in some of the six cases described on saking palsy ("…the speech was very much interrupted…", "…a similar affection of the speech, when the tongue thus outruns the mind, is termed volubility.", "…but was continually checked by the impediment in his speech, and the difficulty which his hearers were put to." Parkinson, 1817).

- Laryngeal visualization: The invention of laryngoscopy in the late 19th century by Manuel Patricio García (Radomsky, 2005), a Spanish professor of operistic singing style and vocal pedagogist in London, allowed for direct visualization of the larynx and vocal folds by means of a handle-stuck mirror. This simple device enabled clinicians to observe structural and functional abnormalities that could contribute to voice quality alterations.

- Objective acoustic analysis: In the early-20th century, with the invention of the phonography by T. A. Edison predating on an earlier design by Édouard Léon Scott de Martinville, the field of acoustic phonetics emerged, facilitating objective measurement of various voice parameters. Researchers began using spectrograms and other acoustic analysis techniques to quantify characteristics such as fundamental frequency, intensity, and spectral properties of the voice.

- Perceptual assessment: Perceptual evaluation methods were developed to assess voice quality. The Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) protocol, introduced in the late 20th century (Kempster et al., 2009), provided a standardized framework for rating vocal characteristics, including roughness, breathiness, strain, and overall severity.

- Quantitative assessment: Advances in technology led to the development of computer-based tools for voice analysis. Quantitative measures, such as jitter, shimmer, harmonics-to-noise ratio (HNR), and cepstral analysis, became increasingly utilized for objective assessment of voice quality.

- Multidimensional Approaches: Modern voice quality evaluation incorporates multidimensional approaches that combine perceptual, acoustic, and physiological measures. This integrated approach acknowledges the complexity of voice production and the interplay between various factors affecting voice quality.

- Advanced Instrumentation and Automation: Recent advancements in digital signal processing, machine learning, and artificial intelligence have paved the way for automated voice analysis systems. These systems can extract and analyze a wide range of acoustic and perceptual features, providing objective and efficient assessment of voice quality.

At this point, it is important to note that voice quality evaluation remains an active area of research, with ongoing efforts to refine assessment methods, establish normative data, and improve the clinical application of voice analysis tools.

1.5.1 Preexisting knowledge on PD manifestations on speech

The instrumental evaluation of speech and phonation using instrumental acoustic analysis allowed the definition of the examination frameworks to be capitalized when more powerful computer-based methods were available. In what follows, the focus of this subsection will concentrate on the acoustic analysis of PD speech, benefiting from translational methodological practice from the field of voice quality analysis in laryngeal and organic pathologies affecting phonation.

Early research:

- Longemann et al. (1978) studied the frequency of occurrence of speech and voice symptoms in 200 Parkinson patients by two expert listeners from high-fidelity tape recordings of conversational speech samples and readings of the sentence version of the Fisher-Logemann Test of Articulation Competence.

- Ramig and Ringel (1983) described the effects of aging on basic features of phonation (fundamental frequency, jitter, shimmer, and range) from a sample of 48 men. This can be considered a historical precedent regarding the use of acoustic analysis in the characterisation of ageing voice, which is not a study on PD properly, but given the relationship between some of the characteristics of ageing and PD speech, it can be considered a precursor study from samples of connected speech and sustained vowel production.

- Ringel and Chodzko-Zajko (1987) reported results from a longitudinal study involving 200 elderly male subjects to further investigate the influence of the subjects' physiological status on their phonatory and auditory performances on voice fundamental frequency, duration, intensity, jitter and shimmer, harmonics/noise ratios, and listener perceptions of vocal quality; measures of auditory sensitivity, discrimination, and acoustic immittance; and measures of hemodynamic, pulmonary, metabolic, and biochemical function to understand the basic mechanisms that underly control of the laryngeal mechanism.

- Caliguri (1989) addressed the question of whether or not speaking rate influences articulatory hypokinesia in dysarthria associated with Parkinson's disease. Analyses of parkinsonian speech samples revealed mean speaking rates consistent with normal controls. The results provided a physiologic basis for the perception of hypokinetic dysarthria in Parkinson's disease and suggest that speaking rate may be an important control variable contributing to articulatory hypokinesia in Parkinson's disease.

- Illes (1989) reported results from an analysis of the temporal form, syntactic form, and lexical form of spontaneous language production of early and middle stage Alzheimer's, Huntington's, and Parkinson's patients, showing that the language structure was disrupted in each disease, but in different ways.

- Ackermann and Ziegler (1991) studied acoustic speech analysis of sentence utterances to provide information on speech tempo and accuracy of articulation from twelve patients with idiopathic Parkinson's disease.

- Forrest and Weismer (1995) presented results from lower lip and jaw movement from parkinsonian dysarthric and age-matched, neurologically normal speakers during the production of alternating stress contrasts. Discrete measures of movement, including displacement amplitude, peak velocity, the relation of amplitude to peak velocity, and movement durations were compared across groups for stressed and unstressed syllables.

- Hertrich and Ackermann (1995) collected electroglottographic and acoustic recordings during sustained vowel production in men and women with PD. They concluded that PD seems to have a differential impact on phonation in men and women determined by the sexual dimorphism of laryngeal structures.

- Weismer and Wildermuth (1998) studied formant trajectories in three groups of individuals with neurogenic speech disorders, as well as a group of normaI older speakers. Results indicated that there are certain disorder-specific characteristics reflecting the classic pathophysiologies of the individual diseases.

- Kegl, Cohen, and Poizner (1999) reported a number of studies on the articulatory consequences of Parkinson's disease (PD) in the spoken and signed modalities to highlight the commonalities and distinctions between the two modalities of speech and sign that will allow to better understand the impingements of PD on language production in general.

More recent research:

The expansion of speech and phonation instrumental analysis experienced a strong push forward with the development of computer-aided applications, offering new and more detailed findings and insights. A classical reference in instrumental measurements of speech and voice in clinical practice is due to Baken and Orlikoff (2000).

- Goberman and Coelho (2002) reviewed the literature pertaining to PD and the speech dysfunction typically associated with it, including the effects on respiration, phonation, articulation, resonance, and prosody, and effects of treatment with the drug L-Dopa on Parkinsonian speech. This work was extended in Goberman, Coelho, and Robb (2002).

- Liotti et al. (2003) provided a study on PD hypophonia based on neural correlate assessment by positron emission tomography.

- Harel, Cannizaro, and Snyder (2004a), and Harel et al. (2004b) described the acoustic characteristics of PD speech in prodromal and incipient stages and their use as potential biomarkers to monitor disease progression and treatment evaluation.

- Ho, Bradshaw, and Iansek (2008) conducted a careful study on the effects of levodopa on the speech of treated PD patients.

- Tsanas et al. (2010a, b and c) published some of the first studies in the use of speech processing from remote sensors in the telemonitoring of PD. Consequent research may be found in Tsanas et al. (2010d and 2011). These studies were a fundamental support of A. Tsanas' PhD thesis (Tsanas, 2012), considered a specific indispensable reference in the field since on. These studies had a continuation in Tsanas et al. (2012), Tsanas (2013), Tsanas and Gómez (2013), and Tsanas et al. (2014).

- Skodda, Visser, and Schlegel (2011) give a description of the characteristics of vowel articulation in terms of speech rate and intonation variablity by PD patients.

- Rektorová et al. (2012) provided a functional description of the neuroanatomy of phonation in PD patients.

- Skodda, Grönheit, and Schlegel (2012) propose the use of vowel articulation features as a biomarker of PD progession. This study is further extended in Skodda, et al. (2013).

- Hanratty et al. (2016) analysed the use of glottal source features to characterize PD speech.

- Ricciardi et al. (2016) studied the relationship between speech disturbances in PD and gait on a data sample from 43 speakers of Italian.

- Godino et al. (2017) offered a very complete study on idiopatic PD speech including kinematic biomarkers.

- Mekyska et al. (2018) studied the relationship between hypokinetic dysarthria and gait disorders in PD.

- Pinho, et al. (2018) presented a review on the effects of levodopa on PD speech.

- Gillivan-Murphy, Miller, and Carding (2018) aimed to evaluate voice tremor in people with PD (pwPD) and a matched control group using acoustic analysis, and to examine correlations with voice disability and disease variables.

- Further work on PD using statistical analysis to quantify symptom severity may be found in Tsanas (2019), Tsanas and Arora (2019), Tsanas and Arora (2020), Tsanas and Arora (2021), Tsanas, Little and Ramig (2021), and Tsanas and Arora (2022).

1.5.2 Speech based discrimination between PD and HC

Since the identification of the intrinsic characteristics of speech affected by PD attempts have been undertaken to use them as differentiation tool between healthy controls and PwP. The differentiation stems from a loss in overall movement control, this in turn causes: monotone pitch, reduced loudness, imprecise articulation, altered rate of speech and breathy or hoarse voice quality.

Early research

- Robbins, Logemann, and Kirshner (1986) used videofluoroscopy to examine movement patterns during swallowing and speech production in six parkinsonian subjects and six age-matched controls. Duration of velar movement during speech production significantly differentiated the groups, reflecting reduced range of velar motion.

- Ludlow, Connor, and Bassich (1987) compared patients with Parkinson's Disease (PD) and Huntington's Disease (HD) on speech timing tasks. The results suggested that PD and HD patients are not impaired in speech planning or initiation, but have poor control over the duration of speech events.

- Forrest, Weismer, and Turner (1989) published an interesting study on the kinematic, acoustical, and perceptual analysis of PD speech on PD patients and normative aging participants, considering jaw displacements and velocities, and lip movement amplitude and velocity. Acoustically, the Parkinsonian subjects had reduced durations of vocalic segments, reduced formant transitions, and increased voice onset time compared to their age-paired healthy controls.

- Zwirner, Murry, and Woodson (1991) studied, five parameters of phonatory function (jitter, shimmer, signal-to-noise ratio, fundamental frequency, and standard deviation of fundamental frequency) from samples of sustained phonation in three neuropathological groups (Parkinson, Huntington, cerebellar ataxia) and a normal control group to assess the use of acoustic measures in differential diagnosis. The results indicated that perturbation measures of the neuropathological groups showed a higher degree of variability compared to normative subjects.

- Svensson, Henningson, and Karlsson (1993) conducted a kinematic analysis of vertical jaw movements during speech was performed by using an optoelectronic technique on nine individuals with Parkinson's disease (PD) and nine normal control subjects, matched for sex and age. Significant group differences were found for all kinematic measures during a syllable repetition task, as well as for the total dysarthria test scores and certain individual test items.

- Jiménez et al. (1997) quantified several acoustic features of the voice in 22 PD patients and 28 age and sex-matched controls using The Computerized Speech Lab 4300 program (Kay Elemetrics) on two seconds of a sustained /a/ and a sentence. Measures included fundamental frequency (F0), frequency perturbation (jitter), intensity perturbation (shimmer), and harmonic/noise ratio (HIN) of the vowel /a/, and frequency and intensity variability of a sentence, phonational range, dynamic range at the natural frequency, maximum phonational time and ratio. When compared to controls, PD patients showed higher jitter and shimmer, lower ratio, and lower frequency variability of the sentence in the microphonic signal and reported a higher frequency of presence of low voice intensity, monopitch, harshness, voice arrests, and tremor. This study was further extended on 41 PD participants (Gamboa et al. 1997), showing similar conclusions.

- Louis, E. D., Klatka, L. A., Liu, Y., and Fahn, S. (1997). Comparison of extrapyramidal features in 31 pathologically confirmed cases of diffuse Lewy body disease and 34 pathologically confirmed cases of Parkinson's disease. Neurology, 48(2), 376-380. https://doi.org/10.1212/WNL.48.2.376.

- Le Dorze et al. (1998) The realization of prosody (speech rate, fundamental frequency, intonation) was investigated in a group of 10 individuals with Parkinson's disease and a group of 10 individuals with Friedreich's ataxia. Data from these two neurologically disordered groups were compared to individuals without neurological impairment. Both neurologically impaired groups retained some aspects of normal speech prosody, while other aspects were affected to a significant degree. The prosodic characteristics of speakers with Parkinson's disease were distinct from those of speakers with Friedreich's ataxia. These results were interpreted in terms of prosodic competence and prosodic performance.

- Holmes et al. (2000) examined the acoustic and perceptual voice characteristics of patients with Parkinson's disease according to disease severity. The perceptual and acoustic voice characteristics of 30 patients with early stage PD and 30 patients with later stage PD were compared with data from 30 normal control subjects. In comparison with controls both early and later stage PD patients' voices were characterized perceptually by limited pitch and loudness variability, breathiness, harshness and reduced loudness. Acoustically, the voices of both groups of PD patients demonstrated lower mean intensity levels and reduced maximum phonation frequency ranges in comparison with normative data. Data also suggested that the PD patients' voices were characterized by excess jitter, a high-speaking fundamental frequency for males and a reduced fundamental frequency variability for females. Tremor was the sole voice feature which was associated only with later stage PD.

More recent research:

- Yunusova et al. (2008) reported the characteristics of articulatory movements from vowels in speakers with dysarthria compared with healthy controls.

- Gillivan-Murphy (2013) conducted a study in her PhD thesis on voice tremor in 30 persons with PD 'off-medication' and 28 age-sex matched neurologically healthy controls evaluated for voice tremor features using acoustic measurement, auditory perceptual voice rating, and nasendoscopic vocal tract examination. Speech intelligibility, severity of voice impairment, voice disability and disease variables (duration, disability, motor symptom severity, phenotype) were measured and examined for relationship with acoustic voice tremor measures.

- Belalcazar-Bolaños et al. (2015) proposed the estimation of different glottal flow features considering nonlinear behavior of the vocal folds to evaluate the discrimination capability of eight nonlinear dynamic features. The experiment included the five Spanish vowels uttered by 50 People with PD (PPD) and 50 Healthy Controls (HC).

- Tykalová et al. (2020) compared speech disorder between patients with the postural instability/gait difficulty and tremor-dominant motor phenotypes of PD. Speech samples were acquired from a total of 63 participants, 21 with postural instability, 21 tremor-dominant, and 21 healthy controls. Quantitative acoustic vocal assessment of 12 unique speech dimensions related to phonation, vocal tremor, oral diadochokinesis, articulation, prosody and speech timing was performed. The study demonstrated that speech disorder reflects the underlying motor phenotypes. Vocal tremor appeared to be an isolated phenomenon that does not share similar pathophysiology with limb tremor.

- Ozbolt et al. (2022) proposed a summary of methodological issues to be considered takint into account that in PD characterization from speech unaccounted covariates in methodology, experimental design, and data preparation resulted in overly optimistic results employing sustained vowels, such as including record-wise fold creation rather than subject-wise; an imbalance of age between PD participants and healthy controls; using too small corpora; etc. In their study they perform several experiments isolating each issue to measure its influence on three different corpora. Results suggest that each independent methodological issue analysed has an effect on classification accuracy.

1.5.3 Speech in prodromal PD

It is possible to detect prodromal Parkinson's disease through speech. Recent studies have shown that machine learning algorithms can predict Parkinson's disease with a high degree of accuracy. These findings suggest that speech analysis, particularly when combined with advanced machine learning techniques, can be a valuable non-invasive method for the early detection of Parkinson's disease.

- Ho et al (1998) classified speech impairment in 200 patients with PD into five levels of overall severity and described the corresponding type (voice, articulation, fluency) and extent (rated on a five-point scale) of impairment for each level from two-minute conversational speech samples. Parameters of voice, fluency and articulation were assessed by two trained-raters. Voice was found to be the leading deficit, impaired to a greater extent than other features in the initial stages. Articulatory and fluency deficits manifested later, articulatory. At the profound impairment stage articulation was the most frequently impaired feature, drawing parallels with deficits of motor performance in gait and handwriting.

- Postuma et al. (2012) published a dedailed description on the relationship between prodromal PD development and REM sleep behaviour disorder.

- Rusz et al (2013) explored the use of vowel articulation dispersion as a biomarker of prodromal PD.

- Orozco et al. (2015) presented a study on the use of different phonation features (stability and periodicity, noise-to-signal, spectral density, and nonlinear dynamics) for the detection of laryngeal, functional, and neurological disorders using an SVM as a classifier.

- Orozco et al. (2016) published a multi-lingual (Spanish, Czech, and German) approach to PD detection using an SVM as a classifier.

- Mu et al. (2017) used cluster analysis to search for subtypes from a large, cohort of Parkinson's disease patients across all motor stages, using motor features (bradykinesia, rigidity, tremor, axial signs) and rater-based non-motor symptom scales.

- Arora, Baghai-Ravari, and Tsanas (2019) This study presented a statistical framework to account for variations in phonetic backgrounds in telephone-quality voice analysis for PD detection. The statistical framework for analyzing voice was based on 307 dysphonia measures that quantify different properties of voice impairment, such as breathiness, roughness, monopitch, hoarse voice quality, and exaggerated vocal tremor, were computed. Feature selection algorithms were used to identify robust parsimonious feature subsets, which were used in combination with a random forests (RFs) classifier to accurately distinguish PD from HC.

- Kadiri, Kethireddy, and Alku (2020) proposed to use cepstral coefficients derived from the single frequency filtering (SFF) method for the detection of PD. SFF has been shown to provide higher spectro-temporal resolution compared to the short-time Fourier transform. The study used the PC-GITA database, which consists of speech from speakers with PD and healthy controls (50 males, 50 females). The proposed detection system was based on the i-vectors derived from SFFCCs using SVM as a classifier.

- Moro et al. (2021) conducted a comprehensive review to identify the most common features and machine learning methods used in detecting and assessing the severity of PD by phonatory and articulatory aspects of speech and voice to provide a broad overview on the evidence that articulatory and phonatory aspects of speech and voice are meaningful for the automatic detection and severity assessment of PD. A historical perspective of publications in the field since 1956 to 2020 is given.

- Šimek and Rusz (2021) aimed to investigate the voice changes via the CPP measures in the idiopathic rapid eye movement sleep behaviour disorder, and recently diagnosed and advanced-stage Parkinson's disease patients across noise-free and noisy environments. The sustained vowel phonation, reading of passages, and monologues of 60 early-stage untreated PD, 30 advanced-stage Parkinson's disease, 60 participants affected by REM beharvior disorder, and 60 healthy control participants were evaluated.

- Madruga, Campos, and Pérez (2023) analyzed the effects of recording device mismatch in PD speech classification. Multicondition training was proposed to improve robustness against mismatch in an experiment on 30 PD patients and 30 healthy controls. Three vocalizations of sustained /a/ were recorded using different devices. Acoustical features were extracted and averaged per patient and recording device. Machine learning was used to distinguish healthy from PD patients by using different combinations of train-test smartphones.

- Zhang et al. (2023a) proposed a method to distinguish PD patients from healthy controls, combined with the idea of formal structure analysis according to the direction information statistically obtained in a sub-region of the spectrogram to describe the correspondence between energy points and their direction attributes to obtain the coupling information between the direction attributes in the formal context. The number of connected domains in indicate the degree of nodal coupling is used as input to multiple classifiers for validation purposes.

- Zhang, Lin, and Xue (2023b) proposed a methodology for feature extraction from PD speech based on fractional Fourier transform to obtain the spectrograms at different orders. The energy variation information in the spectrograms at each order is estimated, and converted through a mapping relationship between energy points and directional attributes. The connected component features are fed into different classifiers, such as linear regressors, SVMs, random forests, and multilayer perceptrons.

1.6 Objectives

The disabling nature of PD poses significant challenges for PwP healthcare providers, support staff, and caregivers. These challenges can be broadly categorized into two key areas:

- The first area pertains to the understanding and management of disease manifestation, encompassing aspects such as early diagnosis, tracking disease progression, and assessing the efficacy of medication.

- The second area of focus revolves around therapeutic interventions and improving the quality of life for PwP. This includes the development of neuroprotective therapies aimed at slowing down disease progression through personalized rehabilitation strategies to address the specific requirements of an individual.

The primary focus of this study is to develop and validate a functioning biomechanical model of speech production based on the temporomandibular system. Substantial effort has been dedicated to comprehending the underlying biological processes, investigating the dynamics of the jaw-tongue system, and examining associated correlates using speech-derived features, 3D acceleration, and sEMG, this integration of additional signals collected concurrently with speech may provide new additional insights into the pathophysiology of PD. This modelization methodology extends beyond PD, encompassing other NDs as well as the fields of linguistics and speech therapy, where a similar approach might produce positive results.

The rationale for conducting this study lies in the potential to establish a reliable validation benchmark for remote telemonitoring of NMDs solely based on the acoustic analysis of speech. By employing statistical machine learning techniques, the analysis of various biomarkers can be translated into a readily understandable set of indicators, suitable for routine utilization in clinical settings as a diagnostic support tool.

As reflected by a conscientious analysis of the state-of-the-art, it may be concluded that computer-assisted characterization of PD based on acoustic analysis has reached a maturity state. To a large extent, these studies do not provide the same level of insights that mechanistic models can provide, i.e., models that build on the physical principles of voice production to characterize the underlying vocal production mechanism and PD-related pathology (Duffy, 2013). Further progress will need to dive into mechanistic models explaining the neuroacoustical foundation underlying neuromotor control to provide new insights into the underlying neurophysiological processes, which in turn might inspire further signal processing algorithms for the characterization of speech signals.

The main aim of the present PhD thesis is to describe some exploratory studies in PD phonation and articulation which may open the possibility of opening new hypotheses to add new knowledge on the mechanisms underlying altered speech production in NMDs (Darbin and Montgomery, 2022).

The study aims to validate a temporomandibular biomechanical model of speech production by understanding the underlying biological processes and the associated correlates which could serve as biomarkers in NMDs, using speech-derived features, 3D acceleration, and sEMG.

1.7 Hypotheses

- Speech-derived correlates can be validated from features estimated from sEMG and accelerometry.
- The manifestation of PD can be analysed by studying the frequential traces on the different EEG bands of the glottal signals.
- Interpretable first-principle models provide novel insights into the underlying phenomenon to be characterised.

1.8 Summary of contributions

This doctoral research aims to address a specific gap in the study of neuromotor speech disorders. While speech affected by PD has traditionally been characterized using acoustic features such as bandwidth, vowel space area, spectral components, activity indicators, duration, energy, formants, intensity, Linear Predictive coefficients, Mel Cepstrum coefficients, pitch, zero crossings, and speaking rate, this Ph.D. takes a novel approach. Rather than interpreting these indicators as mere features, it seeks to draw conclusions about the underlying biomechanical systems based on their behaviour.

The research involves modelling speech behaviour using first-principle models and applying statistical analysis and information theory to describe these behaviours.

The aim of this Ph.D. was never to attempt maximizing classification scores but to provide explainable interpretable features and methods. To this end, by characterizing speech production and observing the pathological elements of speech, inferences can be made about the biomechanical elements at fault. By having a deep understanding on the driving neurological features responsible for controlling the system, the standing hypothesis is that it would be possible to backtrace the dysfunction to the brain areas at fault, setting speech up for as a reliable exploratory tool that is non-invasive, inexpensive and easy to process in the clinician's diagnostic toolset..

The present thesis summarizes the research activities conducted since my engagement in the DTP PM of UoE in 2019, with special emphasis on the definition and design of three experimental frameworks around the characterization of the glottal source in phonated speech, and on the projection of acoustical features into articulation kinematics. The related research activity is briefly described as follows:

- A methodology for PD characterization based on the amplitude distributions of glottal source and flow as features has been proposed and tested to differentiate PD from HC phonation. This work is described in detail in subsections 5.1.1 (experimental design), 6.1.1 (results), and 7.1 (discussion). It has been published in Gómez-Rodellar et al. (2019c) and Gómez-Rodellar et al. (2020a).

- A methodology for PD characterization based on biomechanical features of the vocal folds during phonation has been devised and tested on PD participants submitted to rTMS. This work is described in detail in subsections 5.1.2 (experimental design), 6.1.2 (results), and 7.2 (discussion). It has been published in Gómez-Rodellar et al. (2021d), Gómez-Rodellar et al. (2022b), and Gómez-Rodellar et al. (2023).

- A neuromechanical model of jaw-tongue articulation in Parkinson's disease speech has been designed and tested on diadochokinetic tests from PD and HC participants. Besides, individual mandibular motor actions have been estimated on speech articulation features using this model. This work is described in detail in subsections 5.2 (experimental design), 6.2 (results), and 7.3 (discussion). It has been published in Gómez-Rodellar et al. (2019a), Gómez-Rodellar et al. (2019b), Gómez-Rodellar et al. (2020b), Gómez-Rodellar et al. (2021a), Gómez-Rodellar et al. (2021b), Gómez-Rodellar et al. (2022a).

The following is a detailed and commented list of the publications leading to the above-mentioned activities in journals and conference papers:

Journal publications as first author:

Gómez-Rodellar, A., Palacios, D., Ferrández-Vicente, J. M., Mekyska, J., Álvarez-Marquina, A., and Gómez, P. (2020a). A methodology to differentiate Parkinson's disease and aging speech based on glottal flow acoustic analysis. International Journal of Neural Systems, 30(10), 2050058. https://doi.org/10.1142/S0129065720500586.

This journal paper extends the results from previous conference papers on PD and aging phonation using a larger database.

Gómez-Rodellar, A., Gómez-Vilda, P., Palacios, D., Rodellar-Biarge, V., Nieto, V., Álvarez-Marquina, A., and Tsanas, A. (2021a). A Neuromotor to Acoustical Jaw-Tongue Projection Model with Application in Parkinson's Disease Hypokinetic Dysarthria. *Frontiers in Human Neuroscience*, *15*, 622825.

This journal paper proposes the study of the neuromotor activity of the masseter-jaw-tongue articulation during diadochokinetic exercising to establish functional statistical relationships between surface Electromyography (sEMG), 3D Accelerometry (3DAcc), and acoustic features extracted from the speech signal.

Gómez-Rodellar, A., Tsanas, A., Gómez, P., Palacios, D., Rodellar-Biarge, V. and Álvarez-Marquina, A. (2021b). Acoustic to Kinematic Projection in Parkinson's Disease Dysarthria. Biomedical Signal Processing and Control 66 102422. https://doi.org/ 10.1016/j.bspc.2021.102422.

This journal paper extends the results from previous conference papers on acoustical signals to kinematic features, using improved models and larger databases.

Gómez-Rodellar, A., Mekyska, J., Gómez-Vilda, P., Brabenec, L., Šimko, P., and Rektorová, I. (2023). A Pilot Study on the Functional Stability of Phonation in EEG Bands After Repetitive Transcranial Magnetic Stimulation in Parkinson's Disease. International Journal of Neural Systems, 2350028.

This journal paper shows the feasibility of estimating EEG-related NMA on the extrapyramidal neural pathways.

Other journal publications as co-author:

Gómez-Vilda, P., Galaz, Z., Mekyska, J., Ferrández-Vicente, J. M., **Gómez-Rodellar, A.**, Palacios, D., Smekal, Z., Eliasova, I., Kostalova, M., Rektorová, I. (2019c) Vowel Articulation Dynamic Stability Related to Parkinson's Disease Rating Features: Male Dataset, Int. Journal of Neural Systems 28(2), 1850037 (13pages). https://doi.org/10.1142/S0129065718500375.

This journal paper presents the results of using articulation kinematic features in the detection of PD on a moderate-size database of male participants.

Gómez-Vilda, P., **Gómez-Rodellar, A.**, Ferrández-Vicente, J. M., Mekyska, J., Palacios, D., Rodellar-Biarge, V., Álvarez-Marquina, A., Smekal, Z., Eliasova, I., Kostalova, M., Rektorová, I. (2019a) Neuromechanical Modelling of Articulatory Movements from Surface Electromyography and Speech Formants. International Journal of Neural Systems, 29:2, 1850039, doi: 10.1142/S0129065718500399.

This journal paper shows the validation of the articulation to kinematic projection model using surface electromyography recordings.

Gómez-Vilda, P., Mekyska, J., **Gómez-Rodellar, A.**, Palacios, D., Rodellar-Biarge, V., Álvarez-Marquina, A. (2019b) Characterization of Parkinson's disease dysarthria in terms of speech articulation kinematics. Biomedical Signal Processing and Control, 52, 312-320, doi: 10.1016/j.bspc.2019.04.029.

This journal paper presents an early elaborated version of the articulation to kinematic projection model using accelerometry.

Conference publications as first author:

Gómez-Rodellar, A., Tsanas, A., Gómez, P., Palacios, D., Álvarez-Marquina, A., Martínez. R. (2019a) A Neuromechanical Model of Jaw-Tongue Articulation in Parkinson's Disease Speech. Proc. of MAVEBA 19; 25-28, Firenze University Press, December 17-19.

This conference paper is an early version of the jaw-tongue articulation model to project acoustical signals into kinematic features, extending previous work modelling the jaw-tongue biomechanical system to further investigate neuromotor activity in muscular activity during certain speech gestures to model hypokinetic dysarthria in Parkinson's Disease (PD) patients.

Gómez-Rodellar, A., Palacios, D., Mekyska, J., Álvarez-Marquina, A., and Gómez, P. (2019b) Comparing Parkinson's Disease Dysarthria and Aging Speech using Articulation Kinematics, Proc. 12th International Joint Conference on Biomedical Engineering Systems and Technologies, F. Putze, A. Fred and H. Gamboa (Eds.), SCITEPRESS, Lisbon, Portugal 52-61. https://doi.org/10.5220/0007355700520061.

In this conference paper the results from comparing speech from PD and aging voice participants showed some of the confounding factors to be taken into account.

Gómez-Rodellar, A., Palacios, D., Ferrández-Vicente, J. M., Mekyska, J., Álvarez-Marquina, A., and Gómez, P. (2019c) Evaluating Instability on Phonation in Parkinson's Disease and Aging Speech, Lecture Notes on Computer Science, 11487(2) 340-351. https://doi.org/10.1007/978-3-030-19651-6_33.

This conference paper summarizes the results from comparing laryngeal biomechanical features obtained from PD and ageing voice with a reference normative population.

Gómez-Rodellar, A., Tsanas, A., Gómez-Vilda, P., Álvarez-Marquina, A., and Palacios-Alonso, D. (2020b). Individual Mandibular Motor Actions Estimated from Speech Articulation. In LREC 2020 Language Resources and Evaluation Conference 11-16 May 2020 (p. 74).

This study aims to compare the articulation characteristics of a person with Parkinson's Disease (PD) with the articulation characteristics of a healthy person on the neuromotor principles of speech production. The study methodology is based on the recording of speech as a vehicular signal accompanied by other multimodal traits associated, such as the surface Electromyography in the masseter and the acceleration in the chin.

Gómez-Rodellar, A., Mekyska, J., Brabenec, L., Simko, P., Rektorová, I., Gómez, P., and Tsanas, A. (2021d). Longitudinal Effect of Repetitive Transcranial Magnetic Stimulation on Phonation in a Patient with Parkinson's Disease: A Case Study. Claudia Manfredi (Ed.), Models and Analysis of Vocal Emissions for Biomedical Applications: 12th International Workshop, December, 14-16, 2021, Firenze University Press (www.fupress.com), pp. 157-160.. https://doi.org/10.36253/978-88-5518-449-6.

This work describes a case study exploring the longitudinal effect of repetitive Transcranial Magnetic Stimulation (rTMS) on hypokinetic dysarthria in a patient with Parkinson's Disease (PD).

Gómez-Rodellar, A. and Tsanas, A. (2021c). F0 Estimation in Irregular Vocal Emissions using Ridge Detection Methods. Claudia Manfredi (Ed.), Models and Analysis of Vocal Emissions for Biomedical Applications: 12th International Workshop, December, 14-16, 2021, Firenze University Press (www.fupress.com). https://doi.org/10.36253/978-88-5518-449-6.

This study analyses the F0 estimation using artificially generated [a:] vowels by employing exploratory functions (kernels) to analyze the repetitive structure found in the Auto Correlation Function (ACF).

Gómez-Rodellar, A., Gómez-Vilda, P., Ferrández-Vicente, J., and Tsanas, A. (2022a). Characterizing Masseter Surface Electromyography on EEG-Related Frequency Bands in Parkinson's Disease Neuromotor Dysarthria. In International Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 219-228). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-06242-1_22.

This study aims to evaluate the behaviour of facial muscles' activity estimating the entropy of their surface electromyographic (sEMG) activity during the production of diadochokinetic speech tests.

Gómez-Rodellar, A., Mekyska, J., Gómez-Vilda, P., Brabenec, L., Simko, P., and Rektorová, I. (2022b). Evaluation of TMS Effects on the Phonation of Parkinson's Disease Patients. In International Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 199-208). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-06242-1_20.

This paper is devoted to describing the potential beneficial effects of rTMS on the phonation stability of Parkinson's Disease Patients (PDPs).

Other conference publications as co-author:

Álvarez-Marquina, A., **Gómez-Rodellar, A.**, Palacios-Alonso, D., Mekyska, J., Tsanas, A., Gómez, P., and Martínez, R. (2020). Parkinson's Disease Glottal Flow Characterization: Phonation Features vs Amplitude Distributions. In BIOSIGNALS (pp. 359-368). https://doi.org/10.5220/0009189403590368.

This work resorts to theoretical modelling of glottal signals under the main known causes affecting phonation quality, which are closure deficits during the phonation cycle.

Gómez-Vilda, P., **Gómez-Rodellar, A.**, Palacios-Alonso, D., and Tsanas, A. (2021). Performance of monosyllabic vs multisyllabic diadochokinetic exercises in evaluating Parkinson's disease hypokinetic dysarthria from fluency distributions. In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies—BIOSIGNALS* (pp. 114-123).

The present work aims to explore the timely responsive performance of two of these exercises (a monosyllabic [ta] vs a multisyllabic [pataka]) when uttered by Parkinson's Disease participants compared to controls.

Álvarez-Marquina, A., **Gómez-Rodellar, A.**, Gómez-Vilda, P., Palacios-Alonso, D., and Díaz-Pérez, F. (2022). Identification of Parkinson's Disease from Speech Using CNNs and Formant Measures. In *International Work-Conference on the Interplay Between Natural and Artificial Computation* (pp. 332-342). Cham: Springer International Publishing.

Through the present work the use of machine learning-based technologies, more specifically the Convolutional Neural Networks (CNNs) and the direct application of formant features extracted from sustained phonations of vowel /a/ are proposed.

Gómez-Vilda, P., **Gómez-Rodellar, A.**, Palacios-Alonso, D., Álvarez-Marquina, A., and Tsanas, A. (2022). Characterization of Hypokinetic Dysarthria by a CNN Based on Auditory Receptive Fields. In *International Work-Conference on the Interplay Between Natural and Artificial Computation* (pp. 343-352). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-06242-1_34.

This study aims to evaluate the processing of speech from people diagnosed with Parkinson's Disease using Convolutional Neural Networks (CNN) towards characterizing speech articulation kinematics to explore differences between Healthy Controls (HC) and PD participants with Hypokinetic Dysarthria (HD), using Auditory Receptive Fields (ARFs) in the convolutional layers.

Gómez-Vilda, P., Mekyska, J., Brabenec, L., Šimko, P., Rektorová, I., **Gómez-Rodellar, A.**, and Rodellar-Biarge, V. (2023a). Description of PD Phonation in Terms of EEG-Related Frequency Bands. In Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2023) - Volume 4: BIOSIGNALS, pages 226-233. https://doi.org/10.5220/0011669100003414.

The present study concentrates on analyzing and comparing the phonation behaviour of two cases before (pre-stimulus) and after (post-stimulus) ten sessions of rTMS treatment, to assess the extent of changes in their vocalization from the EEG-band description of glottal biomechanical features.

Gómez-Vilda, P., **Gómez-Rodellar, A.**, Palacios, D., and Tsanas, A. (2023b). Evaluating the Performance of Diadochokinetic Tests in Characterizing Parkinson's Disease Hypokinetic Dysarthria. In Biomedical Engineering Systems and Technologies: 14th International Joint Conference, BIOSTEC 2021, Virtual Event, February 11–13, 2021, Revised Selected Papers (pp. 102-119). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-20664-1_6.

The present work aims to explore the performance of tests consisting of a monosyllabic repetition [...tatata...] vs a multi-syllable one [...pataka...]) in the characterisation of PD speech.

1.9 Thesis structure

The manuscript is organized in the classical IMRAD structure (Sollaci and Pereira, 2004). Chapter 1 an introduction, definition of motivations, statement of the problem, report of the state-of-the-art, working hypotheses, and general overview of research framework. Chapter 2 offers a detailed description of the fundamentals of speech production, describing its physiological, functional, and neurological structures, and giving a brief reference to neuromotor alterations due to PD. Chapter 3 concentrated on describing speech neuromotor disorder assessment foundations. Chapter 4 is devoted to describing the algorithmic methods behind neuromotor disorder assessment. Chapter 5 summarizes the experimental design to provide the materials for three studies based on the use of glottal flow, vocal fold biomechanics, and jaw-tongue kinematics to assess speech alterations due to PD. Chapter 6 is intended to present the results of these studies. Chapter 7 concentrated on the discussion of the results presented to describe contributions, insights, limitations, and open objectives to new lines of study. Chapter 8 summarizes the findings, reflections, and conclusions derived from the study.

CHAPTER 2

2 Fundamentals of speech production

Speech production comprises physiological, anatomical, and cognitive processes, involving the coordination of various structures and mechanisms within the human body to produce and articulate sounds, which convey meaning according to a pre-established communication code. Speech is often confused with voice, and they often are used interchangeably. Voice has to do with sound production while speech is the articulated outcome; for example, a problem with speech production would be hoarseness while a problem with speech production would be hypernasal-sounding words. Although phonation is the basis of voiced speech, it is not present in voiceless speech sounds. This fact establishes an essential distinction between voice and speech: voice need not be always associated with speech (consider cough, or laughter), and speech need not be always voiced (consider whispering).

Speech production description might consider the following key elements:

1. Respiration: Speech production begins with respiration, where the air is drawn into the lungs and exhaled in a controlled motion to produce voiced and unvoiced speech.

2. Phonation: This refers to the generation of sound by the vocal folds (also known improperly as vocal cords) located in the larynx. As air moves through the vocal folds, they vibrate, generating a pulsed periodic pattern that travels toward the upper oral and nasal cavities.

3. Articulation: As the glottal signal travels outward, it is modified by the shape of the Oro-Naso-Pharyngeal Tract (ONPT), this airway is constricted or expanded through articulation gestures (static or dynamic) of various modifiable structures, such as the lips, tongue, teeth, and palate. These gestures create different speech sounds or phonemes. In voiceless speech, articulation organs produce similar modifications in the turbulent flow of air constituting the unvoiced source of speech to a given extent.

4. Resonance: As the glottal signal (in voiced speech) or the turbulent airflow (voiceless sounds) travels through the ONPT parts of the source pressure signal is reflected backwards generating standing waves. These waves are a well-understood phenomenon in physics, they present a series of local maxima and minima spaced at fixed intervals. The resonance properties of these cavities shape the temporal (onset, nucleus, and decay) and spectral (harmonic display) qualities of the sound, resulting in distinct vocal productions.

5. Prosody: Although prosody is not directly derived from the biomechanical aspects of speech production it plays a major role regarding speech communication. Prosody encompasses the rhythm, intonation, and stress patterns in speech. It involves variations in fundamental frequency, loudness, and duration that convey emotional or grammatical information. The variations in fundamental frequency and loudness introduce important meta-information to speech, such as modulating emphasis in declarative or interrogative statements. In tonal languages, important interpretation aspects are embedded in the fundamental tone. Prosody plays a crucial role in conveying meaning and adding nuance to spoken language.

The production of speech involves complex cognitive processes, including language planning, motor control, and monitoring. The brain coordinates and regulates the complex sequential and precise movements required for speech production, ensuring that the intended message is accurately conveyed.

Understanding the fundamentals of speech production is essential for speech-language pathologists, linguists, and researchers to diagnose and treat speech disorders, study language acquisition and development, and improve understanding of human communication.

2.1 Physiology

The physiology of the human speech production system is constituted by the respiratory, phonatory, and articulatory structures, and the CNS and PNS areas and pathways controlling each of these structures, as represented in Figure 2.1.

2.1.1 The respiratory system

Inhalation and expiration is a natural and autonomous process ensuring the necessary amount of oxygen to keep the metabolic processes responsible for energy production in cellular mitochondria. A secondary function allows the production of harmonic (quasi-periodical) voiced sounds by the vibration of the vocal folds, or frictional (turbulent) unvoiced sounds by disordered airflow. The main muscle controlling respiration is the diaphragm, a strong vault-like muscle separating the thoracic cavity from the abdominal space (Pickering and Jones, 2002); the activity of this muscle is driven by a branch of the vagus nerve. During inhalation, this muscle flattens and tenses creating extra space in the thoracic cavity, which is filled with fresh air. During expiration, the relaxation of the diaphragm, combined with the tensioning of the inter-costal muscles results in a reduction of the thoracic cavity, forcing air outward.

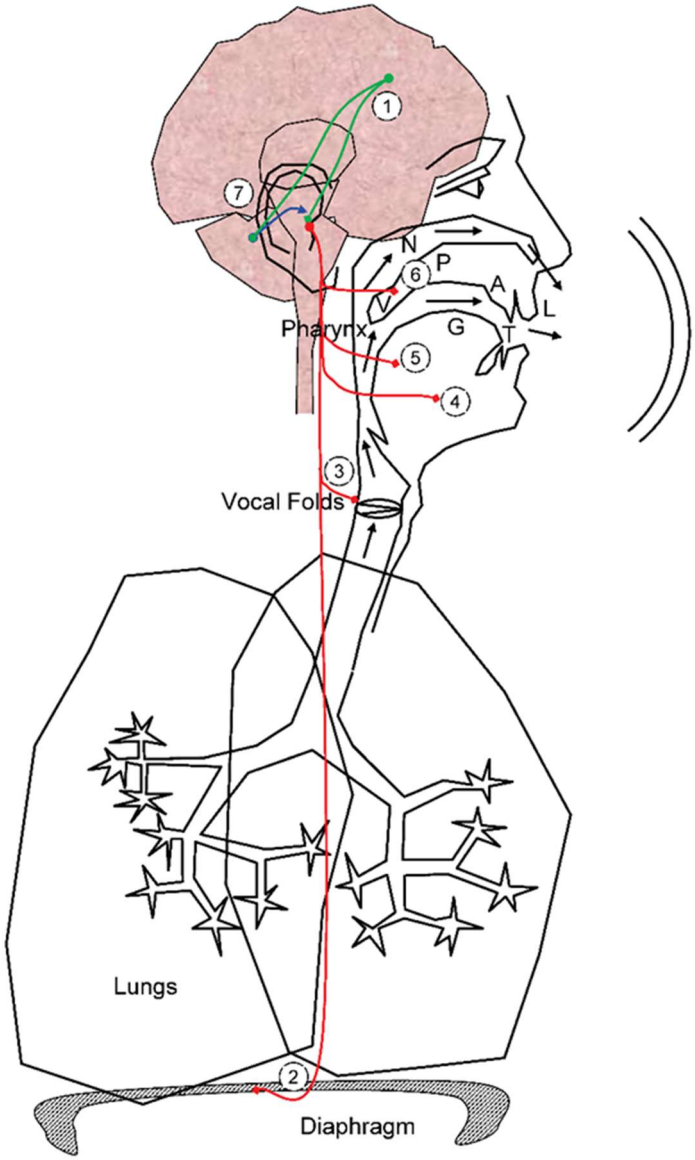

Figure 2.1 Structural schema of the speech and phonation production system.
Description: 1) direct pathway from cortical areas controlling phonation and articulation to subthalamic nucleus; 2) muscular structures controlling inhalation and expiration, comprising mainly the diaphragm and the intercostal muscles; 3) vocal folds and supporting cartilages and laryngeal muscles; 4) mandibular and other extrinsic muscles responsible of raising and lowering the jaw-tongue complex; 5) intrinsic lingual muscles responsible of stretching, enlarging, flattening or rounding the tongue; 6) velopharyngeal muscles blocking or opening the nasopharyngeal passage; 7) indirect neural connections involving BG red).

The muscular and skeletal walls of the thoracic cavity and the bronchial air ducts are relatively flexible, and together with the natural compressibility of air allow the creation of a surplus pressure when the air passage at the vocal folds is stopped. This ability explains the important subglottal pressure build-ups during vocal fold contact, helping the injection of airflow pulses during phonation.

2.1.2 The phonation system

The subglottal pulmonary structures (trachea, bronchi, alveoli) constitute a septic-free space, where any extraneous substances or matter, such as dust, water, saliva, or food might induce dangerous infectious and strong inflammatory responses. To prevent this risk to vital functions, the biology of mammals designed a safety system, consisting of a series of cartilages and muscle structures, responsible for generating an immediate response through a reflex circuit composed of afferent neurons in the laryngeal space acting directly on efferent neurons in the bulbar midbrain, protracting muscular and conjunctive tissular structures at both sides of the laryngeal space, known as vocal folds, which produce an immediate blocking of the air passage (the glottis). Besides, the diaphragm and intercostal muscles are promoted to produce an important supraglottal pressure build-up enough to overcome the vocal fold tenseness, producing a sudden violent airflow (cough) to remove any of the extraneous substances, which is perceived as a notorious and worrying sound. These two mechanisms have been adapted by many animals to voluntarily produce voiced sounds and are the basis of voiced speech as well.

The mechanical foundation behind their function is relatively simple: with previously stored air in the lungs, a simple and voluntary gesture of the laryngeal nerves brings the vocal folds together (adduction) closing the glottis and the airflow.

Immediately after, the relaxation of the diaphragm and the contraction of the inter-coastal muscles promote a moderate pressure build-up in the lung-bronchi-trachea space. This build-up forces both vocal folds to separate (abduction), and a puff of air is expelled. The burst of gas from the subglottal cavity causes a drop in the subglottal pressure, and the reactive muscular forces acting on the vocal folds force their adduction again. As the balance between subglottal pressure build-up and vocal fold occluding tenseness is subtle and prone to voluntary control, the duration, intensity, and repetition rate of the adduction-abduction process is easily controlled to produce repetitive airflow releases and stops. The release of an airflow burst produces an increment in the supraglottal cavity pressure, and its constriction produces a drop in the same pressure. As it happens that the pressure increments and drops are proportional to the rate of flow changes, positive and negative dynamic sound pressure waves are induced in the glottis, which propagate through the ONPT. Given that the flow decay rate during adduction is much larger than the flow increment during abduction, the negative sound pressure waves are prevalent as the dominant sound excitation expressing phonation. This sound pressure wave at the glottis is known as the glottal source.

2.1.3 The articulatory system

An oversimplification of the speech production process assumes many times that the glottal source is the only way of generating a signal to excite the vocal tract and experience the resonant phenomena, that is, enhancing and attenuating parts of its original power spectrum to produce vowel-like timbres. This assumption, the standard for voiced sounds, does not cover all possibilities as stated previously voiceless sounds are a consequence of a turbulent source.

This is a relevant factor as the alternation between voiced sounds, unvoiced sounds, and stops, is an intrinsic characteristic of articulatory speech dynamics. The section of the upper ONPT may be modelled by the action of different muscles that increase or decrease its section and elongation; as an example, the nasal cavity can be connected or disconnected to the oral cavity by the velum, a set of structures controlled by the stylopharyngeal muscle, which may retract the velum to the back pharyngeal wall, blocking airflow to the nasal cavity. When connected to the oral cavity the nasal tract imprints a specific signature to the vocalization power spectrum, known as 'hypernasality', which is a gesture necessary to articulate nasal sounds as [m, n, ɲ, ŋ]. The oral cavity can be subject to important changes in section and longitude due to the action of several groups of muscles, namely, the jaw rising by the masseter, temporalis, medial pterygoid, lateral pterygoid, and digastric, which act together with gravity to lower the jaw. The tongue is a complex structure of vessels and muscles, including a set of extrinsic muscles, such as the styloglossus, and the intrinsic muscles, such as the inferior and superior longitudinalis, and verticalis (Sanguinetti, Laboissière, and Payan, 1997). The lip radiation point may be modelled by the orbicular muscles to produce oval or round lip terminations, buccination, and enlargement. All these groups are controlled by independent neuromotor pathways derived from hypoglossal (XII cranial nerve). These muscular structures are summarized in Figure 2.2.

2.1.4 The nervous system

The nervous system is classically divided into two parts: the central nervous system (CNS) and the peripheral nervous system (PNS). The CNS consists of the brain and spinal cord, while the PNS consists of all the nerves that branch out from the spine, ending at specific muscles, near or distal (see Figure 2.3).

The PNS is further divided into two parts: the somatic nervous system (SNS) and the autonomic nervous system (ANS). The SNS controls voluntary movements and reflexes, while the ANS controls involuntary functions such as heart rate, digestion, and breathing[4].

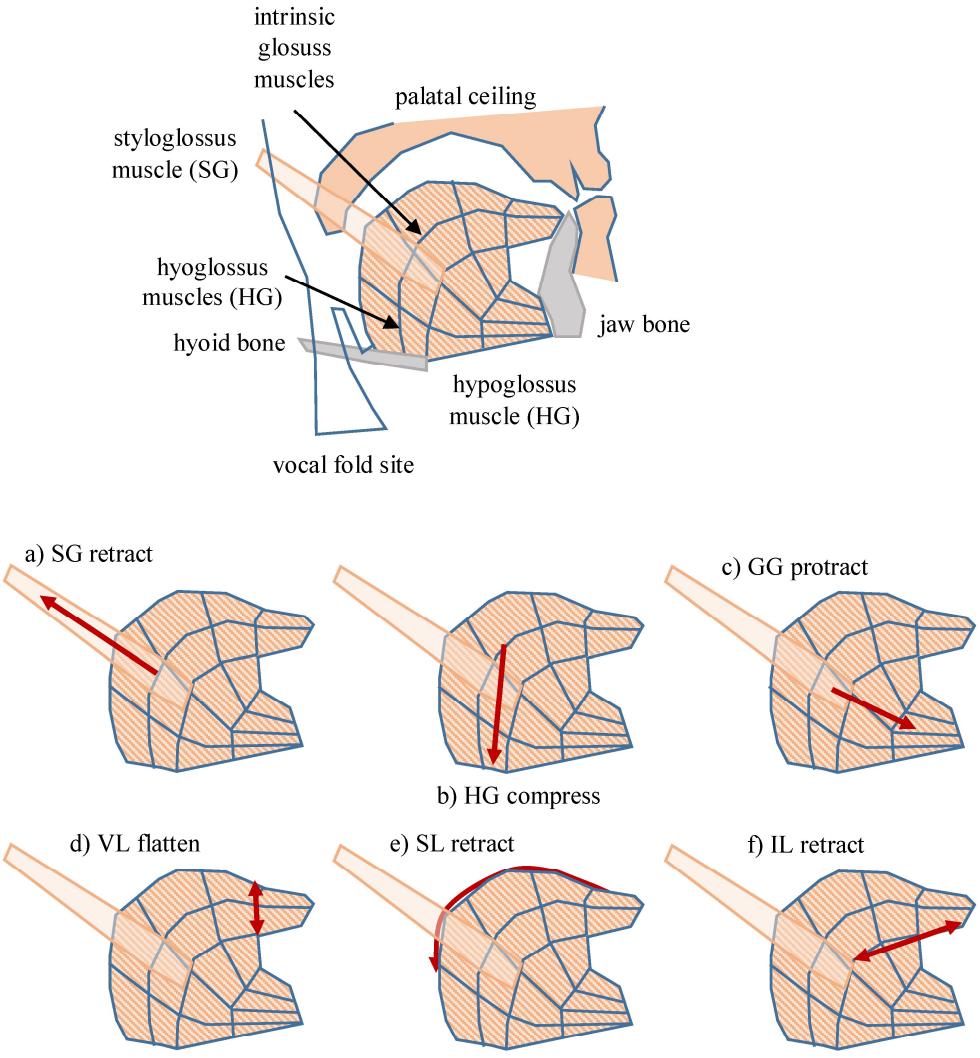

Figure 2.2 Synoptic representation of the extrinsic and intrinsic muscles involved in tongue movement and shaping.
Description: a) extrinsic styloglossus; b) extrinsic hyoglossus; c) extrinsic genioglossus; d) intrinsic transversal; e) intrinsic superior longitudinal; f) intrinsic inferior longitudinal. (adapted from Sanguinetti, Laboissière, and Payan, 1997)

The CNS is responsible for processing information received from the PNS and sending out instructions to the body muscles. It is also responsible for higher functions such as thought, memory, and emotion (Kandel et al., 2013).

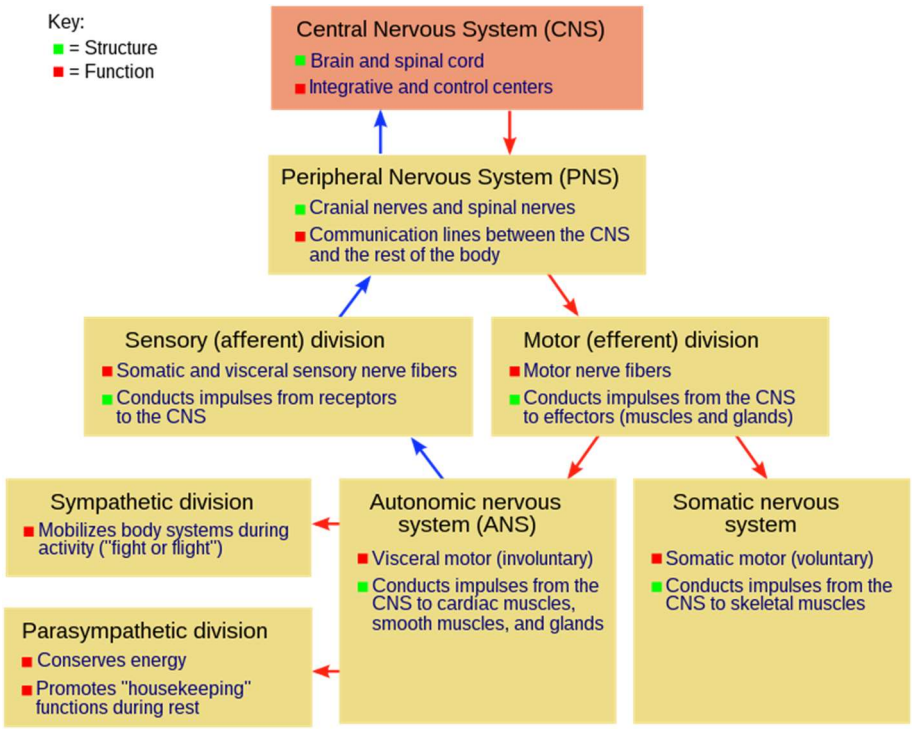

Figure 2.3 Relational and functional description of the nervous system and its parts[5]

2.1.5 Structure and function

The main structure of the CNS is the brain, split into two hemispheres, left, and right, strongly intercommunicated. The left and right hemispheres are responsible for controlling different tasks and behaviours (known as brain lateralization). The left hemisphere is dominant regarding language, logic, and math abilities.

[4] https://www.ncbi.nlm.nih.gov/books/NBK542179/
[5] Source: https://commons.wikimedia.org/wiki/File:NSdiagram.svg, Licence: Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2. Retrieved 2023/10/02.

The right hemisphere is associated with creative activities, being dominant in artistic and musical performing, and intuition.

Classically, the CNS is divided into the following seven parts (see Figure 2.4 and Table 2.1).

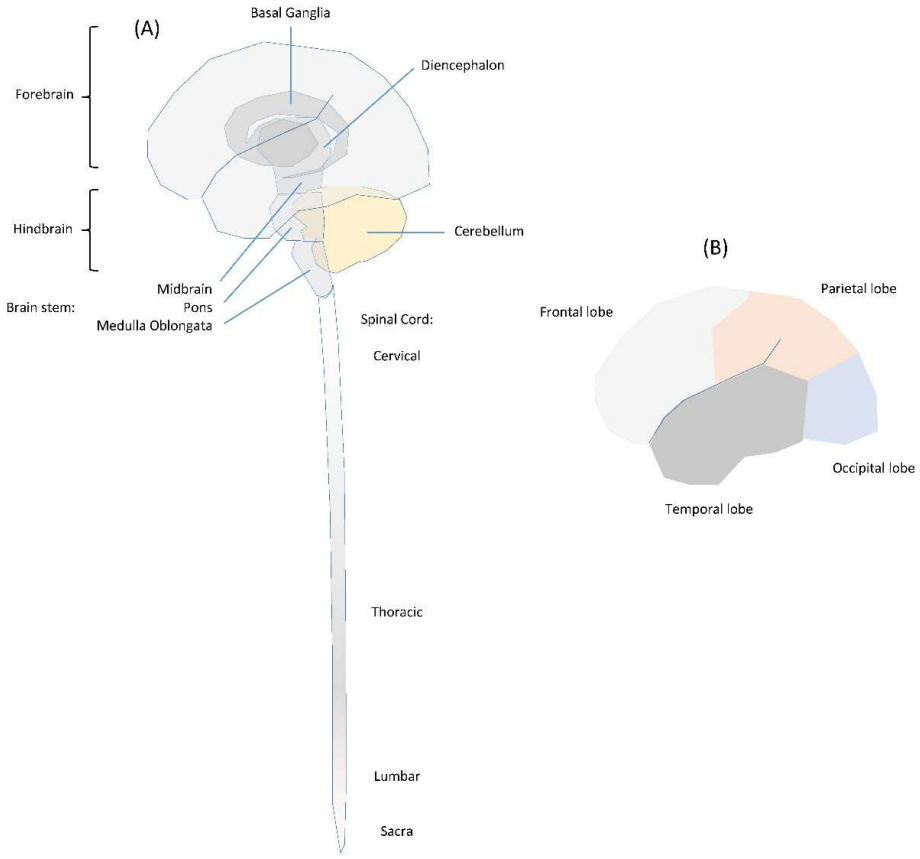

Figure 2.4 Synoptic representation of different CNS structures and cortical lobes.
Description: A: The CNS is divided into seven main parts; B) The four lobes of the cerebral cortex (adapted from Kandel et al., 2013)

Certain elements have been left out of this description, such as the limbic system and reticular formation, as the purpose of this section is to provide an overview of the main brain structures, an exhaustive anatomical description being beyond the scope of this work. As previously introduced, cerebral tissue is composed of neurons, each characterized by three principal components: a cell body (soma), an axon, and dendrites.

The axon functions as the conduit for transmitting electrical nerve impulses away from the soma, while dendrites serve as receptors for signals originating from neighbouring neurons. Critically, neurons possess the capacity to establish connections with a multitude of other neural cells through intercellular communication occurring within specialized synaptic gaps.

Table 2.1 CNS subdivisions.

Section	Description
Cerebral cortex	The brain's outer layer known for its folds (gyri) and grooves (sulci), which is visible upon removing the skull and membranes. This area, also called grey matter, contains neuronal cell bodies and other components, contrasting with the underlying white matter made up of myelinated axons. The cortex's convolutions increase its surface area, allowing for a higher neuron concentration. It's divided into four lobes—frontal, parietal, occipital, and temporal—each responsible for specific functions. The frontal lobe handles voluntary movements, cognitive functions, and language; it includes areas for speech production like Broca's area. The parietal lobe processes sensory information, the occipital lobe manages visual data, and the temporal lobe focuses on auditory information crucial for speech comprehension.
Basal Ganglia	The BG are found within the cerebral white matter, being composed of the caudate nucleus, putamen, and *globus pallidus*. These structures form the pallidum and the striatum. The BG control and coordinate muscle movements (Lanciego, Luquín, and Obeso, 2012).
Diencephalon	The interbrain, situated between the BG and the midbrain, includes the thalamus, hypothalamus, epithalamus, and subthalamus. The thalamus acts as a relay station for sensory and motor signals to and from the spinal cord, medulla oblongata, and cerebellum, and processes sensations like pain and temperature. The hypothalamus controls eye movements and auditory responses, maintains homeostasis by regulating internal balance against external changes, and oversees vital functions such as heart rate, blood pressure, hunger, thirst, body temperature, and hormone release, linking the nervous system with the endocrine system.
Midbrain	The midbrain being one of the smallest parts of the CNS, roughly 2 cm in length plays a vital role in movement planning and execution, excitation, motivation, and habituation. Dopamine is produced in a region referred to as *substantia nigra*, a structure darker than neighbouring areas due to the high levels of neuromelanin in dopaminergic neurons.
Pons	It is a part of the brainstem, connecting the medulla oblongata and the thalamus, responsible for relaying NMA from the motor cortex to the cerebellum, medulla, and thalamus, acting as a distribution hub.

Section	Description
Medulla oblongata	The medulla oblongata, located at the base of the brainstem, controls vital autonomic functions such as breathing, heart rate, blood pressure, and digestion. It uses chemoreceptors to monitor respiratory activity and adjust breathing rates based on blood acidity. Additionally, it serves as a control center for cardiovascular and vasomotor functions, managing reflexes like vomiting, swallowing, coughing, and sneezing.
Cerebellum	The cerebellum is responsible for coordinating smooth and voluntary movements. It consists of three lobes: anterior, posterior, and flocculonodular. The cerebellum is connected to other brain parts via cerebellar peduncles. The superior cerebellar peduncle links it to the midbrain, aiding in limb coordination. The inferior cerebellar peduncle connects to the medulla, involving proprioceptors for balance and posture. The middle cerebellar peduncle, a pathway from the pons, carries information about voluntary motor actions to the cerebellum. The cerebellum works with the cerebral cortex to interpret commands and send signals to the motor cortex, ensuring precise muscle contractions for coordinated movement.
Spinal cord	The cerebellum facilitates the coordination of smooth, voluntary movements and is divided into three lobes: anterior, posterior, and flocculonodular. It is interconnected with the brain via cerebellar peduncles. The superior cerebellar peduncle connects it to the midbrain for limb coordination, the inferior cerebellar peduncle links to the medulla for balance and posture, and the middle cerebellar peduncle conveys voluntary motor action information from the pons. The cerebellum collaborates with the cerebral cortex to process instructions and communicate with the motor cortex, ensuring precise and coordinated muscle movements.

The intricacy of neuronal interactions within the cerebral structures is marked by nonlinear behaviours, feedback loops, and multiple interconnections spanning various cerebral regions. Neurons can be broadly arranged into two primary categories: primary neurons, and secondary or motor neurons. A salient distinguishing feature between these neuronal types lies in the presence or absence of Schwann cells. Primary neurons, generally devoid of Schwann cells, are principally responsible for processing and transmitting signals within the central nervous system and are indispensable for cognitive and sensory processes. They are primarily located within the cortical regions, constituting the grey matter. Conversely, secondary or motor neurons are characterized by the encasement of their axons by Schwann cells.

These neurons are responsible for conveying signals from the central nervous system to muscles and glands, thus facilitating motor functions and physiological responses. Their primary location is situated post-thalamus, with somas clustered into nuclei. These aggregations serve as the origination points for major nerves governing specific body regions. These nerves exit the brainstem at various junctures, collectively forming the principal neural conduits of the body. Figure 2.5 gives an anatomic representation of these structures. Table 2.2 presents the nomenclature to address each nerve branch.

Sensory information travels from the PNS to the spinal cord before reaching the brain. This information ascends upwards using first, second, and third-order neurons. First-order neurons receive impulses from skin and proprioceptors and send them to the spinal cord. They connect with second-order neurons in the dorsal horn and send impulses to the thalamus and cerebellum. Third-order neurons relay sensory neural activity to the somatosensory portion of the cortex. Somatosensory sensations inform about pressure, pain, temperature, and other body senses. Motor information travels from the CNS to lower motor neurons. These efferent neurons instantiate muscle movement. The relationship between movement and neuron activation is not a straightforward matter, as there exists inhibitory circuits that instead of activating a muscle, do precisely the opposite and impede its activation.

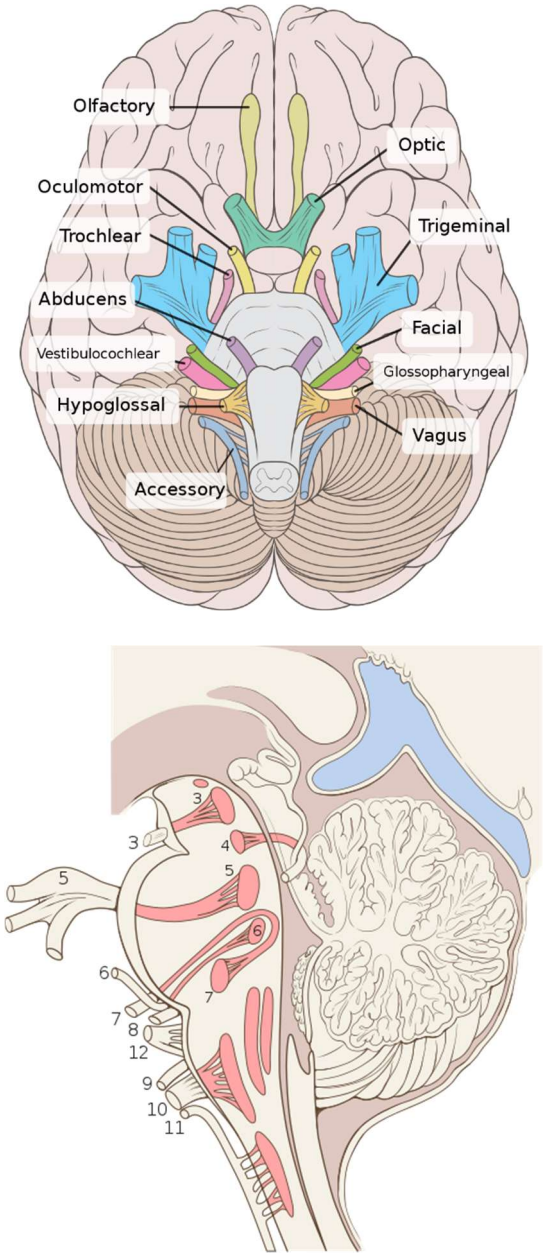

Figure 2.5 Synoptic representation of the different nerves (left)[6] and locations of the STNs (right)[7].

[6] Source: https://commons.wikimedia.org/wiki/File:Brain_human_normal_inferior_view_with_labels_en.svg, Patrick J. Lynch, medical illustrator, Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2. Retrieved: 2023/10/02.

[7] Source: https://commons.wikimedia.org/wiki/File:Brain_stem_sagittal.svg, (Authorizing free share and remix). No changes made with respect to the original file licensed under the Creative Commons Attribution-Share Alike 3.0 Unported. Retrieved 2023/10/02.

Table 2.2 Pathways of cranial nerves leaving the skull.

Location	Nerve	Numeration
cribriform plate	Terminal nerve	0
cribriform plate	Olfactory nerve	I
optic foramen	Optic nerve	II
superior orbital fissure	Oculomotor	III
	Trochlear	IV
	Abducens	VI
foramen rotundum	Trigeminal (ophthalmic)	V1
	Trigeminal (maxillary)	V2
foramen ovale	Trigeminal (mandibular)	V3
stylomastoid foramen	Facial nerve	VII
internal auditory canal	Vestibulocochlear	VIII
jugular foramen	Glossopharyngeal	IX
	Vagus	X
	Accessory	XI
hypoglossal canal	Hypoglossal	XII

Movement tends to be managed by the activation of agonist-antagonist muscles; to produce fine-tuned and accurate motion muscle activation has to be precisely managed and timed, through the activation or inhibition of the respective muscles. This process is linked to fine-tuning movement and precision control that is managed by structures in the thalamus and the BG; these structures gather information from the proprioceptive areas of the cortex and modulate a response, this matter is discussed later.

2.2 Main nervous system dysfunctions

The intricate and delicate structures composing the nervous system are fragile and work under specific mechanical and immunological protections. The skull and spinal bone are rigid protections against mechanical stress and contusions.

The brain-blood barrier is designed to keep the liable neuron structures protected from chemical and microbiological noxious agents. Despite the multiple barriers and protections, many different factors may alter the normal functionality of specific brain areas, leading to subsequent alterations that may be externally observed. The nature of these dysfunctions is varied and is due to many factors, as summarized in Table 2.3.

Table 2.3 Most frequent nervous system disorders.

Disorder	Description
Alzheimer's disease (AD)	It is the most common ND, a type of dementia in which brain cells and neural connections begin to degenerate and die. This condition results in loss of memory, cognitive and physical decline, and difficulty in swallowing. AD is progressive, with conditions worsening over time. Clinical inspection found aggregations of beta-amyloid plaques and neurofibrillary tangles made of tau protein within the neurons in AD patients. It seems that these plaques and tangles induce the death of brain cells and are formed as a consequence of the misfolding of proteins within the neuronal network.
Amyotrophic lateral sclerosis (ALS)	Also known as Lou Gehrig's disease, affects motor neurons that control voluntary and involuntary movements. The ultimate cause of ALS is not known, and unfortunately, there is no cure. Scientists believe that cell death is related to the excess amount of extracellular glutamate in ALS patients. Riluzole, which can disrupt the formation of glutamate, is used to slow down the progression and reduce the painful symptoms of ALS.
Broca's aphasia	Also known as expressive aphasia, it is caused by a stroke, brain tumour, or brain trauma that causes irreversible damage to Broca's area. People affected by Broca's aphasia experience difficulty in producing speech. They can comprehend speech and know what they want to express, although they are unable to produce the required movements to communicate, producing a severe disconnect between thought and language capabilities.
Cerebrovascular accidents (CVA)	Cerebrovascular accidents, also known as strokes, occur when a section of the brain is not able to receive enough oxygenated blood, leading to hypoxia that causes tissues in the brain to die, with the consequent loss of brain function. Generally, strokes are caused by blood clots travelling from elsewhere in the body to a cerebral vessel in the brain clogging it. The consequences of the stroke are dependent on the area of the brain deprived of blood flow. Some patients may experience left-sided paralysis, while others might produce slurred speech. When facing a CVA, time is crucial. Some of the interventions might try to break down the clot with adequate medication or attempt surgical removal. The severity of symptoms directly correlates to how long the tissue's oxygen supply has been cut off.
Huntington disease	It is a hereditary, progressive brain disorder that is caused by a mutation in the huntingtin gene, HTT. This causes the huntingtin protein to accumulate in the brain cells, which eventually leads to cellular death. Initially, Huntington's disease causes chorea, involuntary jerking, and hand-flapping movements. As the disease progresses, cognitive decline occurs. Fatality takes place approximately within 15 years of diagnosis.

Disorder	Description
Multiple sclerosis	Multiple sclerosis is an autoimmune disease, in which the body's immune system attacks the myelin proteins of the central nervous system, disabling the functions of myelinated neurons which are responsible for the communication between the CNS and the PNS. MS has a high prevalence in young adults and shows up as pain, weakness, and vision and coordination loss. Medication is used to depress the body's immune system and can help control the adverse effects of this disease.
Parkinson's disease (PD)	It is a nervous system disorder that results in the deterioration of dopamine-releasing neurons in the midbrain (*substantia nigra pars compacta*). As a consequence of the drop in dopamine levels, tremors, unsteady movements, and loss of balance are some of the most prevalent symptoms that severely condition the living standards of the persons affected. PD is a progressive disease, many patients experience bradykinesia, stiffness, and a mask-freezing face as symptoms progress. While no definite cure exists for the disease, the severity of the symptoms can be controlled to a certain extent by medication. Levodopa is a dopamine precursor drug that can be metabolized into dopamine for CNS use. Deep brain stimulation can provide a degree of improvement, although it does not prevent the progression of the disease.
Poliomyelitis	It is a consequence of an inflammation of the spinal cord due to the poliovirus. It devastates the neurons in the ventral horn of the spinal cord leading to paralysis. The infection of the poliovirus is preventable through vaccination.
Spinal cord traumas	Symptoms of spinal cord injuries are dependent on where the injury takes place. Sensation can be damaged when injury affects the sensory tracts. However, if the ventral roots or ventral horns are damaged, paralysis occurs. There are two causes for motor paralysis; flaccid, when nerve impulses do not reach the intended muscles; without stimulation, the muscles are unable to contract and become spastic, where the motor neurons undergo irregular stimulation, causing involuntary contraction depending on where the injury has taken place. Paraplegia, paralysis of the lower limbs, occurs when the spinal cord is interrupted between T1 and L1. Quadriplegia, paralysis of all limbs, is a result of an injury in the cervical region, and hemiplegia where there is a paralysis of half of the body, as a result of unilateral interruption.
Traumatic brain injury (TBI)	It may happen when there is a severe blow or jolt to the head or body; a foreign object that goes through brain tissue can also cause traumatic brain injury. This inevitably leads to the disruption of normal brain activity. TBI symptoms can vary depending on the severity of the injury. A concussion can cause temporary dizziness or loss of consciousness, while a contusion causes lasting neurological damage. Contusions to the brain stem may cause a coma. TBI can cause subdural or subarachnoid haemorrhage and cerebral oedema. When the brain tissues suffer stress or trauma, the blood vessels in the brain might break, causing blood to pool, increasing intracranial pressure, and compressing the brain tissue. In the most severe cases, the brain might be contracted onto the spinal cord damaging it, leading to hampered autonomic nervous system functions, or even functional loss.
Wernicke's aphasia	It occurs most commonly as a result of a haemorrhagic or ischemic stroke affecting the left middle cerebral artery blood supply capability. In Wernicke's aphasia, a person can speak clearly and produce speech. However, their speech holds no meaning. Additionally, they also experience difficulty with understanding language.

2.3 Neuromotor functional description of speech production

The intricate process of speech production is a complex interplay of neural and biomechanical structures and organs, a detailed exploration of which would necessitate an entire dedicated chapter. In this section, a concise, high-level functional overview of the processes and structures involved in activating the necessary muscles for speech generation is offered. While this content may seem out of place in a technical thesis, it is vital for understanding subsequent sections, because certain elements within the speech signal have direct connections to specific muscular activations, and therefore are consequently influenced by specific regions of the brain, nerves, and other anatomical structures.

The neuromotor elements responsible for controlling speech production involve a complex network of brain regions and neural connections that work together to generate and coordinate the movements required for speaking. This process allows for the transformation of thoughts, feelings, and emotions into spoken language (cognitive-linguistic processing), and the respective muscles located in the vocal tract are activated, including the tongue, lips, jaw, and vocal folds (neuromuscular execution), through the control of neuromuscular execution the selection, sequencing, and regulation of sensorimotor tasks, at appropriate co-articulated times, durations, and intensities (motor speech planning, programming, and control), accordingly with Duffy (2013). The primary neuromotor pathway involved in speech production is known as the corticobulbar tract. This tract originates in the Primary Motor Cortex (PMC), located in the frontal lobe of the brain, specifically in the region known as the precentral gyrus.

The precentral gyrus contains specialized areas called the motor speech cortex or Broca's area, which is crucial for the planning and execution of speech motions. From the motor cortex, the corticobulbar tract sends descending fibres that pass through the internal capsule, a dense bundle of nerve fibres located deep within the brain. These fibres then project to the brainstem, specifically to the cranial nerve nuclei, which mark the beginning of the nerves directly involved in speech production. In the brainstem, the corticobulbar fibres synapse with the motor neurons of the cranial nerve nuclei that innervate the muscles of the vocal tract. The cranial nerves most relevant in speech production include the hypoglossal nerve (cranial nerve XII) for the tongue muscles, the facial nerve (cranial nerve VII) for the muscles of the lips, and the trigeminal nerve (cranial nerve V) for the jaw muscles (refer to Figure 2.5 for an anatomic description). In addition to the corticobulbar tract, there are other neural pathways involved in speech production. These include the corticospinal tract, which controls voluntary movements of the body, including the respiratory muscles needed for breathing during speech. The corticospinal tract originates from the same motor cortex region as the corticobulbar tract but descends further down the spinal cord to innervate muscles throughout the body. Moreover, there are extensive connections between the motor speech cortex and other brain regions involved in language processing, such as the arcuate fasciculus that connects Broca’s area (motor) with Wernike’s area (auditory processing and language comprehension). These connections integrate language processing and motor planning, allowing for responsive, fluent, and coordinated speech production. Overall, the neuromotor pathways controlling speech production involve a complex interplay between the motor cortex, brainstem nuclei, and cranial nerves. These pathways ensure the precise and coordinated movements of the vocal tract muscles necessary for producing intelligible speech. Duffy (2013) in chapter 2 page 76 describes the complexity of these interactions as follows:

"The motor system, of which the speech motor system is a part, contains the complex network of structures and pathways that organize, control, and execute movement. It resides at all levels of the nervous system and mediates many activities of striated and visceral muscles. An appreciation of its organization and basic operating principles is necessary to understand normal speech production and MSDs. The motor system can be subdivided in many ways. Unfortunately, categorizing the components of a complex, integrated, and incompletely understood system inevitably results in some ambiguity, overlap, and confusion. Nonetheless, it would be impossible to develop an understanding of the speech motor system without parsing it in some way. The motor system can be organized purely by anatomy or according to its functions. Because functional labels contribute to an understanding of what the components do." A functional description of the complex speech production system starts with the instantiation of speech from many different cognitive areas in the brain, including visual, conceptual, and long-short memory mapping, which structures them into emotional expressions at a semantic level. These structures are transformed into semantic items including object and concept names, actions, and names of actions, which are syntactically built into ordered sequences of sounds. These activate Broca's area, and premotor, motor, and supplementary motor areas from gesture mapping retrieved from the hippocampus to the fine motor control, including the thalamic centres, the BG, and the cerebellum, to express direct NMA on the extrapyramidal pathways. The phrenic nerve activates inspiration and expiration movements on the diaphragm and intercostal muscles to press airflow through the larynx, inducing the vibration of the vocal folds when emitting voiced sounds, or the turbulence in constricted passages (vocal folds, false vocal folds, velopharynx, tongue dorsum, teeth, lips) when emitting voiceless sounds.

The ONPT tract is shaped, by mandibular muscles rising and lowering the jaw, lip rounding, opening and closing by orofacial muscles, extending and retracting the velum by velopharyngeal muscles, as well as shaping the tongue by extrinsic and intrinsic muscles, conditioning the pharyngeal and oral cavities by jaw, tongue, and palate gestures.

The sound waves, either organized as a harmonic spectrum (voiced sounds) or as an unorganized power distribution (voiceless sounds) propagate through the ONPT, experiencing reflections as the tract presents changes in its transversal section following wave propagation, the reflections being stronger where the section changes are more abrupt, the reflection function being proportional to the derivative of the transversal section logarithm, as it will be exposed in subsection 3.3. The forward and backward propagation of sound waves is explained by a complex convolutional model, determining the presence of standing waves for certain frequencies in the tube, following a resonant accumulation of energy. These frequencies are known as formants, because their presence in the energy spectrum, either as harmonic lines, or as power bands, are especially detected and encoded in the auditory system when speech is perceived (Greenberg et al., 2004; Gómez-Vilda, Ferrández-Vicente, and Rodellar-Biarge, 2013). Changes in the configuration of the ONPT produced by articulation gestures (tongue, mandible, and lip configurations), will result in changes in the shape and distribution of formants (peak position and bandwidth), therefore, the information conveyed by formants might be exploited to infer the position and movement of articulation organs, using articulation kinematic projection models (see subsection 4.2).

2.4 Neuromotor description of the speech production system

As introduced in Section 2.1 of this same chapter, speech production has three distinct components, respiration, phonation, and articulation. Although both are governed by NMA, the neural pathways that innervate each of the required muscles are different. Each process plays a completely different role in the production of the final speech signal. The following subsections provide a description of the neuromotor control processes and the different muscles involved in each task.

2.4.1 Phonation neuromotor control

The anatomical regions involved in regulating phonation can be summarized as follows, drawing from the works of Jürgens (2002), Brown et al. (2009), Rektorová et al. (2012), and Duffy (2013): The Primary Cortex Vocalization Area (PCVA), plays a central role in directly activating the laryngeal muscles through phonatory motoneurons, constituting the direct activation pathway. Additionally, a separate neuromotor control feedback pathway, also modulated by the PCVA, traverses through the pontine grey, cerebellum, and ventrolateral thalamus, ultimately providing feedback to the PCVA. Furthermore, the PCVA extends another control pathway through the putamen, subsequently progressing to the *substantia nigra*, medullar reticular formation, nucleus ambiguous, and solitary tract nucleus, ultimately reaching phonatory motoneurons, thus forming the indirect pathway. Other circuits provide motor control gating functions, through the anterior cingulate cortex and the periaqueductal grey, to the medullary reticular formation, and similarly as in the indirect pathway, to the nucleus ambiguous, solitary tract nucleus, and phonation motoneurons. Auditory feedback is provided by acoustically perceived sensation from the superior temporal cortex (Wernicke's area) to the lateral prefrontal cortex, supplementary motor area, and PCVA.

Summarizing the direct activation pathways, the PVCA includes the laryngeal motor cortex (Brown et al., 2009), premotor cortex, supplementary motor area, and cerebellar lobule VI. Secondary areas comprehend the cingulate motor area, the ventral nuclei of the thalamus, the putamen, the frontal operculum, and the anterior insula (Dietrich et al., 2020).

The functional relationship among them is still a matter of ongoing research, although there is consensus that as far as vocalization is concerned, the direct pathway is mainly controlled by the Ventromedial Central Sulcus Peak (VmCSP) corresponding to Broadmann area 4p, and the Dorsolateral Peak (DsP) in area 6, both on the precentral gyrus next to articulator control areas (Dietrich et al., 2020). Specifically, Rödel et al. (2004) reported selective stimulation of the vocal fold tensor (cricothyroid muscle: CTM) or relaxer (thyroarytenoid muscle: TAM) using transcranial magnetic stimulation. In this sense, Brown et al. (2009) suggested that both areas mentioned before may have functional roles representing different muscles in the Laryngeal Biomechanical System (LBMS) because both muscular systems are innervated respectively by the external Superior Laryngeal Nerve (SLN), and the Recurrent Laryngeal Nerve (RLN). Therefore, a differential NMA (dNA) may be defined as the separated agonist-antagonist activity of neuromotor pathways controlling phonation, which may be indirectly estimated from the kinematics of specific trajectories of the effector being controlled. In the case under study, the agonist-antagonist direct NMA of the *musculus vocalis* is known to be controlled by the vocal fold tensor (cricothyroid) and the relaxer (thyroarytenoid). The working assumption in the present study will consider that the dNA projected to the lower neuromotor units through the direct activation pathway of the LBMS is mainly related to VmCSP and DsP activity. The majority of the research conducted in this project has primarily focused on sustained vowel phonation, specifically the open vowel [a:].

Consequently, the involvement of secondary areas related to the indirect activation pathway becomes more pertinent in dynamic speech tasks, but these aspects are not within the scope of this study. The structures mentioned above, along with their respective roles in phonation control, are summarized as follows in Figure 2.6.

The cricothyroid joint, schematically represented in Figure 2.6 allows the external elongation of the vocal fold by the stretching and activation of the CTM compatible with an increase of tension (Hammer, et al. 2010). The external branch of the SLN is the smaller of the two superior laryngeal nerve branches. It descends to the region of the superior pole of the thyroid and travels medially along the inferior constrictor muscle. The external branch of the SLN innervates the CTM, which is the only tensor of the TAM.

The recurrent laryngeal nerve innervates the TAM, which combined with the arytenoid muscles (lateral and oblique) contributes to adduction, abduction, and relaxation of the vocal folds. The direct pathway from the laryngeal cortex (VmCSP and DsP) to CTM and TAM is the primary neuromotor pathway responsible for the activation of the cricothyroid joint, and the tension and relaxation of the *musculus vocalis*.

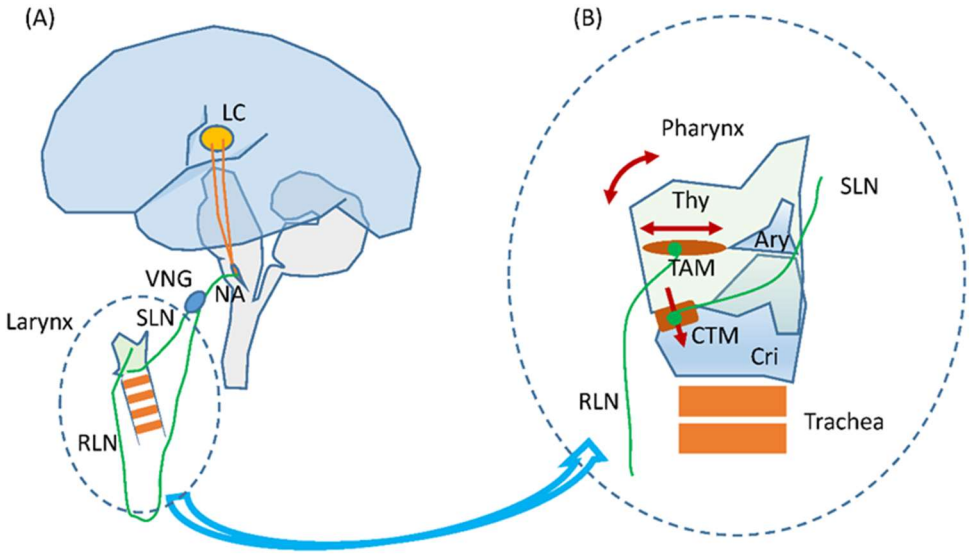

Figure 2.6 Synoptic representation of the cricothyroid joint and associated neuromuscular structures and function.

Description: Structures and function assuming that the cricoarytenoid system is in adduction. (A) Cortico-bulbar-laryngeal direct neuromotor pathways. LC: laryngeal cortex. NA: nucleus ambiguus in medulla oblongata. VNG: extracranial vagus nerve ganglia. SLN: superior laryngeal nerve. RLN: retro-laryngeal nerve. (B) Detailed view of the thyroarytenoid joint: Thy: thyroid cartilage. Ary: arytenoid cartilages. Cri: cricoid cartilage. TAM: thyroarytenoid muscle. CTM: cricothyroid muscle. The role that CTM and TAM will play in the cricothyroid joint is summarized by the swinging represented by the curved double-arrow line, resulting in the stretching, and shortening of the vocal folds during adduction (double-arrow straight line on top of TAM).

2.4.2 Articulation neuromotor control

The neuromotor system controlling articulation is responsible for coordinating the precise movements of the muscles involved in shaping the ONPT to produce speech sounds, including at least the oropharyngeal, mandibular, lingual, and orofacial subsystems.

The primary neuromotor pathway responsible for speech articulation is the corticobulbar tract. This pathway originates in the primary motor cortex, specifically in the region known as the precentral gyrus, which contains specialized areas such as the primary motor cortex and the motor speech cortex (Broca's area), involved in planning and executing the movements associated with speech articulation.

The corticobulbar fibres travel from the cortex through the internal capsule, a dense bundle of nerve fibres located deep within the brain, then project to the brainstem, specifically to the cranial nerve nuclei that innervate the muscles responsible for speech articulation. The corticobulbar fibres synapse in the brainstem with the motor neurons of the cranial nerve nuclei, particularly the hypoglossal nucleus (cranial nerve XII), facial nucleus (cranial nerve VII), and trigeminal nucleus (cranial nerve V), please refer to Figure 2.1, Figure 2.5, and Table 2.2 Pathways of cranial nerves leaving the skull. for an anatomical description. These cranial nerves control the muscles of the tongue, lips, jaw, and face which are essential for precise speech articulation. Additionally, there are other neural pathways involved in speech articulation. The cerebellum plays a crucial role in fine-tuning motor movements and coordinating the timing and accuracy of speech articulatory gestures. The cerebellum receives inputs from the motor cortex and integrates sensory information to modulate and adjust motor commands during speech production. Besides, the BG contribute to the control of speech articulation by influencing movement initiation, coordination, and timing. These structures interact with the motor cortex and help regulate the selection and execution of appropriate speech-motor actions.

According to its importance regarding one of the studies included in the present thesis, a detailed description of the masseter activity in controlling the mandible will be discussed in what follows. The masseter control complex is synoptically represented in Figure 2.7.

In the case of interest for the mentioned study, the motor end plates of the FCPs innervate the masseter fibres producing muscle contraction. The NMA exciting masseter is induced from direct and indirect pathways.

The fine control of a muscle movement requires a certain degree of feedback. This is provided by sensory pathways (in green) consisting of neurons activated by spindles (terminal sensors detecting fibre stretching) attached to the muscles, providing proprioceptive sensing to the LMNs in two ways.

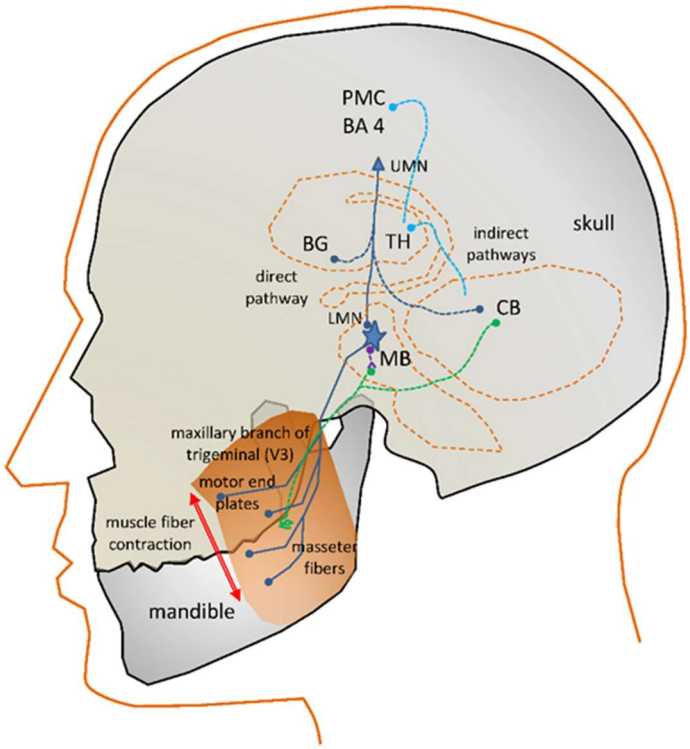

Figure 2.7 Neurophysiological description of the masseter articulation motor system.
Description: PMC: primary motor cortex; BA4: Brodmann area 4; UMN: upper motor neuron; BG: basal ganglia; TH: thalamus; CB: cerebellum; LMN: lower motor neuron; MB: midbrain.

A direct feedback loop is provided by inhibitory interneurons. A more complex feedback loop connects sensory units with the BG and the cerebellum CB. These structures serve feedback information to the motor and frontal cortices, as well as to the Lower Motor Neurons (LNMs) (blue lines). The BG control circuit assists the PMC in accurate and fine-motor speech programming. The CB control circuit coordinates PMC motor planning from proprioceptive information.

2.5 The BG neuromotor feedback circuitry

The BG are located deep in the brain hemispheres, integrated by the striatum (*caudate nucleus* and *putamen*) and the *lentiform nucleus* (*putamen* and *globus pallidus*). The *substantia nigra* and the subthalamic nuclei in the midbrain involved in the indirect activation pathways are nearby structures strongly connected to the BG. The cortical, thalamic, and *substantia nigra* provide inputs to the *striatum*, as well as the premotor cortex in the frontal lobe. The *striatum* provides inputs to the *substantia nigra* and *globus pallidus*, and the *globus pallidus* projects to the thalamus, subthalamic nucleus, red nucleus, and reticular structure in the brainstem. These connections compose different feedback circuits. The main output of the BG is found in the *globus pallidus*. These circuits regulate muscle tone, postural adjustments, goal-directed activities, grading force, amplitude, and duration of movements, and adjust movements to speech production constrictions, supporting learning, preparation, and instantiation of movements. The influence of the BG circuits on speech production is through its direct inhibitory action on cortical motor areas, damping, and modulating NMA from the cortex which could be excessive to produce fine and precise movements. HD would be a consequence of imprecise or excessive control effect, leading to dampened movements. The main cause of this disorder seems to be associated with dopamine imbalance, although its deficit does not completely explain speech deficits (Duffy, 2013). The deterioration of dopamine-supplying neurons in *substantia nigra* results in a reduction of this neurotransmitter, and the BG circuit functionality is distorted.

A more detailed description of the circuitry involving the BG is summarized in the representation shown in Figure 2.8 (Schulz et al., 2005; Blitzer et al., 2009; and Brittain and Brown, 2014).

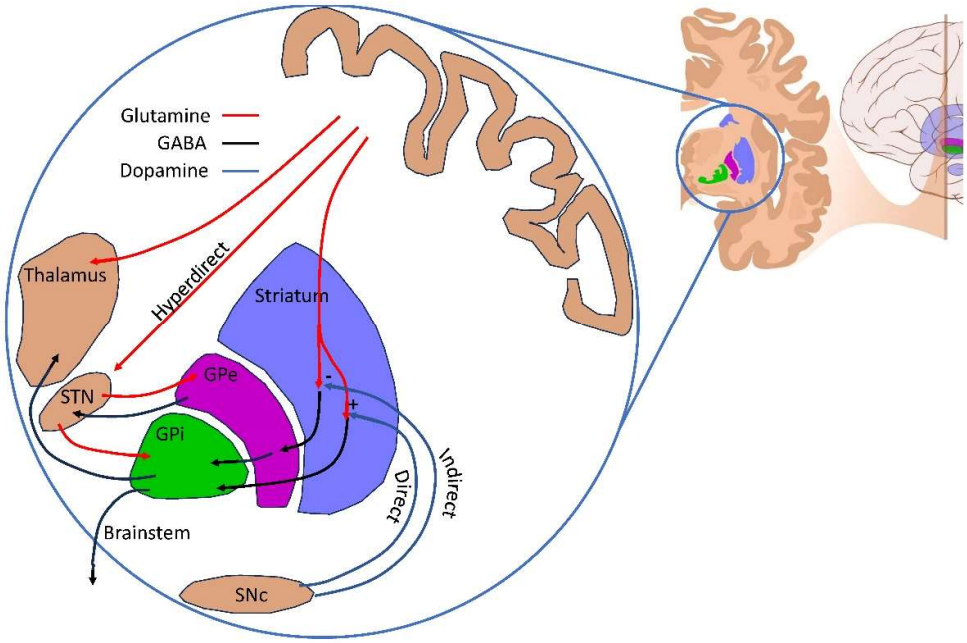

Figure 2.8 Summarized representation of main circuital relationships between the BG and the cortex. Description: GPe: *globus pallidus pars externa*; GPi: *globus pallidus pars interna*; STN: subthalamic nucleus; SNc: *substantia nigra pars compacta*[8] (adapted from Brittain and Brown, 2014).

According to Brittain and Brown (2014), motor projections from cortical areas take place at the striatum, with separate circuits projecting through *globus pallidus* and *substantia nigra* to thalamic and reticular brainstem nuclei. The dominant synaptic inputs to the striatum are the corticostriatal pathways, where cortical afferents innervate the output cells of the striatum (medium spiny neurons labelled indirect and direct). From the *substantia nigra pars compacta* dopaminergic ascending inputs innervate the middle spiny neurons, which can be divided into those which express dopamine D1-class receptors and project to the *globus pallidus pars interna* (GPi), and those which express dopamine D2-class receptors and project to the *globus pallidus pars externa* (GPe).

[8]https://commons.wikimedia.org/wiki/File:Basal_Ganglia_Anterior_Unlabeled.jpg. Licensed under the Creative Commons Attribution-Share Alike 3.0 Unported. Retrieved 2023/10/02.

Dopamine release in the striatum excites the direct pathway, while simultaneously inhibiting the indirect pathway. These pathways form the basis of the classic direct and indirect model of BG function. The excitatory subthalamic nucleus (STN) of the BG network also receives a direct cortico-subthalamic projection (the hyperdirect pathway).

2.6 Speech neuromotor disorders

2.6.1 Phonation disorders

The anomalous behaviour of vocal fold vibration patterns resulting from pathological origin produces important changes in the glottal source pattern, allowing the clinician to trace possible etiological hypotheses during voice quality assessment based on acoustical analysis. The multiple causes inducing anomalous vocal fold vibration may be associated with the following generic groups:

1) Organic origin. The causative circumstances might be due to malformations, or specific lesions affecting the vocal folds, the supporting cartilages, or the muscles activating the adduction, contact, and abduction of the vocal folds, either of endogenous or iatrogenic origin (Dworkin and Meleka, 1997). Some of the different lesions may include nodules, polyps, cysts, oedema, carcinomas, or chronic laryngitis, a sort of permanent inflammation of the vocal folds and surrounding tissues. Iatrogenic lesions may include sulcus, most of the time because of compromised vocal fold surgery. All these lesions modify the pattern of the glottal source in one or another way.

2) Functional origin. When the vocal folds and their supporting and activating structures do not present any kind of lesion or apparent physiological alteration, although their function shows clear anomalous behaviour, most of the time due to voice abuse or inadequate phonation gestures (postural, muscular) hampering a proper phonation practice, producing a clear dysphonic voice, which could be corrected by adequate speech therapy.

3) Neurological origin, when because of lesions or disorders in the CNS or PNS the NMA of the vocal folds the phonation shows a clear disordered behaviour. These anomalies might be produced by lesions at the cortical centres controlling phonation, at the BG and other midbrain structures, or the extrapyramidal neural pathways innervating the larynx. Sometimes, surgery compromising laryngeal nerves might result in partial or complete vocal fold paralysis, as in the radical treatment of thyroplastias.

All these causes might produce different alterations in normal phonation, due to strong modifications to vocal fold biomechanics and associated kinematics. A brief explanation might be derived from the synoptic description given in Figure 2.9.

The most semantic patterns, explaining the pathological behaviour of the vocal fold vibration are produced by glottal gaps, and unexpected defects in the adduction, contact, and abduction of the vocal folds. It is well known that the presence of glottal gaps has a strong influence on the quality of phonation (Chen et al., 2011). These gaps might appear because of organic laryngeal lesions, but also because of disordered neuromotor control of the larynx complex neuromuscular system, according to the synoptic description given in Figure 2.9.

Points a), b), and c), show the usual vocal fold configurations during normal functionality. In a) both vocal folds are represented in the inspiration-expiration position, allowing airflow from and to the lungs under the action of the diaphragm and intercostal muscles.

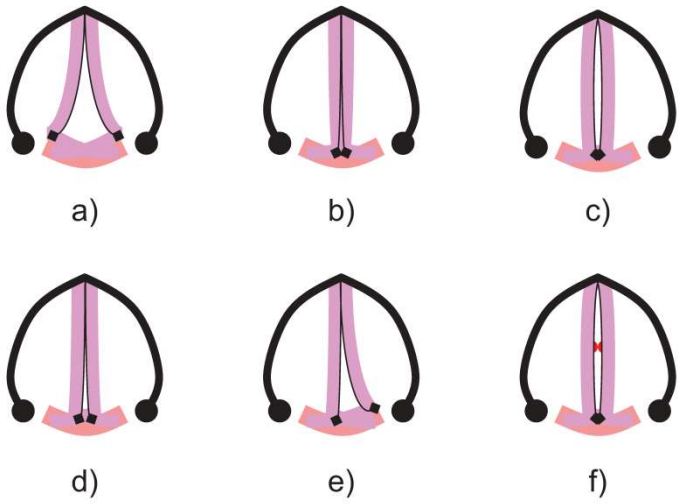

Figure 2.9 Influence of the vocal fold normal and altered function regarding phonation.
Synoptic description: Influence of the vocal fold normal and altered function regarding phonation (synoptic description): a) normal configuration during inspiration and expiration (respiratory function); b: normal configuration during vocal fold contact, either static (blocking position) or during the closed phase; c) normal configuration during the maximum aperture during the open phase; d) permanent contact gap at the arytenoid end of the vocal folds; e) right vocal fold paresis; f hourglass-type of contact defect due to some medial lesion on one or both vocal folds.

In b) the vocal folds are shown during the contact phase when the airflow to the lungs and mouth is completely stopped. This configuration may be an involuntary reflex to prevent the intake of extraneous matter either in solid or liquid form, from the supraglottal cavity to the trachea and bronchi, or because of voluntary phonation activity, where the vocal folds are closed before being forced open by the subglottal pressure before the initiation of the opening phase. In c) both vocal folds are shown during the maximum open phase when the subglottal pressure has forced their separation (abduction) during the open-close cycle. At this point, as the glottal flow is near its maximum, the difference between the subglottal and the supraglottal pressure will be almost null, and the elastic forces acting on the vocal folds will initiate their approximation to a new contact phase (adduction).

Points d), e), and f) show different configurations of anomalous glottal function. In d) there is not complete contact between both vocal folds during any phase of the glottal cycle, either because of an anatomical alteration or as a result of muscle asthenia (affecting either the *musculus vocalis* or the transversal and oblique laryngeal muscles). In this case, the pressure build-up and decay will be less acute, and the MFDR will be less strong, affecting the harmonic display of the glottal source in the frequency domain. A higher volume of air will be lost during the phonation cycle with a lower phonation efficiency, and a need for more frequent air intake will be necessary, accompanied by more respiratory fatigue. This case is also represented in the left picture of Figure 2.10.

Figure 2.10 Videoendoscopic images of the vocal folds taken with stroboscopic light.
Description: Contact phase between vocal folds showing contact defects. Left: bilateral nodules, including a contralateral hematoma on the left vocal fold. Centre: Reinke's Edema in the right vocal fold. Right: polyp on the right vocal fold.

In e) both vocal folds show different dynamic activity, the right-hand side one presenting some difficulty completing the adduction phase, as a result of a lesion affecting the arytenoid cartilages or associated muscles, or the laryngeal nerves responsible for neuromotor activation of these muscles. Both vocal folds will show an asymmetric vibration, appearing as differences between neighbour glottal cycles (short-term jitter and shimmer). The MFDR will worsen as well. The harmonic display will show less slender and lower harmonics, and inter-harmonics will also be present.

This will be one of the consequences shown by the defect in the central picture of Figure 2.10 (hourglass shape). In f) it may be seen that both vocal folds cannot come to complete contact on the anterior and posterior sides (epiglottic and arytenoid) due to a lesion on the medial part of the vocal folds. As a result, the MFDR will worsen, and there will be evident signs of incomplete contact during the closed phase. In this case, when the speaker tries to compensate for the deficient contact, overwork with sporadic openings will appear, and a turbulent airflow will be present. This situation corresponds to the right-hand side picture of Figure 2.10.

The different gap defects can be classified according to the glottal cycle phase they are affecting, as contact defects when they affect the complete closure of the glottal flow, or as permanent when there is no complete contact at any time during the glottal cycle. They will produce the following effects on the phonation quality:

1) Rough phonation, when there is a clear non-periodical repetition pattern in the glottal cycle, due to asymmetric vocal fold vibration. Distortion features, such as jitter (irregular cycle duration), or shimmer (irregular cycle amplitude) will appear because of a different supraglottal pressure configuration in each glottal cycle.

2) Airy or breathy phonation due to airflow escape generates fast gas jets through the contact defects, which will produce strong turbulence on impinging over slower gas masses within the supraglottal cavity. An increment on the non-harmonic part of the frequency spectrum of the glottal source will become evident.

3) Asthenic phonation shows a reduced harmonic display due to permanent contact defects, allowing the continuous airflow escape associated with a small phonation efficiency. The MFDR will also worsen.

4) Strained phonation, when the speaker tries to compensate for some of these defects raises the tension of the vocal folds, producing a worsening of lesions, and generating irregular airflow escapes.

Summarising, the signature of phonation alterations in the glottal source may be divided into two main groups:

- Glottal gap defects due to abnormal adduction, abduction, and contact, which might appear during a specific interval of the phonation time or be permanent. These defects might be the result of laryngeal lesions, functional improper voice user, and less frequently of neuromotor origin.
- Asymmetrical vibration of both vocal folds, mainly due to unilateral laryngeal lesions, iatrogenic causes secondary to surgery or intensive care intubation or NMDs.

In any case, one of the main problems posed by phonation alterations is that their aetiology is uncertain most of the time, their laryngeal, functional, or neuromotor origin being an important confounding factor, especially in ageing voice, which produces many of the observable correlates present in NMDs (Orozco-Arroyave et al., 2015). Interestingly, their relevance in the characterization of PD was assessed in early observations by Perez et al. (1996).

2.6.2 Hypokinetic dysarthria

After introducing the necessary neurologic background, the following section describes how the different neuromotor disorders will affect the speech production apparatus. Motor speech disorders are the result of specific problems affecting some of the described direct or indirect pathways of activation, or the muscle fibres.

These disorders are commonly referred to as dysarthrias, in the particular case of PD HD related to the pathological behaviour of the complex BG control circuit. This phenomenon manifests all through the components of the phonation system that are responsible for the neuromechanical control of speech: respiration, phonation, and articulation. The term "hypokinetic" is employed to describe movements characterized by weakness, limited range, and rigidity, which in turn can convey a sense of speech that is monotonous, lacking inflexion, and devoid of expression. The control circuit of the BG exerts inhibitory influence over the premotor cortex (PMC) regions to regulate cortical activity, citing Duffy, 2013 (Chapter 7, page 344): *"The primary influence of the basal ganglia control circuit on speech is through its connections with motor areas of the cerebral cortex. Its influence on the cortex appears inhibitory; that is, it damps or modulates cortical output that would otherwise be in excess of that required to accomplish movement goals. In hypokinetic dysarthria, this damping effect is excessive."*. An excessive level of inhibition within this circuit can lead to HD, as it will inevitably produce delays in activation or excessive blockage leading to inoperability of the agonist-antagonist dynamic. Most motor problems related to the BG motor circuit have an origin or relate to some inoperability of neurotransmitters. Specifically, dopamine deficits due to the progressive death of dopaminergic neurons in the *substantia nigra pars compacta* (located in the MB). This is the main element that leads to PD neuromotor symptoms: "The substantia nigra is the origin of the nigrostriatal pathway, which travels to various structures within the basal ganglia… The dopamine deficiency in this nigrostriatal pathway and the basal ganglia account for most of the typical features of PD. Once the brain is no longer able to compensate for this dopamine loss, some effects can occur. Typical symptoms include muscle rigidity, akinesia, bradykinesia, and tremor…" (Goberman and Coelho, 2002).

More specifically; “The essential neuropathological changes in PD are a loss of melanin-containing dopaminergic neurons in the *substantia nigra pars compacta*… This results in a dysfunction of the basal ganglia circuitry, which is an integral part of cortico-basal ganglia-cortical loops that mediate motoric and cognitive functions” (Harel et al., 2004b).

The consequence of the neuro-pathophysiological abnormalities that affect the control or execution of movements might lead to weakness, spasticity, incoordination, involuntary movement, and excessive, reduced, or unstable muscle tone. Specifically, dysarthria has a neurologic origin, and it is a disorder of movement, according to Duffy (2013). This author excludes the scope of dysarthria to disregard various other neurological speech disorders, including apraxia of speech, stuttering, palilalia, echolalia, mutism, foreign accent syndrome, and aprosodia. Additionally, cognitive-linguistic disorders such as aphasia and akinetic mutism, along with musculoskeletal defects stemming from an injury, disease, congenital conditions, ageing, or insufficient personal care (such as tooth loss), which encompass dysphonias linked to head and neck lesions, vocal misuse, or hormonal imbalances, excluded from the purview of dysarthria. Furthermore, psychogenic disorders like schizophrenia and depression, in conjunction with age-related alterations in pitch, voice quality, stability, loudness, breathing, fluency, and prosodic variations, are considered potential comorbidity confounders that warrant consideration but are not encompassed within the concept of dysarthria.

Motor speech disorders have been classically studied in many ways, which can be grouped into perceptual and instrumental headings. The perceptual methods rely on auditory perception, being the reference standard for clinical differential diagnosis, severity assessment, disorder handling and treatment, and longitudinal evolution assessment. Visual and tactile inspection at rest, during non-speech activity, and speech utterance are the components of motor speech inspection.

The instrumental analysis is less widely used in clinical evaluation, due to insufficient normative methods, clinicians limited experience with acoustical instrumentation and lack of enough explanatory evidence to support the interpretability of the results. The instrumental methods are classified as acoustical, physiological, and visual imaging based;

- The acoustical methods estimate the frequency, intensity, and temporal evolution of the speech signal, and are related to the auditory-perceptual assessment because they are used as a validation support. Machine learning methods have been used to reinforce the task of dysarthric assessment, although they cannot be granted the ultimate word because of the explainability, interpretability, and causability problems common to automatic computer-aided methods. Nevertheless, they help in the quantification, description, and understanding of motor speech disorders, providing quantitative, confirmatory, and accurate support to perceptual judgments, especially assessing speech rate, voice breathiness, tremor, fluency freezing and blockage events, pitch and loudness estimation, hyper- and hypo-nasality assessment, articulation precision estimation, diadochokinetic test evaluation, etc. The capacity of acoustic analysis to visualize speech is an important help in estimating stability, monitoring speech competence deterioration, and visual feedback during rehabilitation and speech therapy.

- The physiological methods are conceived to assess the functionality of the structures generating speech. They are designed to evaluate muscle contraction, movement of speech structures, the relationship between the musculoskeletal speech levels, the temporal parameters relating central and peripheral neural and biomechanical functions, and the relationship between CNS structures during planning, programming, and executing speech tasks. Physiological analysis methods are based on functional magnetic resonance (fMRI), positron emission tomography (PET), single photon emission computerized tomography (SPECT), electroencephalography (EEG), transcranial magnetic stimulation (TMS), magnetoencephalography (MEG), electromyography (EMG), etc.
- The visual imaging methods are based on videofluoroscopy, nasoendoscopy, laryngoscopy, and videostroboscopy. They are used in assessing tasks related to speech production, such as swallowing, velopharyngeal competence, or laryngeal functionality assessment, among others.

Following Duffy (2013), meaningful concepts related to the characterization of motor speech disorders are age at onset, course, site of lesion, neurologic diagnosis, pathophysiology, speech components involved, severity, and perceptual characteristics. The major types of speech motor dysarthrias are flaccid (lower motor neuron), spastic (bilateral upper motor neuron), ataxic (cerebellar), hypokinetic (BG control circuits), hyperkinetic (BG control circuits), unilateral upper motor neuron, and mixed of some of the precedent ones.

2.6.3 PD HD

HD is a speech motor disorder stemming from deterioration or damage of the BG leading to loss of motor control. It manifests primarily at the respiratory, phonatory, and articulatory levels, exerting a notable influence on aspects such as phonation, articulation, and prosody. The nomenclature "Hypokinetic Dysarthria" is attributed to the condition's hallmark characteristics of diminished movement range and loss of muscular force. These attributes are distinctive hallmarks of BG pathology, of which Parkinson's disease (PD) stands as a prototypical example, albeit not the sole condition displaying such features. PD may present motor speech signs related and unrelated to speech production. According to Duffy (2013), the list of speech and speech-related characteristics of HD would be grouped under the categories of perceptual (phonatory and respiratory, such as reduced loudness and utterance length; articulatory, such as phoneme repetition, palilalia, rapid, blurred, or galloping alternate motion rate, and prosodic, such as reduced stress, mono-pitch, mono loudness, unexpected stops and pauses, short speech bursts, variable rate, over-accelerated rate in segments and in a whole utterance), physical (frozen facial mask, tremor in jaw, lips, and tongue, reduced range in alternate motion, head tremor) and proprioceptive (patient complaining on difficulty in rising loudness, controlling speech rate, mumbling, stuttering, difficulty in starting speech, stiff lip sensation).

The proper identification of which acoustic and physiological findings are to be considered in distinguishing HD from other types of dysarthrias and ageing speech is a critical aspect in the characterization of PD HD. Respiratory issues are a common occurrence in individuals with HD, and in some cases, they can lead to fatal outcomes. These issues arise from several factors, including reduced chest movement, diminished muscle strength, irregularities in breathing patterns, and a decrease in vital capacity.

Of particular note is the relationship between these factors and their impact on speech loudness and prosody, which is a significant observation in this context. This relationship can be effectively assessed through various means, such as the estimation of vowel duration and prolongation, measurement of airflow, determination of the number of syllables between sequential inspirations, and the evaluation of increased respiratory rate during spontaneous speech and reading, among other methods (Hlavnička et al., 2017).

Phonatory issues in individuals with HD exhibit distinct characteristics that are perceptually notable. In males, a prominent feature is an elevation in the fundamental frequency, while in some cases for women, a reduction in fundamental frequency is observed, with severity increasing as the disease progresses. Additionally, vocal intensity is reduced across syllables, during vowel prolongations, and in alternate phonation gestures. There is also difficulty in adjusting loudness appropriately based on the listener's distance, and loudness decay becomes more pronounced when performing concurrent visual or manual tasks. Variability in phonation fundamental frequency and loudness is compromised during various tasks, such as spontaneous speech, reading, word and sentence imitation, and emotional expression, resulting in reduced distinctiveness.

Voice tremor within the range of 4-7 Hz is another significant feature, although it is not universally present, potentially being more frequent in later stages of PD. Flutter tremor within the range of 9-11 Hz has been reported specifically in HD and flaccid dysarthria, serving as a distinguishing characteristic from other dysarthrias. Maximum phonation time is sensitive to changes within the same individual as PD progresses, potentially serving as a metric for quantifying disease evolution, although it may not be useful for initial detection of the disorder.

Perturbation features like jitter, shimmer, and cepstral peak prominence (CPP) are sensitive but not exclusive to PD, and they are associated with reduced short-term neuromotor control of vocal fold abduction and adduction, as well as contact defects leading to turbulent airflow perceived as breathiness. Deficient motor control of phonation can result in slow vowel initiation and termination, along with coordination issues between voicing and articulation gestures. Increased glottal gaps, observed in most patients, have been associated with breathiness, incomplete adduction, and asymmetric vocal fold vibration due to laryngeal muscle rigidity and atrophy, although some of these observations may also be linked to ageing.

In terms of articulation, problems in HD may relate to velopharyngeal dysfunction, where velar movements may be reduced or irregularly executed due to inadequate coordination between antagonist muscle pairs, resulting in hypernasality perception. Imprecise articulation place, and reduced movement range or speed (indicating deficient kinematic competence), could also stem from articulatory muscle rigidity and limited motion range. Uncoordinated jaw opening and closing may occur due to hypokinesia and rigidity, resulting in reduced formant dynamic extension, especially affecting the second formant, and restricted vowel space with centralization, which are associated with HD.

CHAPTER 3

3 Speech production acoustic models

This chapter, presents a series of models that represent the phonation and articulation aspects of speech production, considering the speech signal as the outcome of a sequential filtering of an input signal composed of a series of pulses, ending with the way the stationary wave inside the vocal tract is radiated to the environment.

3.1 Speech production model

The speech production model, upon which this work relies, was formulated by Gunnar Fant (1960, 1981). Fant's theory represents the production of the speech signal as the interplay between distinct subsystems, the GS is generated as the source of the system, it propagates through the ONPT and it emanates from the lips. In our specific context, this ONPT is represented under the assumptions delineated in subsection 3.2. Figure 3.1 depicts Fant's model, which comprehensively incorporates the two primary biological aspects of speech production discussed in the preceding chapter: phonation and articulation. This investigation is specifically focused on examining both phonation and articulation as two distinct parts of speech production. It is noteworthy that in the context of sustained vowel production, the role of radiation is relatively minimal, in contrast to scenarios such as singing or acting, where performers are required to project their voices to suit the requirements of their performance.

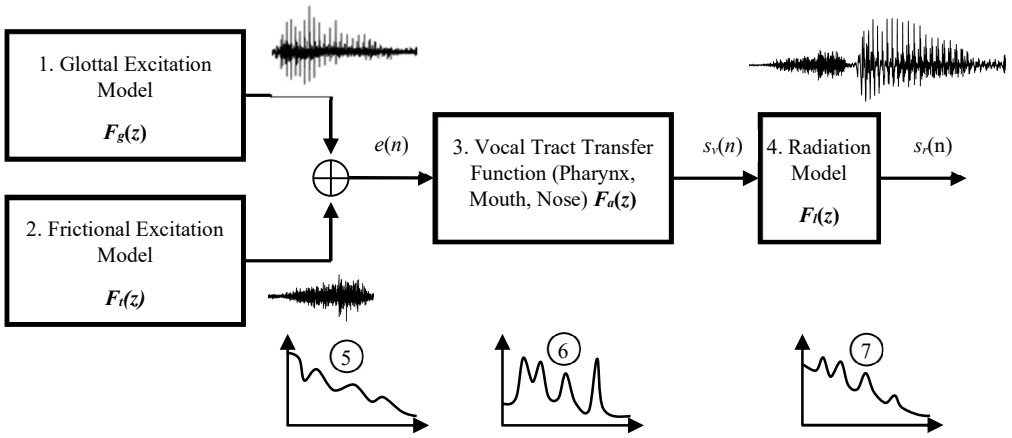

Figure 3.1 Fant's model for the production of speech.
Description: 1) Glottal source excitation model for voiced speech sounds (Liljencrants-Fant); 2) Turbulent excitation model for unvoiced speech sounds; 3) ONPT transfer function $F_a(z)$ equivalent filter (upper respiratory ways); 4) Radiation equivalent model filter; 5) Power spectrum of the excitation signal e(n); 6) ONPT equivalent filter transfer function in the discrete frequency space; 7) Power spectrum of the speech acoustic signal $s_r(n)$ radiated from the lips (assuming mouth radiation). $F_g(z)$: Spectrum of the glottal excitation signal in the discrete frequency space; $F_t(z)$: Id. of the turbulent excitation signal; $s_v(n)$: vocal signal before lips; $s_r(n)$ acoustic signal radiated; $F_l(z)$: Radiation model equivalent transfer function in the discrete frequency space.

The working model in the present study, as given in Deller, Proakis, and Hansen (1993, pg. 340), assumes that the GS signal is generated by the glottal excitation model, a train of deltas $\delta(n)$ being the input to the system; this series of pulses are then filtered by the glottal source filter and transformed into the GS pulses $s_g(n)$, first block of Figure 3.2. The series of glottal pulses travels through the ONPT tract, which acts as a filter (represented by the second block of Figure 3.2). Depending on the position, aperture, and elongation of the structures that compose the ONPT a series of resonant frequencies arise or diminish to build up the vocal signal $s_v(n)$. Assuming than the radiation point of the system is the lips, the acoustic signal, when projected outside the lips is modified by the lip radiation model, producing the microphonic speech signal $s_r(n)$, this being the signal that can be recorded and processed.

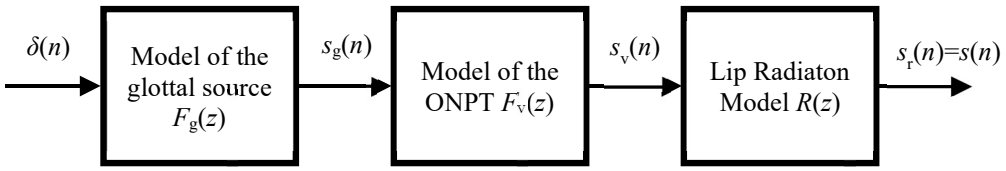

Figure 3.2 Voice production model.
Description: The model assumes that the glottal source is produced by a regular train of δ pulses, which are filtered by the glottal source generating model $F_g(z)$ resulting in a train of glottal source pulses $s_g(n)$. The transfer function $F_v(z)$ of the ONPT will enhance (formants) or attenuate (anti-resonances) the harmonic contents of the glottal source to produce the vocalized signal $s_v(n)$, which will be recorded as a radiated signal $s_r(n)$, which is the speech signal recorded by a microphone $s(n)$.

This abstraction model introduces the series of conceptual connections and changes that the signal undergoes, but to properly analyse the acoustic signal the methodology requires a model with direct physical interpretation, this conceptualization being introduced in the following subsections.

3.2 Phonation model

The subsystems introduced in the previous section are modelling different physiological structures (lungs, bronchi, trachea, larynx, pharynx, nasopharynx, and oral and nasal cavities) with variable acoustical properties, under neuromotor control of specific nerves and brain areas, such as those represented in Figure 2.1. Concerning phonation, Rothenberg's model, as represented in Figure 3.3, transforms the different parts involved in the production of phonation into a series of electromechanical elements that represent the physical properties of the subsystems involved in speech production. These different components aim to represent the oscillatory, inertial, impedance, capacitance, pressure build-up, flow, and propagation phenomena involved in the process of generating phonation. Since this approach aims to solely model phonation, the effects of the ONPT or the radiation considerations are simplified to give an overview of the production of the GS signal.

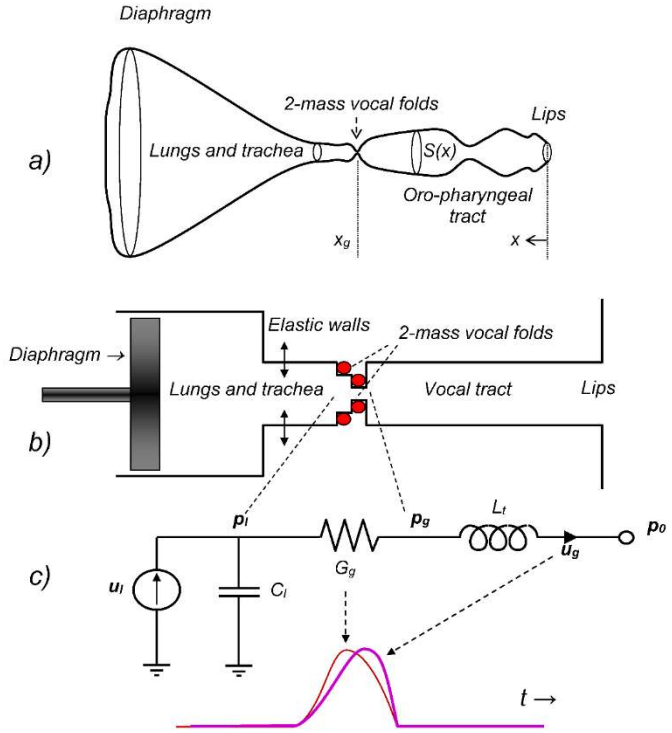

Figure 3.3 Rothenberg's model of phonation.
Description: a) physiological representation; b) acoustical simplified model; c) electromechanical equivalent model. In b) the acoustic pathways of the ONPT are idealised and simplified as a constant section tube, the pressure exerted by the diaphragm forces a flow of air through the glottis (x_g) putting the vocal folds (represented as two moving lumped masses) into vibration. The two masses represented in each vocal fold may move transversal to the x-axis with relative independence, reproducing the mucosal wave movement. When put in contact they may stop the airflow. In c) the action of the diaphragm on lungs is represented as a flow generator u_l, injecting a constant flow, which in part is stored in the yielding elastic walls (represented as a capacitor C_l) when there is a glottal stop, and injected through the glottis as u_g. The action of the moving masses in the glottis is represented by a variable conductance (G_g) oscillating between a minimum and a maximum value following the specific vocal fold vibration pattern. The action of the tubular vocal tract is represented by an inertial inductance L_t, ignoring the yielding properties of its elastic walls. The pressure at the lips is the atmospheric quiescent pressure p_0, dynamically considered null. As a result of the glottal flow injection, a pressure build-up is generated in the subglottal side (p_l), and as a result of a variable flow through L_t a pressure build-up is generated also in the supraglottal side (p_g). The dynamic component of this pressure $p_s=p_g-p_0$, concerning the quiescent atmospheric pressure p_0 can be considered a good representation of the glottal source.

Rothenberg's model is based on the following two main assumptions:

- The ONPT is modelled as the minimal acoustic tube that least interferes with the outward propagation of the GS, showing the most uniform transversal area, to be associated with the production of the schwa [ə] sound that corresponds to the minimal constriction to airflow. The distances along this rectified axis will be denoted by x, with origin x=0 at the lips, moving inwards towards the glottis (x=x_g).
- The ONPT is considered a single, uninterrupted cylindrical tube. This assumption assumes a total obstruction of the nasal tract by the palatal vellum, a structure responsible for regulating airflow into the nasal cavity. Within this framework, the vocal tract is regarded as a continuous, undivided cavity without any branching. Consequently, it is assumed that the vellum is closed, allowing air pressure waves to propagate through the vocal cavity. This condition will exclude the modelling of nasal consonants [m, n, ŋ] and nasalised vowels (symbolically [ʊ̃], those uttered with the nasopharyngeal passage open).

Under these conditions, according to Rothenberg's Model (Rothenberg, 1973, Koc and Ciloglu, 2016), the three following subsystems are responsible for phonation:

- The respiratory cavity (diaphragm, lungs, bronchi, and trachea), stores and produces a pressure build-up p_l on the subglottal side of the vocal folds during the closed phase of the glottal cycle.

- The glottal subsystem, represented by a vocal fold electromechanical conductance G_g. The pressure difference between the subglottal $p_l=p(x_l,t)$ and supraglottal $p_g=p(x_g,t)$ sides of the vocal folds injects an airflow $u_g=u(x_g,t)$ through the glottis, known as the glottal flow (GF):

$$u(x_g,t) = G_g(p(x_l,t) - p(x_g,t)) \tag{3.1}$$

- The ONPT, which in Rothenberg's Model is represented by a single tube of uniform section as an electromechanical inertial parameter L_t, subject to the contour conditions at $x=0$ (open space) and $x=x_g$ (glottis) where the following relationships apply:

$$p_s(t) = p(x_g,t) - p(0,t) = L_T \frac{\partial u(x_g,t)}{\partial t} \tag{3.2}$$

An extended description of the details above exposed can be found in Álvarez-Marquina et al. (2020). The vibrations of the vocal folds produce pressure changes in the supraglottal side of the vocal folds. The difference between the supraglottal side pressure $p(x_g,t)$ and the quiescent pressure at that point (p_0,t) is known as the GS, given in (3.2). Therefore, the GS is proportional to the derivative of the glottal flow. The GS may be estimated from the speech signal using model inversion by adaptive linear prediction (Deller, Proakis, and Hansen, 1993; Fu and Murphy, 2006; Gómez-Vilda et al., 2009; Drugman, Bozkurt, and Dutoit, 2012; Alku et al., 2019), as it is described in subsection 4.1.6.

In what follows, the capability of Rothenberg's model to reproduce some of the most relevant features of the glottal flow and source when different glottal apertures are assumed is shown from Figure 3.5 to Figure 3.8. These effects are obtained assuming a glottal aperture with a generic aspect as the one shown in Figure 3.4.

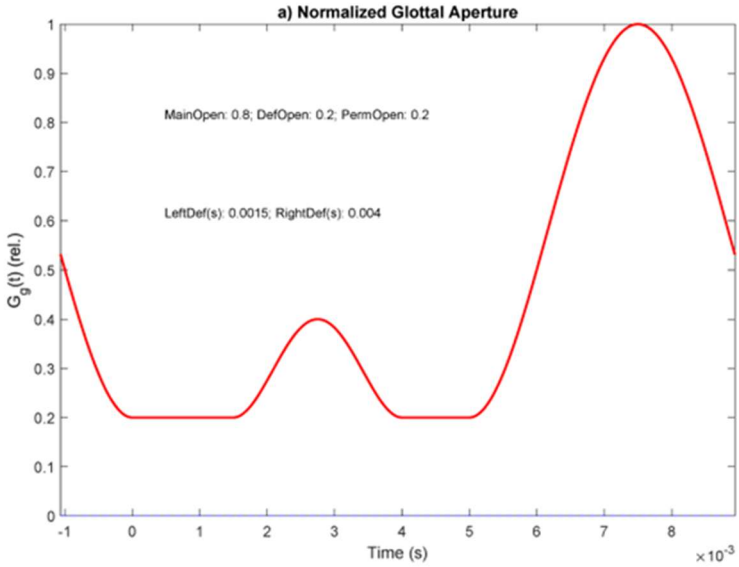

Figure 3.4 Glottal aperture function.
Description: Glottal aperture function $G_g(t)$ assumes permanent and temporary contact gap defects. A permanent gap of 0.2 relative units concerning the maximum aperture is present over the whole cycle. An anticipated (improper) contact defect opening superimposed on the permanent gap is seen between 1.5 and 4 ms with a maximum value of 0.2 (rel. units), followed by a proper opening between 5 and 9 ms with a maximum value of 0.8 (rel. units).

If the vocal folds show any irregularity in the obstruction of airflow they may produce defects in phonation, these irregularities can be of the following two kinds: either due to imperfect aperture, closure, and contact, or due to asymmetric vocal fold kinematics. Imperfect closure produces irregular flow injections to the supraglottal cavity, which are associated with irregular glottal source pulses. This situation is represented in the glottal flow profile shown in Figure 3.4, where an irregular glottal flow can be appreciated as a permanent air escape (permanent contact gap) at 0.2 times the maximum glottal flow injection, and by an irregular air release between 1.5 and 4 ms. due to a spontaneous contact defect. The combination of these two cases (deficient spontaneous contact, and permanent contact gap) will produce most of the expected irregularities in the glottal source and flow patterns, as shown in Figure 3.5 to Figure 3.8.

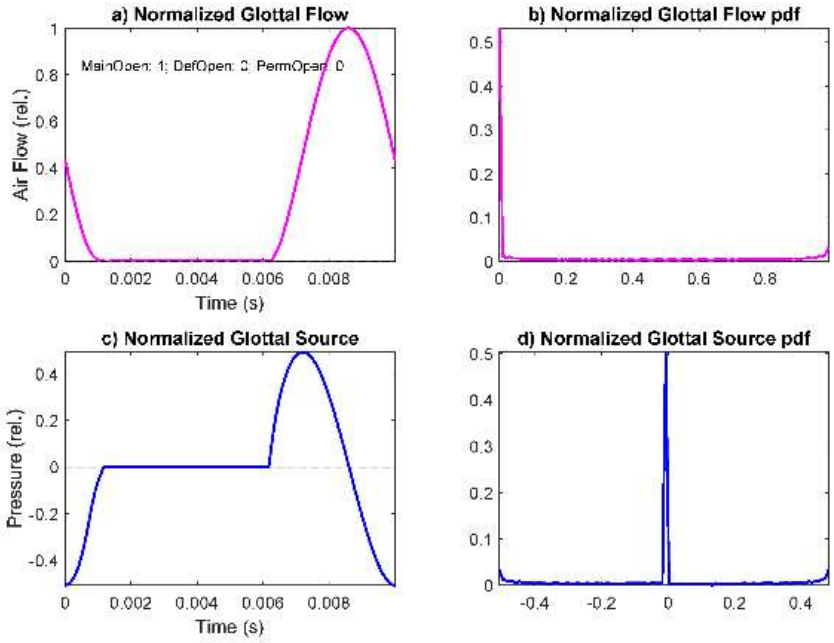

Figure 3.5 Glottal source and flow patterns and their amplitude distributions corresponding to an ideal behaviour with maximum relative amplitude, null contact defect, and null permanent gap.
Description: a) glottal flow; b) amplitude distribution of the glottal flow; c) glottal source; d) amplitude distribution of the glottal source. The glottal source reproduces de ideal L-F profile.

Figure 3.5.a shows an idealized situation where the glottal flow follows a perfect opening-closing pattern (neither permanent escape nor spontaneous opening). The associated normalized glottal source in Figure 3.5.b shows pressure rise and drop profiles aligned with the maximum flow declination rate of the glottal flow. After the pressure drop produced by vocal fold adduction, recovery is observed towards a null dynamic pressure extending till the instant when the vocal folds start a new opening cycle. After a maximum pressure build-up aligned with the maximum flow slope, a decay in the glottal source is observed till the next maximum flow declination rate instant.

Figure 3.6 shows the case of a permanent gap defect producing a constant opening, it may be seen that the glottal flow remains uninterrupted during the whole phonation cycle. Following a decline to a minimum resulting from the adduction process, a consistent increase in flow persists until the point at which the actual opening commences.

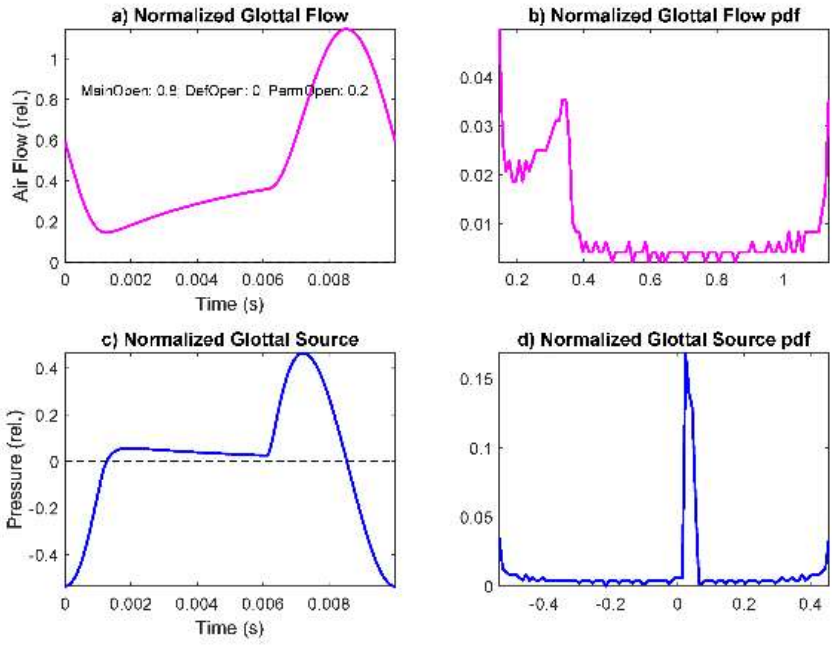

Figure 3.6 Glottal source and flow patterns and their amplitude distributions corresponding to a relative glottal aperture of 0.9, null contact defect, and a permanent gap of 0.2.
Glottal source and flow patterns and their amplitude distributions corresponding to a relative glottal aperture of 0.9, null contact defect, and a permanent gap of 0.2: a) glottal flow; b) amplitude distribution of the glottal flow; c) glottal source; d) amplitude distribution of the glottal source. The glottal flow does not stop at any time, and it follows an almost straight ascending line during the contact phase.

Figure 3.7 illustrates a scenario where a spontaneous opening occurs during the contact phase, leading to an extemporaneous release of airflow in the middle of the glottal cycle. The improper opening takes place between 2 and 5 ms, and as a result, the normalized glottal source profile presents a replica in a small scale of the main L-F pattern.

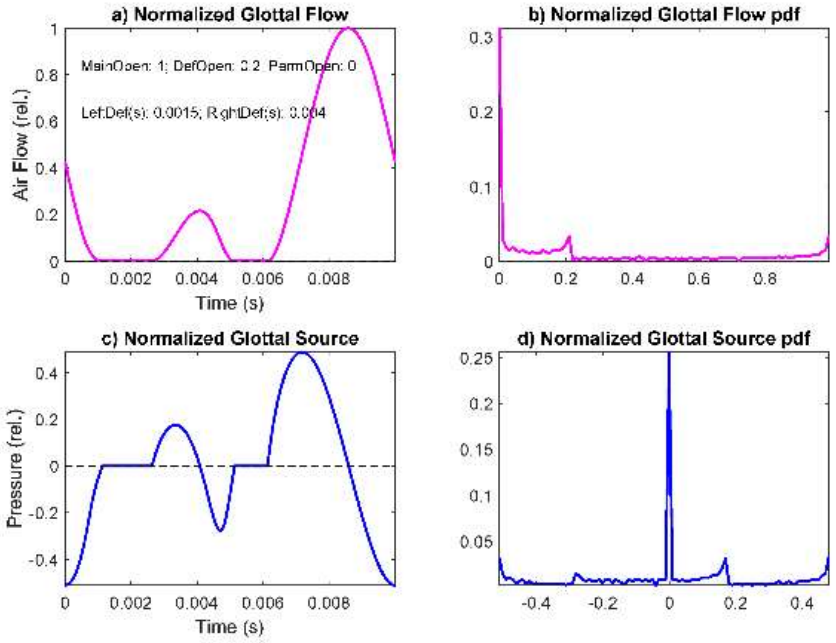

Figure 3.7 Glottal source and flow patterns and their amplitude distributions corresponding to a relative glottal aperture of 1, a contact defect of 0.2, and a null permanent gap.
Description: a) glottal flow; b) amplitude distribution of the glottal flow; c) glottal source; d) amplitude distribution of the glottal source. The glottal flow decays twice, and the contact defect produces a small-scale wavelet of the main L-F pattern on the glottal source.

Finally, Figure 3.8 shows a case that is a combination of a permanent and a spontaneous defect that affects the opening and closing cycles, as the one presented in Figure 3.4. It may be observed that the effects on the glottal source are a combination of the results produced by each perturbation separately, such as a constant flow increment with the insertion of the improper flow escape, a wavelet superimposed on the general L-F pattern, and a jump of the quiescent pressure above the null value.

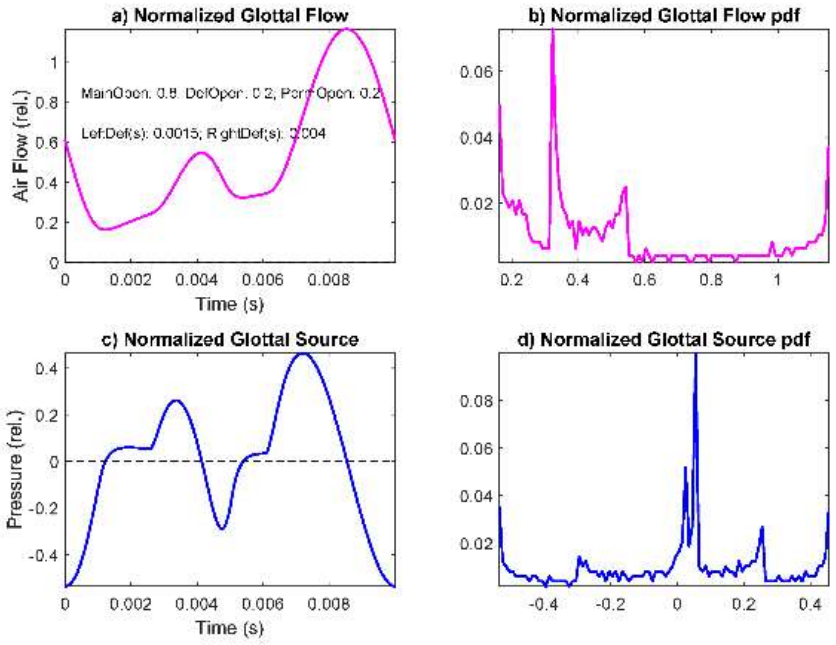

Figure 3.8 Glottal source and flow patterns and their amplitude distributions corresponding to a relative glottal aperture of 0.8, a contact defect of 0.2, and a permanent gap of 0.2.
Description: Glottal source and flow patterns and their amplitude distributions corresponding to a relative glottal aperture of 0.8, a contact defect of 0.2, and a permanent gap of 0.2: a) glottal flow; b) amplitude distribution of the glottal flow; c) glottal source; d) amplitude distribution of the glottal source. This case is a combination of those in Figure 3.6 and Figure 3.7. The glottal flow does not stop at any time, and it follows an almost straight ascending line during the contact phase, whereas the glottal source reproduces a small wavelet of the main L-F pattern.

The question now is to examine to what extent the patterns produced by Rothenberg's model are realistic representations of specific glottal patterns. The answer is provided in Figure 3.9, where a real estimation from a male case showing pathological behaviour, and a case simulated by the model are shown for comparison.

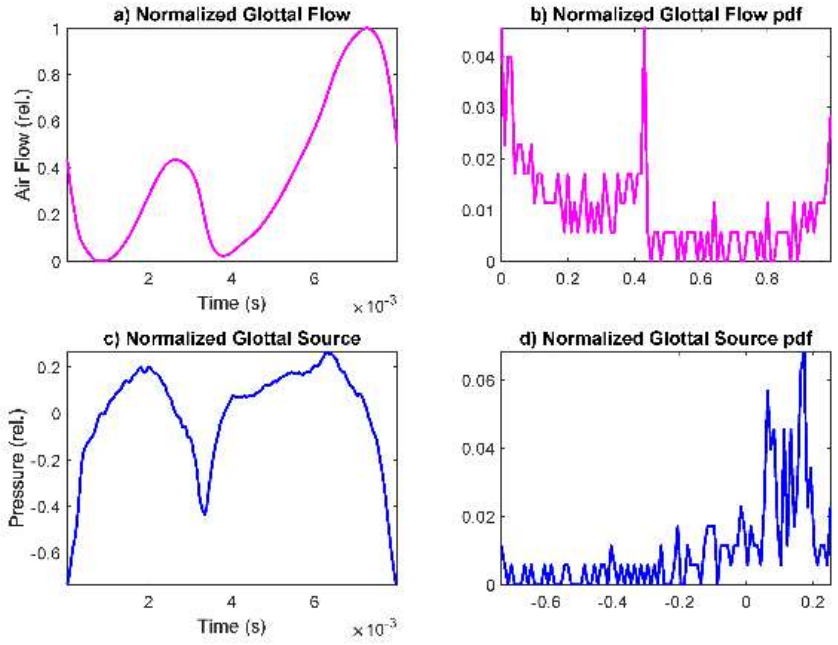

Figure 3.9 The glottal source and flow estimated by inverse filtering from a real case corresponding to a male participant with a clear spontaneous contact defect.

The results of simulating the behaviour of the glottal flow and source with Rothenberg's model may be seen in Figure 3.10.

The comparison of the glottal flow and source from a real case obtained by ONPT inversion, and the results of simulating a permanent and spontaneous contact defect with Rothenberg's model show important similarities, which avail both the process of glottal source estimation and the simulation of a real phonation model.

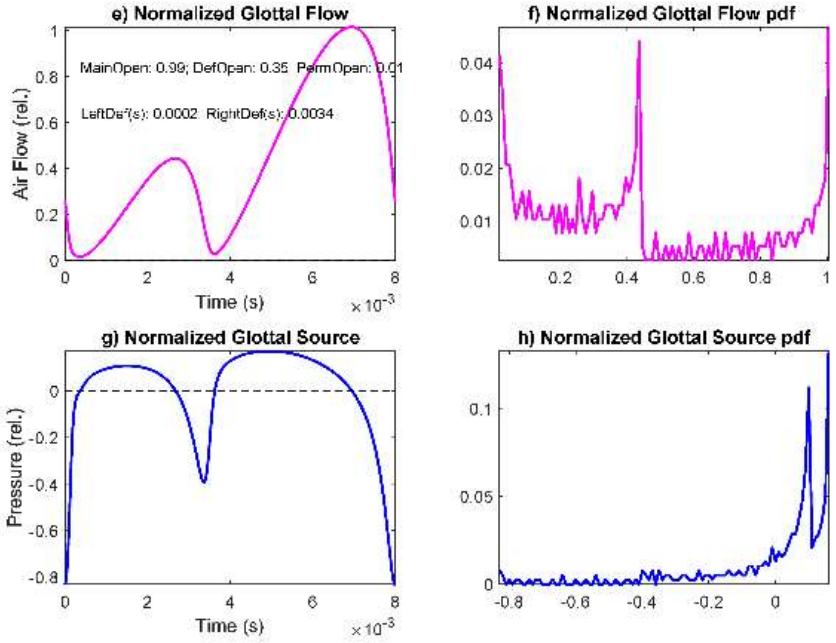

Figure 3.10 Replication of the general pattern behaviour of the case presented in Figure 3.9.
Description: The glottal source pattern corresponds to a main opening of 0.99, a contact defect of 0.35, and a permanent gap of 0.01.

3.2.1 Excitation model

The spectral density of the glottal pulse presents a profile approximately following a curve inverse to frequency, $1/\omega$, which may be associated with a generating model given by the flow graph in Figure 3.11. This excitation model is a first approximation, that corresponds to a first-order IIR filter modelling the delayed decay of the signal input.

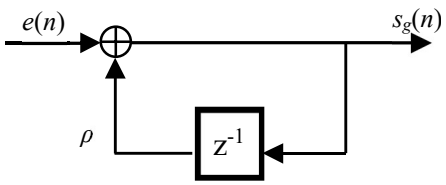

Figure 3.11 Flowgraph corresponding to the model generating the glottal pulse in the discrete-time domain.
Description: $e(n)$ is the signal input (pulse δ-train excitation), $s_g(n)$ is the glottal signal output, and ρ is the feedback coefficient of the system.

The mentioned frequency behaviour corresponds with a profile of $1/\omega$ in the Laplace domain, which is related to the integral of the input:

$$s_{gn} = e_n + \rho s_{gn-1} \tag{3.3}$$

with associated transfer function in the discrete domain given by:

$$F_g(z) = \frac{1}{1 - \rho z^{-1}} \tag{3.4}$$

For $\rho < 1$ but in proximity to unity, this corresponds to a straightforward formulation of the integral of e_n using the first-order finite difference method, incorporating a leakage factor to potentially offset any biasing effect in the integration. It is crucial to note that this model serves as an initial approximation to the generation model of the glottal pulse. The glottal source itself does not exhibit minimum phase characteristics, yet this model can provide an initial approximation of the glottal pulse. This estimate can be subsequently used in iterative refinements for evaluating the ONPT Transfer Function, as elaborated upon in later sections.

3.2.2 Radiation model

As it will be referred to in subsection 3.3.1 the vocal tract may be modelled as a finite number of concatenated tubes of varying sections, with the endpoint at the lips, where the last section of the tube is connected to the open space as shown in Figure 3.12.

The pressure wave coming from the inner vocal tract reaching the lips (vocalized speech) is referred to as s_{ln}. This wave is mostly reflected in the mouth opening due to the abrupt change in section, this causes a reflection due mainly to the change in section and conditions.

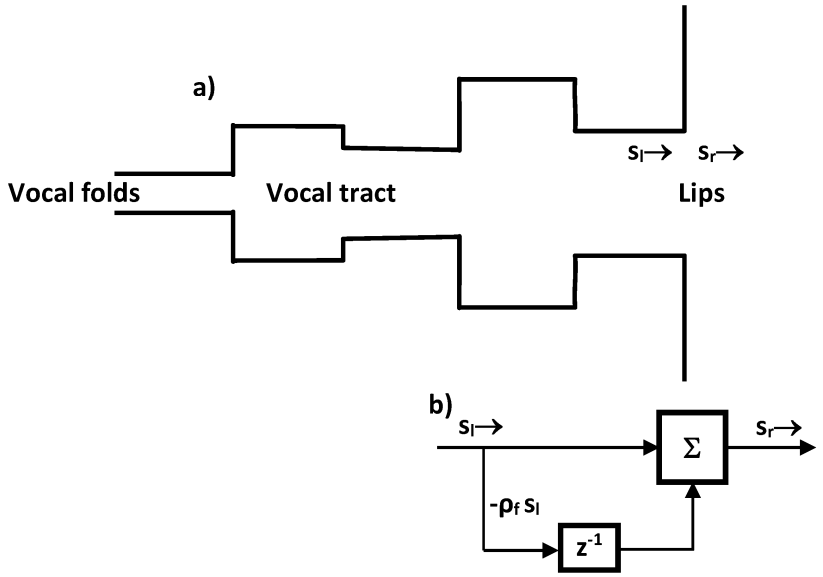

Figure 3.12 Tubular section of the vocal tract and flow graph corresponding to the termination conditions at the lips.
Description: s_l and s_r are the vocalized and radiated speech signals, respectively. ρ_f is the reflection coefficient at the radiation end.

This effect is encapsulated by the *reflection coefficient* from the lips to the open space ρ_f (Deller, Proakins and Hansen, 1993, pp 180-187), which may be expressed as:

$$\rho_f = \frac{Z_f - Z_l}{Z_f + Z_l} = \frac{\frac{\varrho c}{S_f} - \frac{\varrho c}{S_l}}{\frac{\varrho c}{S_f} + \frac{\varrho c}{S_l}} = \frac{S_l - S_f}{S_l + S_f} \tag{3.5}$$

S_l being the equivalent section of the lip aperture, S_f being the equivalent area of the open space, ϱ being the density of air, and c being the speed of sound in air. As $S_l \ll S_f$, this reflection coefficient is close to -1. The transfer function towards the open space is given by $1 - \rho_f z^{-1}$, to produce the output s_{rn}:

$$s_{rn} = s_{ln} - \rho_f s_{ln-1} \tag{3.6}$$

This transfer function behaves quite differently in the frequency domain depending on the value of ρ_f for open or closed tube terminations (open: $S_l \ll S_f$ or closed: $S_l \gg S_f$). For open termination if $\rho_f \to -1$:

$$R_l(z) = \frac{S_r(z)}{S_l(z)} \approx 1 + z^{-1} \tag{3.7}$$

it may be shown that:

$$|R_l(\omega)| \to \sqrt{2 + 2\cos\omega} \tag{3.8}$$

which has a maximum value of 2 for $\omega = 0$ and two null values for $\omega = \pm\pi$. This behaviour corresponds to a two-sample average, with low-pass filtering properties, reducing high frequencies, which gives the speech a spectral tilt. This effect reinforces the $1/\omega$ behaviour of the glottal source generating model given in (3.3), and explains the need for pre-emphasis compensation as a previous step to the ONPT inversion, as it will be commented on in Chapter 4.

3.3 Articulation model

The ONPT is conceptually represented as a series of interconnected cylindrical tubes with varying cross-sectional shapes, approximating the time-varying cross-section of the ONPT during speech production (see Figure 3.13). Depending on the dimensions and lengths of these cylindrical structures, a range of resonant frequencies emerges and evolves as these cavities dynamically change section and shape with the movement of the articulation organs (velum, tongue, jaw, orofacial, etc.). To effectively model this complex, time-varying cavity-like structure using a concatenation of rigid-walled cylindrical tubes, several critical assumptions must be considered.

The most pertinent ones are:

- The complex 3D structure of the ONPT is assumed to be extended on a single longitudinal axis (straight unfolding).
- The cavities of the ONPT are symmetrical concerning the longitudinal axis.
- The irregular cross-section is represented by a circular section of equivalent area normal to the propagation of the sound wave.
- The ONPT is considered a single cavity, with no derivations. This assumption excludes nasals and nasalized vowels (see Figure App.1 showing the four different configurations of the ONPT in Appendix III). The valid configuration would correspond to an open vocal tract and a closed nasal tract at the velopharyngeal switch (VT: O; NT: C).
- The time-varying conditions of the articulation studied might be considered quasi-stationary compared with the duration of the time windows used in the estimations. This condition is fulfilled for time windows under 10 ms (Huang, Acero, and Hon, 2001).

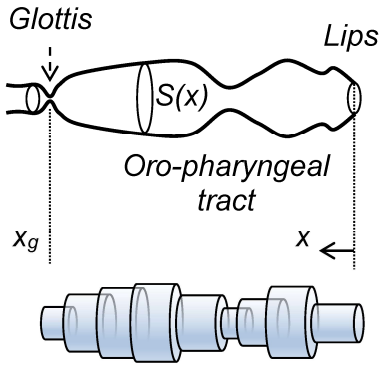

Figure 3.13 Structure of concatenated tubes representing the vocal tract.

Therefore, for each section, there will be a forward propagation wave that when reaching a section change will be in part reflected backwards into the section, and transmitted in part to the next section. Therefore for each section interconnection, this interaction can be modelled as forward and backwards transmitting branches, and forward-to-backward and backward-to-forward reflection branches, the propagation delay being represented by z^{-1} boxes, as depicted in Figure 3.14, showing two connected tube sections (Atal and Hanauer, 1971; Wakita, 1973).

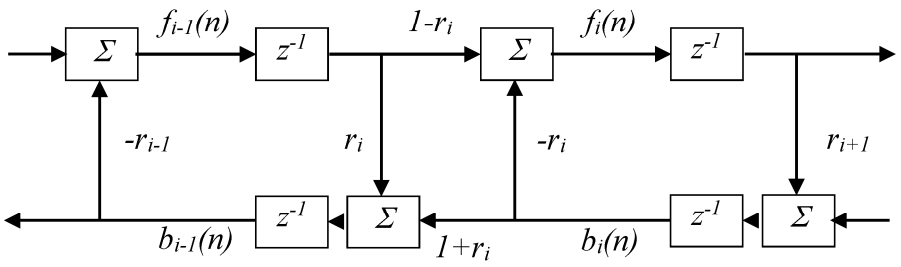

Figure 3.14 Vocal tract equivalent filter explaining concatenated tube propagation.
Description: Forward and backward propagation waves at stage 1: $f_i(n)$ and $b_i(n)$. Reflection coefficient at stage i: r_i (adapted from Atal and Hanauer, 1971).

The relationship among the forward and backward propagating waves in one of the sections (from Figure 3.14) will be the following:

$$\begin{aligned} f_n^i &= (1-\rho_i) f_{n-1}^{i-1} - \rho_i b_n^i \\ b_n^{i-1} &= \rho_i f_{n-2}^{i-1} + (1+\rho_i) b_n^i \end{aligned} \tag{3.9}$$

where the reflection coefficient will be:

$$\rho_i = \frac{Z_i - Z_{i-i}}{Z_i + Z_{i-i}} = \frac{\frac{\varrho c}{S_i} - \frac{\varrho c}{S_{i-1}}}{\frac{\varrho c}{S_i} + \frac{\varrho c}{S_{i-1}}} = \frac{S_{i-1} - S_i}{S_{i-1} + S_i} \tag{3.10}$$

S_{i} and $S_{\mathrm{i}-1}$ being the equivalent areas of two connected tubes, ϱ being the density of air, and c being the speed of sound in air. It will be straightforward to reconstruct the surface profile of the vocal tract from (3.10) by the following recursion:

$$S_i = S_{i-1_f}\frac{1-\rho_i}{1+\rho_i} \tag{3.11}$$

For normalisation purposes, it will be assumed that the initial section (oral opening at lips) will be $S_0 = 1$.

3.3.1 Two-section transmission model

It will be seen that the transfer function of such a system between the vocal cords and the lips when terminated under the radiation conditions, assuming that it is composed of two tubes with perfect impedance adaptation at the excitation side (vocal folds), as summarised in Figure 3.15 behaves as an *all-pole function*.

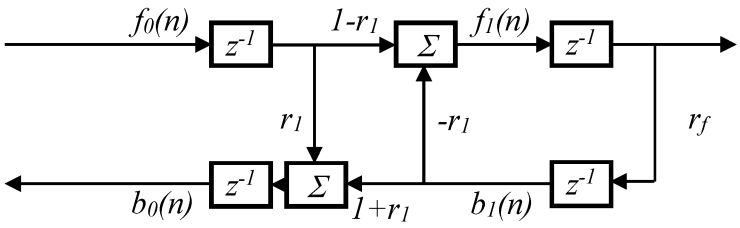

Figure 3.15 Vocal tract equivalent filter explaining concatenated tube propagation. Adapted from Atal and Hanauer (1971).

The following relations will hold from the flowgraph in Figure 3.15:

$$\begin{aligned} S(z) &= \left(1-\rho_f\right)z^{-1}F_1(z) \\ F_1(z) &= (1-\rho_1)z^{-1}F_0(z) - \rho_1 B_1(z) \\ B_1(z) &= z^{-1}\frac{\rho_f}{1-\rho_f}S(z) \end{aligned} \tag{3.12}$$

Now, applying the chain rule to the last expressions:

$$S(z) = \frac{(1-\rho_1)(1-\rho_f)z^{-2}}{1+\rho_1\rho_f z^{-2}} F_0(z) \tag{3.13}$$

Therefore, the transfer function between the glottal pressure wave $p_{\text{sn}} = f_{0n}$ will be expressed as:

$$F_2^{vt}(z) = \frac{S(z)}{F_0(z)} = \frac{(1-\rho_1)(1-\rho_f)}{z^2+\rho_1\rho_f} \tag{3.14}$$

which is a *two-pole* function. When inserting a new section, it may be shown that the new transfer function may be expressed as:

$$F_3^{vt}(z) = \frac{(1-\rho_1)(1-\rho_2)(1-\rho_f)z^{-3}}{1+\rho_1\rho_2 z^{-1}+\rho_2\rho_f z^{-2}+\rho_1\rho_f z^{-3}} \tag{3.15}$$

which is also an *all-pole* function. The generalisation of the transfer function for any order under the conditions expressed above may be given as an *all-pole system* of order *K*:

$$F_K^{vt}(z) = \frac{G_K z^{-K}}{\sum_{i=0}^{K} w_i\, z^{-i}} \tag{3.16}$$

where G_K is a gain factor (scale factor), $z^{-\text{K}}$ is a delay of *K* samples accounting for the travel time required for the wave to travel all along the vocal tract, and the denominator is a *K-order* polynomial with weights related to the *reflection coefficients* of the structure. This *all-pole system* may be inverted using an equivalent order *prediction-error filter* if the order is known (which is not an exact task and may be inaccurate if the sound is nasalised), or by a *Wiener filter* if the order is not known[9].

[9] In relation to this, it must be said that the single duct vocal tract is not valid for certain sounds when the nasal tract is connected to the vocal tract. In this case, more complicated *pole-zero* models must be used, and their inversion will not be feasible by *prediction-error filters* of finite order. The *Wiener filter* would still be valid, as it can approximate the inverse of any transfer function provided the depth of the *equivalent prediction-error filter* is large enough.

CHAPTER 4

4 Algorithmic methods

This chapter provides a comprehensive overview of the methodologies employed throughout this study, categorizing them based on their specific applications into three primary groups: phonation estimation, articulation estimation, and supporting methods. Within the first two categories, the methodologies are primarily concerned with the isolation of relevant characteristics from the speech signal. In the realm of phonation, the primary objective is the extraction of critical GS information from the speech signal, involving the removal of the effects of the ONPT. Conversely, in articulation analysis, the focus is on estimating the contribution of the ONPT tract, by tracking the temporal changes in its profile and the resonant frequencies (formants) associated with these changes. Finally, the supporting methodologies encompass various secondary aspects, such as evaluation, comparison, statistical analysis, and transversal vs longitudinal assessment.

As introduced in the previous chapters, phonation and articulation are key characteristics of speech production that require specific processing tools and methods. As introduced in the previous section, determining the contribution of the ONPT is key for both approaches. The multitube lossless model (as an extension of Rothenberg's model, following Deller, Proakis, and Hansen (1993, pp. 174-192) represents the ONPT as a series of concatenated cylindrical sections that act as acoustic cavities that filter the GS. Thus, these series of reflections, feedback, and transference phenomena can be modelled using a lattice filter derived from Levinson-Durbin recursion (Levinson, 1946; Durbin, 1960; Wiener, 1964; Deller, Proakis, and Hansen, 1993, pp. 297-307).

This method is referred to as vocal tract inversion by linear predictive filtering and is the cornerstone for the estimation of phonation and articulation alike, both methods branching after this same methodology.

4.1 Phonation model estimation

The goal when estimating phonation is to remove the effects of the ONPT and reconstruct the main features of the GS. By using linear predictive coding, the pattern GS emerges presenting a series of specific landmarks that serve to estimate other phonation-related signals and biomechanical correlates.

4.1.1 Vocal tract inversion by linear predictive filtering (LPF)

The antecedents of linear predictive filtering applied to the modelling of the vocal tract can be traced back to the pioneering works of Itakura and Saito (1968). A good reference work on linear prediction can be found in Makhoul (1975). A very detailed description of linear predictive inverse filtering is given in Deller, Proakis, and Hansen (1993). This last reference will be used together with the seminal work of Alku (1992) for deriving the specific methodology used in the present study.

4.1.2 Compensation of the Radiation Model

The joint effects of GS build-up and the low-pass behaviour of the radiation model introduce a spectral tilt in speech propagating to the lips, enhancing lower with respect to higher frequencies. This combined effect of the sometimes called “glottal formant” which can be modelled by a pole on the real axis, may create instabilities during the ONPT inversion process, and substantially alter the pattern of the GS cycle (Deller, Proakis, and Hansen, 1993, pp. 192-197). To compensate for these undesired effects, a first-order prediction-error filter lattice, such as the one shown in Figure 4.1 may be used.

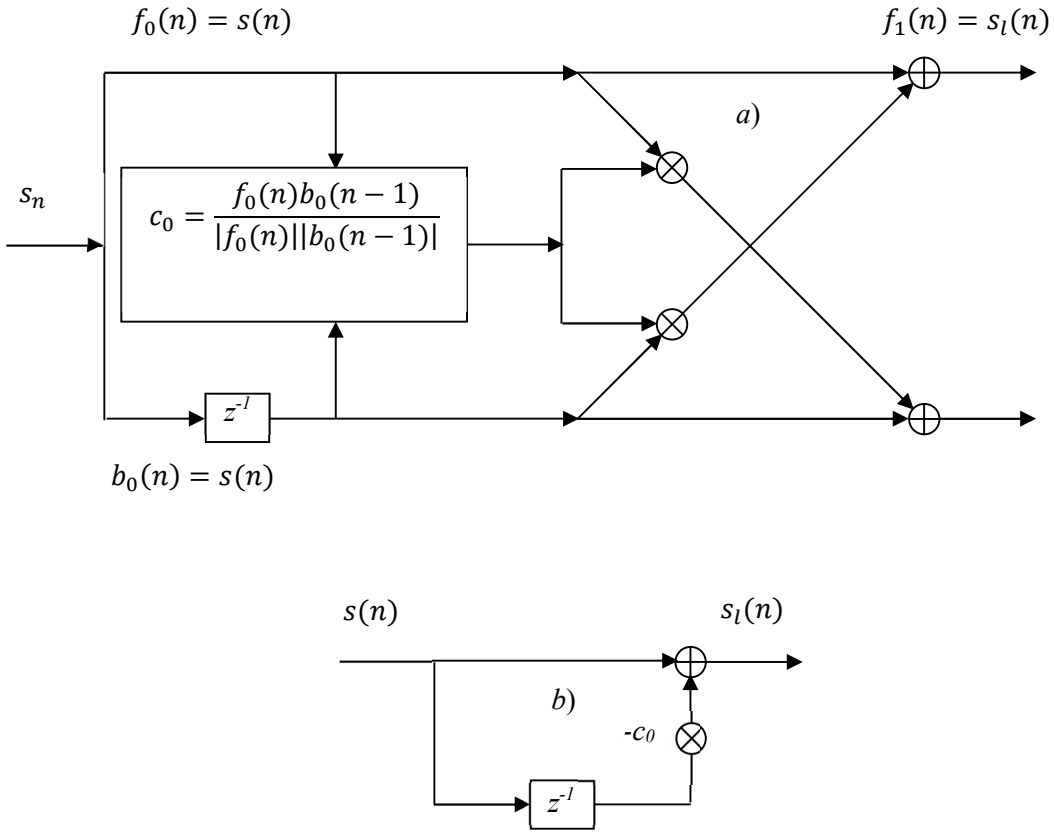

Figure 4.1 First-order prediction error filter lattice and its equivalent FIR Filter.
Description: a) First order prediction error filter lattice; b) its equivalent FIR Filter for cancelling the lip radiation effect. The pivoting reflection coefficient c_0 estimated in (a) is used in the differentiator with leakage 1-c_0 working as a pre-emphasis filter (b).

This setup operates as a FIR filter recursion given by:

$$f_k(n) = f_{k-1}(n) + c_{k-1}b_{k-1}(n-1) \tag{4.1}$$

With the initial conditions of $k = 1$ and $c_0 = -\rho_f$ (first reflection coefficient). Provided that:

$$f_0(n) = b_0(n) = s(n) \tag{4.2}$$

the lattice operates as a first-order differentiator:

$$s_l(n) = f_1(n) = s(n) - \rho_f s(n-1) \tag{4.3}$$

with transfer function given by:

$$H_1(z) = 1 - \rho_f z^{-1} \tag{4.4}$$

cancelling the first-order pole introduced by the radiation effects as discussed above.

4.1.3 Compensation of the excitation model

In order to estimate the frequency response of the ONPT, the effects of the GS have to be removed. The GS low-frequency behaviour if not corrected for will show a $1/\omega$ spectral tilt when estimating the response of the ONPT. The inversion of the system may be instrumented using:

$$e(n) = u(n) - \gamma u(n-1) \tag{4.5}$$

which corresponds to the first difference filter given by:

$$H_g(z) = 1 - \gamma z^{-1} \tag{4.6}$$

Corresponding with a first-order predictor with a coefficient given by $-\gamma$, similarly as in the radiation compensation filter. This procedure is only approximated, as it can be shown that the glottal pulse is not a minimum phase signal. Nevertheless, this problem may be solved by iterative approaches, refining the compensation after a few iterations.

4.1.4 Estimation of the vocal tract transfer function

The generating model for the speech trace given in Figure 3.2 may be inverted by the structure exposed in Figure 4.2. Under the assumption that the operations of the various blocks are commutative, it becomes feasible to group the GS and the radiation model together. The effects on the spectrum of this grouping can then be inverted using a second-order prediction-error filter.

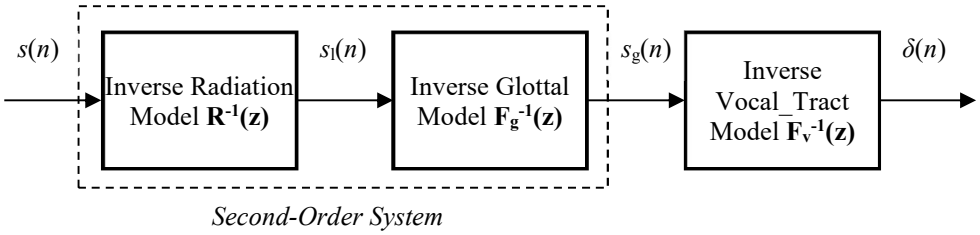

Figure 4.2 Combined effects of the inverse radiation transfer function and the inverse first-order glottal model.
Description: The combined effects of the inverse radiation transfer function $\mathbf{R}^{-1}(z)$ and the inverse first-order glottal model $\mathbf{F_g}^{-1}(z)$ can be implemented as a second-order lattice predictor acting as a pre-processing stage before the vocal tract model inverter $\mathbf{F_v}^{-1}(z)$. Correspondingly, as defined before, $s(n)$, $s_l(n)$, $s_g(n)$, and $\delta(n)$ are the speech signal at the microphone, the estimation at the lips (pre-radiation), the glottal signal, and the generating delta train, respectively. The spectral whitening properties of the whole inversion system are manifested in producing a white spectrum signal $\delta(n)$ from spectrally coloured speech $s(n)$.

The equivalent transfer function of this second-order inverse filter will given by:

$$\begin{aligned} H_{lg}(z) = R^{-1}(z)F_g^{-1}(z) &= \left(1-\rho_f z^{-1}\right)(1-\gamma z^{-1}) \\ &= 1-\left(\rho_f+\gamma\right)z^{-1}+\rho_f\gamma z^{-2} \end{aligned} \tag{4.7}$$

where $\mathrm{R}^{-1}(\mathrm{z})$ and $\mathrm{F_g^{-1}(z)}$ are the respective inverse transfer functions of the lip radiation model and the GS generation model. This second-order inverse filter may be implemented by a second-order prediction-error lattice. The output of this filter $s_v(n)$ is an estimation of the signal contributed by the vocal tract model:

$$s_v(n) = s(n) - \left(\rho_f+\gamma\right)s(n-1) + \rho_f s(n-2) \tag{4.8}$$

thence, taking each signal in its vector form:

$$\begin{aligned} \boldsymbol{s_v} &= \{s_v(n)\}; \\ \boldsymbol{s} &= \{s(n)\} \end{aligned} \tag{4.9}$$

and $\boldsymbol{h_g}$ being the impulse response of the second-order prediction error lattice implementing the filter:

$$\boldsymbol{h_g} = \{1, -\rho_f+\gamma, \rho_f\}; \tag{4.10}$$

the filtering process may be expressed as a convolutional operator[10]:

$$\boldsymbol{s_v = s * h_g} \tag{4.11}$$

Similarly, the vocal tract inverse model might be seen as a Wiener filter, reducing $s_v(n)$ to a signal with white power spectral distribution:

$$\boldsymbol{s_v * h_v = \delta} \tag{4.12}$$

where $\boldsymbol{h_v}$ is the vector form of the Wiener filter impulse response, and $\boldsymbol{\delta}$ is the vector form of a train of delta impulses.

As it was discussed before, the Wiener filter may be implemented by a prediction-error lattice filter (PELF) with a dimension K large enough to reduce the output power spectrum to a flat (white) spectrum signal as considered adequate enough for the estimation purposes assumed.

[10]Throughout this section and the following ones, the sequence notation (e.g. $s(n)$) and the vectorial one (e.g. **s**) will be used interchangeably, for this last is better suited to express convolutional relationships in a more compact way.

4.1.5 Glottal source estimation

With what has been commented on, the complete system to reconstruct the glottal pulse u_{n} from the speech trace, which may be represented in Figure 4.3.

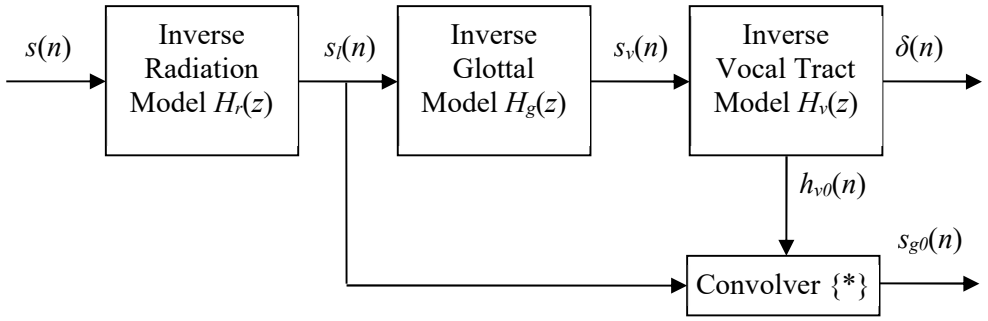

Figure 4.3 Estimation of the glottal residual by deconvolution with the impulse response of a de-glottalized vocal tract model.
Description: The transfer functions of the different deconvolution filters are the inverse convolver transfer functions of the generation models: $H_r(z)=R^{-1}(z)$; $H_g(z)=F_g^{-1}(z)$; $H_v(z)=F_v^{-1}(z)$.

The production process of $s_l(n)$ described in Figure 3.2 may be expressed in compact convolutional form as:

$$\boldsymbol{s} = \left\{\{\boldsymbol{\delta}_n * \boldsymbol{f_g}\} * \boldsymbol{f_v}\right\} * \boldsymbol{r} = \{\boldsymbol{f}_g * \boldsymbol{f_v}\} * \boldsymbol{r} = \boldsymbol{s_l} * \boldsymbol{r} \tag{4.13}$$

where $\boldsymbol{f_g}, \boldsymbol{f_v}$, and $\boldsymbol{r}$ are the impulse responses of the filters $F_{\mathrm{g}}(z)$, $F_{\mathrm{v}}(z)$, and $R(z)$, respectively. Thence, by application of the commutative and associative properties of the convolution operator, taking $\boldsymbol{h_r}$ as the impulse response of $H_{\mathrm{r}}(z)$:

$$\boldsymbol{s_l} = \boldsymbol{s} * \boldsymbol{h_r} = \{\boldsymbol{s_l} * \boldsymbol{r}\} * \boldsymbol{h_r} = \boldsymbol{s_l} * \{\boldsymbol{r} * \boldsymbol{h_r}\} \tag{4.14}$$

as the operators $\boldsymbol{r}$ and $\boldsymbol{h_r}$ are inverse by definition to each other concerning the convolution operation. Besides, as:

$$\boldsymbol{s_l} = \boldsymbol{f_g} * \boldsymbol{f_v} \tag{4.15}$$

convolving this signal with the impulse response of the inverse vocal tract model $\boldsymbol{h_{v0}}$:

$$\boldsymbol{s_l} * \boldsymbol{h_{v0}} = \{\boldsymbol{f_g} * \boldsymbol{f_v}\} * \boldsymbol{h_{v0}} = \boldsymbol{f_g} * \{\boldsymbol{f_v} * \boldsymbol{h_{v0}}\} \cong \boldsymbol{f_g} = \boldsymbol{s_g} \tag{4.16}$$

thus defining a procedure to estimate the glottal residual $\boldsymbol{s_g}$ from the speech trace at the lips $\boldsymbol{s_l}$. A more accurate reconstruction of the GS is obtained if the first estimation is used to better determine the vocal tract transfer function from the speech trace at the lips as shown in Figure 4.4.

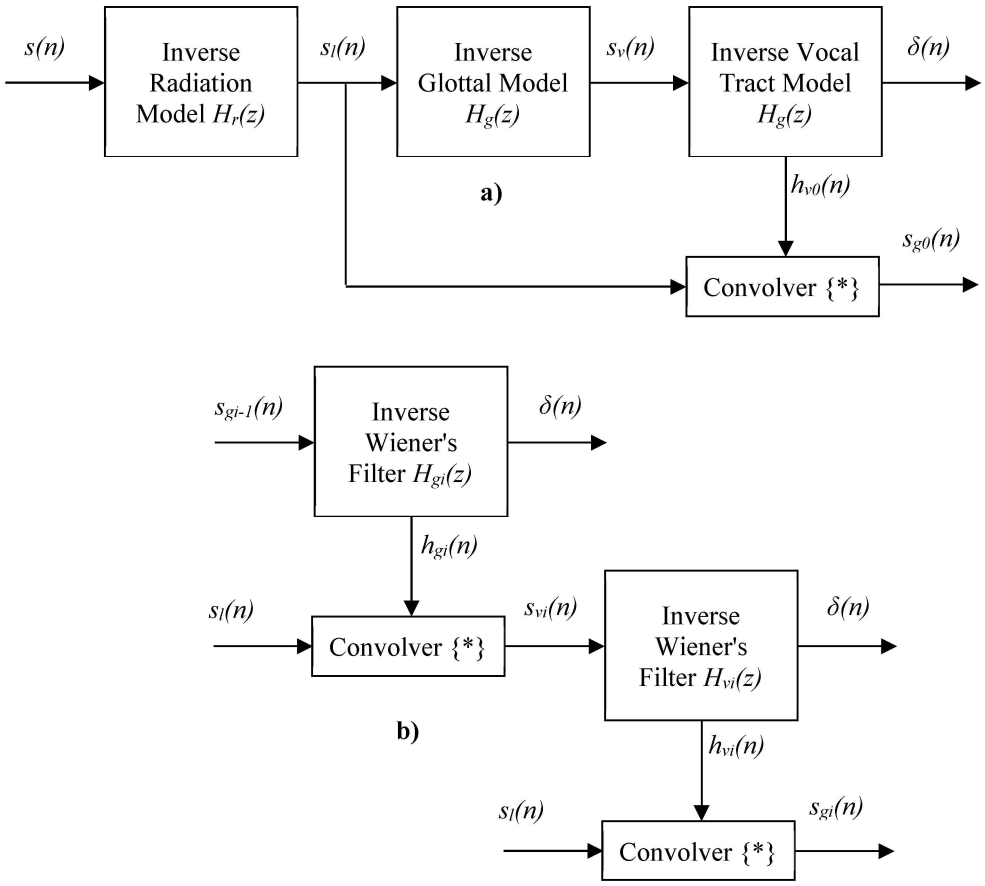

Figure 4.4 Structure of filters to support the joint estimation of the ONPT transfer function and the glottal residual.
Description: a) joint estimation of the ONPT transfer function $H_g(z)$ and glottal residual $sg_0(n)$ by an initial deconvolution step of the phonation signal $s_l(n)$ using the impulse response of the ONPT estimated on the glottal-compensated phonation signal; b) iteration chain extending the modelling of the glottal residual to deconvolve the radiation compensated phonation signal to model the ONPT transfer function and impulse response, which is used in deconvolving the radiation compensated phonation again to produce an updated version of the glottal residual. Description: $h_{gi}(n)$ – impulse response of the Inverse Glottal Wiener's Filter at i-th iteration; $h_{vi}(n)$ – impulse response of the Inverse Vocal Wiener's Filter at i-th iteration; $s_{gi}(n)$ – estimation of the glottal residual at i-th iteration; $s_{vi}(n)$ – estimation of the de-glottalized speech at i-th iteration.

The operation of the iterative system is the following: In the first step, the *i-1*th version of the glottal pulse $s_{gi-1}(n)$ is Wiener-inverse filtered, reducing it to a white process $\delta(n)$.

The coefficients of the equivalent Wiener filter $h_{gi}(n)$ constitute the impulse response of such filter, and when convolved with the radiation-compensated speech $s_l(n)$ produce the de-glottalised speech $s_{vi}(n)$. This signal is also Wiener-inverse filtered, reducing it to a white process $\delta(n)$. The coefficients of the equivalent Wiener filter $h_{vi}(n)$ when convolved with the radiation-compensated speech eliminate the influence of the vocal tract, reducing it to the *i-th* estimation of the glottal residual $s_{gi}(n)$. Usually, a two-step iteration suffices to produce a reliable estimation of the glottal pulse in most real cases.

4.1.6 Iterative joint estimation of the GS and ONPT transfer function

The structure of a Wiener Filter implemented by a PARCOR lattice and its associated convolver (see Figure 4.3 and Figure 4.4) to remove the influence of the transfer function estimated by inverse filtering may be integrated into a single structure as put forth in Figure 4.5.

As shown for example in Deller, Proakis, and Hansen (1993) the residual error from a lattice filter may be seen as the output of an all-pole filter inverse to the lattice input trace. This result allows us to jointly build the inverse impulse response to $\mathbf{s_v}$ by a lattice, reducing this signal to a white series, and at the same time convolves the associated signal $\boldsymbol{s_l}$ with the same inverse impulse response using a paired lattice (lower) which uses the same reflection coefficients estimated in the driving lattice (upper), as shown in Figure 4.5 (joint process estimation).

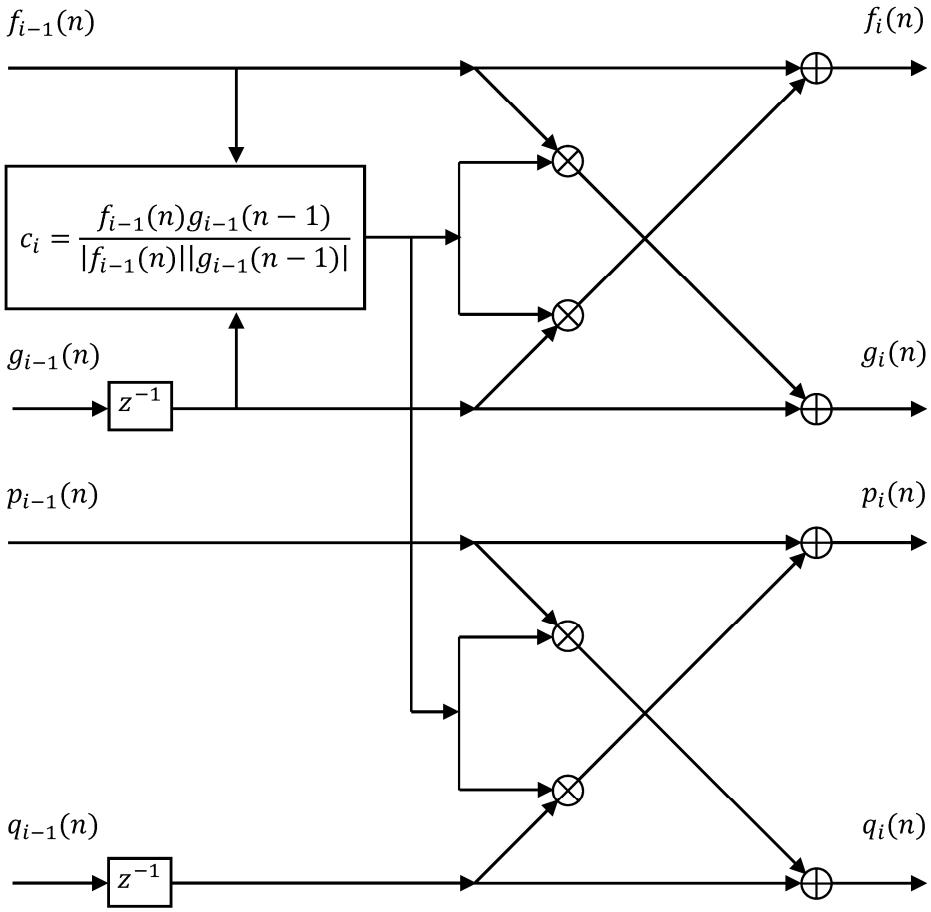

Figure 4.5 Structure of a lattice-ladder filter and deconvolver.
Description: $f_i(n)$ and $g_i(n)$ are the lattice-equivalent main acoustic tube forward and backward propagating errors of the Inverse Wiener's Filter, associated the propagation waves inside the tube at the i-th lattice stage; $p(n)$ and $q_i(n)$ are the backward and forward propagating errors in the paired lattice implementing the Convolver; c_i is the i-th pivoting coefficient (lattice) as well as the i-th reflection coefficient (acoustic tube).

The functionality of the system is the following:

1) The phonation signal $\boldsymbol{s}$ will be processed by an inverse radiation model $\boldsymbol{h_l}$, following (4.14), having in mind the radiation place (oral, nasal) as well as the distance to the capturing device (microphone), the channel, and the effects of both factors, producing a radiation-compensated signal $\boldsymbol{s_l}$.

2) This signal is processed by a mirror filter to neutralize the influence of the GS by an initial filtering model $\boldsymbol{h_{g0}}$ following (4.11), producing a partially de-glottalized phonation signal $\boldsymbol{s_{v0}}$.

3) This new signal is treated by inverse filtering, to produce an inverse model of the ONTP impulse response $\boldsymbol{h_{v0}}$.

4) This inverse model is applied by another mirror filter on the original phonation signal $\boldsymbol{s}$ to generate a glottal residual $\boldsymbol{s_{g0}}$.

5) This signal is used to start a new iteration building a new inverse glottal model $\boldsymbol{h_{g1}}$ which is an update of $\boldsymbol{h_{g0}}$, by repeating steps 2)-5) the number of times considered sufficient to produce an accurate estimation of the glottal residual $\boldsymbol{s_{gI}}$.

The iteration of steps 2)-5) may be represented in a circular diagram as given in Figure 4.6.

Given the relevance of the glottal signals in the estimation of phonation quality, an example of the estimation of the glottal residual and glottal flow is given in Figure 4.7.

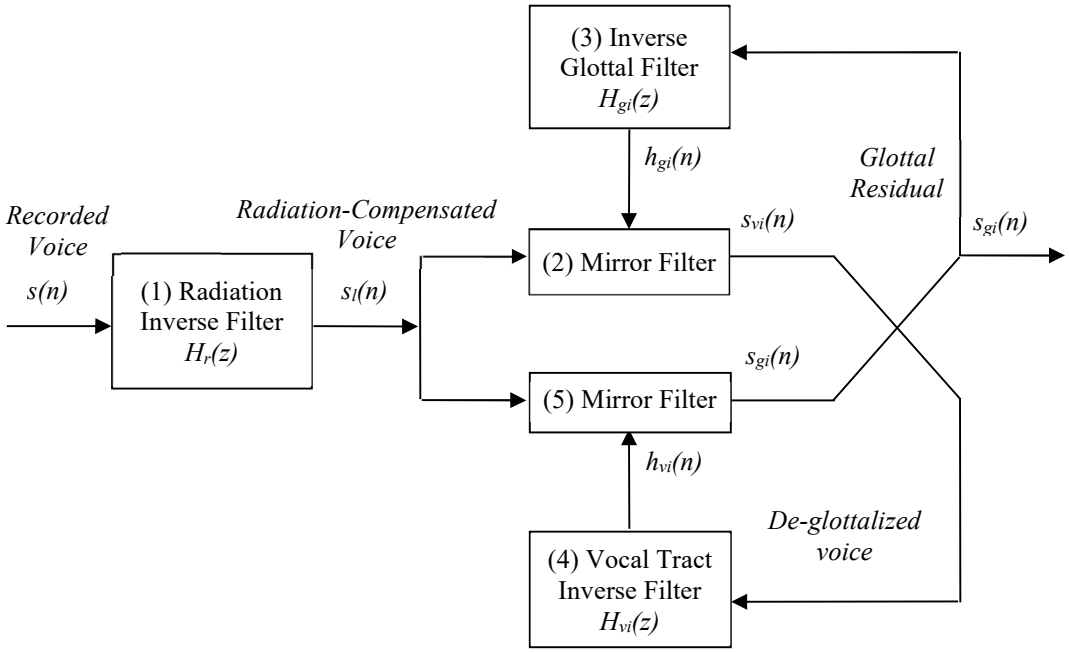

Figure 4.6 Circular diagram of the iterative procedure to jointly estimate the ONPT and the glottal residual.
Description: Iterative procedure to jointly estimate the ONPT $H_{vi}(z)$ and the glottal residual $s_{gi}(n)$ from the phonation signal $s(n)$, presented unfolded in Figure 4.4.

This methodology may be traced back to the work of P. Alku (1992), conveniently modified to estimate the biomechanical parameters of a 2-mass vocal fold model (Gómez-Vilda et al., 2007; Gómez-Vilda et al., 2009), and incorporates the following main features:

- Antiradiation filter compensation. It is a well-known fact (Huang, Acero, and Hong, 2001) that the acoustic wave once radiated from the emission place (mouth, cranial bones, chest) experiences a frequency decay of 1/ω, which may be compensated by a first-order filter (pre-emphasis), which could be implemented as a simple first-order difference. For optimal filtering, in our case, a first order LPF predictor is used, because it corrects optimally the frequency tilt using the first reflection coefficient in the predictive inversion chain.
- Inverse ONPT transfer function estimation by linear predictive filtering (LPF) from the speech de-glottalised signal.

- Cancellation of the vocal tract transfer function by a ladder filter (mirror filter) to produce an estimation of the glottal residual signal.
- Inverse glottal filter estimation by LPF from the glottal residual signal.
- Glottal filter transfer function cancellation by a ladder filter (mirror filter) to produce an estimation of de-glottalised voice.

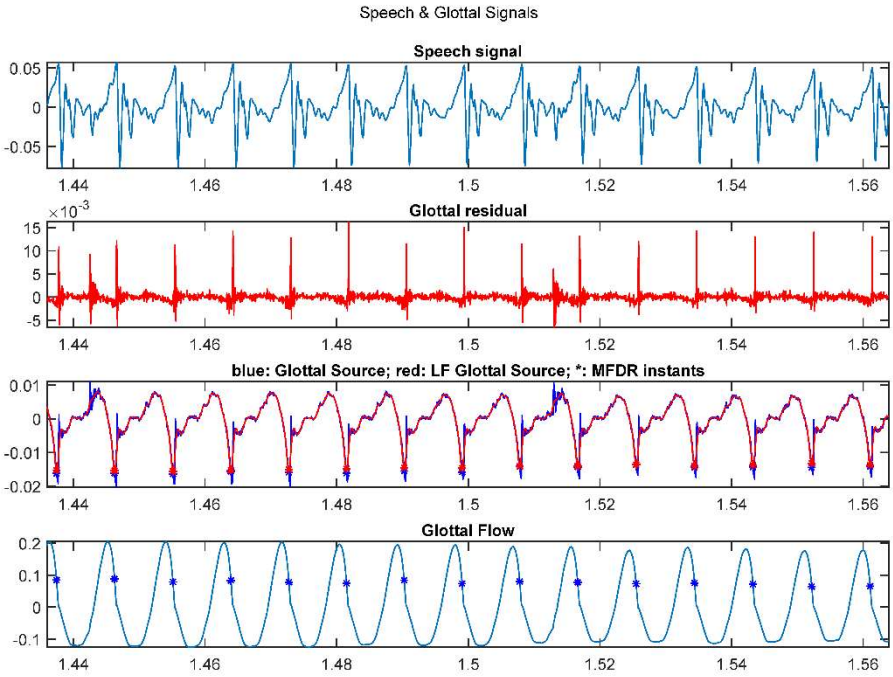

Figure 4.7 Glottal signals from a sustained emission of vowel [a:] uttered by a male normative speaker. Description: a) segment of vowel recording $s(n)$; b) glottal residual $s_g(n)$ after inverse filtering (blue: original integration, red: low frequency filtered); c) GS $p(x_g,n)$, estimated from integrating the glottal residual; d) glottal flow $u(x_g,t)$, estimated from integrating the GS. The MFDR is marked with stars ().

The radiation-compensated phonation signal $s_l(n)$ is shown in Figure 4.7.a, as a series of pseudo-periodic phonation cycles marked by negative spike-like saliences, followed by bouncing oscillations corresponding to the presence of formants (resonances of the ONPT). Once processed following the procedures illustrated in Figure 4.6, a spike-like train of impulses ($\boldsymbol{\delta}$) is produced as a glottal residual $s_g(n)$ (Figure 4.7.b) where all predictable frequency components have been removed by inverse filtering.

The glottal residual is integrated by a leakage and de-trend algorithm to remove slow-moving shifts, producing the GS signal at the glottis $p_g(x_g, n)$ given in Figure 4.7.c, which remembers the general pattern of the phonation signal, reproducing the classical Liljencrants-Fant profile (Fant, Lijencrants, and Lin, 1985), which will be explained later on. This same signal is integrated again to generate the glottal flow, as seen in Figure 4.7.d. This is an estimation of the air escape through the glottis, its aspect being that of a triangular shape, showing a moderate slope-up leading edge from a minimum, a strong rise to a maximum, and a faster slope-down to a new minimum. The more relevant features of this signal are the initial slope-up related to the permanent gap, and the final slope-down, associated with the MFDR (maximum flow declination rate).

4.1.7 Detailed description of the GS

The typical GS cycle follows a specific pattern described by Rothenberg's glottal flow commutation model discussed in subsection 3.2, producing an ideal L-F profile as the one depicted in Figure 3.5.c (corresponding to a null permanent and a null contact gap). That idealised behaviour is in part reflected in the detailed profile shown in Figure 4.8, which corresponds to one glottal cycle in Figure 3.5.c.

The upper plot (a) shows a normative GS profile, characterised by the following intervals:

1) Recovery Interval 0-t_R, where the dynamic supraglottal pressure returns to its quiescent value (null, corresponding to the atmospheric pressure) as a consequence of the strong descent experienced at the MFDR due to the decay in the airflow produced by the vocal fold adduction process.

2) Quiescent interval t_R-t_O, corresponding to the contact phase, where the dynamic supraglottal pressure is almost zero and the glottal flow remains at a minimum.

3) Flow injection Interval t_O-t_M, where the dynamic supraglottal pressure experiences an increment to a maximum because of the vocal fold abduction process when the glottal flow rises faster.

4) Dynamic supraglottal pressure decay interval t_M-t_T, when the glottal flow experiences the fastest descent (MFDR), due to a decline in flow injection because of the vocal folds reaching a maximum aperture, initiating the adduction process. This effect limits the rate of flow increment and conditions its downfall. The MFDR is the instant when the glottal flow decays at its fastest rate, according to Titze and Palaparthi (2016), and it is considered the end of the GS cycle. It must be stressed that this instant is not the minimum of the glottal flow, which is reached at t_R. The MFDR signals the end of a GS cycle and the start of a new one.

According to Fant's theory on the GS the sharpness and amplitude of the MFDR are the reference features to produce a good harmonic display in the frequency domain (number and sharpness of the harmonics found in the power spectrum of the GS), and a direct causability factor in the production of a large cepstral peak prominence (CPP), in alignment with Burk and Wats (2019), and Šimek and Rusz (2021).

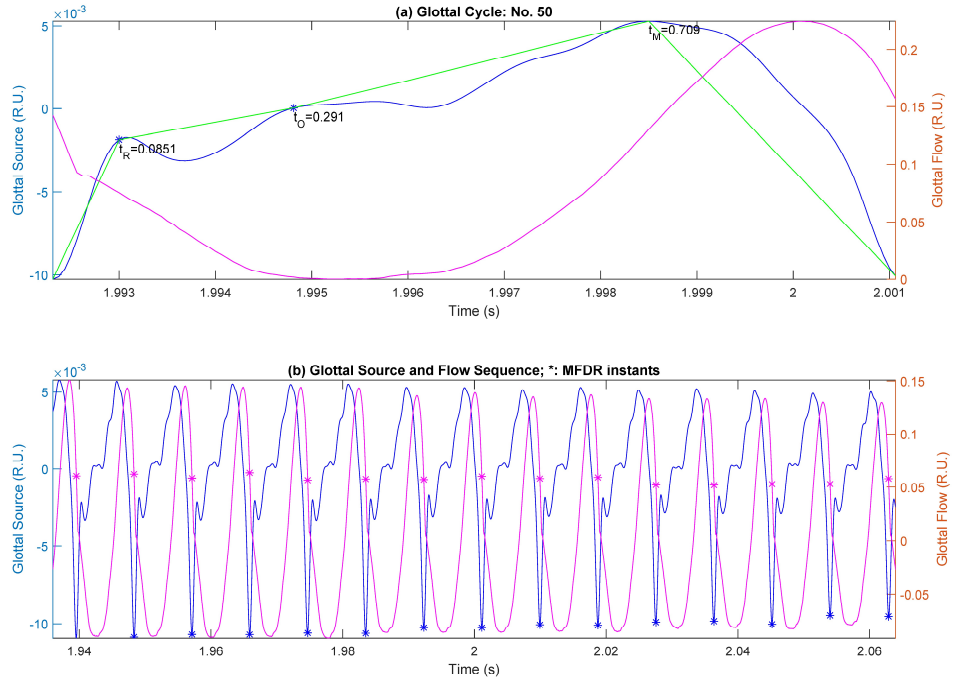

Figure 4.8 Characteristics of a real L-F phonation cycle from a normative male participant (HUGMM). Description: a) detailed GS cycle between two neighbour MFDR instants corresponding to the same phonation signal analysed (blue: Glottal Source; magenta: Glottal Flow; green: support profile). b) sequence of glottal cycles. Time interval description: tR – recovery instant; tO – opening instant; tM: maximum supraglottal pressure instant. Blue: GS; green: glottal flow; purple: reference straight segments used in estimations. R.U.: relative units.

4.1.8 Biomechanical analysis of phonation

It is a well-known fact that PD produces important alterations in neuro-motor activity affecting the muscular tone, and resulting in symptoms such as spasms, rigidity, bradykinesia, unbalance, and tremors, among others (Christine and Aminoff, 2004). This altered neuromuscular activity affects not only to limbs but the phono-respiratory system as well, with instability in phonation being observed as dystonia and tremor in voice, which may be referred to as biomechanical instability of phonation (BIP).

The study of BIP or phonation hypokinesia and dyskinesia, is based on the biomechanical description of the larynx (Titze and Story, 2002). The body-cover 2-mass model (Berry, 2001), reproduced in Figure 4.9, is especially adequate for this purpose.

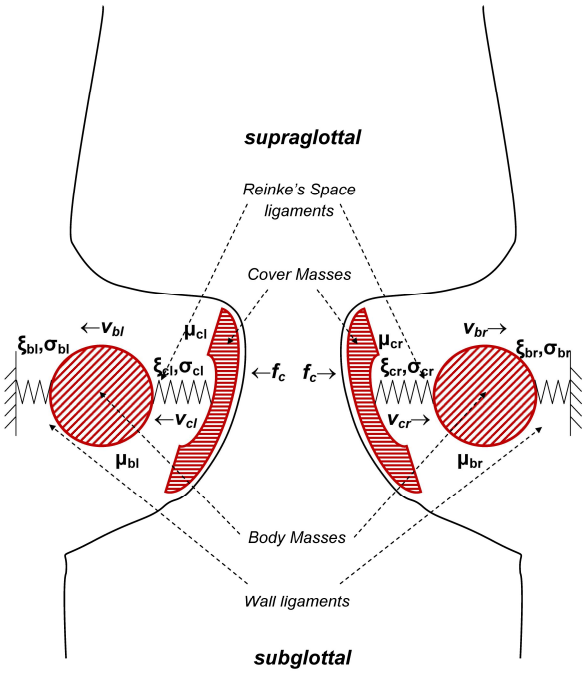

Figure 4.9 Vocal fold 2-mass biomechanical model used in the study
Description: The body and cover structures (diagonal and horizontal line textures), behave as dynamic masses, corresponding to *musculus vocalis* and *lamina propria*. The visco-elastic wall links and Reinke's space ligaments behave as damped springs.

It has a simple description, given by a pair of dynamic equations, coupled through the force f_c resulting from trans-glottal pressure (difference of pressures between the subglottal and the supraglottal sides of the vocal folds), and the transversal velocity of the body and cover mass of the left vocal fold (v_{bl} and v_{cl}). The parameters μ_{bl} and μ_{cl} are the dynamic masses of the left vocal fold body (*musculus vocalis*) and cover (*lamina propria*) given in grams per unit length: g.cm^{-1}. ξ_{bl} and ξ_{cl} are the dumped elastic springs acting between the body and the walls of the thyroid cartilage, and the body and cover masses, respectively.

Their elastic modules are given in N.cm^{-1}. On their turn, σ_{bl} and σ_{cl} are the viscous dissipative parameters explaining the loss of energy experienced by the vocal fold during phonation, given as N.s.cm^{-1}. With these premises, the left and right vocal fold dynamic relationships are:

$$\begin{aligned} f_c &= \mu_{cl}\frac{\partial v_{cl}}{\partial t} + \xi_{cl}\int_{\zeta=t_0}^{t}(v_{cl}-v_{bl})d\zeta + \sigma_{cl}v_{cl} + \mu_{bl}\frac{\partial v_{bl}}{\partial t} + \xi_{bl}\int_{\zeta=t_0}^{t} v_{bl}d\zeta + \sigma_{bl}v_{bl} \\ f_c &= \mu_{cr}\frac{\partial v_{cr}}{\partial t} + \xi_{cr}\int_{\zeta=t_0}^{t}(v_{cr}-v_{br})d\zeta + \sigma_{cr}v_{cr} + \mu_{br}\frac{\partial v_{br}}{\partial t} + \xi_{br}\int_{\zeta=t_0}^{t} v_{br}d\zeta + \sigma_{br}v_{br} \end{aligned} \quad (4.17)$$

The details of the model description can be found in Titze and Story (2002) and are the basis of the further model developments presented in what follows.

4.1.9 Estimating vocal fold biomechanical parameters

Once the GS has been reconstructed using the methods described in 4.1.5, its dynamic behaviour is matched with estimations derived from vocal fold modelling (Berry, 2001) by fitting its power spectral density with the model prediction, as it may be shown that the power spectral density of the GS is strongly conditioned by the biomechanical parameters of the vocal fold models (Story and Titze, 1995). This methodology may be used in the characterisation of the pathologic behaviour of a specific speaker's voice or in the biometric characterisation of the speaker.

The estimation of the elastic stiffness of the vocal fold body and cover is carried out in the following steps:

1. The given phonated speech segment under analysis (voiced speech, preferably an open vowel) is inverted using the already described signal processing methods (Alku, 1992; Deller, Proakis, and Hansen, 1993; Gómez-Vilda et al., 2009) to remove the effects of the supra-laryngeal resonant cavities (ONPT).

2. As a result, an estimate of the supra-glottal pressure just at the point where the glottal flow is injected, known as the GS will be obtained:

$$p_g(t) = \int_{-\infty}^{t} s_r(\zeta)d\zeta \tag{4.18}$$

3. It may be shown (Gómez-Vilda et al., 2009) that the power spectral density of the GS $|P_g(\omega)|^2$ given by

$$\begin{aligned} S_g(\omega) &= |P_g(\omega)|^2; \\ P_g(\omega) &= \int_{-\infty}^{\infty} p_g(t)e^{-j\omega t}dt \end{aligned} \tag{4.19}$$

can be associated with a functional derived from expressions (31) in terms of the model parameters as $T_g(\omega, \mathbf{B})$, where $\mathbf{B} = \{\mu_{cl}, \xi_{cl}, \sigma_{cl}, \mu_{bl}, \xi_{bl}, \sigma_{bl}\}$ is the n-tuple biomechanical feature set. Under certain conditions both functions can be approximated using parameter variation to minimize the following error function:

$$L(\omega, \boldsymbol{B}) = \oint_{2\pi} \left(\|P_g(\omega)\|^2 - \|T(\omega, \boldsymbol{B})\|^2 \right) d\omega \tag{4.20}$$

4.1.10 Average acoustic wave and mucosal wave correlate

The biomechanical parameter estimations are based on the separation of the GS into two components, referred to as the average acoustic wave (AAW) and the mucosal wave correlate (MWC).

The AAW, as it will be later explained, can be seen as the body dynamic component, because it may be associated with the one-mass/one-spring equivalent model of the vocal fold body.

The residual left when removing the AAW from the GS signal is designed as the MWC, as it can be associated with higher-order oscillation modes of the vocal folds related mainly to the dynamic behaviour of the vocal fold cover. Both signals can be considered correlates to the body and cover dynamics and will be referred to as such.

For such, the GS $p_g(t)$ is decomposed into two parts, the AAW, given as $p_a(t)$, and the MWC, given as $p_m(t)$:

$$p_g(t) = p(x_g(t)) = p_a(t) + p_m(t) \tag{4.21}$$

The AAW is a term coined by Titze (1994b, pg. 17, exp. 21-22) to refer to the low-frequency contents of the signal under analysis. The AAW is defined as a rectified sinusoid with the same duration ($T/2$) as the phonation cycle to be matched:

$$p_a(t) = P_{a0}\sin(\omega t); 0 \leq t \leq T/2\,;\ \omega = \frac{2\pi}{T} \tag{4.22}$$

where the amplitude P_{a0} may be evaluated by minimising the mean square of the difference between the AAW and the GS over the time window of the *k-th* phonation cycle:

$$\min_{P_{a0}}\left\{\int_{T_k/2}\left[p_g(t) - p_a(t)\right]^2 dt\right\} \rightarrow P_{a0} = \frac{\int_{T_k/2} p_g(t)\sin(\omega t)dt}{\int_{T_k/2}[\sin(\omega t)]^2\, dt} \tag{4.23}$$

In this way, the AAW would represent a second-order system response (one mass + one spring) associated with the vocal fold body. Therefore, the AAW is dominated by the dynamics of the vocal fold body, and MWC is mainly contributed by the dynamics of the vocal fold cover. This study was developed in detail by Gómez-Vilda et al. (2007).

Incidentally, it may be said that in tense phonation the decay of the GS during the closing phase will mimic the shape defined by Figure 3.5.c very closely as it is shown in Figure 4.10, explaining the rich semantics of the AAW which can be exploited for the interpretation of the nature of the pathologic behaviour expressed in the GS.

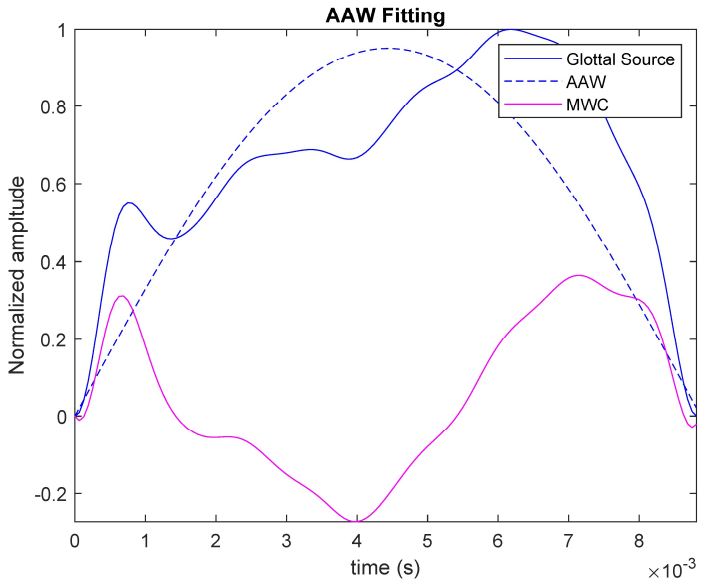

Figure 4.10 Example of the AAW and the MWC.
Description: AAW and MWC estimated from fitting the GS to the vocal fold ideal first vibration mode, corresponding to a normative male participant. The minimum of the MWC serves as a good estimator of the opening instant. This pattern is typical in healthy adult male participants.

4.1.11 Estimation of the AAW biomechanics

The numerical estimation of the biomechanical parameters and particularly, the stiffness induced by the neuro-motor activity on the thyroarytenoid, transversal, and oblique laryngeal muscles controlling phonation, can be carried out using different approaches. A simple yet efficient method consists of first estimating the parameters associated with the biomechanics of the vocal fold body using the average acoustic wave (the first mode of the GS, see Berry, 2001). The mucosal wave correlate (related to the upper modes of the GS) is used to estimate the biomechanical parameters of the vocal fold cover.

The functional $T_b(\omega, \mathbf{B})$, where $\mathbf{B} = \{\mu_b, \xi_b, \sigma_b\}$ is to be matched to the power spectral density of the AAW $|S_a(\omega)|^2$ will be defined on the electromechanical equivalent of the first-order vocal fold body model given in Figure 4.11.

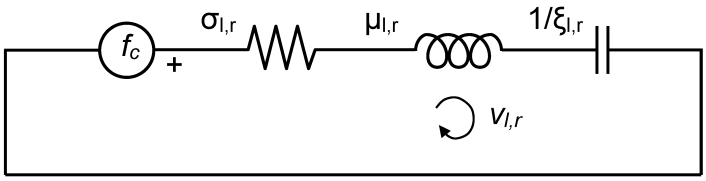

Figure 4.11 Electromechanical equivalent of the vocal fold first vibration mode.
Description: Vocal fold first vibration mode (AAW) where the effects of higher vibration modes (MWC) have been neglected.

The work hypothesis is based on the assumption that the AAW is determined by the fold body dynamic component, therefore the power spectral density of the AAW is directly related to the square modulus of the input admittance derived from the 1-mass model in Figure 4.11 as given by relationship between the vocal fold mass velocity and the lateral force resulting from the pressure difference between the subglottal and supraglottal ridges of the glottis (trans-glottal pressure):

$$T_b(\omega) = |Y_b|^2 = \left|\frac{V_b(\omega)}{F_c(\omega)}\right|^2 = [(\mu_b\omega - \xi_b\omega^{-1})^2 + \sigma_b^2]^{-1} \tag{4.24}$$

where ω is the angular frequency in rad.s^{-1} and μ_b, ξ_b and σ_b are respectively the parameters associated with the lumped mass, elasticity, and dissipative losses of the 1-mass model when only the low-order vibration mode of the vocal fold body is taken into account following the dimensional reduction of the Story-Titze model (Story and Titze, 1995).

The robust estimation of the model parameters is based on the selection of two points on the power spectral density of the AAW, these being $\{T_{b1}, \omega_1\}$ and $\{T_{b2}, \omega_2\}$. The lumped body mass may be estimated then as:

$$\mu_b = \frac{\omega_2}{\omega_2^2 - \omega_1^2}\left(\frac{T_{b1} - T_{b2}}{T_{b1}T_{b2}}\right)^{1/2} \tag{4.25}$$

The selection of the most adequate points for $\{T_{b1}, \omega_1\}$ and $\{T_{b2}, \omega_2\}$ is highly related to the accuracy and robustness of the estimation procedure. A good candidate for $\{T_{b1}, \omega_1\}$ is the position of the main (resonant) peak in the amplitude of the power spectral density of the dynamic correlate:

$$\omega_r^2 = \frac{\xi_b}{\mu_b} \tag{4.26}$$

A good candidate for and $\{T_{b2}, \omega_2\}$ is the position of the third harmonic from the peak position, as the time series shows odd symmetry. These two points have shown to be robust enough in all the cases studied, although some other possibilities are functional if combined properly.

Once the mass has been estimated, the elastic parameter (body stiffness) ξ_b can be obtained from the precise determination of the position of the maximum associated with the resonant peak, this being $\{T_{br}, \omega_r\}$

$$\xi_b = \mu_b \omega_r^2 \tag{4.27}$$

The parameter of body losses can be estimated (but for a scale factor G_b) as:

$$\sigma_b = \frac{G_b}{\sqrt{T_{br}}} \tag{4.28}$$

where T_{br} stands for the value of the square modulus of the input admittance in expression (4.24) at the frequency of resonance ω_r associated to the first maximum in the GS power spectral density.

The evaluation methodology must first produce a very accurate estimation for f_0, which is used to evaluate ω_r. This leads to the determination of the mass from (4.28) and the losses and stiffness from (4.28) and (4.27), respectively. The stiffness of the vocal fold body is the most relevant biomechanical parameter because it is directly related to the neuron firing rate acting on the laryngeal muscles (Jürgens, 2002; Brown et al., 2009; Ludlow, 2005 and 2015), and retains neurologic disease behaviour in marks of hypo- and hypertension, as well as in tremor. Important correlates quantifying neuro-degenerative behaviour in speech are thus vocal fold stress, as well as its statistical dispersion.

Similar derivations may be defined for the biomechanical parameters of the vocal fold cover using in its case the spectral density of the MWC, as the influence of the body dynamics has been removed implicitly on separating the AAW from the GS, reducing the problem to a single mass model. In this way, the application of the same methodology to the cover biomechanics may follow essentially the same steps in a similar way.

An example of the AAW spectral power $|P_a(\omega)|^2$ matching the trans admittance function using this methodology is shown in Figure 4.12.

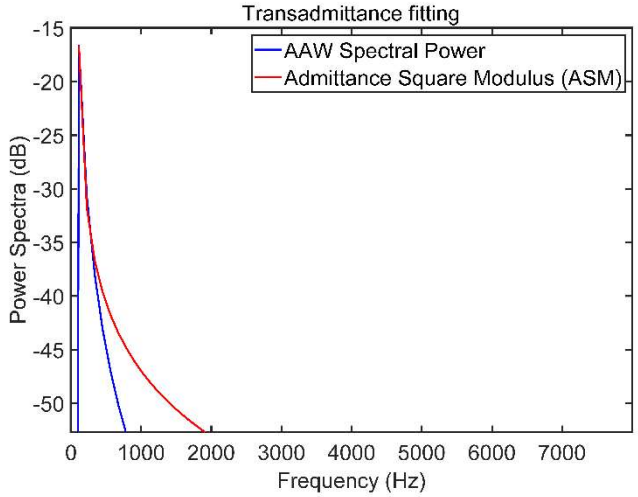

Figure 4.12 Fitting the power spectrum of AAW to an electromechanical equivalent model.
Description: Results of fitting the power spectrum of AAW with the results of the direct estimation of the electromechanical equivalent given in expressions (4.24), (4.25), (4.27), and (4.28).

An example of the estimation of the vocal fold tension for the phonation segment shown in Figure 4.7.a is given in Figure 4.13.

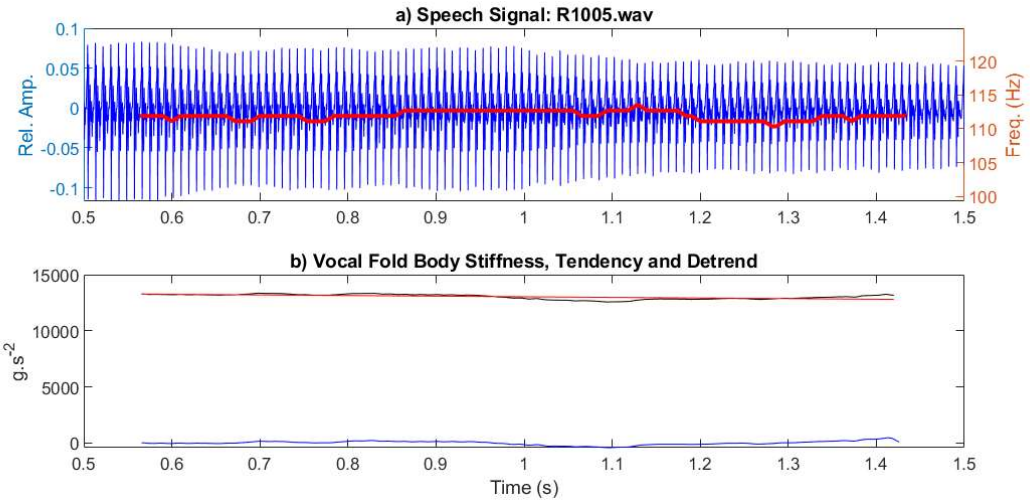

Figure 4.13 Example of the vocal fold tension estimated on a phonation segment.
Description: Vocal fold tension estimated on a phonation segment taken from the same participant as shown in Figure 4.7: a) phonation signal (in blue) and its fundamental frequency (in red); b) vocal fold tension (in black), its trend (in red), and its detrended estimation (in blue). The detrend shows a slight decay from the first estimations to the last ones.

4.2 Articulation model estimation

4.2.1 The neuromechanical jaw-tongue model

The present study is based on a simplified jaw-tongue articulation model (Gómez-Rodellar et al., 2019a) which is known to be representative of PD dysarthria (Gómez-Rodellar et al., 2019b). It allows to create a relationship between acoustic and kinematic variables relating the first two formants $\mathbf{F}=\{F_1, F_2\}$ to the horizontal and vertical coordinates $\mathbf{S}=\{x_r, y_r\}$ of the joint Jaw-Tongue Reference Point (P_{rJT}) in the sagittal plane. The centre of moments for the biomechanical system, consisting of the maxillary bone, tongue, and associated facial tissues, is represented by this point (see Figure 4.14).

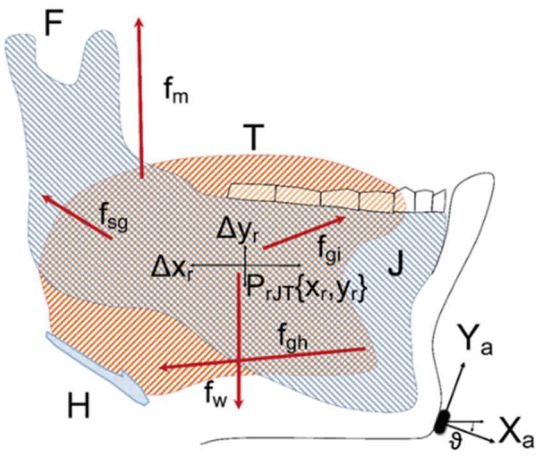

Figure 4.14 Synoptic representation of the jaw-tongue kinetic structure.
Description: H: hyoid bone; J: mandible (dash grey); T: tongue (dash orange). $P_{rJT}(x_r, y_r)$: jaw-tongue reference centre of moments. $\{\Delta x_r, \Delta y_r\}$: displacements of the centre of moments relative to the reference. f_m, f_{gh}, f_w, f_{sg}: extrinsic forces acting on the jaw-tongue structure (masseter, geniohyoid, gravity, styloglossus). f_{gi}: intrinsic force acting on the tongue. $\{X_a, Y_a\}$: coordinates in the sagittal plane of an external accelerometer fixed to the chin. ϑ: angle between the accelerometer axis X_a and the reference sagittal axis x.

The model assumes that it can be established a Linear Time-Invariant (LTI) relationship between the P_{rJT} sagittal coordinates and the relative displacements of the first two formants, which may be summarized as:

$$\begin{aligned} \Delta\mathbf{S} &= \mathbf{W} \times \Delta\mathbf{F}; \\ \mathbf{W} &= \{w_{ij}\}_{i=1,2}^{j=1,2} \end{aligned} \tag{4.29}$$

where $\Delta = \{\Delta s_1, \Delta s_2\}$ is the vector of the horizontal and vertical displacements of P_{rJT} in the time domain, which may be obtained from the rotation and integration of the tangential and normal acceleration components $\{X_a, Y_a\}$ measured by the accelerometer fixed on the chin, $\mathbf{W}$ is a 2x2 matrix expressing the LTI projection model, which will be referred to as the acoustic-to-kinematic projection, and $\Delta\mathbf{F}$ is the relative displacement in the frequency of the first two formants concerning their means in the time domain, as the first two formants are strongly associated with articulation kinematics (Gómez-Rodellar et al., 2019a), defined as:

$$\Delta\mathbf{F} = \{\mathbf{F}_1 - mean(\mathbf{F}_1), \mathbf{F}_2 - mean(\mathbf{F}_2)\} \tag{4.30}$$

4.2.2 Formant estimation

As it was commented in subsection 4.1.6, the methodology used for inverse filtering, produced the simultaneous estimation of the GS and the ONPT transfer function as an all-pole structure (in the case of a pure oral phonation, when the nasopharyngeal pathway is occluded by the stylo-pharyngeal muscles):

$$H_{vK}(z) = \frac{G_v}{\sum_{i=0}^{K} h_{vk} z^{-i}} \tag{4.31}$$

where K is the all-pole order, $h_{v0} = 1$, and G_v is a scale factor.

This transfer function could be represented by an equivalent all-pole function in terms of the zeros of the denominator polynomial:

$$H_{vK}(z) = \frac{G_v}{\prod_{i=0}^{K}(z - z_i)} \tag{4.32}$$

which correspond to K/2 pairs of complex conjugate zeros in case of K being a pair, and to (K-1)/2 pairs of complex conjugate zeros and a real zero in case of K being odd. This property may be exploited to extract a real zero with the purpose of de-trending the behaviour of the ONPT transfer function, contributing to enhancing radiation compensation effects.

Each one of these complex conjugate zeros represents a pole in the transfer function, which may be seen as frequency places where the transfer function amplitude on the unity circle is enhanced as a result of resonance effects:

$$H_{vK}(z = e^{j\omega}) = \frac{G_v}{\prod_{i=0}^{K}(e^{j\omega} - r_i e^{j\varphi_i})} \tag{4.33}$$

where r_i and φ_i are the modulus and the angular frequency of each zero, represented as:

$$z_i = r_i e^{j\varphi_i} \tag{4.34}$$

where the contribution of each associated pole to the modulus of the transfer function on the unity circle is proportional to $1/(1 - r_i)$, at the frequency given by $F_i = f_s\,\varphi_i/\pi$, f_s being the sampling frequency.

Different approaches to the estimations of the zeros of a polynomial in the complex plane may be used to find the polynomial roots (Edelman and Kostlan, 1995; Epperson, 2013; García and Díaz, 2014). The MATLAB function used in the present study is based on the companion matrix method of Jenkins and Traub (Epperson, 2013, pg. 518).

Examples of the first two formant estimations from a normophonic speaker and a PD speaker are shown in Figure 4.15.

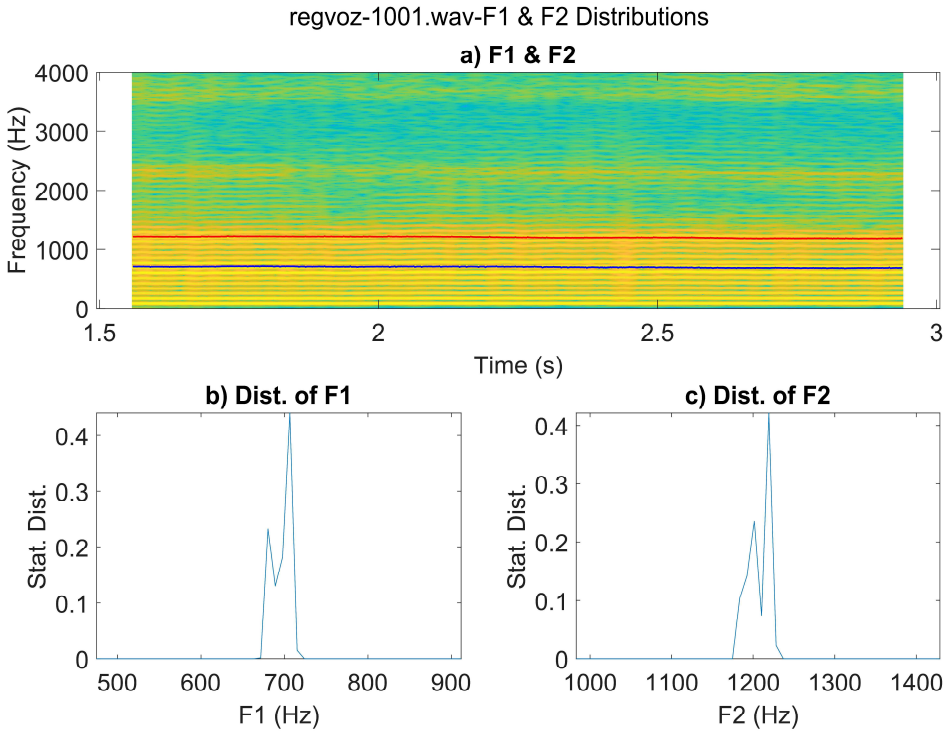

Figure 4.15 Estimations of the first two formants on the sustained utterance of the vowel [a:]. Description: Example from a normative adult male speaker, following the methodology for polynomial-zero estimation by Jenkins and Traub (Epperson, 2013, pg. 518) and the modulus-angle pattern (4.34).

4.2.3 Mapping formant dynamics to articulation kinematics

Speech Articulation is determined by the movement of the jaw, tongue, lips, and velopharyngeal tissues (Dromey, Jang, and Hollis, 2013; Whitfield and Goberman, 2014). Specifically, vowels are defined by certain articulation gestures as the open-close (also low-high attending to lower jaw position relative to the upper jaw), front-back, and round-oval configurations, determining acoustic features perceived as formant positions.

This gesture-acoustic association is usually represented on a vowel polygon (see Figure 4.16).

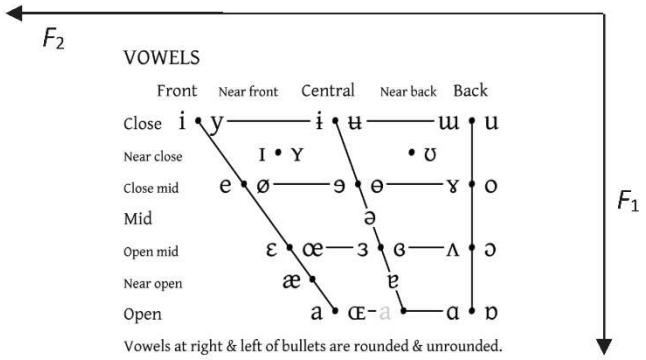

Figure 4.16 Multilingual Vowel Set (IPA).
Description: The vertical and horizontal axes represent the first (F_1) and second (F_2) formants, respectively (reversed axes). The feature close/open corresponds to jaw position high/low.
International Phonetic Association (IPA): http://www.internationalphoneticassociation.org

The open-close gesture is mainly dominated by the jaw, predominantly affecting the first formant F_1 (pulling up the jaw is the dominant gesture in the phonation of [i:] and [u:], whereas relaxing down jaw is the gesture to phonate [a:]). The front-back gesture is mainly controlled by the tongue position, affecting the second formant F_2 (pushing the tongue forward is the gesture for [i:], pulling it back results in [u:]). This is an oversimplification of what is a more complicated relationship between articulation gestures and formant positions, but it will be the starting point for defining kinematic correlates of phonation in this study. The articulation gesture of the jaw has been studied to relate it to acoustic features, as the first two formants (F_1, F_2).

4.2.4 The vowel triangle

The vowel triangle is a simplification of the multilingual vowel set, in which only the extreme positions of the triangle are considered, usually marked by the positions of [u:, a:, i:], because these vowels are easily reproducible by most of the speakers in almost any language. A very meaningful feature related to the articulation span that a speaker can attain is derived from the area enclosed by such a profile, as shown in Figure 4.17.

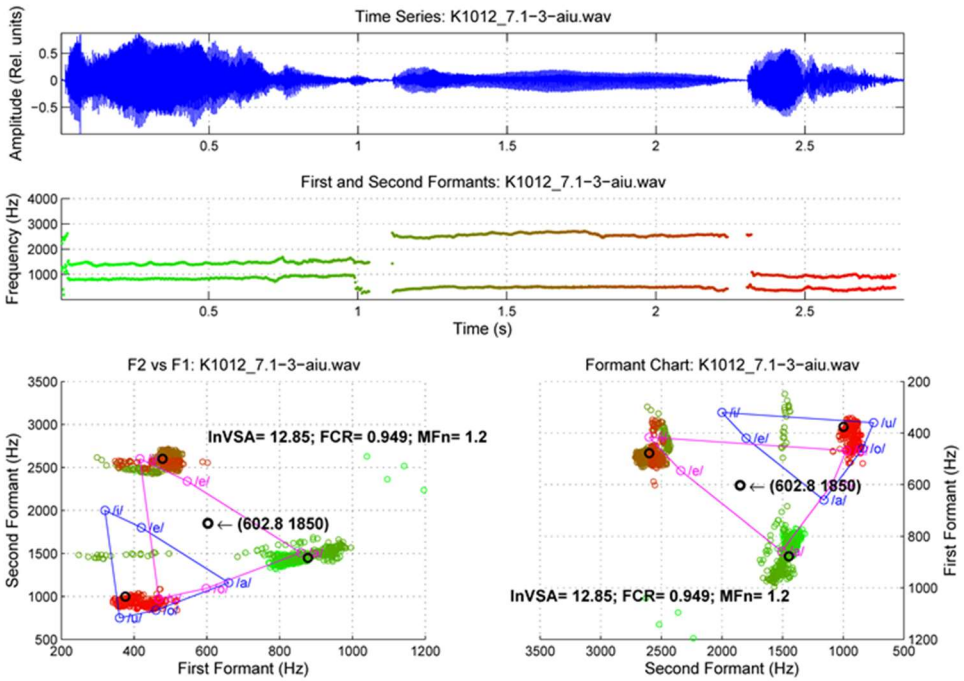

Figure 4.17 Example of the vowel triangle and the vowel space area (VSA) from a sequence [a:, i:, u:]. Description: Top: speech signal. Middle: first two formants from LPC spectral estimation. Bottom left and right: formant projection on the vowel triangle in F2 vs F1 (left) and in F1 vs F2 (reverted) as it is traditional in Linguistics (right). Black circles give the vowel centroids and the vowel triangle centre of gravity to evaluate the lnVSA, the FCR, and the MF (see text), which are shown superimposed on the vowel triangle. The male and female reference vowel triangles for Spanish are given in blue and purple. Colouring is used to signal time positions in the speech sequence, from green (beginning) to red (end).

Figure 4.17 reproduces an utterance by a normative female speaker reproducing the sequence [a:, i:, u:] (top), sampled at 8 kHz, from which the vector {F1, F2}(middle) has been estimated every 2 ms using a 9-order LPC adaptive lattice filter.

Each position in the bottom plots is signalled by a circle, being coloured according to its time position in the sequence, from green (beginning) to red (end).

4.2.5 Features based on vowel distribution.

The following two metrics (Kent and Kim, 2003; Sapir, et al., 2010) have been classically used to quantify the articulation span (VSA: vowel space area) relative to the triangle centre of gravity (FCR: formant centralization ratio), accordingly to the following formulation:

$$lnVSA = ln\left\{\frac{|F_{1i}(F_{2a} - F_{2u}) + F_{1a}(F_{2u} - F_{2i}) + F_{1u}(F_{2i} - F_{2a})|}{2}\right\} \quad (4.35)$$

$$FCR = \frac{|F_{1u} + F_{1i} + F_{2u} + F_{2a}|}{F_{1a} + F_{2i}} \quad (4.36)$$

where F_{1u}, F_{1a}, F_{1i}, F_{2u}, F_{2a}, and F_{2i} are respectively the first and second formants of the corner-cluster data-centroids relative to [u:], [a:], and [i:], respectively; and ln{·} stands for the natural logarithm. Additionally, the modulus of the frequency span in both formants may be used as another meaningful feature:

$$MFn = ln\left\{\left(\frac{F_{1a} - min\{F_{1i}, F_{1u}\}}{mean\{F_1\}}\right)^2 + \left(\frac{F_{2i} - F_{2u}}{mean\{F_2\}}\right)^2\right\} \quad (4.37)$$

where $man\{F_1\}$ and $mean\{F_2\}$ are the averages of the first and second formant estimations over the utterance, excluding silences. This last feature expresses the ability of the speaker to reproduce a wide formant span.

4.2.6 Features based on formant dynamics

The features defined in the previous subsection refer to static positions derived from a set of recordings trying to fix a wider vowel span on the vowel triangle. They do not explore the dynamic characteristics of articulation during vowels, glides, or other articulated phonated utterances. To help introduce the kinematic concept behind formant variability during non-stable utterances, an Articulation Kinematic Model (AKM) is proposed based on Figure 4.14. The ensuing study is focused on the dynamic tracking of the kinematic activity of the jaw-tongue reference point (P_{rJT}), which may be defined as a hypothetical point in the sagittal plane (x: caudal-rostral; y: dorsal-ventral). As it may be seen in Figure 4.14, the P_{rJT} is a hypothetical point $\{x_r, y_r\}$ where the sum of the different forces is null (masseter: f_m, stylo-glossus and genio-hyoglossus: f_{sg} and f_{gh}, genio-glossus: f_{gi}, and the gravity: f_w). The force exerted by the masseter f_m will pull up the low mandible acting as a third-order lever. Other acting muscles are the styloglossus, genio-hyoid, and glosso-intrinsic, acting on the jaw-tongue concerning the reference point. The AKM is integrated by the jaw (J) and tongue (T) and the facial tissues attached to them. The dynamics of this system (Hannam et al, 2008) may be approximated by a third-order lever, its fulcrum (F) being attached to the skull, articulating movements on the sagittal plane (x, y). Gravity acts as a constant force downwards (f_w). The articulation gesture will determine the position of a hypothetical reference point in the jaw-tongue centre of masses (P_{rJT}), attached to the jaw joint point F. The acoustic features $\{F_1, F_2\}$ may be associated to the reference point coordinates $\{x_r, y_r\}$ as in (4.38), assuming that the system is linear and time-invariant and that a one-to-one association between articulatory gestures and acoustic features is possible (Dromey, Jang, and Hollis, 2013). The position of the P_{rJT} will change in time under the action of the forces mentioned, modifying the resonant properties of the oral cavity, and producing dynamic changes in formants.

The work hypothesis considers that the changes in the first two formants F_1 and F_2 can be related to the AKM dynamics as by:

$$\begin{bmatrix} F_1(t) \\ F_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_r(t) \\ y_r(t) \end{bmatrix} \tag{4.38}$$

where a_{ij} are the parameters relating P_{rJT} to formant values. Under these assumptions, a relationship between the dynamic components of articulation and acoustics may be derived from (4.38):

$$\begin{bmatrix} \Delta F_1(t) \\ \Delta F_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \Delta x_r(t) \\ \Delta y_r(t) \end{bmatrix} \tag{4.39}$$

The utility of these relationships is conditioned by the possibility of estimating the set of parameters a_{ij}, as it will discussed in subsection 5.2.

Assuming now the invertibility and time invariance of (4.39) the following relationship could be established:

$$\begin{bmatrix} dx_r(t)/dt \\ dy_r(t)/dt \end{bmatrix} = \mathbf{W} \begin{bmatrix} dF_1(t)/dt \\ dF_2(t)/dt \end{bmatrix}; \quad \mathbf{W} = \mathbf{A}^{-1} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \tag{4.40}$$

With these expressions in mind, it will be possible to define the absolute kinematic velocity (AKV) of the reference point (P_{rJT}) as:

$$|v_r(t)| = \left[\left(\frac{dx_r(t)}{dt} \right)^2 + \left(\frac{dy_r(t)}{dt} \right)^2 \right]^{1/2} \tag{4.41}$$

which may be rewritten as:

$$|v_r(t)| = \left[H_1 \left(\frac{dF_1(t)}{dt} \right)^2 + H_2 \left(\frac{dF_2(t)}{dt} \right)^2 + H_{12} \frac{dF_1(t)}{dt} \frac{dF_2(t)}{dt} \right]^{1/2} \tag{4.42}$$

where H_1, H_2 and H_{12} are quadratic forms of w_{ij}:

$$H_1 = w_{11}^2 + w_{12;}^2 H_2 = w_{21}^2 + w_{22;}^2 H_{12} = w_{11}w_{12} + w_{21}w_{22} \tag{4.43}$$

If it happens that $w_{11} \ll w_{12}$ and $w_{21} \ll w_{22}$, $H_1 \approx w_{11}^2$; $H_2 \approx w_{22;}^2$, and $H_{12} \ll H_1, H_2$. On the other hand, if the time derivative amplitudes of F_1 and F_2 are the same order of magnitude, the third term $H_{12}\, dF_1(t)/dt\, dF_2(t)/dt$ might be removed from the expression at almost no cost in terms of information loss. In this way, an estimation of the AKV may be produced exclusively in terms of formant dynamics:

$$|v_r(t)| \approx \left[H_1 \left(\frac{dF_1(t)}{dt} \right)^2 + H_2 \left(\frac{dF_2(t)}{dt} \right)^2 \right]^{1/2} \tag{4.44}$$

Reliable estimates for w_{ij} can be obtained from articulations involving changes in the positions of the reference point showing predictable dynamic changes. A very relevant statistical feature to describe articulation kinematics can be defined from the probability distribution of the AKV in (4.44), directly estimated as its normalized amplitude histogram over a given number bins (N) between 0 and 50 cm.s^{-1} as:

- F_1 and F_2 are estimated each 2 ms using an adaptive linear predictor at an implicit frequency resolution of 500 Hz.
- The AKV's ($|v_r(t)|$) is estimated from (4.44).
- An N-bin histogram of counts by amplitudes is built from each subject's AKV. The interval covered for speeds is [0, $|v_r|_{max}$], with $|v_r|_{max}$=50 cm.s^{-1}, and N=400, therefore each bin size is Δb_k=[$|v_r|_{max}/N$]=0.125 cm.s^{-1} wide.
- The following histogram of counts is built for each bin b_k=$k\cdot\Delta b_k$:

  if $b_{k-1} \leq |v_r(t)| < b_k$ *then* c_k=c_{k+1}

  where c_k is the number of counts for bin b_k.

- Count histograms c_k ($0 \leq k \leq N$) are normalized to their total number of counts $C_t = \Sigma b_k$ ($0 \leq k \leq N$), therefore they could be considered approximate estimators of probability density functions $p_k = c_k / C_t$.

Thence $p(|v_{rk}|) = p_k$ will be an estimate of the AKV probability density function. This feature has proven to be quite relevant in differentiating dysarthric from normative speech (Gómez-Vilda et al., 2017a).

4.3 Supporting methods

4.3.1 Performance evaluation

The main objective of this study regarding voice and articulation quality analysis is to create a methodological framework for establishing comparisons to be applied in the evaluation of dysarthria treatment performance, as it is directly linked with comparing the specific estimations produced during point-like assessment sessions within two possible scenarios: transversal or longitudinal. In the first case, features are evaluated against a statistical data framework produced on estimations obtained from normative or control sets of participants (inter-participant). In the second case, features are evaluated against a statistical data framework obtained from previous recording sessions out of the same participant (intra-participant).

The most important question when conducting data evaluation is the experimental design, which must start as a search for an answer to a specific research question, therefore, a working hypothesis must be set forth before formulating and answering this question. Participants potentially affected by the disorder under study provide the target dataset. Participants unaffected by the disorder will provide the control dataset. Consequently, an experimental sample set must be defined, including target and control participants, establishing the inclusion and exclusion criteria for such.

Control participants may be drawn from the same demographic conditions as target participants. Sometimes, a third dataset provided by normative participants is also included in the study as a golden standard. Normative participants are subjects not only unaffected by the disorder but also presenting the best performing features, according to their age and health conditions, usually drawn from mid-age adults in good general health conditions regarding inclusion and exclusion criteria.

Another important consideration to be taken into account is the specific statistical framework for the evaluations to be carried on. In this case, the methodological issues to be taken into account for the specificity of the analysis are the type of features to be evaluated, either continuous or categorical character, the data sample size, and the transversal or longitudinal character of the evaluation sessions.

4.3.2 Three-way comparisons

The usual methodology for detecting pathological behaviour is based on the comparison of a given set of features from a target participant against two distributions, namely from an assumed pathological dataset, and from a control dataset. This methodology presents important undesirable side effects, as the control dataset is usually integrated by participants matched in age. In the case of PD, the main incidence for the first diagnosis is in the interval from 55 to 65 years old, the peak of the distribution being around 65-75 years old. This means that the speech of age-matched controls may be affected by ageing, and potential effects in phonation, articulation, and fluency might be expected. The introduction of a normative database from the mid-age participants has been proposed for studies on phonation (see subsection 6.1 and Appendix I.7) to grant a background reference database to compensate for the co-morbidity effects of ageing on phonation.

This methodology is standard in the forensic analysis of speech (Taroni et al. 2006, Gómez-Vilda et al., 2012), where the samples under assessment are named "the targets", the control reference samples are named "the suspects", and the reference normative database is named the "normative background model". The determination of the logarithmic likelihood of a target sample being attributed to a given suspect sample is based on the probability of matching the target with the suspect distribution relative to the probability of the target concerning the normative background model. As Gaussian distributions are defined by exponentials of normalized distances between samples and averages, their logarithms may be associated with these compensated distances. Therefore, the comparisons are based on three distances, the distance from the target dataset to the normative dataset, the distance of the target dataset to the control dataset, and the distance between the control dataset concerning the normative dataset. In this way, the alterations potentially affecting the control dataset due to ageing are also taken into account in the study, to compensate for co-morbidity effects. This situation is illustrated in the representation of Figure 4.18.

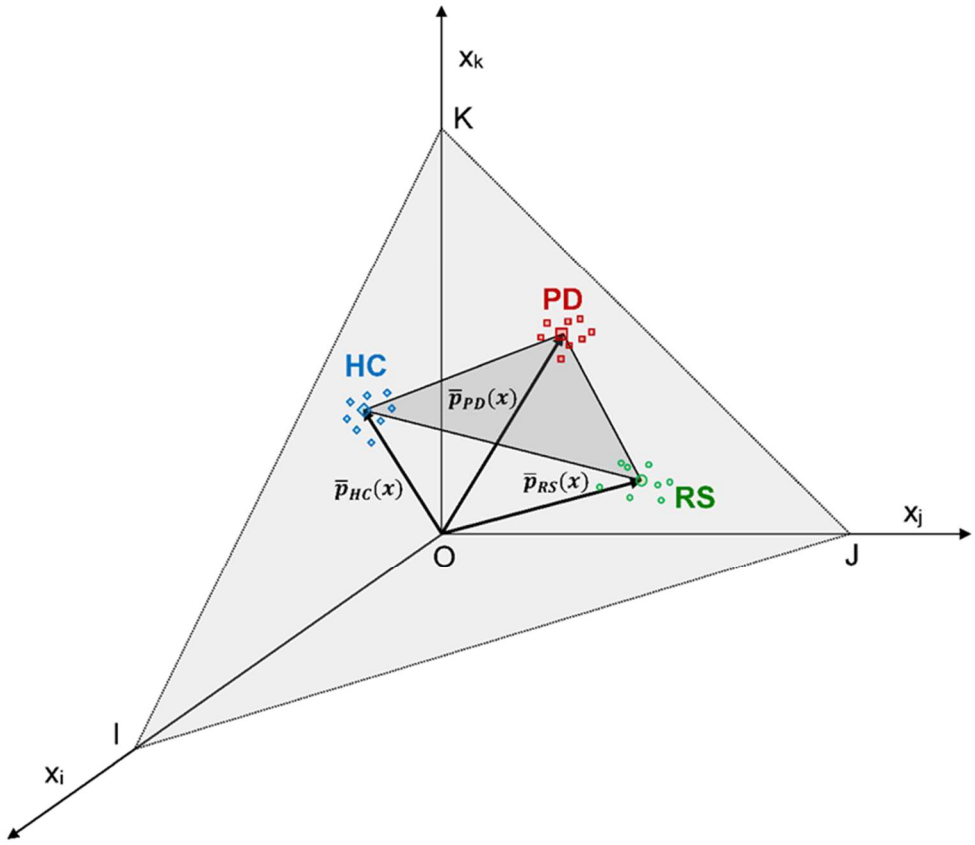

Figure 4.18 Synoptic representation of three-way comparisons on the PD-HC-RS plane. Description: The sample comparisons defined in the manifold {x} (three-dimensional in this example) may be reduced to the plane where the triangle HC-PD-RS is defined. PD: Parkinson's Disease dataset; HC: age-matched healthy control dataset; RS: normative background reference dataset.

In a multidimensional manifold of features $\{\mathbf{x}_i\}$ (in this case exemplified by a three-dimensional case, for easy visualization) a target dataset PD represented by its centroid $\bar{p}_{PD}$ is to be compared against a control dataset HC represented by its centroid $\bar{p}_{HC}$, relative to the reference (normative) dataset RS, represented by its centroid $\bar{p}_{RS}$. Independently of the dimensionality of the feature space, these three centroids will define a triangle within a plane. The decision to assign the target sample to the control sample or to the normative sample depends on the distance between the control and the normative samples.

If this distance is large, it would imply that the controls are potentially affected by co-morbidity factors, therefore, the decision favouring a non-pathological classification would be safely taken only if the distance from the target to the normative samples is smaller than the distance from the target to the control set. If the distance from the control sample to the normative sample is small the decision could be taken as a simpler two-band comparison between the target and the control datasets.

4.3.3 Statistical analysis

The following approaches were used, depending on each specific analysis:

- Age-matching between the HC and PD datasets was assessed using the non-parametric Mann-Whitney U-test on the null hypothesis of the same medians between same-gender populations.
- The descriptive statistics (degrees of freedom, mean, standard deviations, three quartiles, skewness, and kurtosis) of each JSD dataset (MNS, MHS, and MPD for males, and FNS, FHC, and FPD for females) are estimated. These statistics give a clear picture of each distribution regarding normality.
- Besides, each JSD dataset has been tested for normality using Lilliefors tests (Abdi and Molin, 2007) including Monte Carlo simulations granting a p-value standard error under 0.001, to complement the descriptive statistics before mentioned.
- The JSD datasets are compared in pairs to determine their differentiation capability as a complementary study to the classification. The null hypothesis of equal means between each distribution pair under comparison was evaluated on two-tailed tests for Student's-t (parametric) and for Kolmogorov-Smirnov (non-parametric), and on equal medians for Mann-Whitney U (non-parametric).

4.3.4 Mutual information

Regarding the studies included in the present manuscript, the classification methods proposed used either SVMs or the MI between paired GFADs taken as probability densities $p(x)$ and $q(x)$, defined in the positive part of the real axis ($x \geq 0$), as the glottal flow is a positive definite function. The normalized MI between two given probability density functions may be estimated by Jensen-Shannon Divergence (JSD), see Lin (1991), Endres and Schindelin (2003), and Cover and Thomas (2012):

$$D_{JS} = \frac{D_{KL}\left(p(x)\middle|m(x)\right) + D_{KL}\left(q(x)\middle|m(x)\right)}{2}; \quad m(x) = \frac{p(x) + q(x)}{2} \tag{4.45}$$

The variable x represents the normalized GF amplitude ($0 \leq x \leq 1$) and D_{KL} is the Kulback-Leibler Divergence between the two distributions (Salicrú et al., 1994; Georgiou and Lindquist, 2003), defined as:

$$D_{KL}(p(x)|q(x)) = \int_0^{\infty} p(x) abs\left\{log\frac{p(x)}{q(x)}\right\} dx \tag{4.46}$$

Jensen-Shannon's Divergence is symmetrical for $p(x)$ and $q(x)$: ($D_{JS}\{p(x), q(x)\} = D_{JS}\{q(x), p(x)\}$), and it is normalized on the interval [0, 1].

4.3.5 Classification methods

Three different classification methods are used in this study. The first method is based on the three-way comparison of JSDs from each target sample *i*, to the respective average distributions (see subsection 4.3.2):

$$\begin{aligned} D_{i|MNS} &= D_{JS}\left(p_i(x)\middle|p_{MNS}(x)\right); \\ D_{i|MHC} &= D_{JS}\left(p_i(x)\middle|p_{MHC}(x)\right); \\ D_{i|MPD} &= D_{JS}\left(p_i(x)\middle|p_{MPD}(x)\right) \\ D_{i|FNS} &= D_{JS}\left(p_i(x)\middle|p_{FNS}(x)\right); \\ D_{i|FHC} &= D_{JS}\left(p_i(x)\middle|p_{FHC}(x)\right); \\ D_{i|FPD} &= D_{JS}\left(p_i(x)\middle|p_{FPD}(x)\right) \end{aligned} \tag{4.47}$$

where $p_{MNS}(x)$, $p_{MHC}(x)$, $p_{MPD}(x)$, $p_{FNS}(x)$, $p_{FHC}(x)$ and $p_{FPD}(x)$ are the averages of the male and female distributions regarding their respective data sets. $D_{i|MNS}$, $D_{i|MHC}$, and $D_{i|MPD}$ are the distances of the GFAD of sample i to each average GFAD of the male datasets. $D_{i|FNS}$, $D_{i|FHC}$, and $D_{i|FPD}$ are the respective distances to the female datasets. A naïve Bisector Criterion (BiCr), will classify a subject i as non-pathological if

$$\begin{aligned} D_{i|MPD} &> D_{i|MNS}; male\ subsets \\ D_{i|FPD} &> D_{i|FNS}; female\ subsets \end{aligned} \tag{4.48}$$

and as pathological otherwise.

The second method is based on the hierarchical clustering (HiCl) of each sample's JSD concerning the PD and NS average distributions. Hierarchical clustering is an unsupervised methodology based on ordering a dataset of observations using a dissimilarity measure between all pairs of observations (James et al., 2013, pp. 390-396).

The input to HiCl for the male datasets is the n-tuple $\{D_{i|MNS}, D_{i|MHC}, D_{i|MPD}\}$, and in the female datasets the corresponding n-tuple is $\{D_{i|FNS}, D_{i|FHC}$ and $D_{i|FPD}\}$. The types of linkage used are the shortest distance (simple) and farthest distance (complete). The number of clusters is fixed adaptively, to produce the following orderings:

- Male Dataset: $Z_m = \{\text{HiCl}(D_{i|MPD}, D_{i|MNS})\}$
- Female Dataset: $Z_f = \{\text{HiCl}(D_{i|FPD}, D_{i|FNS})\}$

where Z_m and Z_f are the indices classifying each subset member as non-pathological (CN) or pathological (CP). The clustering results are presented as dendrograms. The classification methods BiCr and HiCl are summarized in Figure 4.19.a.

The third classification method is based on the GFAD as a main feature, and an SVM as a classifier to distinguish between PD and HC phonation (see Álvarez-Marquina et al., 2020), according to the following protocol (see Figure 4.19.b):

- Feature selection using ReliefF (Robnik-Šikonja and Kononenko, 2003) was applied to each GFAD dataset (MPD vs MHC in males, and FPD vs FHC in females). In the male datasets, the GFADs of the MPD set are confronted with the MHC set: $\{p_i|p_{MPD}|p_i|p_{MCH}\}$. In the female datasets, FPD GFADs are confronted with the FHC: $\{p_i|p_{FPD}|p_i|p_{FCH}\}$. The number of neighbours was varied between 1 and 50. Subsets of N features ($15 \leq N \leq 120$) were selected in descending order following the ranking provided by ReliefF.

- Afterwards, each feature subset pair (MPD vs MHC in males, and FPD vs FHC in females) was fed to an SVM with a Gaussian radial basis function kernel (RBF) following Chang and Lin (2011). Cross-validation was used on all the datasets distributed in 10 groups (10-fold cross-validation) testing the combinations of the SVM grid space parameters (**C**, **γ**), as C=[2^{-3}, 2^{-2},..., 2^{12}] and γ = [2^{-1}, 2^{-2}, ..., 2^{-10}]. Sensitivity (STV), specificity (SPC), and accuracy (ACC) for each best-performing subset of *N* features were estimated on the average of 1000 runs over the 10 groups (1000 runs of the 10-fold cross-validation).

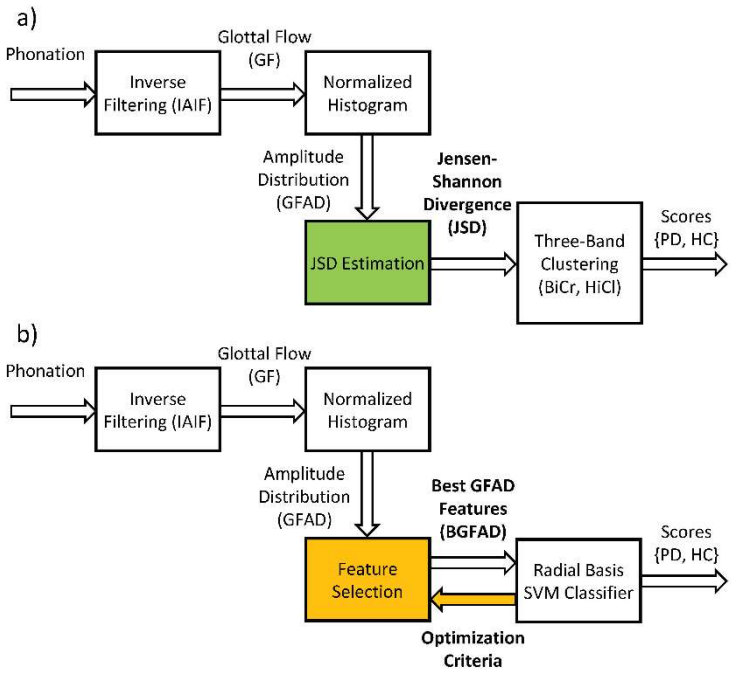

Figure 4.19 Comparison between data flows in hierarchical clustering and SVM classification. Description: a) hierarchical clustering and b) SVM classification. The main difference between both methods is the use of JSD (green box) in BiCr or HiCl or specific features selected from the GFAD (yellow box) producing the reduced feature subset BGFAD used in SVM classification. Whereas JSD preserves information contents, BGFAD is a reductionist version of GFAD in terms of information contents.

In all cases, the classification performance was estimated according to the standard definition for STV, SPC, and ACC, as well as the factor F1[11]:

$$\begin{aligned} STV &= \frac{TP}{P} = \frac{TP}{TP + FN} \\ SPC &= \frac{TN}{N} = \frac{TN}{TN + FP} \\ ACC &= \frac{(TP + TN)}{TP + FP + TN + FN} \\ F1 &= \frac{2\,TP}{2\,TP + FP + FN} \end{aligned} \tag{4.49}$$

where TP is true positives, TN is true negatives, FP is false positives and FN is false negatives.

4.3.6 Transversal vs longitudinal assessment

The specificities of data analytics in the medical field are quite different than in other areas. The main axiomatic consideration regarding this field can be summarized in the sentence "There are no diseases, only sick individuals" on the best tradition of Richard Koch's ethics of Medicine (Töpfer and Wiesing, 2005a and 2005b). According to this observation, «Diagnosis is not the assignment of a term of a species to a patient's disease: this would not do justice to the individuality of a clinical manifestation and would fail to provide a reason for individual therapy. Nevertheless, the terms assigned to diseases, although fictitious, are not useless, but assist in differentiating various phenomena» Therefore, the orientation of statistical studies must be quite different either if the target is centered on the disease, or if it is centred on the patient.

[11] The F1 score is a single metric that takes into account both precision and recall (sensitivity) of a diagnostic test or model. It is the harmonic mean of precision and recall, providing a balanced evaluation when there is an uneven class distribution. The F1 score ranges from 0 to 1, with 1 being the best possible value (perfect precision and recall).

This dichotomy is addressed by Rose (2001) as «Aetiology confronts two distinct issues: the determinants of individual cases, and the determinants of incidence rate. If exposure to a necessary agent is homogeneous within a population, then case/control and cohort methods will fail to detect it: they will only identify markers of susceptibility. The corresponding strategies in control are the "high-risk" approach, which seeks to protect susceptible individuals, and the population approach, which seeks to control the causes of incidence». The approaches to the study of PD disorder should find two different orientations: the statistical comparison of common features from a population of affected participants with the same features from a cohort of non-affected control individuals (transversal approach), and the statistical comparison of specific features known to be meaningful and statistically relevant from affected participants each on their own at different stages of the disorder progress in the timeline (longitudinal approach). The transversal approach should be designed to provide information on common features altered as a result of the disease for instance, in establishing the susceptibility (disorder-centred). The longitudinal approach should be designed to monitor the progress of the disorder on each patient taken individually, either because of pharmacological, surgical, or rehabilitative treatment (patient-centred). The selection of paired control participants and normative references is critical in the transversal approach, whereas it is less meaningful in the case of longitudinal studies. The transversal approach requires the design of the normative and control sets, in the sense that the normative set is to be configured by mid-age participants and declared free of any pathological profile after clinical inspection. The number of participants should be large enough to fulfil the limit theorem.

The paired control set would require a minimum similar number of participants in the age range of the PD participants, which means that the influence of potential co-morbidities due to ageing will be present in both sets, making it difficult to establish clear cuts between some meaning features, which will affect their statistical relevance. Inclusion and exclusion criteria and demographic descriptions will constitute a necessary requirement to provide a clear framework for the study. A strong condition regarding transversal studies is gender separation because phonation and speech are behavioural traits highly impacted by the impact of sexual hormones, therefore, gender separation is mandatory because the larynx is the second more sexually hormone-affected structure in the human body (Hertrich and Ackermann, 1995; Abitbol, Abitbol, and Abitbol, 1999; Inamoto et al., 2015; Cirillo et al., 2020). The longitudinal approach will require a large number of estimations from the same participant in the timeline, which means frequent inspections and exhausting recording sessions, the normative and control sets being irrelevant, as well as demographic information. The statistical methods associated with the study are also different in both approaches, regarding transversal studies, parametrical tests being acceptable once the conditions of normativity are assessed on the features involved. If this would not be the case, non-parametric approaches could be used instead. In the case of longitudinal studies, having in mind that the availability of frequent tests in the timeline is cumbersome and exhausting for persons affected by reduced mobility, the use of mutual information tests for pairwise comparison would be preferable.

Hypothesis casting is also strongly dependent on the character of both approaches. In the case of transversal studies, the susceptibility of a given set of features concerning environmental, social, ageing, or other grouping conditions is a clear objective. In the case of longitudinal studies, the susceptibility of a given set of features concerning treatment protocol specificities would be the main target.

4.3.7 Specific character of the studies presented

The final objective of any feature classification study is to provide the highest performance scores regarding sensibility, specificity, and accuracy in the classification tasks. Not being far from that intentionality, the orientation of the present studies is essentially different, as they are oriented to explore deeper aspects of the inner explicability, interpretability, and causability beyond mere performance score assessment.

In this sense, it is to be recognized that there has been a lot of interest in developing signal-processing approaches to mine speech data, extract dysarthria and dysphonia measures, and employ statistical machine-learning algorithms for biomedical speech applications (Tsanas, 2013; Brabenec et al., 2017; Arora and Tsanas, 2021; Arora et al., 2021; Tsanas, Little, and Ramig, 2021). However, to a large extent, these studies do not provide the same level of insights that mechanistic models can provide, i.e., models that build on the physical principles of voice production to characterize the underlying vocal production mechanism and PD-related pathology (Duffy, 2013). Exploring a mechanistic model can provide new insights into the underlying physiological processes, which in turn might inspire further signal processing algorithms for the characterization of speech signals. The present exploratory studies are a first step towards describing sustained vowel phonation and specific diadochokinetic tests recorded from participants with PD (PwP), age-matched Healthy Control participants (HC), and normative mid-age Reference Participants (RSPs) in terms of the Glottal Flow Amplitude Distributions (GFAD), Vocal Fold Body Stress (VFBS), and Absolute Kinematic Velocity (AKV) to gain an insight on possible different behavioural properties depending on age, disorder, and gender, would allow a future wider study on more populated databases and classification methods.

To achieve this, specific estimations of the GFAD, VFBS, and AKV in terms of amplitude distributions of tensor and relaxer NMA are used as potential predictors of dysarthric behaviour during the utterance of sustained vowel and diadochokinetic phonations from PwP and HCs. This would open the opportunity of using phonation in regular clinical practice to help in better detection, assessment, and monitoring of PD and other neurodegenerative disorders, as well as in advancing neurolinguistics studies.

CHAPTER 5

5 Experimental studies based on phonation and articulation models

This chapter is devoted to the application of the methods described in the previous section to several studies regarding PD characterization, providing an exemplification of the methods introduced previously and how they can be used to characterize or represent the manifestation of PD. In this examination, the first characterization of PD speech is based on the perspective of glottal biomechanics. This approach seeks to isolate the distinctive phonation patterns associated with PD, shedding light on how the biomechanics of the vocal folds are altered in PwP. In a second study, the APM deciphers the subtleties of articulatory speech movements in PD in the upper jaw, tongue, and pharynx structures responsible for controlling the airflow through the upper respiratory cavities. Together, these studies offer a comprehensive understanding of PD speech from two different perspectives: biomechanics and articulatory behaviour.

5.1 Characterization of PD speech from glottal biomechanics

This subsection delineates two distinct approaches to characterizing phonation in individuals with PD. This chapter is part of the work published in Gómez-Rodellar et al. (2019c, 2020a, and 2023). All the databases used and processed can be referred to in Appendix I. The outcomes of these studies will be presented in subsection 6.1.

There are two ways glottal biomechanics have been characterized. The first approach is based on the GFAD (Glottal Flow Amplitude Distribution), such as the glottal flow and source.

The second approach is based on the EEG-band decomposition of the VFBS (Vocal Fold Body Stress) to differentiate PD phonation from age- and gender-matched healthy controls (HC). These approaches put the focus on the observation of amplitude properties within the GS pattern. As introduced in prior sections, this analysis yields insights into the coordination of muscles involved in vocal cord tension control. By understanding the interplay between the vocal fold tensor (cricothyroid) and relaxer (thyroarytenoid) muscles a distinct speaker profile can be constructed. Using divergence metrics, it is possible to calculate the relative difference between the centroid of a reference group and an individual entry, allowing for a differentiating quantifiable metric.

5.2 Study based on the GFAD

The GS and GF can be understood as the pressure build-up increase in the larynx and the airflow release through the glottis during phonation, respectively. These signals are quasi-periodic, meaning that a similar repeating pattern is present in each phonation cycle, although not identical (the time interval between two neighbouring pulsations of the vocal folds). Variations between periods both in amplitude and in frequency are a measure of stability and regularity which are of great interest for their semantic potential. It is expected that a stable and regular phonation would produce similar signals in neighbouring phonation cycles, assuming they are affected by the same production condition. The more similar the phonation cycles are to each other, the more regular and stable the phonation will be. The alterations to phonation may be of two conditions: either presenting jitter (alterations of the phonation cycle duration), or shimmer (alterations of the signal amplitude), or both of them. Other perturbations affecting the shape of the glottal signals, will appear as deformations on the theoretical L-F time-amplitude pattern.

These characteristics are exploited to distinguish altered from normative phonation using GFAD-based divergence metrics such as Jensen-Shannon's (Lin, 1991, Cover and Thomas, 2012). This metric is based on information theory, providing a higher score the less similar the probability density functions of the signals under test are.

5.2.1 Dataset description

This study is based on the use of two different datasets. The first one is a subset of a speech database (PARCZ) collected at St. Anne's University Hospital in Brno (Czech Republic), containing recordings produced by PD patients of both genders. This database contains also speech recordings and demographic information from age-matched HC subjects. The details of the participants included in the study and the recording conditions and protocol are described in depth in Appendix I.1 and Appendix I.2. A brief summary is provided here to maintain the connecting narrative.

The second dataset is a subset of a speech database (HUGMM) containing vowel emissions from normative participants, collected at Hospital Universitario Gregorio Marañón of Madrid (HUGMM), and it is described in detail in Appendix I.6.

The data selected for the study were recordings of the sustained vowel [a:] from 24 male and 24 female PD participants from the PARCZ database. The age distribution of the PD male dataset was between 49 and 78 years (mean=67.4, standard deviation=9.1). The PD female dataset included participants between 49 and 78 years (mean=66.6, standard deviation=7.2).

Similarly, recordings of the sustained vowel [a:] from 24 male and 24 female HC participants selected from PARCZ. The HC male dataset included participants between 49 and 83 years old (mean = 65.1, standard deviation = 8.9). The HC female dataset included participants between 49 and 78 years old (mean = 62.7, standard deviation = 9.1). The specific demographic and clinical details of both datasets are given in Table App. 1 (Appendix I.2).

In the same way, sustained vowel utterances of [a:] from 24 male and 24 female participants were selected from the HUGMM database (NS male and female datasets). The NS male dataset included participants between 21 and 62 years old (mean = 42.3, standard deviation = 11.2). The NS female dataset included participants between 20 and 59 years old (mean=37.3, standard deviation=11.7). The biometrical information from NS participants is given in Table App. 2 (Appendix I.4).

The statistical significance of the difference between the age ranges of PD vs HC male and female datasets was evaluated under the null hypothesis of equal means using the non-parametric Mann-Whitney U test. The null hypothesis could not be rejected under a 0.05 level, casting p-values of 0.307 and 0.074 respectively for the male and female datasets. The rejection in the case of the female dataset (PD vs HC) is slightly over the limit of 0.05, this being one of the potential study limitations.

The protocol and instrumentation details of the recordings included in the PARCZ subset are described in Appendix I.1. Similarly, the corresponding details from the HUGMM subset have been included in Appendix I.6.

The quality and compatibility of recordings and acoustic analysis were carefully evaluated directly on the glottal signal estimates generated from recordings of both databases, in terms of signal-to-noise levels, bandwidth, and saturation limits to grant the degree of compatibility of the experimental methodology used in the study. No relevant factors affecting the reliability of the study were found.

5.2.2 Methodological procedures

The following study is based on the GFAD of each vowel utterance. These distributions are estimated by inverse filtering as described in the following steps:

- Recordings of the vowel [a:] from Male Normative Subjects (MNS), Female Normative Subjects (FNS), Male Healthy Control (MHC), Female Healthy Control (FHC), Male Parkison's Disease (Male Parkinson's Disease (MPD), and Female Parkinson's Disease (FPD). Each of these samples is low-pass filtered (antialiasing) and down-sampled to 16 kHz. This sampling rate preserves most of the frequency contents of glottal signals, which are known to be under 8 kHz, and equalises the effects of the recording platforms and settings granting comparative signal quality standards between both databases.

- The ONPT transfer function is evaluated by a 20-pole inverse adaptive lattice-ladder filter (Deller, Proakis, and Hansen, 1993) based on iterative adaptive inverse filtering (IAIF: Alku et al., 2019). The lattice structure reconstructs the tube chain, the propagating waves, and the reflection coefficients of the chain, and the modified Rothenberg's model explains the non-linear source-filter interaction. The ONPT transfer function is removed from the spectral contents of the speech signal (inversion), producing a filtering residual. A description of the inversion filter details can be found in Gómez-Vilda et al. (2009).

- The inverse filtering residual is integrated twice to estimate the GS and GF as $p(x_l, t)$, and $u(x_g, t)$, respectively.

- The GFAD is estimated by a normalized 50-bin amplitude histogram of the GF ($H_i = h_i/\Sigma_i h_i$, where h is the amplitude histogram and i is the bin number).

- Once the histogram for each participant is extracted each element can be compared using the divergence metrics defined in subsection 4.3.4, according to expression (4.45). Each data subset (MNS, FNS, MHC, FHC, MPD, and FPD) is clustered around a centroid, estimated as the average histogram of the individual samples within each group. Each instance can be compared to this set of centroids and given a score according to the divergence metric, acting as a distance from the individual instance to the cluster of a particular group (4.47), being: $p_{MNS}(x)$, $p_{MHC}(x)$, $p_{MPD}(x)$, $p_{FNS}(x)$, $p_{FHC}(x)$ and $p_{FPD}(x)$ the divergence between each individual entry to the respective reference centroids. Each one of these clusters provides a position on a hyperplane where each instance is positioned. It is of relevance to mention that all instances were separated according to sex and processed accordingly.

- Using these divergences as pseudo-distances it is possible to provide an estimation for the proximity of a given instance to a group, or as the likelihood of belonging to a particular subset. Once this quantitative element is extracted, several classification methods can be used, as has been commented on in subsection 4.3.5. The simplest classifier is a naïve BiCr which labels each instance with the category of lowest divergence according to expression (4.48). Another classification method is based on the HiCl of each sample's JSD concerning the PD and NS average distributions, ordering the dataset using a dissimilarity measure between all pairs of observations providing a relationship between groups of instances. A third method is based on an SVM classifier to distinguish between PD and HC phonation classifies each instance according to each of these categories. The results of this study are presented in subsection 6.1.1.

5.3 Study based on the VFBS

5.3.1 Working hypothesis

The VFBS may be seen as a correlate of the biomechanical tension on the *musculus vocalis* resulting from laryngeal NMA. The VFBS is cycle-synchronously estimated following subsection 4.1.9, therefore, it is expected that a stable and regular phonation would produce similar VFBS estimates on neighbouring phonation cycles. The more neighbouring estimations resemble each other, the more regular and stable the phonation will be. A very important property of VFBS is its direct relationship with the NMA governing laryngeal nerves (see subsection 2.4.1). This circumstance is exploited to distinguish altered from normative phonation using the EEG-band-related description of the VFBS and Jensen-Shannon divergence (see subsections 4.3.4 and 6.1.2).

This description splits the VFBS profile into each corresponding EEG band. The rationale behind this is that the VFBS is a direct result of NMA, influenced by the tension and stretching of larynx muscles that condition the patterns of the glottal source. Since NMA is the outcome of brain activity, some residual information in the EEG-band range would necessarily be present in the GS, leaving an imprint on the phonation signal. This relationship between the NMA and the VFBS would be based on Cortical Muscle Coupling (CMC: Brown et al., 2009). The GS produces the Vocal Sound Pressure (VSP) that propagates through the ONPT; therefore, a simplified model of the conceptual connection chain can be defined as NMA→VFBS→GS→VSP.

It is a well-known fact that alterations in NMA coherence, as observed through the analysis of the EEG bands, may be present in motor-related disorders (Salenius et al., 2002; Feng et al., 2021; Barios et al., 2021). The working assumption is that alterations in the NMA might be in part modelled from VSP using model inversion, as reverting the before mentioned conceptual connection chain: VSP→GS→VFBS→NMA. This possibility would benefit from ongoing studies on CMC, as it can be used to explore the coupling relationship of different functional frequency bands (Gao et al., 2018; Colamarino et al., 2021).

Therefore, this study proposed using long-lasting utterances of an open vowel as [a:] to conduct the reconstruction of the way-back path from VSP to EEG-related frequency bands to produce functional phonation descriptions in terms of NMA, and to use them in detecting functional phonation improvements after rTMS stimulation on the frontotemporal gyrus, an area of the brain responsible for acoustic proprioception. The rationale being that by stimulating such area local neuron populations, responsible for controlling and monitoring sound perception, become more active leading to better speech production as they have a stronger influence on premotor areas.

The use of maintained phonations of [a:] in speech pathology studies is well suited for functional phonation evaluation, this fact being recognized by its wide application in clinical practice, given that slight variations in NMA will be immediately reflected in GS, and thus, easily tracked and monitored.

Having fixed the conducting narrative justifying the reconstruction process from VSP to NMA, it should be decided which frequency bands would be of higher interest to further explore their application. It would seem reasonable to focus on the activity in the ϑ- and γ-bands following the description of the nonlinear character of motor unit recruitment in muscular agonist-antagonist activation in Darbin and Montgomery (2022), taking into account the direct neuromotor pathways involving the cortex-thalamus-BG (Ctx-Th-BG) circuitry activating the cricothyroid and thyroarytenoid muscles, and the projection of the organized oscillators that explain the ultimate nonlinear character of motor unit activity involved in the laryngeal function, because the ϑ-band seems to be strongly related to unstable nonlinear NMA (Solomon et al., 2017), whereas, on the other hand, the ϑ-γ coupling should be related to the Ctx-Th-BG activity (Aguilera et al., 2022).

A primary objective of the study is to evaluate the functional competence of phonation in TMS participants using the simplified signal model inversion to reconstruct the way-back path as VSP→GS→VFBS→NMA. A secondary objective is based on the existence of strong relationships between muscular contraction under biomechanical drive and neuromotor EEG activity on the brain areas responsible for premotor and motor control (Chiang, Wang, and McKeown, 2012; Gao et al., 2018; Manríquez et al, 2019) in the sense that the laryngeal motor activity is controlled by larynx muscles, inducing the contraction of the thyroarytenoid muscle, estimated on the unbiased VFBS using nonlinear projection actuators transforming neural discharges into muscle contraction.

This modelling would allow reverse system inversion, provided that adequate actuator-inverse operators could be designed based on system identification methodologies. Therefore, it would be possible to advance in the projection of the NMA estimated from phonation biomechanics over the brain area activity, measured by the EEG.

According to the primary objective, the present work is intended to assess the validity of features estimated on the glottal neuromechanics to characterize the stability of phonation in pre-stimulus and post-stimulus vocal emissions from a limited set of PD participants submitted to repetitive transcranial magnetic stimulation (rTMS), using a methodology aligned with the secondary objective.

5.3.2 Dataset description

This study has been conducted on data provided by the Applied Neuroscience Research Group, CEITEC, Masaryk University, Brno, Czech Republic, from PwP participants showing mild to moderate HD directly related to PD, following a program of rTMS, as described in Appendix I.3 and Appendix I.4. All were on stable dopaminergic medication for the duration of the whole study. The study included cases recorded at pre-stimulus and after four post-stimulus sessions spaced in time. This reduced the number of cases to 18 out of the 33 participants included in the original database. Half the participants received an active stimulation, and half submitted to a sham stimulation (see Appendix I.3 for more details). The demographic description of the participants is given in Table 5.1 and Table App. 2 (Appendix I.4). The cohort distributions are broadly similar in terms of UPDRS grade (females: 16.6±4.1; males: 12.3±3.9) and age (females: 74.6±3.0; males: 69.7±8.4).

Table 5.1 Participants' demographic and clinical data.

A: active stimulation; S: sham stimulation; F: Female; M: Male; Y: years. UPDRS-III: Unified Parkinson Disease Rating Scale, section III (motor section).

PwP code (pre)	Active/Sham	Gender	Age (Y)	UPDRS-III
0100	A	F	71	10
0800	A	M	58	9
1100	A	M	73	14
1200	A	M	72	21
1400	A	M	64	10
1600	S	F	79	20
1700	S	M	70	16
1800	S	M	61	9
1900	S	M	77	8
2000	A	F	76	28
2200	S	M	66	13
2300	S	M	55	7
2400	S	M	72	10
2500	S	M	81	14
2600	S	F	73	16
2700	A	M	77	14
2800	A	M	80	15
2900	A	F	74	17

The recording protocol established that each participant had to go through a baseline assessment (pre-stimulus evaluation at the first session: T0) before being submitted to ten stimulation sessions (stimulation process) within two weeks; a follow-up evaluation session two weeks after stimulation (post-stimulus at T1); additional follow-up evaluations around six weeks (post-stimulus at T2), and around ten weeks (post-stimulus at T3). The 18 participants of the subset included in the study submitted also to a fourth post-treatment evaluation session around fourteen weeks after the stimulation process (post-stimulus at T4). The evaluation dates are listed in Table App. 3 (Appendix I.5). Each participant in the study was randomly assigned to active or sham stimulation. Audio recordings of several spoken tests from each participant were taken before (pre-stimulus) and after (post-stimulus).

5.3.3 Methodological procedures

The study was conducted by taking 4 s fragments of selected long emissions of a sustained vowel [a:] (see Appendix I.5), because this interval, selected 2 s after the vowel onset to avoid phonation start transients, provides the maximum duration to prevent fatigue effects on phonation (Manríquez et al., 2019). The stimulation protocol and speech recording conditions are described in detail by Brabenec et al. (2021). The processing methods for the estimation of NMA descriptions consisting of EEG-aligned frequency bands of the biomechanical vocal fold stiffness estimated from speech recordings will be commented on in what follows:

- Fragments of 4 s long from the recordings of the vowel [a:], at a sampling rate of 16 kHz were selected between the time instants at 2 and 6 s, to skip potential vowel onset and decay effects. This sampling rate preserves most of the frequency contents of glottal signals.
- The ONPT transfer function was evaluated by a 24-pole inverse adaptive lattice-ladder filter. The size of the filter was fixed using the Akaike's criterion (Cavanaugh and Neath, 2019) by 1.5 times the sampling frequency divided by 1000, for a sampling frequency of 16 kHz, the size of the filter set 24. The adaptive lattice-ladder inverse filter estimates a prediction-error polynomial reducing the speech segment being analysed $s(n)$ to a residual $r(n)$ by classical deconvolution as $r(n) = h_T(n) * s(n)$, where $h_T(n)$ is the impulse response of the prediction-error polynomial emulating the inverse transfer function of the ONPT, such that $H_T(\omega) = \sum_i h_{Ti} e^{-ji\omega}$, where $j = \sqrt{-1}$, and ω is the angular frequency, therefore the lattice structure reconstructs the tube chain structure of the ONPT, and its associated transfer function is removed from the spectral contents of the speech signal (Gómez-Vilda et al. (2009).

- The GS was estimated in pitch-synchronous cycles (Naylor et al., 2007) by numerically integrating the inverse filter residual.
- The VFBS (ζ), given in N.m^{-1} was estimated from the spectral tilt of the GS adjusted on a 2-mass model of the vocal fold biomechanics (Gómez-Vilda et al., 2009; Meghraoui et al., 2021).
- The VFBS was de-biased and de-trended by a moving-average filter.

The working hypothesis established that VFBS is the direct consequence of the NMA of the cricothyroid and thyroarytenoid muscles. To sustain a given stable phonation frequency F0, a delicate equilibrium between both activations is necessary (Brown et al., 2009). This equilibrium is represented by an average baseline value of VFBS (trend). Oscillations around this trend would reproduce small-signal alterations of the NMA, therefore, a de-trended VFBS would produce a good estimate of neuromotor instability of agonist-antagonist misadjustment. To obtain a frequency-band description of neuromotor instability, the de-trended VFBS was processed by a bank of fifth-order Butterworth band-pass filters tuned at the respective EEG-related frequency bands, (δ: $f \leq 4$Hz; ϑ: 4Hz$<f\leq$8Hz; α: 8Hz$\leq f\leq$16Hz; β: 16Hz$<f\leq$32Hz; γ: $f>$32Hz; μ: 8Hz$<f\leq$12Hz). As a result, a set of de-trended vocal fold stiffness frequency-band time signals $\xi_{ij}^{k}(n)$ is produced, where i=(0,..., I) is the evaluation session index (I=4), j=(1,…, J) is the participant index (J=18), and k=(1,…, K) is the frequency band index (K=6), pointing to the six frequency bands defined above, and n is the time index.

5.3.4 Data analysis

An estimation of the VFBS was produced using the methods described in subsection 4.1.6, as it is assumed to be directly related to the activity of neuromotor areas responsible for laryngeal control during phonation. This signal has been decomposed in frequency bands $\xi_{ij}^{k}(n)$ corresponding to EEG ϑ-γ activity following the working hypothesis defined in subsection 5.1.2.1. The amplitude distributions of each EEG-related frequency are estimated from their histograms. Distributions from post-stimulus recordings were compared with the corresponding ones from pre-stimulus conditions, and two methods were designed to produce explainable interpretations of potential behavioural changes in the phonation function.

The comparison methods proposed were based on log-likelihood ratios and hypothesis tests. The log-likelihood ratio between two given probability density functions (pdfs) $p_i(\xi)$ and $p_j(\xi)$ can be defined as:

$$\lambda(p_i|p_j) = \int_{\xi\epsilon\Omega} log\{p_i(\xi)/p_j(\xi)\}d\xi\,; \qquad (5.1)$$

where Ω is the estimation interval. Besides estimating the similarity between two distributions, this ratio provides the sense of the comparison. Assuming a longitudinal character of the study, the distributions under study $p_i(\xi)$ correspond to the post-stimulus phonation to be compared to the pre-stimulus phonation $p_0(\xi)$, therefore, two possible situations are to be considered, i.e. either $p_i(\xi)$ is more dispersed than $p_0(\xi)$, such as $\mathrm{Var}\{p_i(\xi)\} > \mathrm{Var}\{p_0(\xi)\}$, or that $p_i(\xi)$ is less dispersed than $p_0(\xi)$, such as $\mathrm{Var}\{p_i(\xi)\} < \mathrm{Var}\{p_0(\xi)\}$, where $\mathrm{Var}\{.\}$ is the variance of the distribution.

In the first case, let's assume $\xi \in \Omega_b$ to be the interval where $p_o(\xi) > p_i(\xi)$, and $\xi \in \Omega_a$ to be the interval where $p_i(\xi) > p_o(\xi)$, given that $\Omega_a \subset \Omega$ and $\Omega_b \subset \Omega$ provided that $\Omega_a \cup \Omega_b = \Omega$, and $\Omega_a \cap \Omega_b = \emptyset$. Therefore, from (61(5.1)) it will follow that

$$\lambda(p_i|p_o) = \int_{\xi \epsilon \Omega_a} log\{p_i(\xi)/p_o(\xi)\}d\xi + \int_{\xi \epsilon \Omega_b} log\{p_i(\xi)/p_o(\xi)\}d\xi \qquad (5.2)$$

where:

$$\begin{aligned} I_a &= \int_{\xi \epsilon \Omega_a} log\{p_i(\xi)/p_o(\xi)\}d\xi < 0; \\ I_b &= \int_{\xi \epsilon \Omega_b} log\{p_i(\xi)/p_o(\xi)\}d\xi > 0; \end{aligned} \qquad (5.3)$$

It may be seen that the sign of both integrals is the opposite, therefore it will be expected that in the case of unimodal distributions, when there is a strong difference in variance $\mathrm{Var}\{p_i(\xi)\} \gg \mathrm{Var}\{p_o(\xi)\}$, $p_o(\xi)$ will be much narrower than $p_i(\xi)$, and consequently Ω_b will be much smaller in size than Ω_a, and $\lambda(p_i|p_o) = I_a + I_b < 0$. Conversely, when $\mathrm{Var}\{p_i(\xi)\} \ll \mathrm{Var}\{p_o(\xi)\}$ the opposite condition will prevail and $\lambda(p_i|p_o) > 0$. This is especially evident in normal and quasi-normal distributions.

To put it otherwise, given the properties of probability densities, if $p_i(\xi) > p_o(\xi)$ on the interval $\xi \in \Omega_a$, it will be most likely expected for $p_i(\xi)$ to be narrower (lower variance) than $p_o(\xi)$, or in other words, that the generating process of $p_i(\xi)$ would produce less dispersed outcomes (corresponding to a more stable feature) than that of $p_o(\xi)$. Therefore, it would be reasonable to expect that functional improvements in phonation will produce lower variance post-stimulus frequency band distributions, and therefore, positive log-likelihood ratios, and on the contrary, worsening phonation conditions will produce negative log-likelihood ratios.

In their turn, the significance of pre- and post-stimulus feature distributions was assessed by Mann-Whitney U-tests on the null hypothesis of equal medians, because typically $p_i(\xi)$ distribution patterns might differ from normality. The tests were carried out on EEG-related band-frequency feature samples $\xi_{ij}^k(n)$, therefore, hypothesis tests for a given participant *j* and a given feature *k* would be conducted on each post-stimulus sample (*i*=1…I) concerning the pre-stimulus one (*i*=0), with I=4 being the number of post-stimulus recordings, as:

$$h_{jk} = T_{MW}\{\xi_{ij}^k(n), \xi_{0j}^k(n)\} \tag{5.4}$$

where $T_{MW}\{\xi_{ij}^k(n), \xi_{0j}^k(n)\}$ is the Mann-Whitney U-test between pre-stimulus and post-stimulus samples $\xi_{0j}^k(n)$, and $\xi_{ij}^k(n)$, h_{jk} being the p-value estimated by the test, assuming that ergodic conditions apply (Papoulis, 1991).

The above-described methods allow determining the behaviour of each frequency band based on log-likelihood estimations (5.1) to explore whether these features improve significantly as a result of the intervention by rTMS, with special attention to which frequency bands would be more sensitive to changes in functional behaviour. To get an overview of the general behaviour of potential improvements regarding a specific participant, a global score would be more suitable than a partial one. In this case, the averages over time of each feature sample $\langle\xi_{ij}^k(n)\rangle$ were used to assess functional improvements. As $\xi_{ij}^k(n)$ is an unbiased estimation of vocal fold stiffness, it may be associated with the tremor (oscillating instability) of the vocal fold around a given trend. Its amplitude is expected to be larger the more acute the functional disorder affection of each participant's phonation. Therefore, lower values of $\langle\xi_{ij}^k(n)\rangle$ will be associated with a less unstable phonation, and with larger stimulation beneficial effects.

The following definition was used as a normalized weighted score associated with each frequency band feature per evaluation session:

$$c_{ij}^{k} = \frac{\langle \xi_{ij}^{k}(n) \rangle - \langle \xi_{0j}^{k}(n) \rangle}{\langle \xi_{0j}^{k}(n) \rangle} w_i \tag{5.5}$$

where the weight $w_i = (d_i - d_0)/(d_I - d_0)$ is a normalizing factor to take into account the time interval between each post-stimulus date d_i and the corresponding pre-stimulus date d_0 normalized to the longest interval $(d_I - d_0)$ in days. In this way, long-lasting beneficial effects were given larger importance than short-duration effects. As it may be easily checked, improving phonation features produces negative scores $c_{ij}^{k} < 0$.

Another relevant score was defined on the progression trend of potential phonation improvements, based on the first difference of average estimations between successive recordings, as:

$$d_{ij}^{k} = \langle \xi_{ij}^{k}(n) \rangle - \langle \xi_{i-1j}^{k}(n) \rangle \tag{5.6}$$

therefore, progressive improvements in a given feature would produce also negative scores $d_{ij}^{k} < 0$. Both scores, for relative and progressive improvements, were fused into a single score as $g_{ij}^{k} = c_{ij}^{k} + d_{ij}^{k}$. The total score per participant was defined as the average of all post-stimulus evaluation sessions and all frequency bands:

$$g_j = \frac{1}{(I-1) \cdot K} \sum_{i=1}^{I} \sum_{k=1}^{k} g_{ij}^{k} \tag{5.7}$$

where *I*-1 is the number of post-stimulus evaluation sessions (four in this study), and *K* is the number of frequency bands considered (five in the present case, from δ to γ).

5.4 Characterization of PD speech by the articulation projection model (APM)

5.4.1 Experimental description

In this study on articulation, the biomechanical system of the jaw-tongue defined in subsection 4.2.1 will be used for the characterization of PD HD, to estimate the neuromotor behaviour of the system, and provide specific markers of proper or improper NMA during vowel utterances. This dataset includes speech, accelerometry, and sEMG data from eight Spanish native speakers (four males and four females, stage 2 on H&Y scale) who were recruited from a PD patient association in the metropolitan area of Madrid (Asociación de Pacientes de Parkinson de Alcorcón y Móstoles, APARKAM), as well as from eight HC age-paired volunteers (four males and four females) participating in the study. Figure 5.1 shows the instrumentation used in the simultaneous recording of speech, chin acceleration, and sEMG. The sEMG on the masseter was recorded, as well as the acceleration on the chin, simultaneously with the speech signal during the utterance of specific diadochokinetic exercises, as commented in Appendix I.9.

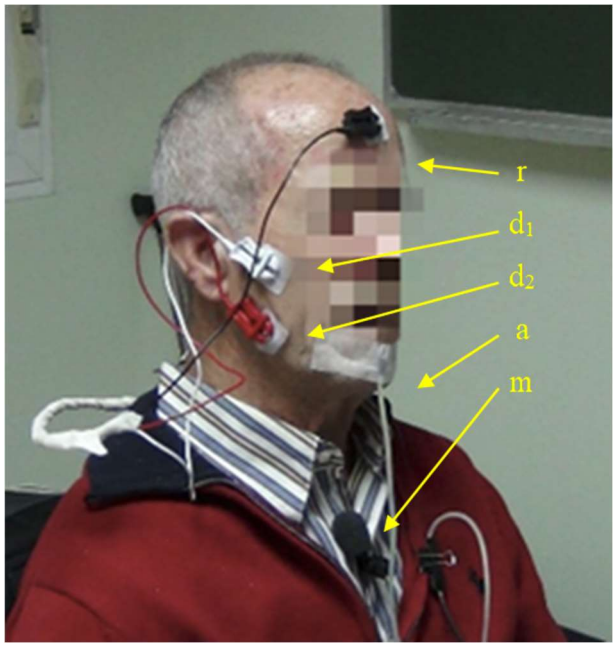

Figure 5.1 Signal acquisition set-up for speech, sEMG, and 3DAcc.
Description: Two sEMG electrodes are placed on the longitudinal ends of the masseter (differential pair: d1, d2) and one on the forehead (reference: r). The 3D accelerometer is fixed to the chin (a). A cardioid clip microphone is attached to the collar (m).

The selection of the masseter as the target muscular structure obeys the following reasons: it is a powerful muscle developing a strong sEMG when contracting, it is accessible (beneath the caudal section of the cheek), it may modify strongly the oral cavity when contracting or relaxing leaving a clear acoustic signature in formants, and its biomechanical activity is well understood. Further details on the demographic information relative to the participants as well as on instrumentation and protocol are provided in Table App. 5 (Appendix I.8).

In the present study only the recordings from the utterance of the diphthong […a→i→a→i...] were used in the estimation of **W** in expression (4.29) by multiple regression because this diphthong produces the widest sweeps of formant dynamical patterns associated to the high-low and forward-backwards displacement of P_{rJT}.

This decision is justified because this sequence presents the advantage of being mainly controlled by the action of the masseter, a powerful muscle producing clear sEMG recordings.

5.4.2 Projecting formant dynamics onto articulation kinematics

The accurate estimation of the first two formant displacements as expressed in (4.30) is essential to the study. The procedures used in formant estimation are based on adaptive linear prediction as described in subsections 4.1.6 and 4.2.2, built on a previous in-depth study (Gómez-Vilda et al., 2019b). The details of format estimation are briefly described as follows:

- The speech signal $\boldsymbol{x}$(n) was bandlimited (low-pass filtered) to 4 kHz by a 4-th order Butterworth filter.
- The speech was divided into consecutive segment windows of 64 ms separated on a 2 ms stride. A Hamming window was used.
- A radiation-compensation first-order high-pass filter with a drop-off coefficient of 0.6 was used to remove radiation effects (see subsections 3.2.2 and 4.1.2).
- The glottal formant was eliminated by a first-order inverse lattice filter (see subsections 3.2.1 and 4.1.3).
- A ninth-order inverse lattice filter was used to estimate the error-predictor polynomial $H_K(z)$.
- The roots z_k of the error predictor polynomial $H_K(z = z_k) = 0$ were estimated.
- The formants were obtained from the positive angles of z_k: $F_k = f_s\,\varphi_k/\pi;\ \varphi_k > 0/\pi$, ($f_s$: sampling frequency).

- The root modules were used as a quality factor for formant selection: $r_k = 0 < |z_k| < 1$.

The purpose of the model described in subsection 4.2 is to allow indirect estimation of the spatial oscillations $\Delta\mathbf{S}$ solely from the dynamics of the recorded signal and the acoustic formants $\Delta\mathbf{F}$ by an acoustic-to-kinematic model described by its weight matrix $\mathbf{W}$ defined in (4.29), which will be the main objective of this study:

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \tag{5.8}$$

The individual weight values w_{ij} are to be estimated from healthy controls and PD participants, to establish possible regression models on the kinematic variables associated with the reference point P_{rJT} exclusively from acoustic estimates ($\Delta\mathbf{F}$), in other words, to establish a methodology for estimating articulatory kinematic features solely from the acoustic speech signal. The methodology proposed is based on solving for the model weights $\mathbf{W}$ using standard optimisation methods to establish the relationship between the observed variables $\Delta\mathbf{S}$ and $\Delta\mathbf{F}$. The problem may be formulated as the minimisation of the cost function C

$$C = \|\Delta\mathbf{S} - \mathbf{W} \times \Delta\mathbf{F}\|^2 \; subject\ to\ \mathbf{W}_{est} = argmin_{\mathbf{W}}\{C\} \tag{5.9}$$

where $\|\cdot\|^2$ denotes the module of a vector.

Given the structural properties of C, the estimation of $\mathbf{W}$ may be decomposed in the independent minimisation of each of its separate components ($C = C_1 + C_2$):

$$C_i(w_{i1}, w_{i2}) = \|\Delta \boldsymbol{s}_i - w_{i1}\Delta\mathbf{F}_1 - w_{i2}\Delta\mathbf{F}_2\|^2; i = 1, 2 \tag{5.10}$$

If expression (5.10) is expanded, it can be easily observed that the partial cost functions depend only on a single row of the matrix $\mathbf{W}$ (5.8). Therefore, the error minimisation problem can be split into minimising each of the partial error functions $C_{\mathrm{i}}(w_{\mathrm{i1}}, w_{\mathrm{i2}})$ for the weights w_{i1} and w_{i2}.

$$[w_{i1}, w_{i2}] = argmin_{[w_{i1}, w_{i2}]}\{C_i\} \tag{5.11}$$

The minimisation methodology is based on an iteration using a gradient descent procedure with a variable step size to estimate each individual weight as

$$w_{ij}^k = w_{ij}^{k-1} - \gamma_i^{k-1}\boldsymbol{\nabla}_i C_i^{k-1};\ i, j = 1, 2 \tag{5.12}$$

where k is the iteration step which can be estimated by the Barzilai–Borwein method (Barzilai and Borwein, 1988)

$$\gamma_i^k = \frac{\langle \boldsymbol{w}_i^k - \boldsymbol{w}_i^{k-1}, \boldsymbol{\nabla}_i C_i^k - \boldsymbol{\nabla}_i C_i^{k-1}\rangle}{\left\|\boldsymbol{\nabla}_i C_i^k - \boldsymbol{\nabla}_i C_i^{k-1}\right\|^2} \tag{5.13}$$

where the weight and the gradient vectors are defined as

$$\begin{aligned} \boldsymbol{w}_i^k &= \left[w_{i1}^k, w_{i2}^k\right] \\ \boldsymbol{\nabla}_i &= \left[\frac{\partial}{\partial w_{i1}}, \frac{\partial}{\partial w_{i2}}\right] \end{aligned} \tag{5.14}$$

Practical convergence is reached after a few iteration steps.

The initial estimation (step k=0) for the weights $\mathbf{W}^0$ is achieved using simple linear regression (James et al., 2017) between the input and output signals of the inverse model:

$$\mathbf{W}^0 = \begin{bmatrix} \boldsymbol{w}_1^0 \\ \boldsymbol{w}_2^0 \end{bmatrix} = \begin{bmatrix} w_{11}^0 & w_{12}^0 \\ w_{21}^0 & w_{22}^0 \end{bmatrix} \tag{5.15}$$

$$w_{ij}^0 = R_{ij}(\Delta\mathbf{s}_i, \Delta\mathbf{F}_j) = \frac{\Delta\mathbf{s}_i \Delta\mathbf{F}_j^T}{\Delta\mathbf{F}_j \Delta\mathbf{F}_j^T};\ i,j = 1,2 \tag{5.16}$$

This estimation process is represented in the regression plots shown in Figure 6.11. It was noticed during the estimation of the initial weights that there seems to be a misalignment between $\Delta\mathbf{s}_\mathrm{i}$ and $\Delta\mathbf{F}_\mathrm{j}$. To improve the estimation of the model matrix $\mathbf{W}$ it may be interesting to reduce this misalignment by maximising the correlation function R_{ij} after introducing a time shift. The relative misalignment may be a consequence of formant insertion dynamics associated with resonance in tubes with losses (this assumption needs further study), and it results in a non-optimal estimation of $\mathbf{W}$. To compensate it, each weight may be re-estimated after the realignment of signals derived from the following optimisation problem:

$$C_i(z) = \|\Delta\mathbf{s}_i(z) - w_{i1}\Delta\mathbf{F}_1(z)z^{-n_{i1}} - w_{i2}\Delta\mathbf{F}_2(z)z^{-n_{i2}}\|^2;\ i = 1,2 \tag{5.17}$$

where $\Delta\mathbf{s}_i(z)$ and $\Delta\mathbf{F}_j$(z) are the z-transforms of $\Delta\mathbf{s}_i$ and $\Delta\mathbf{F}_j$ and n_{i1}, and n_{i2}, are the relative misalignments between each of the components of $\Delta\mathbf{S}$ and $\Delta\mathbf{F}$, given in numbers of samples, assumed to be independent of each other. Similarly, a solution to the problem in (5.8) is sought as:

$$[n_{i1}, n_{i2}]_{min} = argmin_{[n_{i1},n_{i2}]}\{C_i(z)\}; i = 1,2 \tag{5.18}$$

The independent minimisation of $C_1(z)$ and $C_2(z)$ allows estimating the misalignments on ensuring that:

$$n_{ij} = argmax_{n_{ij}}\{|\Delta \mathbf{s}_i(n - n_{ij})\Delta \mathbf{F}_j(n)|\} \tag{5.19}$$

The alignment fitness may be evaluated using the root mean square error between the real displacement and the value predicted from regression for each weight w_{ij} as:

$$\varepsilon_{ij} = \frac{\|e_{ij}\|}{\|\Delta \mathbf{s}_i\|};\ e_{ij} = \Delta \mathbf{s}_i - w_{ij}\Delta \mathbf{F}_j;\ i,j = 1,2 \tag{5.20}$$

The results of applying these methods to the dataset described in Appendix I.8 are presented in Chapter 6 and discussed in Chapter 7.

CHAPTER 6

6 Results

The outcomes of the studies conducted in the previous chapter are presented in the following sections along with a description of the results, from the overview of the methods, outcomes, and estimations presented in Chapter 5.

6.1 Dysphonia assessment based on the GFAD

The results presented in this subsection correspond to the experimental design described in subsection 5.1.1. The aim of this study focused on utilizing the GFAD as a long-term time-domain average feature of the glottal signals to characterize the pathological phonation resulting from PD HD. To achieve this, the GFAD was extracted from a participant's sample listed in Table App. 1 (Appendix I.2). The J-S divergence was estimated following the procedure described in subsection 5.1.1. The average GFADs of the male and female datasets (avMNS, avMHC, avMPD, avFNS, avFHC, and avFPD) are represented in Figure 6.1 as an illustrative example. It may be seen that normative distributions are accumulated mainly on both extreme amplitude values, whereas the PD and HC datasets tend to concentrate on mid-amplitude values. The JSD between the average GFADs of the PD, HC, and NS datasets, from(4.45) and (4.46) are given in Table 6.1 Relative JSD between each pair of average GAFDs.

Table 6.1 Relative JSD between each pair of average GAFDs.

Distance between average datasets	JSD
$D_{MPD\|MNS}$	0.204
$D_{MHC\|MNS}$	0.211
$D_{MPD\|MHC}$	0.051
$D_{FPD\|FNS}$	0.262
$D_{FHC\|FNS}$	0.299
$D_{FPD\|FHC}$	0.066

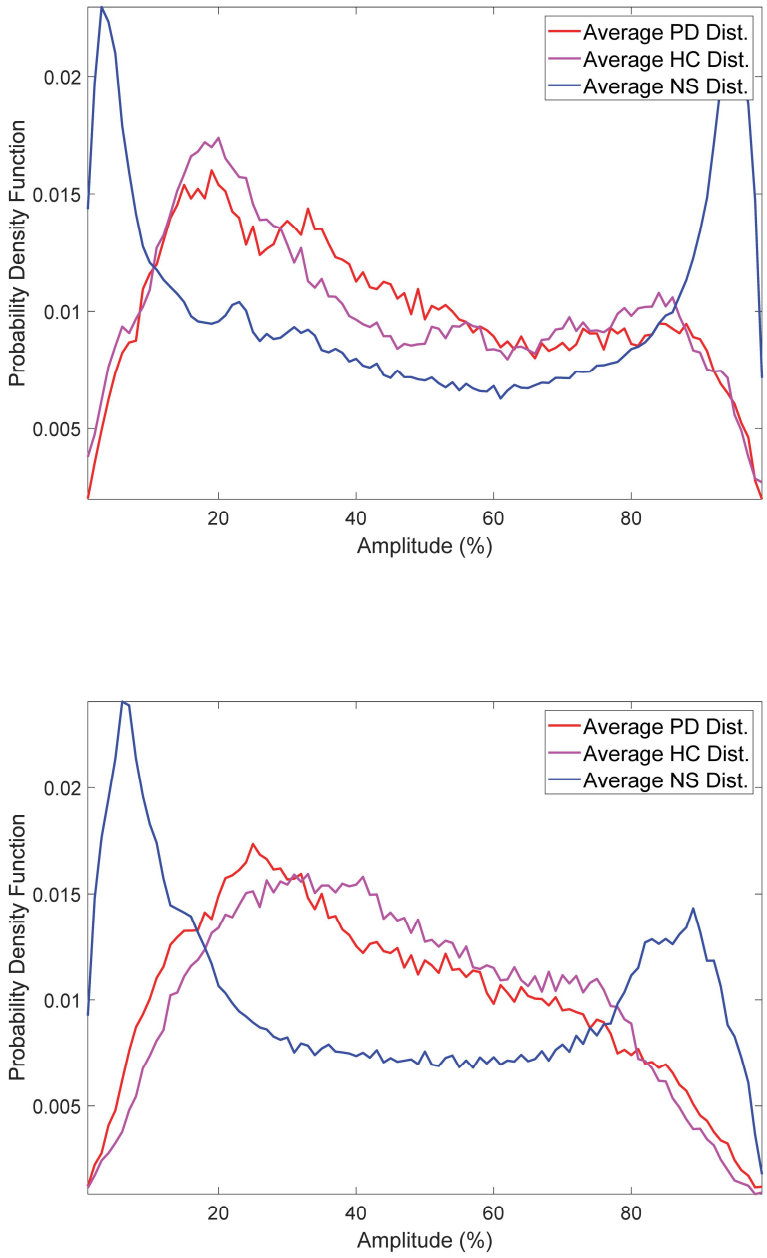

Figure 6.1 Average GFADs of the male and female datasets.
Description: Top: average GFAD of the male datasets in terms of the normalized GF amplitude: PD (red line), HC (purple line), and NS (blue line). Bottom: Idem of the female datasets.

The JSD between each participant's sample GFAD and each dataset's average distribution is given in Table App. 6 (Appendix II).

6.1.1 Three-way comparisons

The procedure implemented for deciding on the normal or altered phonation condition in this case was based on the fundamentals exposed in subsections 4.3.2 and 4.3.5. The JSDs estimated for each sample in the study as given in Table App. 6 (Appendix II) are represented on the PD-HC-RS plane as in Figure 6.2. These representations make use of an algebraic property allowing to project a larger-dimension manifold on lower dimensions. This approach represents an N sample histogram vector as a single point on the PD-HC-RS plane. The point coordinates are given by the JSD of this vector concerning the reference HC-RS ones. In this way, each discrete GFAD distribution given by a 50-dimension vector is represented as a position in the representation plane. This method considers that the temporal GAFD signal is ordered by amplitude positions, and the resulting histogram vector provides a profile of how frequently a given amplitude position is seen along a given phonation cycle, estimating the quasi-likelihood of finding the signal in that point. In the representation of Figure 6.2 built following the methodology explained in subsection 4.3.2, two vectors are used as a reference, corresponding to the distance between the averages of the subsets PD and NS ($D_{MPD|MNS}$ and $D_{FPD|FNS}$) represented by a segment on the abscise. This segment and the two segments corresponding to the distance from each sample *i* to the avPD and the avNS define a triangle (three-distance comparisons), the average distance from avPD to avNS ($D_{MPD|MNS}$ and $D_{FPD|FNS}$) being one of its sides. Each sample *i* defined by a tuple $\{D_{i|MNS}, D_{i|MPD}\}$, or $\{D_{i|FNS}, D_{i|MFD}\}$ will be labelled with a different symbol depending on its respective subset (NS: green balls, HC: blue diamonds, PD: red squares). In this way, any sample can be displayed as an individual position in a 2D plot on the PD-HC-RS plane. The distances between each sample and the respective avNS and avPD centroids determine which samples are farther apart from the respective averages.

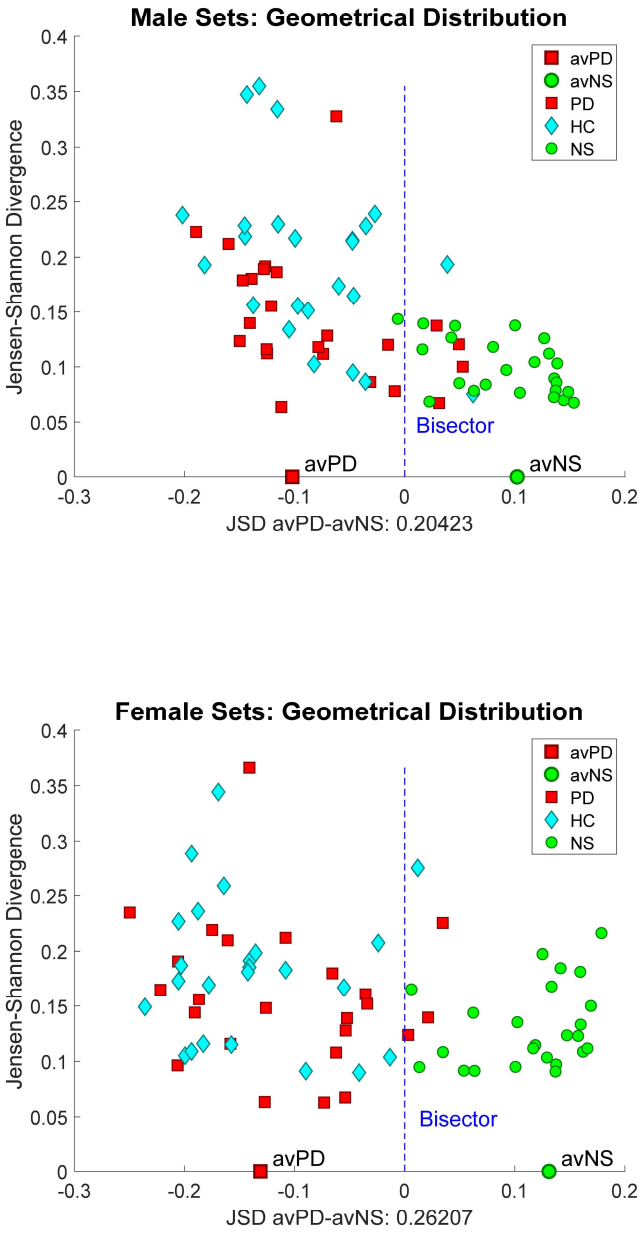

Figure 6.2 Three-distance plots of each PD sample for male HC and NS averages. Description: a) JSD to avHC and avNS. b) Idem for the female sets. Red squares: PD samples. Blue diamonds: HC samples. Green bullets: NS samples. Blue dash line: equidistance bisector.

In this way, each sample is plotted on the left or right-hand side of the bisector orthogonal to the abscise through the midpoint between avPD and avNS (blue-dash line). The classification of each sample as pathological or non-pathological follows the expressions (4.48), comparing the sample JSD to the avPD with its respective one to the avNS, as

given in Table App. 6 (Appendix II). The confusion matrices produced using this classification method (HiCl) are given in Table 6.2.

Table 6.2 Confusion matrices for the male and female subsets separated by the bisector criterion.

MCP and FCP: male and female samples classified as pathological. MCN and FCN: male and female samples classified as non-pathological. Male non-pathological dataset: MNS∪MHC. Male pathological dataset: MPD. Female non-pathological dataset: FNS∪FHC. Female pathological dataset: FPD. TP: true positives; TN: true negatives; FP: false positives; FN: false negatives.

Cla.\Sets	**MNS**	**MHC**	**MPD**
MCP	1 (FP)	22 (FP)	20 (TP)
MCN	23 (TN)	2 (TN)	4 (FN)
Cla.\Sets	**FNS**	**FHC**	**FPD**
FCP	0 (FP)	23 (FP)	21 (TP)
FCN	24 (TN)	1 (TN)	3 (FN)

The table is to be read as classifications (Cla.) vs datasets (Sets). For instance, the results in row MCP are classified as pathological 1 sample from MNS (false positive), 22 samples from MHC (false positives), and 20 samples from MNS (true positives). Subsequent rows can be interpreted the same way.

6.1.2 Clustering analysis and threshold

As introduced in section 4.3.5, data points can be grouped into clusters according to some distance or proximity metric. In this setting the JSD of each sample to each subset centroid is taken as the proximity metric of each sample to the rest. The classification of each sample into pathological or non-pathological is based on the definition of a threshold level to select those branches of the hierarchical tree best aligned with any one of the options. In doing so, two possibilities were considered: *single* (minimal inter-cluster dissimilarity) and *complete* (maximal inter-cluster dissimilarity). Regarding the combination of JSDs, the following settings were considered: *raw* (no combination), *diff* (subtraction of the distances to avPD and avNS), or *reldiff* (subtraction of each pair of distances to avPD and avNS divided by their sum).

The possibility of referencing the study to avNS vs avHC was also tested. The number of clusters considered was $\{3 \leq n \leq 8\}$, the types of linkage tested were {*single*, *complete*}, the combinations tested were {*raw*, *diff*, *reldiff*}, and the reference settings {*avNS vs avPD*, *avNS vs avHC*} were also checked. The best results from 6 x 2 x 3 x 2 = 72 combinations per gender are presented in Figure 6.3 and Figure 6.4.

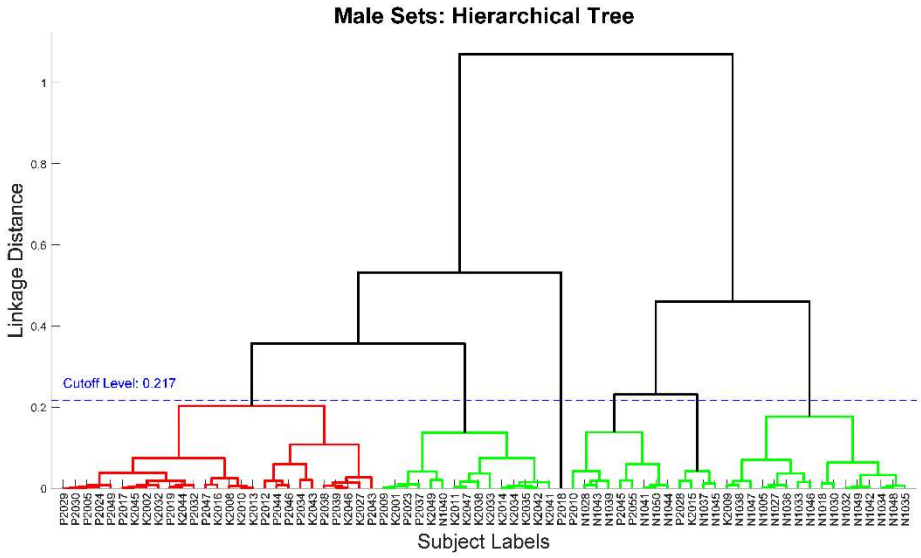

Figure 6.3 Hierarchical tree of male speaker samples classified by the three-distance JSD to their average normative and pathological distributions.
Description: green: non-pathological clusters, red pathological clusters. The vertical axis plots the linkage distance between the cluster and sub-cluster members. The horizontal axis gives each participant code. The linkage threshold level (cutoff) is shown in dash blue. Configuration settings: six clusters, *complete* linkage, and *reldiff* combination. Sample labels: N: normative; K: control; P: pathological.

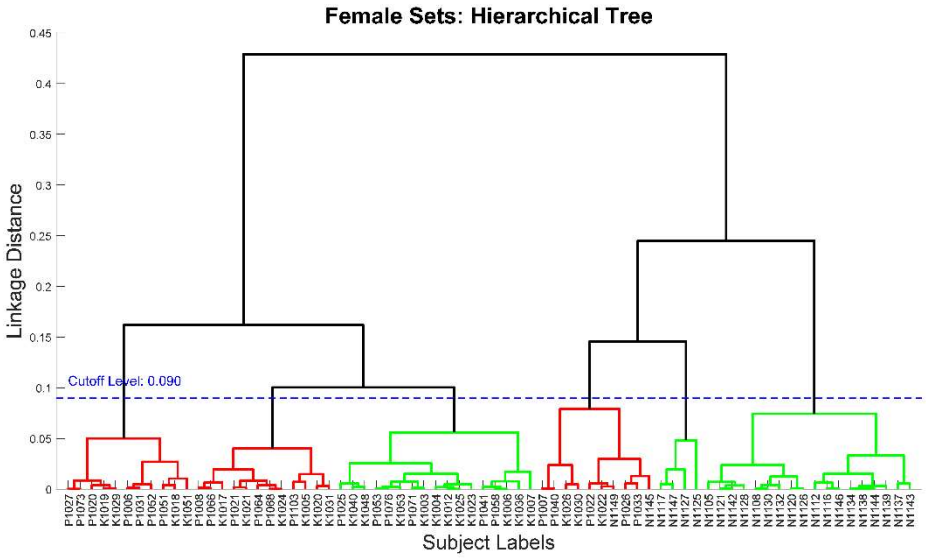

Figure 6.4 Hierarchical tree of female speaker samples classified by the three-distance JSD to their average normative and pathological distributions.
Description: green: non-pathological clusters, red pathological clusters. The vertical axis plots the linkage distance between the cluster and sub-cluster members. The horizontal axis gives the code of each participant. The linkage threshold level (cutoff) is shown in dash blue. Configuration settings: six clusters, complete linkage, and diff combination. Sample labels: N: normative; K: control; P: pathological.

The confusion matrices from hierarchical clustering are given in Table 6.3.

Table 6.3 Confusion matrices for the male and female subsets separated by hierarchical clustering.

MCP and FCP: male and female samples classified as pathological. MCN and FCN: male and female samples classified as non-pathological. Male non-pathological dataset: MNS∪MHC. Male pathological dataset: MPD. Female non-pathological dataset: FNS∪FHC. Female pathological dataset: FPD. TP: true positives; TN: true negatives; FP: false positives; FN: false negatives.

Cla.\Sets	**MNS**	**MHC**	**MPD**
MCP	0 (FP)	21 (FP)	17 (TP)
MCN	24 (TN)	3 (TN)	7 (FN)
Cla.\Sets	**FNS**	**FHC**	**FPD**
FCP	2 (FP)	13 (FP)	18 (TP)
FCN	22 (TN)	11 (TN)	6 (FN)

6.1.3 Differentiation of PD from HC using an SVM classifier

The hierarchical clustering method is not a reliable classification tool, showing difficulty in the separation of PD from HC datasets, because as ageing induces similar effects to PD on the glottal signals there is substantial overlap between groups. To investigate if there are other more efficient separation methodologies, an SVM-based classifier was proposed, as described in subsection 4.3.5. The results of the SVM classification corresponding to the subset of N features producing the best scores in terms of STV, SPC, and ACC are given in Table 6.4.

Table 6.4 SVM classification scores (%)

MPDvsMHC: between male PD and HC subsets. FPDvsFHC: Idem between female PD and HC subsets. STV: sensitivity. SPC: Specificity. ACC: Accuracy.

Gender	**Datasets**	**STV**	**SPC**	**ACC**
Males	MPDvsMHC	93.4	96.1	94.8
Females	FPDvsFHC	97.8	86.6	92.2

6.1.4 Comparing PD, HC, and NS by their respective JSDs

The results produced by BiCr and HiCl show that HC samples are not much different from PD samples in terms of their JSD to the avNS and avPD references. This finding is to be interpreted in the sense that HC GFADs may be as irregular as PD ones.

The question now is how different the JSD estimates of each dataset are in statistical terms to allow subset separation under statistical relevance. As commented in subsection 4.3.3 the first requirement is to assess the normality of the distributions to be tested. This condition can be inferred from the descriptive statistics of each dataset given in Table 6.5.

Table 6.5 Descriptive statistics of JSD from each dataset.

df: degrees of freedom; μ: mean; Q1: first quartile; Q2: second quartile (median); Q3: third quartile; skw: skewness; kur: kurtosis.

Dataset	df	μ	Q1	Q2	Q3	skw	kur
MNS	24	0.11	0.09	0.10	0.13	0.85	2.73
MHC	24	0.28	0.23	0.26	0.33	-0.01	2.83
MPD	24	0.24	0.16	0.25	0.30	-0.20	1.99
FNS	24	0.14	0.11	0.13	0.16	0.67	2.39
FHC	24	0.33	0.30	0.33	0.39	-0.47	2.77
FPD	24	0.29	0.22	0.28	0.36	0.33	1.97

Besides, each dataset has been checked for normality using the two-sided Lilliefors test (Abdi and Molin, 2007) on the null hypothesis of normality at a significance level of 0.05 using a Monte Carlo simulation granting a standard error under 0.001. The results of the tests are given in Table 6.6.

Table 6.6 Results from Lilliefors test on normality.

df: degrees of freedom; Statistic: value under test; p-value: the probability of erroneously rejecting the null hypothesis; RNH: null hypothesis reject/no reject, Y/N.

Dataset	df	Statistic	p-value	RNH
MNS	24	0.157	0.104	N
MHC	24	0.109	0.632	N
MPD	24	0.122	0.446	N
FNS	24	0.159	0.110	N
FHC	24	0.143	0.221	N
FPD	24	0.151	0.152	N

MHC and MPD are the data sets more aligned with the normality hypothesis, whereas MNS and FNS are the less aligned ones. Having these conditions in mind, the results of confronting each dataset JSD to each other under the hypothesis of equal means (two-tails) are given in Table 6.7.

Table 6.7 Estimated p-values from inter-subset two-tail tests.

t-St: Student's t-test; KS: Kolmogorov-Smirnov; MW: Mann-Whitney. The cases not rejecting the null hypothesis at α=0.05 are in bold. MPD vs MNS: pathological vs normative (males). MHC vs MNS: control vs normative (males). FPD vs FHC: pathological vs control (males). FPD vs FNS: pathological vs normative (females). FHC vs FNS: control vs normative (females). FPD vs FHC: pathological vs control (females).

Datasets	t-St	KS	MW
MPD vs MNS	<0.001	<0.001	<0.001
MHC vs MNS	<0.001	<0.001	<0.001
MPD vs MHC	**0.101**	**0.622**	**0.220**
FPD vs FNS	<0.001	<0.001	<0.001
FHC vs FNS	<0.001	<0.001	<0.001
FPD vs FHC	**0.101**	**0.109**	**0.110**

It may be seen that the three tests reject the null hypothesis comparing PD and NS male and female datasets. Similarly, the null hypothesis is rejected by the three tests in the case of HC and NS datasets, both for males and females. However, none of the three tests rejected the null hypothesis in comparing HC and PD datasets. These results, together with the difficulties found by BiCr and HiCl to differentiate HC and PD datasets, allow us to conclude that both HC and PD present more similarity between themselves than their NS counterparts. A possible explanation for this finding may be the effect of medication, provided that the PD participants were all in an ON state by dopaminergic effects at the time that recordings were taken. Besides, ageing would induce another approaching effect of HC to PD as far as phonation is concerned. To take into account the effects of medication, a correlation study between LED and JSD estimations per participant could be conducted, given that the Levodopa Equivalent Dose (LED) administered to each patient was known (see the LED column in Table App. 1, Appendix I.2). The results of correlating LED and JSD of PD patients (Spearman) were $\rho = -0.22$ (males) and $\rho = -0.33$ (females). These findings suggest that medication might have a mild tendency to enhance phonation similarities between individuals with Parkinson's disease (PD) and healthy controls (HC).

Table 6.8 summarizes the results from the three classification methods presented to have a general impression for a concluding comparison. The SVM methodology produces the best separation results, as it is commented in subsection 7.1.

Table 6.8 Classification scores (%).

BiCr: Bisector Criterion. HiCl: Hierarchical Clustering. SVM: Support Vector Machine. The results from BiCr and HiCl are produced from the recounts of TP, FP, TN, and FN cases given in Table 6.2 and Table 6.3 after applying (4.49). The results from the SVM classification are directly taken from Table 6.4.

Method	Datasets	STV	SPC	ACC
BiCr	MPDvsMNS	83.3	52.1	62.5
BiCr	FPDvsFNS	87.5	52.1	63.9
HiCl	MPDvsMNS	70.8	75.0	75.0
HiCl	FPDvsFNS	75.0	68.7	70.8
SVM	MPDvsMNS	93.4	96.1	94.8
SVM	FPDvsFNS	97.8	86.6	92.2

6.2 Dysphonia assessment based on the VFBS

The results presented in this subsection correspond to the experimental design described in subsection 5.1.2, in which a further step forward has been taken by estimating the VFBS, a highly semantic biomechanical correlate of phonation.

To give a better description of the methodological procedures implicit in this study, an example from a sustained emission of the vowel [a:] during 4 s by one of the participants actively stimulated with rTMS is shown in Figure 6.5. The VFBS was estimated from the glottal source using the methods described in subsection 5.1.2. This specific example is included as a prototype to describe the speech processing protocol, and it is not to be taken as a mark of generalized behaviour, but as a particular phenomenological description of the examination procedures conducted on each phonation analysed to be considered in detail as a singular case of study.

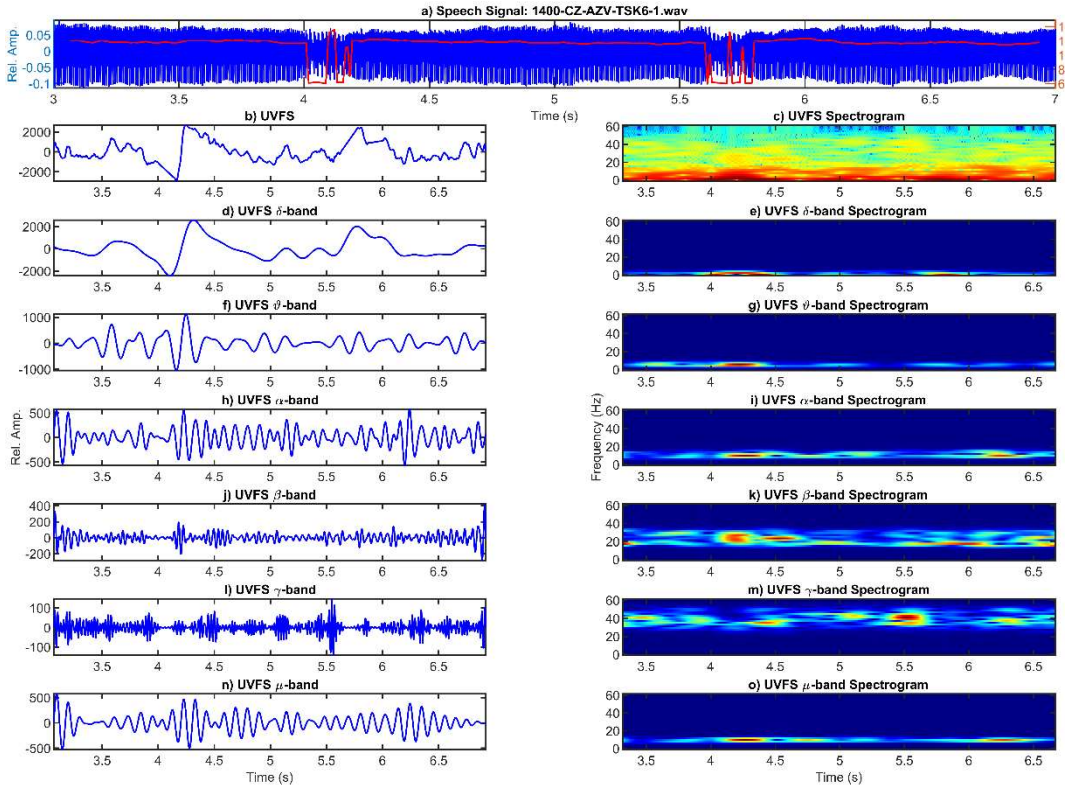

Figure 6.5 EEG-band description of a 4 s segment of phonation from the pre-stimulus recording of active case 1400 during the utterance of a sustained vowel [a:].
Description: a) original speech signal, with the F0 line superimposed in red; b) estimation of the unbiased VFBS (UVFS); c) its logarithmic power spectrogram; d-e) activity on the δ-band and its linear spectrogram; f-g) id. on the ϑ-band; h-i) id. on the α-band; j-k) id. on the β-band; l-m) id. on the γ-band; n-o) id. on the μ-band. The activity in the β-band is especially relevant following the incident in the interval 4.0-4.2 s. Clarification note: the labelling "Rel. Amp." in template h) applies to all the left-hand side vertical templates, from b-n), whereas the label "Frequency (Hz) in template i) applies to all right-hand side vertical templates, from c-o).

The ϑ-band and γ-band frequency distributions, corresponding to the best-behaving active case (1400) are presented in Figure 6.6.a) and b). A similar set of ϑ-band and γ-band distributions, corresponding to the worst-behaving sham case (2200) is shown in Figure 6.6.c) and d).

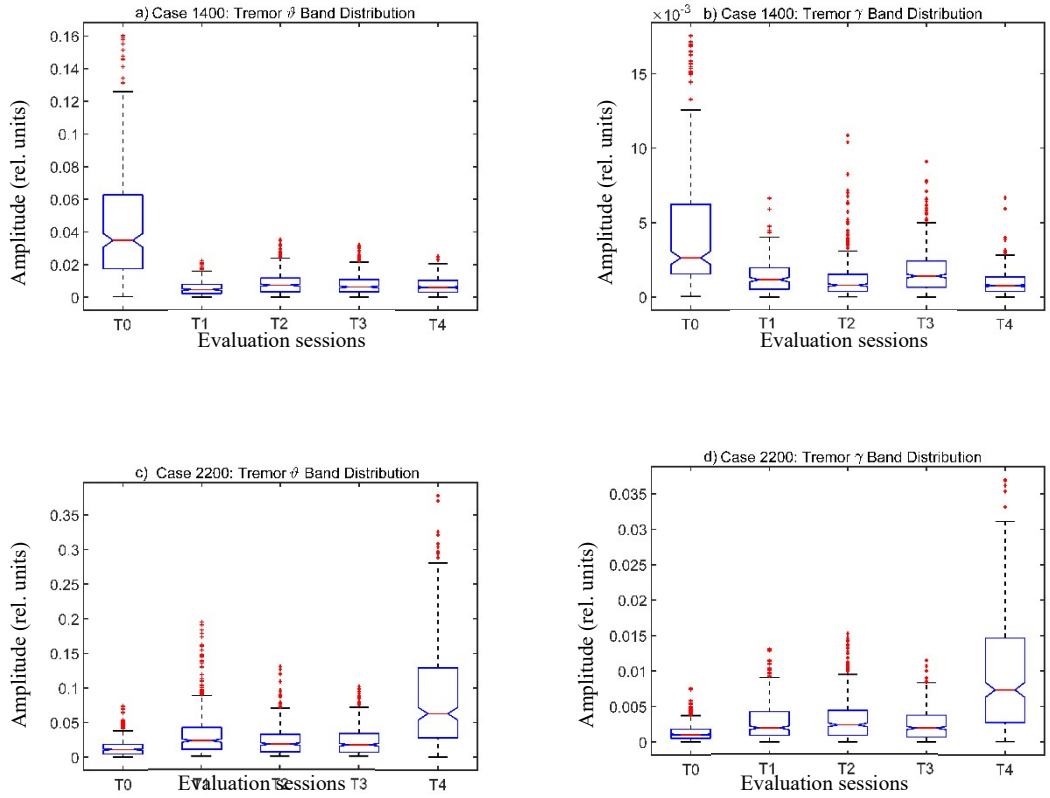

Figure 6.6 Tremor amplitude distribution boxplots for the best (active 1400) and worst (sham 2200) behaving cases.
Description: a) case 1400 ϑ-band; b) case 1400 γ-band; c) case 2200 ϑ-band; d) case 2200 γ-band. When two relevant features related to phonation stability are considered, such as the medians of the tremor amplitudes, and their dispersion, measured by the interquartile range, it may be seen that in the active stimulation case (1400) a strong reduction in amplitude and dispersion is observed in the evaluations T1-T4 following the pre-stimulus evaluation session (T0) in both bands considered (ϑ and γ), whereas case 2200 experiences a clear deterioration in subsequent evaluation sessions (T1-T4) concerning the pre-stimulus one (T0).

The results of evaluating the log-likelihood ratios between the pre-stimulus (T0) and the four post-stimulus on the frequency bands following expression (5.1) are given in Table 6.9. The results of the corresponding Mann-Whitney tests following expression (5.4), availing the relevance of the log-likelihood ratios are given in Table 6.10, and the global scores from (5.7) are given in Table 6.11. The null hypothesis assumed that the change in the medians before and after stimulation would not differ under a statistical significance level of α=0.05.

Table 6.9 Log-likelihood ratios (LLR) from all participants on the ϑ-band and γ-band activity from comparing pre-stimulus (T0) to post-stimulus recordings (T1-T4).
Values showing improvement (λ>0) are noted in bold. The codes of the active (A) and sham (S) cases are given in the left-most column.

Part. T0-Code	T1 LLR ϑ	T2 LLR ϑ	T3 LLR ϑ	T4 LLR ϑ	T1 LLR γ	T2 LLR γ	T3 LLR γ	T4 LLR γ
0100 (A)	-0.323	-1.825	-0.014	-0.102	-0.055	-0.843	**0.217**	**0.113**
0800 (A)	**0.156**	**0.160**	**0.241**	-0.002	**0.224**	**0.306**	**0.316**	**0.310**
1100 (A)	**0.072**	-0.320	**0.093**	**0.214**	**0.109**	**0.189**	**0.095**	**0.249**
1200 (A)	**0.177**	**0.228**	**0.113**	**0.068**	**0.463**	**0.446**	**0.444**	**0.422**
1400 (A)	**0.047**	**0.162**	**0.092**	**0.077**	**0.155**	**0.158**	**0.136**	**0.184**
1600 (S)	-0.367	**0.146**	-0.044	**0.284**	**0.012**	**0.023**	**0.067**	**0.021**
1700 (S)	-0.455	-0.228	-0.686	0.000	-0.009	-0.086	-0.240	-0.049
1800 (S)	-0.836	-0.334	-0.877	-1.108	-0.914	-0.744	-1.439	-0.890
1900 (S)	**0.302**	**0.064**	**0.289**	-0.183	**0.116**	**0.009**	**0.102**	-0.032
2000 (A)	**0.133**	-0.441	**0.254**	**0.113**	-0.079	-0.444	-0.087	-0.030
2200 (S)	-0.575	-0.362	-0.325	-1.317	-0.461	-0.543	-0.365	-1.432
2300 (S)	**0.100**	**0.062**	-0.033	**0.068**	-0.223	-0.323	-0.214	**0.098**
2400 (S)	-0.608	-0.593	-0.333	-0.139	-0.216	-0.115	-0.374	-0.306
2500 (S)	-0.802	-0.867	-0.952	-0.576	**0.075**	-0.064	-0.256	-0.928
2600 (S)	-0.098	**0.198**	-0.029	-0.115	**0.104**	**0.172**	-0.043	-0.068
2700 (A)	-1.029	-0.101	-1.391	-1.906	-0.586	-0.025	-0.442	-0.556
2800 (A)	-0.453	-0.433	-0.946	-0.699	-0.505	-0.036	-0.326	-0.976
2900 (A)	-0.113	**0.203**	-0.220	-0.717	-0.252	**0.222**	-0.252	-0.514

Those tests failing to reject the null hypothesis (p>0.05) are noted in bold, in which case a significant change in phonation stability could not be assumed along the interval of study. It may be seen that this happens for sham cases 1600, 1700, 1900, 2300, 2500, and 2600, whereas it is observed also in the active cases 0100, 2000, and 2700. The results in Table 6.11 avail that improvements in Table 6.9 when (λ>0) are significant for the level of α being considered.

Table 6.10 Results of Mann-Whitney tests (p-values) on equal medians from all participants. Corresponding to comparisons between post-stimulus evaluations (T1-T4) concerning the pre-stimulus one (T0) for the ϑ-band and γ-band tremor amplitudes.

Part. T-Code	T1 pvϑ	T2 pvϑ	T3 pvϑ	T4 pvϑ	T1 pvγ	T2 pvγ	T3 pvγ	T4 pvγ
0100 (A)	<0.001	<0.001	**0.899**	0.006	**0.101**	<0.001	<0.001	<0.001
0800 (A)	<0.001	<0.001	<0.001	0.006	<0.001	<0.001	<0.001	<0.001
1100 (A)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
1200 (A)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
1400 (A)	<0.001	<0.001	<0.001	0.006	<0.001	<0.001	<0.001	<0.001
1600 (S)	<0.001	<0.001	**0.934**	<0.001	**0.172**	0.007	<0.001	0.010
1700 (S)	<0.001	<0.001	<0.001	**0.281**	**0.574**	0.007	<0.001	**0.632**
1800 (S)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
1900 (S)	<0.001	<0.001	<0.001	**0.802**	<0.001	0.006	<0.001	**0.219**
2000 (A)	<0.001	<0.001	<0.001	<0.001	**0.985**	<0.001	**0.730**	0.026
2200 (S)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2300 (S)	<0.001	<0.001	**0.114**	<0.001	<0.001	<0.001	<0.001	<0.001
2400 (S)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2500 (S)	<0.001	<0.001	<0.001	<0.001	0.001	**0.089**	<0.001	<0.001
2600 (S)	0.001	<0.001	0.921	0.004	<0.001	<0.001	**0.078**	0.013
2700 (A)	<0.001	0.018	<0.001	<0.001	<0.001	**0.216**	<0.001	<0.001
2800 (A)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2900 (A)	0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Regarding the optimal separation criterion, the threshold to consider that a specific case experienced a functional improvement was fixed at g_t = **-0.1**, following the criterion of equal error rate detection-error trade-off (Martin et al., 1997). The scores showing functional improvements in phonation stability conditions (producing values under the threshold $g_j<g_t$) are given in bold. The agreement between the nature of the stimulation column (Active/Sham) and the global score column is given in the column to its right (Agreement, assumed to be 0 when active stimulation does not produce improvements, or when sham stimulations do, and conversely, assumed to be 1 if active stimulation produces improvements, and sham does not). The last four columns label the cases where the presumed assumption agrees or disagrees with the observed result. TP: number of cases treated with active stimulation showing functional improvement; TN: number of cases receiving sham stimulation not showing functional improvement; FP: number of cases receiving sham stimulation showing functional improvement; FN: number of cases receiving active stimulation not showing functional improvement.

The six bottom files give the grand totals, the value of the threshold to consider improvement or not, and the detection performance in terms of sensitivity (Sn), specificity (Sp), accuracy (Ac), and F1 score as defined in (4.49).

Table 6.11 Summary of the results listing the global score per participant.
Following expression (69), including all post-stimulus evaluation sessions (T1-T4) and frequency bands (δ, ϑ, α, β, γ, the band μ having been excluded, for being a subset of the α one).

Code	Active/Sham	Gender	Global score (g_j)	Agreement	TP	TN	FP	FN
0100	A	F	0.243	0	0	0	0	1
0800	A	M	**-0.343**	1	1	0	0	0
1100	A	M	**-0.392**	1	1	0	0	0
1200	A	M	**-0.533**	1	1	0	0	0
1400	A	M	**-0.625**	1	1	0	0	0
1600	S	F	**-0.220**	0	0	0	1	0
1700	S	M	0.174	1	0	1	0	0
1800	S	M	1.870	1	0	1	0	0
1900	S	M	-0.091	1	0	1	0	0
2000	A	F	**-0.128**	1	1	0	0	0
2200	S	M	3.374	1	0	1	0	0
2300	S	M	**-0.154**	0	0	0	1	0
2400	S	M	0.336	1	0	1	0	0
2500	S	M	1.528	1	0	1	0	0
2600	S	F	-0.046	1	0	1	0	0
2700	A	M	2.116	0	0	0	0	1
2800	A	M	1.456	0	0	0	0	1
2900	A	F	1.328	0	0	0	0	1
			Total	0.71	5	7	2	4

Threshold (g_t)	-0.1
Sensitivity (St)	0.56
Specificity (Sp)	0.78
Accuracy (Ac)	0.67

6.3 Assessing articulation alterations based on the APM

The results presented in this subsection correspond to the experimental design on the effects of PD on speech articulation, described in subsection 5.2, using the biomechanical system of the jaw-tongue defined in subsection 4.2.1 for the characterization of PD HD, to estimate the neuromotor behaviour of the system, and provide specific markers of proper or improper NMA during vowel utterances.

6.3.1 Data recording examples

Similarly, as in the previous studies being described, a practical example is presented on the proposed signal recording is illustrated, where the speech signal, the sEMG, and the three acceleration channels from two repetitions of the […a→i→a→i...] by a female HC participant (CF1) are shown in Figure 6.7. The sEMG signal has been included in the plots (channel b) to evidence that the acceleration and speech signals are concordant with the action of the masseter.

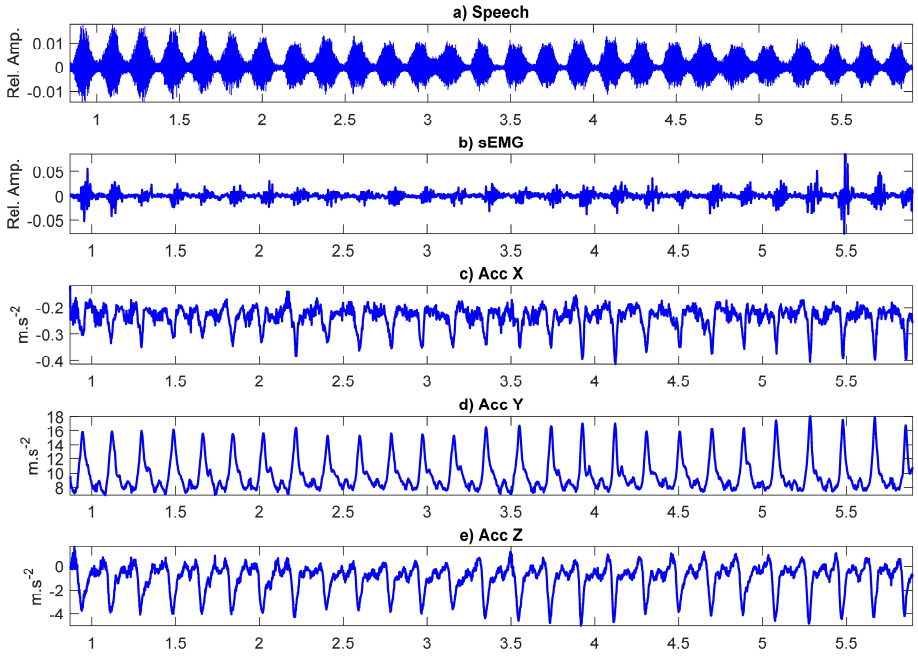

Figure 6.7 Signal acquisition example from the repetition of the phonetic sequence […a→i→a→i…] by CF1.
Description: a) speech signal; b) surface electromyographic signal on the masseter; c) channel X accelerometer signal; d) channel Y accelerometer signal; e) channel Z accelerometer signal.

Similarly, the same set of recordings from one of the PD female participants included in the study (PF1) is shown in Figure 6.8.

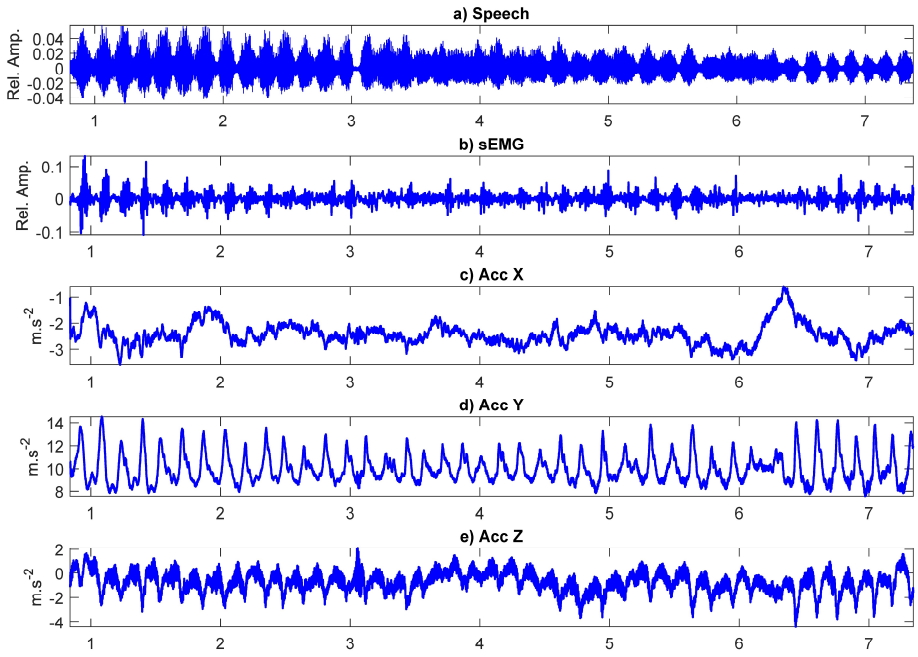

Figure 6.8 Signal acquisition example from the repetition of the phonetic sequence […a→i→a→i…] by PF1.
Description: a) speech signal; b) surface electromyographic signal; c) channel X (Acc); d) channel Y (Acc); e) channel Z (Acc).

The unbiased and smoothened formants extracted from the case shown in Figure 6.7 are to be compared with the jaw-tongue reference displacements obtained after rotation and integration of the acceleration signals (Gómez-Rodellar, et al., 2018). As an example, the estimations of $\Delta\mathbf{S}$ and $\Delta\mathbf{F}$ from CF1 in reference to expression (4.29) are given in Figure 6.9.

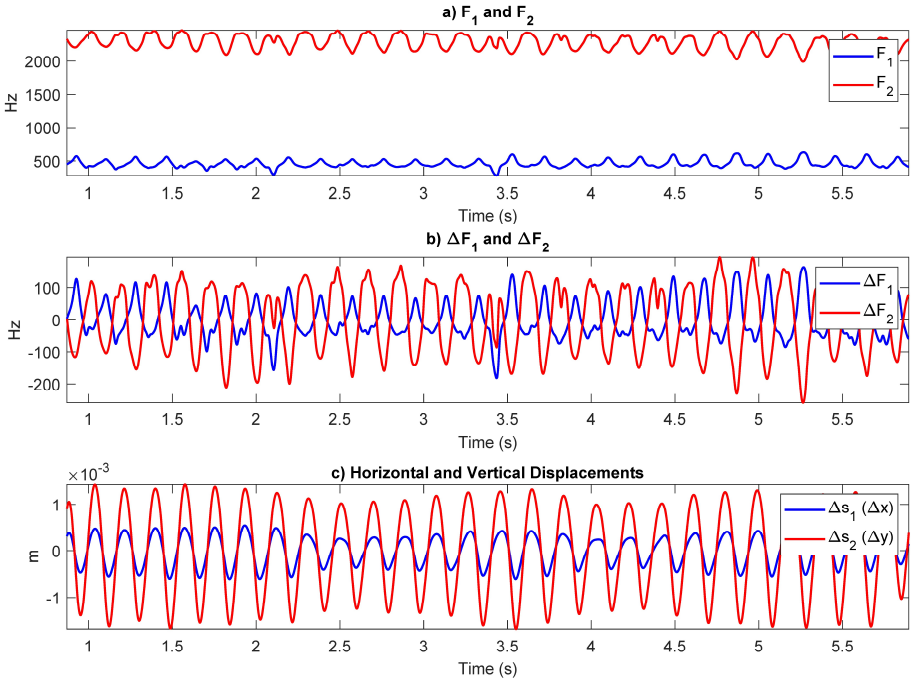

Figure 6.9 Formant deviations and reference point displacements obtained from CM1 (HC participant). Description: a) formants F1 and F2; b) formant deviations **ΔF** ; c) reference point displacements **ΔS**.

The estimations of **ΔS** and **ΔF** corresponding to PF1 are shown in Figure 6.10.

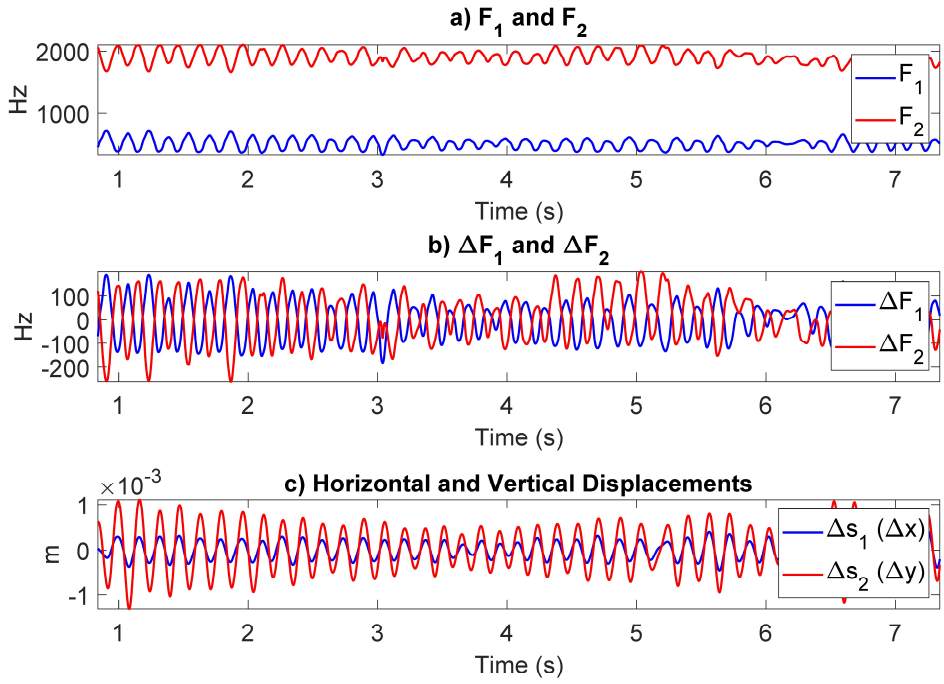

Figure 6.10 Formant deviations and reference point displacements obtained from PF1 (PD participant). Formant deviations and reference point displacements obtained from PF1, corresponding to a PD participant: a) formants F1 and F2; b) formant deviations $\Delta\mathbf{F}$; c) reference point displacements $\Delta\mathbf{S}$.

6.3.2 Weight estimation from linear regression

The initial estimation of $\mathbf{W^0}$ is illustrated using the formant deviations and reference point displacements from the healthy control participant (CF1) shown in Figure 6.9. The scatter plots in Figure 6.11 show the distribution patterns of each pair of $\Delta\mathbf{s}_\mathrm{i}$, related to each pair of $\Delta\mathbf{F}_\mathrm{i}$. A a regression line is fitted to each of these distributions with a structure $\Delta\mathbf{s}_\mathrm{i} = w_\mathrm{ij}\Delta\mathbf{F}_\mathrm{i} + b_\mathrm{ij}$, as printed within each scatterplot. The slope of the regression line is the respective initial weight w_{ij}^0 of matrix $\mathbf{W^0}$.

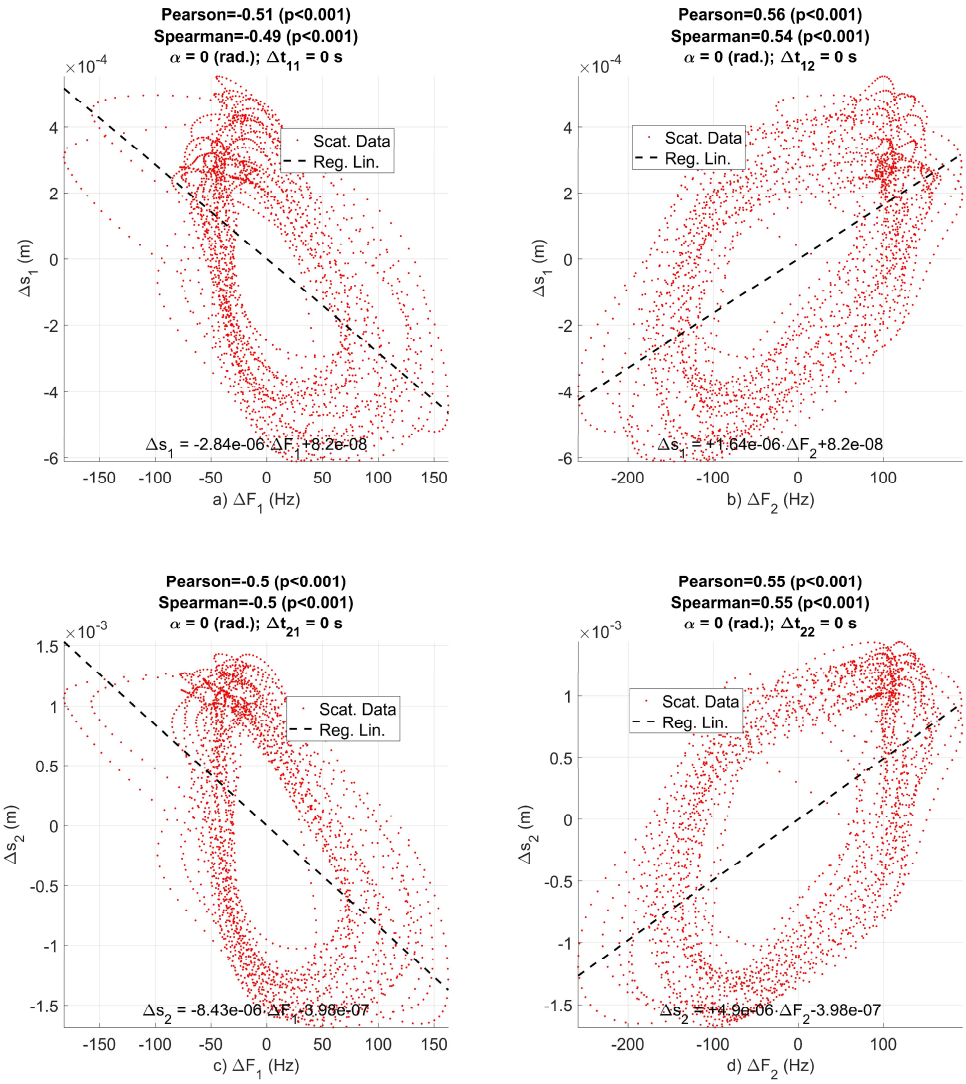

Figure 6.11 Scatter plots and regression results from CF1.
Description: The regression analysis is carried out for each pair of input signals (ΔF_i) and output signals (Δs_j): a) w_{11}^0=-2.84.10^{-6} m.Hz^{-1}; b) w_{12}^0=1.64.10^{-6} m.Hz^{-1}; c) w_{21}^0=-8.43.10^{-6} m.Hz^{-1}; d) w_{22}^0=4.90.10^{-6} m.Hz^{-1}.

This seminal analysis shows what was expected from the hypothesized dynamic relation between formant dynamics ($\mathbf{\Delta F_i}$) and the kinematic outcome ($\mathbf{\Delta s_i}$); the first formant dynamics (ΔF_1) increases when there is a descent and retraction of the P_{rJT}, whereas the value of the second formant dynamics (ΔF_2) is assumed to descend under the same movement conditions (Gómez-Vilda et al., 2019b).

This behaviour is shown in the negative sign of the weights w_{11} and w_{21}, while in the case of w_{12} and w_{22} a positive sign is obtained. This method is applied to the signals from all male participants in the cohort and the healthy control, see Table App. 5 (Appendix I.8) for details, producing the initial results of the acoustic to kinematic projection given in Table 6.12, providing a first inter-participant comparison.

Table 6.12 Male cases: Model weights and correlation coefficients per participant.

P: Pearson; p-values <0.001; × 10^{-6} cm.Hz^{-1}.

Participant Labels	w_{11}*	w_{12}*	w_{21}*	w_{22}*	$P_{\Delta x \Delta F1}$	$P_{\Delta x \Delta F2}$	$P_{\Delta y \Delta F1}$	$P_{\Delta y \Delta F2}$
CM1	-6.08	3.77	-9.90	6.31	-0.61	0.71	-0.57	0.68
CM2	-2.44	1.47	-2.71	1.56	-0.52	0.42	-0.43	0.33
CM3	-4.04	4.92	-3.45	4.34	-0.42	0.43	-0.36	0.38
CM4	-4.37	3.63	-5.65	4.71	-0.60	0.63	-0.63	0.66
PM1	-1.26	0.97	-4.56	3.45	-0.35	0.38	-0.36	0.38
PM2	-2.12	0.78	-1.05	0.45	-0.14	0.10	-0.23	0.19
PM3	-1.41	1.29	-3.17	2.88	-0.71	0.72	-0.84	0.85
PM4	-1.04	0.24	-1.13	-0.18	-0.30	0.09	-0.18	-0.04

The values of the model weights are accompanied by the correlation coefficients (Pearson) between each pair of signals, confirming the relationships expected from the acoustic-to-kinematic projection properties. The same study was conducted on a set of female participants (four HC and four PD participants) summarized in Table 6.13.

Table 6.13 Female cases: Model weights and correlation coefficients per participant.

P: Pearson; p-values <0.001; *× 10^{-6} cm.Hz^{-1}.

Participant Labels	w_{11}*	w_{12}*	w_{21}*	w_{22}*	$P_{\Delta x \Delta F1}$	$P_{\Delta x \Delta F2}$	$P_{\Delta y \Delta F1}$	$P_{\Delta y \Delta F2}$
CF1	-2.84	1.64	-8.43	4.90	-0.51	0.56	-0.50	0.55
CF2	-3.78	2.15	-5.39	3.17	-0.47	0.68	-0.44	0.65
CF3	-2.15	1.19	-7.70	4.51	-0.72	0.73	-0.68	0.72
CF4	-0.28	0.30	-0.67	0.67	-0.32	0.37	-0.49	0.53
PF1	-2.09	1.56	-5.24	4.03	-0.81	0.75	-0.81	0.77
PF2	-1.46	1.15	-1.22	0.91	-0.54	0.52	-0.37	0.33
PF3	-1.88	1.75	-6.54	6.73	-0.24	0.17	-0.32	0.25
PF4	-1.52	0.95	-2.90	1.81	-0.54	0.55	-0.51	0.51

The values of the respective correlation coefficients (Pearson) between each pair of signals are given under the same conditions as described before (see further comments in subsection 7.3).

6.3.3 Weight estimation from regression iteration

Taking the initial weight estimation of $\mathbf{W^0}$ as a starting point, an iterative adjustment has been carried out according to the procedure based on the iterative gradient-descent with variable step size, as described in (5.12) and (5.13). This process aims to find a minimum of the error surfaces corresponding to the partial cost functions $C_{\mathrm{i}}(w_{\mathrm{i}1}, w_{\mathrm{i}2})$. The plots given in Figure 6.12 show this process illustrated for the control participant CF1. The trend of the descent for the pair of weights $\mathbf{w}_{\mathrm{i}}^{\mathrm{k}} = (w_{\mathrm{i}1}^{\mathrm{k}}, w_{\mathrm{i}2}^{\mathrm{k}})$ can be observed, as the estimation for the *k*-iteration is represented as a point on the surface C_{i}.

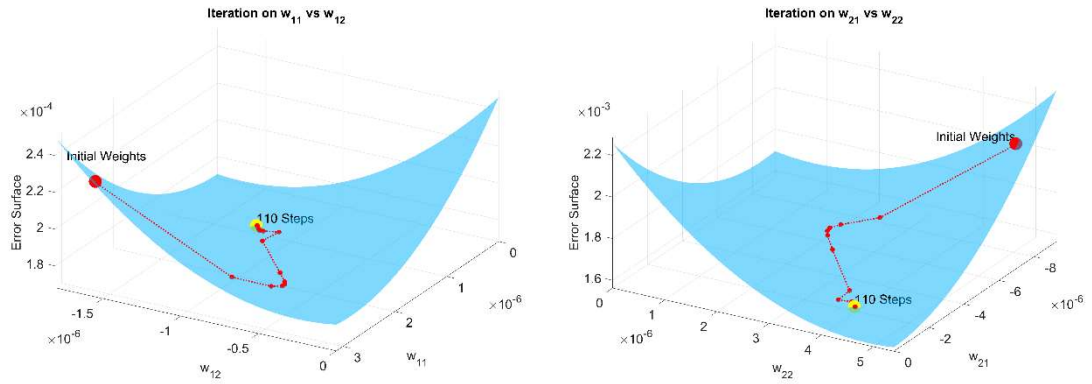

Figure 6.12 Error surfaces corresponding to the iteration process on participant CF1.
Description: Left: $C_1(w_{11}, w_{12})$. Right: $C_2(w_{21}, w_{22})$. The starting position (in red) shows the values of the weights obtained from linear regression $\{w_{11}^0, w_{12}^0\}$ and $\{w_{21}^0, w_{22}^0\}$, whereas the stop position (in yellow) corresponds to the values of the weights after the iteration refinement.

It may be seen that although the error surfaces show similar behaviour to Rosenbrock's function (Rosenbrock, 1960) displaying a kind of *wadi*–shaped shallow valley, the variable step tracker based on the Barzilai–Borwein method is capable of reaching the minimum point of the curve in a reasonable number of iteration steps (110 for this particular case). The new weights after iteration refinement are given in Table 6.14.

Table 6.14 Male cases: model weights, number of iterations, and error reduction.

ΔE in percent; *x10^{-6} cm.Hz^{-1}.

Participant Labels	w_{11}*	w_{12}*	w_{21}*	w_{22}*	No. Ite.	ΔE (%)
CM1	-0.71	-4.11	2.44	7.46	142	**23.7**
CM2	3.86	1.16	-4.85	-1.74	142	10.68
CM3	1.49	-3.32	-0.70	3.59	123	7.37
CM4	1.07	-2.83	-1.17	3.84	142	**21.67**
PM1	-0.25	0.82	-1.30	2.63	111	6.05
PM2	-6.24	-2.31	-2.13	-0.61	193	1.37
PM3	-0.55	0.84	-1.37	1.75	140	**41.58**
PM4	-2.37	-1.20	-4.09	-2.68	101	5.05

The results of the iteration refinement corresponding to the female participants are given in Table 6.15[12].

Table 6.15 Female cases: model weights, number of iterations, and error reduction.

ΔE in percent; *x10^{-6} cm.Hz^{-1}.

Participant Labels	w_{11}*	w_{12}*	w_{21}*	w_{22}*	No. Ite.	ΔE (%)
CF1	-0.82	1.28	-2.27	3.89	110	12.82
CF2	-0.03	2.14	0.31	3.25	108	14.45
CF3	-1.26	0.71	-3.85	3.06	88	**22.75**
CF4	-0.15	-0.43	0.09	0.75	153	11.55
PF1	-1.94	0.14	-4.08	1.03	132	**38.47**
PF2	-0.97	0.47	-1.03	0.19	111	8.07
PF3	-1.80	0.17	-5.69	1.73	77	2.34
PF4	-0.70	0.55	-1.36	1.04	163	13.00

The values of these weights would be the basis of a study towards a definition of a possible unified weight model for HC and PD participants, which is left as a future line.

6.3.4 Time realignment

When comparing the formants and displacements, it was observed that the input $\mathbf{\Delta S}$ and output $\mathbf{\Delta F}$ showed similar patterns (number of cycles and periods), but there appeared to be a misalignment between them. Therefore, a realignment method based on maximizing correlations between $\mathbf{\Delta S}$ and $\mathbf{\Delta F}$ following (5.17)-(5.19) was implemented.

[12]The iterative process is stopped once the gradient variation with respect to the previous iterative step is lower than a given minimum threshold.

This process has been used on the same example being presented. The resulting changes from the initial estimation in the scatter plots and regression analysis can be observed by comparing the plots in Figure 6.11 with those in Figure 6.13.

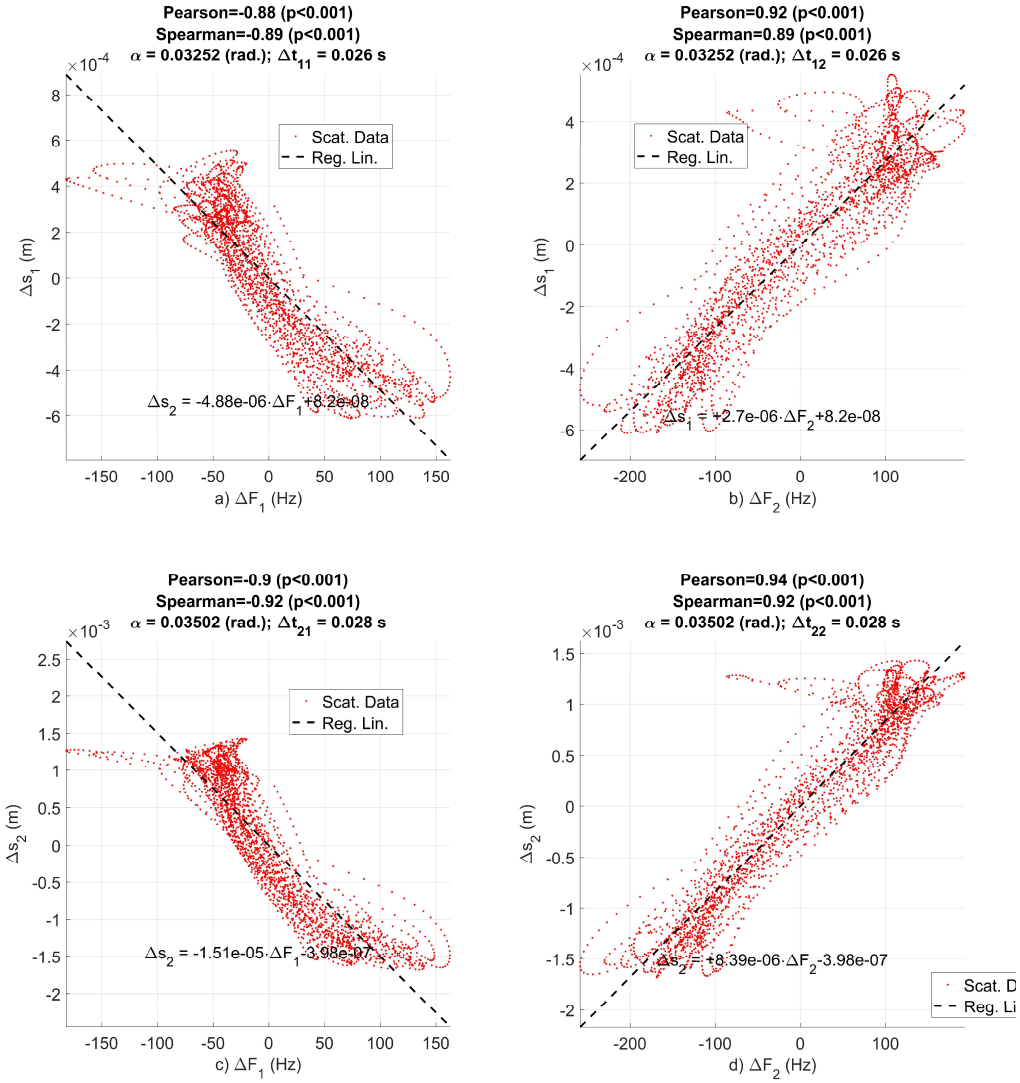

Figure 6.13 Scatter plots and regression results from CF1 after signal realignment. Description: Realignments denoted as $R'(\Delta F_i,\Delta s_j)$: a) $R'(\Delta F_1,\Delta s_1;)$: w_{11}=-4.88.10^{-6}; b) $R'(\Delta F_2,\Delta s_1)$: w_{12}=2.70.10^{-6}; c) $R'(\Delta F_1,\Delta s_2)$: w_{21}=-1.51.10^{-5}; d) $R'(\Delta F_2,\Delta s_2)$: w_{22}=8.39.10^{-6}. The size of the realignment time shift is given as Δt in seconds (Δt_{11}=26 ms, Δt_{12}=26 ms, Δt_{21}=28 ms, Δt_{22}=28 ms). The coefficients w_{ij} are given in cm.Hz^{-1}.

As it may be seen the realignment has reduced sensibly the dispersion of data in the new scatter plots, making the relationship between **ΔS** and **ΔF** more linear, as the dispersion along the perpendicular dimension to the regression line has been reduced. S ee the relative quadratic errors before (Table 6.14 and Table 6.15) and after realignment (Table 6.16 and Table 6.17). The scatter plots and regression analysis from the female PD participant (PF1) are given in Figure 6.14 as a complementary example to be contrasted with Figure 6.13 related to the HC participant.

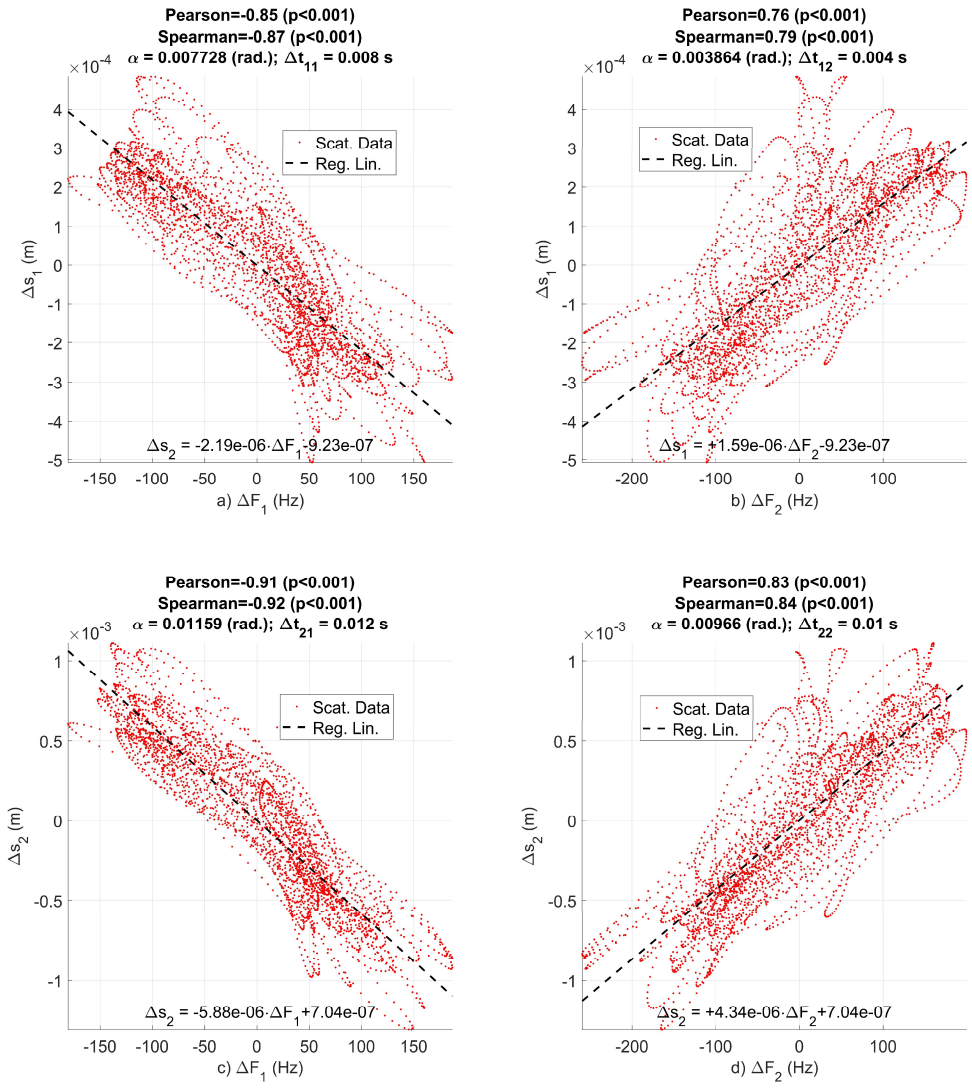

Figure 6.14 Scatter plots and regression results from PF1 after realignment.
Description: a) *R'*(ΔF_1,Δx); b) *R'*(ΔF_2,Δx); c) *R'*(ΔF_1,Δy); d) *R'*(ΔF_2,Δy). The size of the realignment time shift is given as Δt in seconds (Δt_{11}=8 ms, Δt_{12}=4 ms, Δt_{21}=12 ms, Δt_{22}=10 ms).

After realignment, the same cross-correlation analysis as in the one shown in Table 6.12 and Table 6.13 is carried out for the male and female datasets. The results are shown in Table 6.16 and Table 6.17, together with the alignment errors following expression (5.20).

Table 6.16 Male cases: Model weights, correlation coefficients, and relative rms errors after realignment per participant.

P: Pearson; p-values <0.001; ε_r: relative rms error in %; *x10^{-6} cm.Hz^{-1}.

Participant Labels	w_{11}*	w_{12}*	w_{21}*	w_{22}*	$P_{\Delta x \Delta F1}$	$P_{\Delta x \Delta F2}$	$P_{\Delta y \Delta F1}$	$P_{\Delta y \Delta F2}$	$\varepsilon_{r\Delta x \Delta F1}$	$\varepsilon_{r\Delta x \Delta F2}$	$\varepsilon_{r\Delta y \Delta F1}$	$\varepsilon_{r\Delta y \Delta F2}$
CM1	-8.27	4.51	-14.53	7.97	-0.83	0.85	-0.84	0.86	0.56	0.53	0.54	0.51
CM2	-4.19	3.08	-5.75	4.20	-0.89	0.87	-0.90	0.89	0.46	0.49	0.43	0.47
CM3	-8.90	9.83	-9.12	10.12	-0.92	0.86	-0.94	0.88	0.38	0.51	0.34	0.47
CM4	-6.42	4.81	-8.23	6.21	-0.89	0.83	-0.92	0.87	0.46	0.56	0.39	0.50
PM1	-3.15	2.04	-11.08	7.08	-0.87	0.80	-0.87	0.78	0.49	0.61	0.50	0.63
PM2	-14.30	7.24	-4.08	2.05	-0.93	0.89	-0.90	0.85	0.38	0.46	0.45	0.52
PM3	-1.62	1.43	-3.40	3.02	-0.81	0.80	-0.90	0.89	0.59	0.60	0.44	0.45
PM4	-2.81	1.96	-5.15	-3.47	-0.81	0.76	-0.80	-0.73	0.59	0.65	0.60	0.69

Table 6.17 Female cases: Model weights, correlation coefficients, and relative rms errors after realignment per participant.

P: Pearson; p-values <0.001; ε_r: relative rms error in %; *x10^{-6} cm.Hz^{-1}.

Participant Labels	w_{11}*	w_{12}*	w_{21}*	w_{22}*	$P_{\Delta x \Delta F1}$	$P_{\Delta x \Delta F2}$	$P_{\Delta y \Delta F1}$	$P_{\Delta y \Delta F2}$	$\varepsilon_{r\Delta x \Delta F1}$	$\varepsilon_{r\Delta x \Delta F2}$	$\varepsilon_{r\Delta y \Delta F1}$	$\varepsilon_{r\Delta y \Delta F2}$
CF1	-4.88	2.70	-15.05	8.40	-0.88	0.92	-0.90	0.94	0.48	0.40	0.44	**0.33**
CF2	-6.34	2.65	-9.86	4.16	-0.79	0.84	-0.80	0.85	0.61	0.55	0.60	0.52
CF3	-2.24	1.20	-9.00	4.94	-0.75	0.73	-0.79	0.79	0.66	0.68	0.61	0.61
CF4	-0.65	0.63	-1.07	1.03	-0.75	0.78	-0.79	0.82	0.67	0.63	0.61	0.57
PF1	-2.19	1.59	-5.88	4.34	-0.85	0.76	-0.91	0.83	0.53	0.65	0.41	0.56
PF2	-2.12	1.75	-2.62	2.14	-0.79	0.79	-0.78	0.78	0.62	0.61	0.62	0.63
PF3	-5.39	7.75	-14.04	19.95	-0.68	0.74	-0.70	0.74	**0.73**	0.68	0.72	0.67
PF4	-2.31	1.44	-4.86	3.04	-0.82	0.83	-0.85	0.86	0.58	**0.73**	0.58	0.72

A further comparison between the model weights may be carried by normalizing each weight set w_{ij} to its vector norm as $\hat{w}_{ij} = w_{ij}/|w_{ij}|$. The results of the normalization are shown in Table 6.18.

Table 6.18 Weight normalization results.

A scale factor of 10^{-6} cm.Hz^{-1} applies to the whole weight dataset.

Male Set	$\hat{w}_{11}$	$\hat{w}_{12}$	$\hat{w}_{21}$	$\hat{w}_{22}$	Female Set	$\hat{w}_{11}$	$\hat{w}_{12}$	$\hat{w}_{21}$	$\hat{w}_{22}$
CM1	-0.43	0.24	-0.76	0.42	CF1	-0.27	0.15	-0.83	0.46
CM2	-0.48	0.35	-0.65	0.48	CF2	-0.50	0.21	-0.78	0.33
CM3	-0.47	0.52	-0.48	0.53	CF3	-0.21	0.11	-0.85	0.47
CM4	-0.49	0.37	-0.63	0.48	CF4	-0.38	0.36	-0.62	0.59
PM1	-0.23	0.15	-0.81	0.52	PF1	-0.28	0.20	-0.75	0.56
PM2	-0.86	0.43	-0.24	0.12	PF2	-0.49	0.40	-0.60	0.49
PM3	-0.32	0.28	-0.68	0.60	PF3	-0.21	0.30	-0.54	0.76
PM4	-0.40	0.28	-0.73	-0.49	PF4	-0.36	0.23	-0.77	0.48

Mann-Whitney tests between the normalized weights from the HC and the PD samples failed to reject the null hypothesis of equal means $\mu(\widehat{w}_{ij})$ with a p-value of 0.965. A similar test between the normalized weights from the male and female participants failed to reject the null hypothesis of equal means with a p-value of 0.904. These results may indicate that a general model may be built independently of gender and alteration conditions, depending on a generalization relying on a larger sample database. The medians of the normalized weights may serve as a robust estimation of the model weight matrix $\{\widehat{w}_{11}$=-0.39, $\widehat{w}_{12}$=0.28, $\widehat{w}_{21}$=-0.70, $\widehat{w}_{22}$=0.48$\} \times 10^{-6}$ cm.Hz^{-1}. The realignment sample shifts (n_{ij}) expressed as time shifts (in ms) are given in Table 6.19.

Table 6.19 Realignment time stride between ΔS and ΔF per participant in ms.

Most cases are within the range of 26-40 ms, with exceptions highlighted in bold.

Male Set	Δt_{11}	Δt_{12}	Δt_{21}	Δt_{22}	**Female Set**	Δt_{11}	Δt_{12}	Δt_{21}	Δt_{22}
CM1	**18**	**14**	**20**	**16**	CF1	26	26	28	28
CM2	26	28	28	32	**CF2**	**20**	**14**	**22**	**16**
CM3	28	26	30	28	**CF3**	**8**	**4**	**14**	**10**
CM4	30	26	30	24	CF4	28	26	22	20
PM1	28	26	28	26	**PF1**	**8**	**4**	**12**	**10**
PM2	34	36	32	32	PF2	26	28	36	38
PM3	**16**	**14**	**12**	**10**	PF3	28	32	26	30
PM4	32	38	36	-40	PF4	26	26	28	28

6.3.5 Formant Dynamics and Articulation Kinematics

From the regression study results it may be inferred how the different magnitudes (**ΔS** and **ΔF**) relate to each other. Based on these observations a transformation function is defined as **W**, projecting formant dynamics to spatial displacements. An interesting indicator to compare between speakers is to observe the ranges that they can produce in these spaces. Table 6.20 shows the range of variation covered by **ΔS** and **ΔF**. The ranges are estimated by the 0.05 to 0.095 interquartile distance.

Table 6.20 Formant (in Hz) and displacement (in mm) ranges per participant r(·).

Males	***r*(Δx)**	***r*(Δy)**	***r*(ΔF1)**	***r*(ΔF2)**	**Females**	***r*(Δx)**	***r*(Δy)**	***r*(ΔF1)**	***r*(ΔF2)**
CM1	1.37	2.27	115	203	CF1	0.97	2.88	178	307
CM2	0.74	0.81	163	239	CF2	1.34	1.96	169	435
CM3	1.90	1.65	229	187	CF3	0.63	2.33	180	282
CM4	2.60	3.37	308	380	CF4	0.13	0.29	213	225
PM1	0.78	2.8	284	360	PF1	1.00	2.55	246	309
PM2	1.11	0.58	276	495	PF2	1.52	1.24	521	725
PM3	0.84	1.88	255	314	PF3	0.43	1.56	156	118
PM4	0.28	0.22	213	279	PF4	0.99	1.89	324	586

It may be observed that the size of the ranges shows a broad direct relationship between the formant and displacement oscillation ranges. Whether this observation could be the basis for defining new markers of HD is the subject of further study.

CHAPTER 7

7 Discussion

This chapter is devoted to examining the results presented in Chapter 6 regarding the experiments conducted to test the possibilities of glottal signals for describing PD HD by GFAD and VFS, as well as the APM to describe alterations in articulation kinematics from acoustical correlates. The ordered structure of the chapter preserves that of former ones.

In this PhD project, model selection prioritizes interpretability over the black-box versus open-box dilemma. The primary objective is to offer interpretable indicators for knowledge construction, rather than merely maximizing classification scores. This approach involves sacrificing powerful computational tools that are gaining prominence. The task narrows down to selecting an algorithmic architecture to build relationships between elements, emphasizing observation and interpretation of component behavior rather than repetitive solutions.

7.1 Study based on the GFAD

The most immediate observation when examining the glottal flow amplitude distributions shown in Figure 6.1 is that the NS distributions split towards both extremes of the normalized amplitude axis (bimodal behaviour), whereas the PD and HC ones tend to concentrate in the mid-values of the amplitude axis. This behaviour is observed in both gender datasets. The similarity between HC and PD distributions is larger than that between HC and NS or PD and NS. This finding reveals that there might be a factor explaining phonation deterioration by itself besides neuromotor degeneration, similarly, affecting PD and HC participants, which might be associated with the potential effects of ageing.

This finding is also confirmed by the estimates of JSDs between the average GFADs from the HC, PD, and NS datasets, given in Table 6.1. It may be seen that the distances between the average PD and HC datasets ($D_{MPD|MHC}$=0.051 and $D_{FPD|FHC}$=0.066) are smaller than the distances between the average PD and NS or the average HC and NS ($D_{MPD|MNS}$=0.204, $D_{MHC|MNS}$=0.211, and $D_{FPD|FNS}$=0.262, $D_{FPD|FHC}$=0.299).

These observations indicate that the average distributions of HC and PD participants diverge much more than average NS distributions between themselves. As the average HC and PD distributions come from age-matched participants, it may be assumed that the phonation conditions of both HC and PD subsets do not differ much as far as glottal flow patterns are concerned. This fact may question why the level of overlap between HC and PD participants would be larger than expected relative to a normative younger population. This situation is reflected in the geometrical distributions shown in Figure 6.2, where the majority of HC and PD samples overlap with each other and are closer to the average PD centroid than to the average NS one.

Several consequences may be derived from these observations. On one hand, this finding confirms that the glottal flow profile is at least as sensitive to ageing-induced dysphonic conditions as to neuromotor dysfunction (Robnik-Šikonja and Kononenko, 2003; Midi et al., 2008; Belalcázar-Bolaños et al., 2015; Hanratty et al., 2016; and Novotný et al., 2020). On the other hand, this finding brings to our attention the assumption that dopaminergic treatments may induce the regression of the altered glottal function of PD patients to a more normative behaviour which improves their phonation conditions towards the HC baseline (Ho, Bradshaw, and Iansek, 2008; Pinho et al., 2018), a plausible assumption under the point of neuropsychology (Rektorová et al., 2012). This differentiation garners a dedicated experimental setting. This is a bit problematic as there are a series of ethical issues.

Requesting from a volunteer to avoid taking medication is a delicate proposition, as it would require to endure a degree of discomfort, pain and potential fall risk in the service of an exploration that might or might not yield expected or valuable results. It is here where speech shines as a conveyance signal, because an experimental set up where patients could record their speech at home before and after medication would provide an interesting dataset to explore precisely these effects.

A similar conclusion may be derived from the confusion matrices given in Table 6.3, resulting from separating male and female samples according to the criterion expressed in (4.48). It may be seen that the criterion separates NS from PD samples with an accuracy of 89.6% for males and females, although it fails in separating HC from PD, as 22 male and 23 female samples out of a total of 24 HC were misclassified as pathological ones (false positives). This would reduce the overall accuracy of BiCr to 62.5% (males) and 63.9% (females). This insight corroborates the observation regarding the ageing factor behind age-paired HC and PD phonation, to be considered as the main factor behind the degrading of glottal dynamics, either by natural ageing processes or by PD deterioration.

Regarding the results produced by hierarchical clustering, as reported in Figure 6.3 and Figure 6.4 the separation between HC and PD phonation is a little bit better, although not much. Each sample is included either in the non-pathological clusters (in green) or in the pathological ones (in red). The male datasets are separated into six clusters, #1 including 27 samples (16 PD and 11 HC, in red), #2 including 16 samples, (15 NS and 1 HC, in green), #3 including 15 samples (11 HC, 3 PD, and 1 NS, in green), #4 including 9 samples (6 NS and 3 PD, in green), #5 including 4 samples (2 NS, 1 HC, and 1 PD, in green) and #6 including 1 sample (1 PD, in black).

The female datasets are also separated into six clusters, #1 including 18 samples (18 NS, in green), #2 including 17 samples (11 HC and 6 PD, in green), #3 including 12 samples (6 HC and 6 PD, in red), #4 including 11 samples (7 PD and 4 HC, in red), #5 including 10 samples (5 PD, 3 HC, and 2 NS, in red), and #6 including 4 samples (4 NS, in green).

Summarizing the results from the male datasets, 15 out of 24 NS samples are grouped together (#2), but most PD and a large group of HC samples are mixed (#1). Another important cluster (#3) includes a large group of HC samples. Clusters #4 and #5 mix up samples from the three groups. Summarizing the results of the female datasets, 22 out of 24 samples are clustered together (#1 and #6), but PD and HC are mixed up in the remaining clusters, in different proportions. HC samples are a majority in cluster #2, but they are mixed with an equal number of PD in #3 and are a minority in #4 and #5.

Consequently, it may be concluded that most NS and PD samples are correctly classified, but HC samples are included in pathological and non-pathological clusters with no clear differentiation. These results point again to the difficulty of separating HC from PD sets according to phonation conditions. This situation is also reflected in Table 6.3, where the results of the classification in terms of true positives and negatives, and false positives and negatives are summarized. It may be seen that the accuracy of separating PD from NS provided by hierarchical clustering is 85.4% (males) and 83.3% (females). Attending to PS vs NS$\cup$HC (if HC is to be considered non-pathological, a criterion which would need a strong reformulation) the overall accuracy of detection would be 75.0% (males) and 70.8% (females). The differentiation results based on SVM classification, as given in Table 6.4 show much better accuracy in separating PD from HC samples than the previously commented methods, but some questions remain unclear.

It must be stressed that the features selected from the GFAD vectors of each sample follow a strict feature selection process in terms of optimizing classification scores, rendering much better results, but due to the intricacies of input vector component selection, the process, although more efficient, becomes less transparent. Therefore, the suspicion that by using different input data, the selected features would not overlap much in different experiments makes this process much less interpretable.

In turn, the examination of the statistical description of each sample JSD given in Table 6.5 shows that the distributions closest to normality are MHC and MPD. MNS is the dataset appearing farthest apart from normality. Kurtosis reveals that the distributions closer to normality are MHC, MNS, and FHC. Moreover, according to the normality tests shown in Table 6.6, none of the datasets reject the null hypothesis, MHC and MPD being again the ones farthest apart from hypothesis rejection. One of the objectives of the study, as mentioned in subsection 6.1.1.4 was to evaluate potential differences between the JSD distributions in terms of statistical relevance. The results of the tests (parametric and non-parametric) are given in Table 6.7. It may be seen that the three tests reject the null hypothesis of similar PD and NS distributions, and HC and NS distributions, both for males and females. However, the situation is completely different when PD sets are compared with HC sets. In the case of the male and female sets, Student's-t, Kolmogorov-Smirnov and Mann-Whitney U tests fail to reject the null hypothesis under a p-value<0.05. These results point again to the existence of more similarities than expected between HC and PD phonation, in full alignment with three-distance comparisons (BiCr and HiCl). The examination of the correlation between LED and JSD, as commented at the end of subsection 6.1.1.4 pointed to the effects of medication.

Because the PD participants were in an ON state by dopaminergic medication, their effects could drive the phonation neuromotor behaviour of PD participants to be more aligned with that of HC participants (Pinho et al., 2018). This is a controversial issue, as levodopa seems to play an unclear role in the restoration of some motor functions (Ho, Bradshaw, and Iansek, 2008). Apparently, some correlates of speech/phonation could be affected, some others might not, depending on dosage and subject, because this behaviour is probably influenced by individual conditions, and may depend on the progress of PD. In the present case, the values of LED administered to PD participants and the estimated JSDs showed a modest but undeniable degree of correlation, not only in magnitude but what is more importantly, in sense. Larger levels of LED were associated with reductions in JSD. The correlation estimated was larger for female than for male datasets. These results open questions to be taken into account facing future studies, in the sense that new experiments might have to be designed based on specific phonation features to ensure that they are sensitive to these two intertwined problems: effects of ageing on HC and PD participants, and effects of medication on PD participants.

A very important question that must be carefully examined is the inter-linguistic issue raised by the fact that PARCZ and HUGMM databases have been produced by speakers of Czech and Spanish, respectively. For such, it must be taken into account that phonating a sustained vowel is the result of a purely physiological process not conveying any linguistic clue per se. On the contrary, speech production is based on articulating different sounds by the production of dynamic changes on the ONPT (articulation), which involves phonated (voiced) and non-phonated (voiceless) sounds. Under the point of view of speech articulation Czech and Spanish are quite different languages, and treating inter-linguistic databases would require special care (Orozco-Arroyave et al., 2016).

Nevertheless, under the point of view of an open sustained vowel phonation as [a:], there is no meaningful difference between both Czech and Spanish sound emissions. Therefore, as far as the study is based on comparing amplitude distributions of glottal flow patterns produced by larynges sustaining a similar open vowel as [a:], the inter-linguistic issue appears to produce almost no effect on the study results provided that the glottal flow estimations are robust to different channel conditions (Orozco-Arroyave et al., 2016).

The comparison among the classification methodologies given in Table 6.8 shows that hierarchical clustering (HiCl) behaves better than the simple distance difference criterion (BiCr), although both are outperformed by SVM classification. This fact questions how feature selection preserves information contents, an issue that will deserve an independent study per se. To explain this observation, one must consider the difference between hierarchical clustering and SVM data flows given in Figure 4.19. It may be seen that the main difference is the preprocessing of the GFAD before classification. In BiCr and HiCl, the classification feature is the JSD (MI divergence) between each sample GFAD and the respective dataset average. In Figure 4.19.b the process of feature selection keeps only those GFAD features relevant to improving classification performance. Therefore, in this second case, there is an undetermined information content reduction from GFAD, which removes whatever seems to be superfluous (or even counterproductive) for SVM classification. Consequently, most of the classification improvement is due to feature selection, at the cost of information loss. What raises more concerns in this reductionist process is the fact that the selected feature sets for different classification experiments show little intersection (Álvarez-Marquina et al., 2020).

Throughout this discussion it is important to bear in mind that JSD is an estimate of the MI relationships between distributions, therefore its potential capability to explain distribution similarities is undeniable. Summarizing, feature selection could produce exceptionally well-performing classifiers, but at the cost of information reduction with unclear further effects, for instance, in generalization and interpretability (Tennenholtz, Zahavy, and Mannor, 2018; Gómez-Rodellar et al., 2019b). At least, a future study is due in this respect.

The work of Novotný et al. (2020) establishes the possibility of establishing a comparison with the present approach regarding the study of phonation from PD vs age-matched HC participants. The mentioned work shows a similar methodology to the one proposed in the present study based on inverse filtering (IAIF method) to estimate the glottal flow and pressure. In their case, time-domain glottal features are used, as the quasi-open quotient (QOQ), the normalized amplitude quotient (NAQ), jitter, shimmer, and frequency-domain features as the H1H2 factor, the harmonic richness factor (HRF), the harmonic-to-noise ratio (HNR), and cepstral peak prominence (CPP). Their dataset included 40 male PD and 40 male age-matched HC (no female participants). Phonations consisted of sustained utterances of the vowel [a:]. The classification methods used by Novotný et al. (2020) were logistic regression, achieving a sensitivity of 65.8%, specificity of 64.3%, and an AUC of 0.73; and SVM, achieving an STV of 67.5%, an SPC of 62.5%, and an area under the curve (AUC) of 0.78, below those reported by this study in Table 6.4. The lower performance attained by the referred study compared to the present one may be explained by some of the limitations of the features used. It must be pointed out that the features proposed by Novotný et al. (2020) may not be as robust as expected because in deteriorated phonation (ageing, dysphonic or neurodegenerative) glottal patterns might not follow strictly the Liljencrants-Fant (L-F) pattern (Davis, et al., 1996).

Another relevant consideration regarding the study of Novotný et al. (2020) is that it has been carried out on drug-naïve patients which means that symptoms may not be altered by medication, and differentiation regarding age-matched HC subjects is expected to work differently. This question opens a debate about whether voice testing must be conducted during OFF or ON states. Given that many patients may experience difficulties in carrying on simple daily life tasks or suffer pain and distress during the OFF state, studies should be conducted on patients in the ON state to support potential applications of the methodology to patient care, monitoring, and rehabilitation during normal patient activity.

Regarding the similarity between PD and HC phonation, another relevant explanation factor to be further studied is the role played by the loss of elasticity in the vocal folds induced by the decay of elastic proteins present in Reinke's Space (Hidalgo, Gómez, and Garayzábal, 2017), concerning phonation ageing effects both in HC and PD participants. This explanation would deserve further research.

7.2 Study based on the VFBS

As it was stated in the introduction, the study based on the VFBS was intended to characterize the stability of phonation in pre-stimulus and post-stimulus vocal emissions following rTMS by assessing the validity of features estimated on glottal neuromechanics from a limited set of PD participants. The database used is described in Appendix I.3, Appendix I.4, and Appendix I.5. The example depicted in Figure 6.5.a was given to illustrate the protocol followed in the assessment.

This example is especially meaningful, as it shows two events of phonation activity breaks typical in PD dysphonia, seen at intervals 4.0-4.2 s and 5.1-5.2 s, resulting in a muscle tension drop at 4 s appreciated in Figure 6.5.b, which is not that evident at 5.5 s, although the speaker's correction action by tension intensification from proprioceptive perception is quite well appreciated in both cases on the δ-band (Figure 6.5.c). A relaxation towards stable muscle tone is observed between 4.2 and 5.0 s as a follow-up after tension restoration. Intuitively, it could be hypothesized that the δ-band signal would be associated with the group activity of sets of motor units reacting to proprioceptive feedback (coarse tuning), whereas the γ-band could be related to individual motor actions maintaining muscle tone activity as stable as possible according to tonal settings (fine-neuromotor control of the thyroarytenoid muscle tension).

Although the findings presented in Figure 6.5 are rather specific to that particular case, similar F0 blocking events are not infrequent in PD dysarthria. These events and their associated EEG-band activity reveal interesting information, which could help in clarifying the phenomena behind muscular blocking and neuromotor failure (Shi et al., 2022; Chiang, Wang, and McKeown, 2012). The first important observation is that an apparent correction takes place immediately after neuromotor blocking, possibly from the auditory system and cerebellum feedback, as revealed in the δ-band, which shows clear corrective actions to recover muscle tone. The ϑ-band (Figure 6.5.f and g) shows intensified activity before and after the first blocking, and a moderate degree of tremor all over the whole interval studied. Interestingly, the α-band shows a low level of activity before each blocking, which is reactivated immediately after (Figure 6.5.h and i). It is also seen as a ripple in the vocal fold stiffness Figure 6.5.b). A similar phenomenon is observed in the β-band (Figure 6.5.j and k) where a strong burst of activity is seen after the first blocking following an interval of very low activity.

On the contrary, the activity on the γ-band (Figure 6.5.l and m) is very intense before the blocking, to become more diffuse and less intense after the incident. The activity in the μ-band (Figure 6.5.n and o) seems to intensify after each blocking.

As it was mentioned before, bands ϑ and γ are considered especially relevant in the study according to previous knowledge on EEG-related NMA as they seem to bear semantic information in the case of PD (Brambilla et al., 2021). In this sense, it will be of special interest to review the behaviour of the best and worst functional cases comparing pre- and post-stimulus feature distributions in both bands, corresponding to cases 1400 and 2200 shown in Figure 6.6.a-d. The boxplots in Figure 6.6 show the distributions of the tremor amplitude in the ϑ (a) and γ (b) frequency bands from case 1400, corresponding to the same participant (pre-stimulus code T0, post-stimulus codes T1, T2, T3, and T4). Assuming that tremor is associated with phonation instability, a decrement in the tremor amplitude should be considered an improvement concerning the disorder conditions. Therefore, as boxplots T1-T4 (post-stimulus) show a smaller median and a much smaller interquartile dispersion than boxplot T0 (pre-stimulus) in both bands (ϑ and γ), it might be concluded that the post-stimulus tremor is less intense than pre-stimulus, and consequently, that an improvement is observed. The situation concerning Figure 6.6.c and d is just the opposite, as it may be seen that post-stimulus distributions (T1-T4) show larger medians and much larger interquartile dispersion than the pre-stimulus one (T0) in both bands (ϑ and γ), therefore, it might be concluded that the situation, contrary to being improved, has deteriorated substantially.

These two cases showed the largest improvements and the largest deterioration in phonation stability. Whereas improvements seen in 1400 could be attributed to the beneficial effects of rTMS, the increment in instability shown in case 2200 could not be attributed to the effects of non-stimulation, therefore, a possible explanation would be a worsening in the phonation conditions within a short time interval between T0 and T4. Some other circumstances could have influenced the deterioration observed. The same criticism could apply to case 1400, which experienced a strong improvement because recordings at session T0 showed a large instability, which was not shown in T1-T4. Of course, the possible benefits of rTMS might not be the only explanation for such evolution, nor its single cause. Obviously, this question is fully open to discussion.

The information provided by the activity in the ϑ and γ bands as presented in Figure 6.6.a and b for the same case (active stimulation case 1400) reveals an important decay of phonation instability after the stimulation, where the tremor distributions reduce notably their variance and average values, the effect being more clearly observed in the ϑ-band (Figure 6.6.a, T1). This drastic improvement, in value and dispersion, is slightly worsened in the next observations, although post-stimulus tremor amplitudes keep under 1%. This fact may be especially important, as the ϑ-band is classically associated with the so-called "low-frequency tremor" by many studies in voice quality, as it is easily perceived by listening, and it is typically associated with many cases of PD phonation (Brückl, Ghio, and Viallet, 2018; Ibarra-Lecue, Haegens, and Harris, 2022). These findings open the possibility of conducting studies on F0 blocking for further generalization, depending on the dual neuromotor mechanism influencing VFBS; on one hand, the cricothyroid muscle innervated by the inferior laryngeal nerve, and on the other hand, the thyroarytenoid muscle innervated by the recurrent laryngeal nerve (Dietrich et al., 2020).

Pointing to misadjustments in the agonist-antagonist role played by these muscles when failing to ensure a fine F0 tuning. Of course, this assumption would demand further studies to provide generalization insights on the VFBS neuromotor driving functionality.

Table 6.9 presents also interesting results when analysing the phonation instability of each participant before and after stimulation. It may be appreciated that a subset of participants who underwent active stimulation (0800, 1100, 1200, 1400) presents mainly improvements in most of the evaluation sessions on the two bands of reference. Another subset of sham cases (1600 and 1900) experienced unexpected improvements, whereas two other sham cases (1700 and 1800) showed worsening behaviour (quite strong in 1800, possibly pointing out to an extremely stable starting condition in T0). Another subset of cases (2000, 2700, 2800, 2900) experienced slight or no improvements at all. A fourth subset of sham cases showed worsening of phonation quality (2200 and 2400), whereas the remnant sham cases expressed mixed behaviour (2300, 2500, and 2600).

Regarding the results presented in Table 6.10, it may be seen that most of the tests reject the null hypothesis, generally associated with low absolute-value LLR, pointing to the fact that most of the tremor bands examined showed a substantial difference between each first exploration in time (T0) and the subsequent ones (T1-T4) except for low $|\lambda|$. This fact would seem reasonable when observed on actively stimulated cases, under the assumption of tremor improvements as a result of treatment, but should not sound that plausible when sham cases were examined. Nevertheless, test results from the best-behaving cases (0800, 1100, 1200, 1400) rejected the null hypothesis, disregarding any similarity between T0 and T1-T4 distributions (in alignment with what could be expected). Cases 2700 and 2800 rejected also the null hypothesis, although they did not show improvements (misalignment).

Part of the remnant active cases (0100, 2000, 2900) showed similarity on some tests (alignment). In six sham cases, at least one test or more showed possible similarity (1600, 1700, 1900, 2300, 2500, 2600) in alignment with what could be expected. Alternatively, the remnant sham cases (1800, 2200, and 2400) rejected similarity between pre- and any post-stimulus examinations (misalignment).

It is a well-known fact that correlation does not mean causality, therefore, when so many underlying confounding factors are involved tracing a clear relational path is problematic. Consequently, the present study was intended to assess if phonation stability might be reduced after active rTMS cases compared to sham cases. Given the premises of the experimental setup, the most robust comparative metric should be based on correlation. As the longitudinal study aims to track the progressive effects after stimulation, an important robustness factor entails some form of sustained trustability metric. This is because improvement assessment is not based on a single post-stimulus evaluation because a tendency may be inferred from expressions (5.5) and (5.6), summarized in a single score given by (5.7) because each single summarized score is supported by four longitudinal estimations comparing $\langle \xi_{ij}^{k}(n) \rangle$ vs $\langle \xi_{0j}^{k}(n) \rangle$, plus three others comparing $\langle \xi_{ij}^{k}(n) \rangle$ vs $\langle \xi_{i-1j}^{k}(n) \rangle$, having also the time intervals T_i - T_0 into account. Besides, the log-likelihood ratios given in Table 6.9 have been estimated using probability density functions generated from normalized histograms of 100 bins summarizing tremor amplitude time series on a cycle-synchronous basis, each one calculated on around 400 samples from a male voice, and about double size from a female voice. These same distributions have been used in computing MW U-tests to assess log-likelihood ratios' significance.

A reduction in variance following a phonation stability improvement is expected when comparing each single pre-stimulus frequency band of the vocal fold stiffness estimate to the corresponding frequency band estimate from each post-stimulus recording in a longitudinal sequence (intra-participant) as given in Table 6.9 in terms of log-likelihood ratios, validated by corresponding non-parametric Mann-Whitney (MW) U-tests shown in Table 6.10. A transversal (inter-participant) t-test has also been conducted to assess how general improvement scores given in Table 6.11 compare in the context of all participants.

Because the checking for false-discovery cases has been conducted already using MW U-tests (an example of them being given in Table 6.10 validating the results given in Table 6.9 not being produced by chance when their respective p-values are below the confidence level of 0.05), global scores have to be taken as reliable in the best and worst behaving cases. Adding a hypothesis t-test using the global scores in Table 6.11 between the set of actively stimulated cases vs the set of sham cases will produce a p-value of 0.2, which does not reject the equal means between both sets. Does it mean that the functional assessment methodology proposed fails in its objectives? Evidently, not. The most plausible interpretation is that because actively stimulated cases 2700, 2800, and 2900 show a large deterioration after stimulation, the average global score of the active set (0.36) is not far enough from the average global score of the sham set (0.87) considering the two sets overlapping variances (1.18 and 1.61, respectively). This observation does not invalidate the individual results, because they are evaluated longitudinally on their timeline statistics, independently of what any other participants are experiencing, as this transversal comparison points out.

Longitudinal tests would be in full alignment with the old medical lemma “Treat the patient, not the disease”, which is the grounding floor emphasizing the balance of intra-participant over inter-participant studies.

Another possible explanation for this apparent different behaviour could be attributed to the small sample size, as well as to the effects of possible confounding factors affecting the phonation of participants during the tests, such as possible different types of HD involved, and the emotional or comorbid conditions affecting vocal emissions during the evaluation sessions, such as ageing, depression and anxiety, and medication, among them, (Gillivan-Murphy, 2013, Gillivan-Murphy, Miller and Carding, 2018) taking into account the amount of elapsed time between the pre-stimulus and the post-stimulus examinations in all cases, ranging from 93 and 119 days (see Table App. 3 in Appendix I.5), which could introduce important changes in daily life conditions not taken into account, affecting vocal production in one or another way. Given the apparent high sensitivity of frequency bands to varying conditions in vocal fold expressed instabilities, minor changes in these confounding phonation conditions could produce substantial changes in amplitude distributions for inducing null hypothesis rejection. A further extended study should concentrate on comparing each examination from each participant to observe the extent to which improvement or worsening phonation stability conditions could be observed among them.

Table 6.11 presents the summarized overall results indicating which cases showed better phonation stability conditions confronting pre- and post-stimulus examinations taking into account all participants, evaluation sessions, and frequency bands. The global score g_j per participant j is given in the fourth column from the left, considering all post-stimulus evaluation sessions (T1-T4) and five out of the six frequency bands (δ, ϑ, α, β, γ).

The fifth column (Agreement) shows the cases where active stimulation produced a negative global score (reduction of tremor instability, in agreement with expectations), and those cases of sham (non) stimulation which presented an increment of tremor instability (positive scores, showing no improvement). These two subsets configure the true positive and negative cases respectively (columns TP and TN). Those sham cases (non-stimulus), showing improvement (false positives), as well as those active stimulus cases showing deterioration (false negatives), are listed in columns FP and FN, respectively. The total amounts of TP, TN, FP, and FN are given at the bottom of the respective columns (sixth to ninth). Fixing the detection threshold at $\boldsymbol{g_t = -0.1}$ (in the application of the equal-error rate criterion commented in subsection 6.1.2), the sensitivity of the stimulation methodology would be estimated as 56%, its specificity would be around 78%, and its overall accuracy would rise to 67%, corresponding to an F1 score of 0.63. The agreement rate including TP and TN concerning the whole set size would be around 71%. Twelve cases produced results aligned with expectations: five active cases yielded scores under the threshold (improvement in phonation stability), and seven sham cases brought scores over the threshold (worsening in phonation stability). The other six cases produced results contrary to expectations: four active cases yielded scores over the threshold (worsening in phonation stability), and two sham cases brought scores under the threshold (improvement in phonation stability). As a general comment, two out of three cases included in the study behaved according to expectations after having been submitted to the stimulation protocol, including active and sham cases.

These findings support the use of log-likelihood ratios in the evaluation of phonation quality, as well as conducting hypothesis tests, focusing in particular on the ϑ- and γ-bands, as these are believed to summarize well both neuromotor and cognitive activity (Solomon et al., 2017). The correction of this type of activity around phonation motor blocking and disruption could demand coordinated action of motor units spiking at different frequencies (Brambilla, et al., 2021; Numssen et al., 2021; Leodori et al., 2021; Saravanamuttu et al., 2021; Ibarra-Lecue, Haegens, and Harris, 2022). Of course, the clarification of all these observations would require further studies combining other cooperating methods, such as electroencephalography (EEG), although at the cost of complicating signal acquisition and pre-filtering to remove facial muscle activity during speech production, possibly by the use of sEMG.

Although the reach of the study presents evident limitations, such as the relatively small sample size and gender unbalance, some insights on the efficacy of the stimulation methodology and the data analytics methodology used might still be drawn from the results, which had to be taken more as speculative assumptions to open new research lines than proven facts based on exhaustive testing. On one hand, the effects of rTMS on phonation stability offer mixed results. Some of the active cases studied report notorious improvements in the frequency bands studied which could not be explained by a by-chance effect, whereas some other actively stimulated cases do not seem to improve stability, or even show more unstable phonation. Conversely, some sham cases show clear improvements, whereas some others show undeniable worsening phonation. Relating both behaviours to the absence of stimulation is not an easy task, opening an issue for further discussion.

It seems that phonation stability behaviour might be too sensitive to confounding factors such as emotional or comorbid conditions to serve as a unique marker by itself, and it should be combined with other speech-based traits, including articulation acoustic features. Besides, tremors might not be a bi-univocal feature of PD HD (Gillivan-Murphy, 2013; Gillivan-Murphy, Miller, and Carding, 2019). On the other hand, it seems that the study of phonation instability to monitor disruption events in vocal fold stiffness, and its association with EEG-related frequency bands would be fully justified as a powerful introspective methodology to disclose interesting NMA in cortical areas affecting larynx control.

Remarkably, it could be asserted that rTMS seems to have some influence on phonation stability in the cases commented with a low margin of error. The reasons why these beneficial effects are not seen in other active cases can also be asserted with little margin for error. Why stimulation may seem beneficial for phonation stability in some cases, and not in others, is a matter which could not be determined within the framework of this study considering the data available at the moment, and it is to be left for further research.

An important fact to be taken into account in the design of the potential future extension of the present study has to see with the medication condition of patients, in the sense that some of the unexpected behaviour seen in some of the cases included in the study might have been related not exclusively to PD but also to dopaminergic-induced dyskinesia (Filipović et al., 2009).

The methodology designed for this study would be well aligned with relevant research activity on the structural complexity of brain from EEG frequency sub-bands (Ahmadlou and Adeli, 2010), the functional connectivity pattern assessment from EEG signals (Yuvaraj et al., 2016; García et al., 2022), or the detection of movement intention in BCI systems from EEG signals (Karakullukcu and Yilmaz, 2022), among others.

Although the analysis being proposed uses glottal estimates from speech utterances instead of EEG recordings, the tentative findings reported based on previous work (Gómez-Vilda et al., 2019a; Gómez-Rodellar et al., 2021d) might demonstrate differences in functional behaviour based purely on acoustic signal analysis, which could shed light beyond what can be provided by classical acoustic analysis. The ultimate objective driving future extensions of this study would be to reproduce EEG-aligned descriptions of phonation estimated from audio recordings only, which could be used in neurodegenerative speech characterization. The justification for characterizing a biomechanical correlate such as the VFBS by filtering according to the frequency bands of EEG signals, as if it were an EEG channel signal comes from preliminary work done on the NMA of the muscles involved in articulation (Gómez-Vilda et al., 2019a; Gómez-Rodellar et al., 2021d), which is conveyed by acoustic correlates, such as formants and glottal features. This relationship seems natural because the activity of the larynx, oropharyngeal, and facial muscles is the ultimate cause of modulation and framing of the glottal source into acoustically perceived speech. Therefore, if muscles would respond to neuromotor stimulation, and eventually, to brain cortical activity, a way to study the intervening links of interest in modelling possible dysfunctions in the activation chain, would consider the important advancements in modelling and activity coding driving CMC, to convey signal descriptions to a common code.

Given that this common code is already well established as EEG-frequency band descriptions, it seems natural to build biomechanical correlate descriptions in the same coding convention and to benefit from the accumulated knowledge already available in CMC studies.

Other possible methodologies for decomposing and characterizing the biomechanical correlate VFBS derived from the acoustic phonation signal would include Gabor transform (Loh et al., 2021), Wavelet filter banks (Sharma, Patel, and Acharya, 2020), Fourier-based synchro squeezing transform (Karakullukcu and Yilmaz, 2022), transfer entropy (Vicente et al., 2011), or fuzzy synchronization likelihood (Ahmadlou and Adeli, 2017), among others. These studies would require a further extension of the present manuscript out of reasonable limits; therefore, they are considered for future research. Likewise, phonation improvements from rTMS based on comparisons using glottal source features, such as the harmonic-noise ratios, first-second harmonic ratio, cepstral peak prominence, parabolic spectrum coefficient, open and closed quotients, and other indices of voice quality analysis (Mekyska et al., 2015) could enrich the assessment protocol within a future study framework.

7.3 Study based on the APM

The inverse linear model for acoustic to kinematic projection introduced in subsection 4.2 defined the methodological foundation of the experimental study presented in subsection 5.2, producing the results shown in subsection 6.2, from which the following highlights may be summarized:

- The relationship between acoustic to kinematic variables ($\Delta\mathbf{S}$ and $\Delta\mathbf{F}$) has been established and may be explained using the inverse model described in expression (4.29).
- An estimation of the model weights has been carried out using least squares linear regression.
- A gradient-descent method using a variable step size has been used in the iterative refinement of model weights to minimize the error cost function implicit in the inverse model.
- To linearize the relationship between the acoustic and kinematic estimates in the time domain a realignment procedure has been introduced with a considerable improvement of correlation results.

As in the precedent studies, the observations derived from the illustrative data recordings shown in Figure 6.7 and Figure 6.8 from an HC and a PD female participants will be briefly commented on here. These figures present the speech, surface electromyography, and X, Y, and Z accelerations in the accelerometer system of coordinates, which is projected onto the sagittal plane shown in Figure 4.14. The X coordinate in the accelerometer system corresponds with the coordinate normal to the sagittal plane, the Y corresponds with the sagittal y (s_2), and the accelerometer Z corresponds with the sagittal x (s_1).

To interpret the signal sequence, it is convenient to bear in mind that each masseter contraction appears marked by a specific outburst of sEMG activity. This contraction rises the jaw, producing a narrowing of the frontal vocal tract, switching from an open low vowel [a:] to a closed high vowel [i:] provided that no retraction of the tongue is present. This action would correspond with the acoustical gesture [a→i], and it corresponds with subsequent spikes in the accelerometer signals, which are largest in the Y and Z components on the sagittal plane, and much smaller in the X coordinate normal to the sagittal plane.

Correspondingly the amplitude of the speech signal (channel a) shows a strong reduction in amplitude immediately after each sEMG burst, because the activity of the masseter reduces the opening of the radiation end, and less energy is projected outwards, this observation being aligned with what it could be expected.

The main difference found between both figures, corresponding to the HC and PD participants is that the interval cadence and the amplitude and pattern of the sEMG and X, Y, and Z accelerometer signals are more regular in Figure 6.7 (corresponding to the HC female participant) than those in Figure 6.8 (corresponding to the PD female participant). The behaviour presented in both figures cannot be generalized, but it may give a graphical view of the real signals produced by the different sensors. It must be clarified that the sEMG was not used for the estimation of the model weights, only used in this case to check the correspondence between the speech signal and the accelerometer data to help in building on a kinematic explanation of this multimodal representation model, even though the full exploratory capability of sEMG was not exploited in the study being described.

In a first step forward, Figure 6.9 and Figure 6.10 present the results from estimating the acoustic and kinematic variables in the sagittal plane ($\Delta\mathbf{S}$ and $\Delta\mathbf{F}$) from the same participants. Interestingly, more regularity may be observed in the estimates from the HC participant than in those from the PD participant, although this behaviour cannot be generalised. A closer observation of the relationship between acoustic to kinematic variables from the HC participant (CM1) is presented as scatter plots in Figure 6.11, from the regression association of the signals in Figure 6.9 (b and c). It may be observed that all the plots show an eye-like pattern associated with phase shifts between each pair of acoustic and kinematic variables. This is due to time misalignments resulting from formant dynamics and explains the modest values of Pearson's correlation coefficients given in the four rightmost columns of Table 6.12 and Table 6.13.

The relationship between acoustic to kinematic variables ($\Delta\mathbf{S}$ and $\Delta\mathbf{F}$) given in both tables, expressed by the weights estimated from least squares linear regression requires a detailed analysis. The weights w_{11} and w_{12}, relate the first and second formant increments ΔF_1 and ΔF_2 with the horizontal displacement Δs_1. Weights w_{11} are negative and weights w_{12} are positive, relating a forward horizontal displacement of the jaw-tongue reference point with a descent of F_1 and with an ascent of F_2, respectively. Similar relationships may be observed in weights w_{21} and w_{22}, concerning the first and second formant increments ΔF_1 and ΔF_2 regarding the vertical displacement Δs_2. In this case, w_{21} is always negative, and w_{22} is always positive, because the upward movement of the jaw-tongue reference point is related to a descent of F_1 and with an ascent of F_2. This behaviour is aligned with the prediction of the acoustic-to-kinematic projection in the sense that increments of the first formant and decrements in the second formant are associated with the vertical pull-up action of the masseter (negative values of w_{21} and positive values of w_{22}).

A reflection is due at this point concerning the classical convention under the assumption of independent movements of the jaw and tongue in static vowel positions. The real phenomenon is a bit more complicated when it comes to dynamic diphthong movements, as the jaw and tongue cannot be considered moving independently. This is especially so regarding the diadochokinetic exercise used in the study. As examples of non-independent movement, it must be considered that depending on the position of the tongue (back or front), the sole movement of the jaw may produce the diphthong [wa] as in /wah-wah/ when the position of the tongue is back (static), or the diphthong [jeə] as in /yeah/ when the tongue position is front (static). In the first case both F_1 and F_2 ascend to higher values when the jaw descends, whereas in the second case F_1 ascends and F_2 descends when the jaw descends. In both cases, the tongue did not change its position, but both formants moved, as the jaw per se may modify completely the oral cavity, conditioning the movement of both formants. Conversely, should the jaw be kept in a stable medial position, the tongue per se could produce the diphthong [jʊ] like in /you/, where both formants descend from high to low values without the intervention of the jaw. These observations question the conventional view of independent relationships among dynamic formant movements and tongue and jaw positions, showing that the whole configuration of jaw and tongue is responsible for the production of important changes in formant positions, each system independently. In the present study, no independent movement of the jaw and tongue has been assumed.

Regarding the gradient-descent iteration dynamics expressed in Figure 6.12, it may be said that the patterns shown by the error surfaces of E1 and E2 are quite similar, and correspond to a convex surface with a single minimum in the shape of a *wadi*, producing a large descent at the beginning followed by shorter descent steps once the bottom of the *wadi* is approached. This effect may produce some unstable predictions of the step size in(5.12) and (5.13). The shape of this narrow valley distorts the space of solutions, as their geometrical place is the set of possible values of the pairs of coefficients $\{w_{11}, w_{12}\}$ and $\{w_{21}, w_{22}\}$. Slight variations in the estimation conditions may lead to different numerical solutions, all of them sharing the property of producing a quasi-optimal approximation. The shape of this geometrical place is a kind of narrow ellipse, approaching the limit of a straight line: $w_{12} = m_1 w_{11} + b_1$; $w_{22} = m_2 w_{21} + b_2 w_{22}$. The results of the estimation refinements in the model weights after the iteration process, given in Table 6.14 and Table 6.15 are modest, as expressed by the relative error reduction in per cent given in the rightmost columns. Reductions larger than 20% have been highlighted in bold.

The realignment process results, exemplified in Figure 6.13 and Figure 6.14 from a female HC participant (CF1) and a female PD participant (PF1), show a substantial increment in the correlation coefficients at the cost of introducing a relative delay between displacements and formant estimations, which in the case of the HC participant is around 26-28 ms, whereas in the case of the PD participant, it is much shorter (between 4-12 ms). In this second case, substantial increments in the correlation coefficients are also observed. This different behaviour may be explained by resonance dynamics in non-rigid tubes with losses, a hypothesis that would deserve an independent study per se.

Regarding the model weights after realignment, as given in Table 6.16 and Table 6.17 compared to those before realignment in Table 6.12 and Table 6.13, it may be seen that realignment does not change acoustic-to-kinematic projection properties of the model, because displacements and formant oscillations maintain the relative concordance observed before the realignment. A quadratic relative error between the horizontal and vertical reference point displacements estimated from accelerometry, and the regression-predicted values as obtained from expression (5.20) have been calculated for each model weight after signal realignment. These errors are reported in the rightmost columns of Table 6.16 and Table 6.17. These errors are larger for the cases where the distribution of $\mathbf{\Delta S}$ is more dispersed with respect to the regression line $w_{ij}\mathbf{\Delta F}$, and therefore they serve as an indication of the goodness of fit. The best case corresponds to the prediction of ΔS_2 relative to ΔF_1 from CF1 (0.33), and the worst case (0.73) corresponds to PF3 and PF4 (ΔS_1 vs ΔF_1 and ΔS_1 vs ΔF_2, respectively). The HC subset behaves slightly better (0.52 ± 0.09) than the PD subset (0.58 ± 0.10), although given the small sample sizes this observation is not generalizable.

An important remark comes from the observation that no relevant differences may be appreciated in the general pattern of the normalized model weights between HC and PD subsets given in Table 6.18, which allows for the definition of an overall average model, as shown in Section 6.2.4.

The realignment shifts ($\Delta t_{11}, \Delta t_{12}, \Delta t_{21}$, and Δt_{22}) associated with pair-wise weights $\{w_{11}, w_{12}\}$ and $\{w_{21}, w_{22}\}$ shown in Table 6.19 are in most cases within the range from 26-40 ms with some exceptions marked in bold (CM1, PM3, CF2, CF3, and PF1), and do not show relevant intra-speaker differences. All of them are multiples of the formant estimation time sampling rate of 2 ms.

Their origin may be a consequence of algorithmic delays produced during formant build-up because of resonance effects in non-rigid tubes with losses. The narrower the pole bandwidth associated with the formant, the shorter the time interval for the resonance format to grow in amplitude. As pole bandwidths are associated with the viscoelastic properties of the resonant cavities (oro-pharyngeal tract), more rigid and less viscid tissues would produce sharper poles and faster formant build-up, in opposition to more elastic and viscid tissues, producing duller poles and slower formant build-up, explaining the differences found in Table 6.19. It is known that the alterations in the viscoelastic properties of mucosal tissues are due to ageing and living style (loss of elastin and collagen, irritating agents, respiratory diseases, etc.), among other factors (Inamoto et al., 2015). Should this hypothesis be confirmed in a further study, these delays could serve as features of tissue ageing and decay (Hidalgo, Gómez, and Garayzábal, 2017).

The estimations of formant ascents and jaw-tongue reference point displacements ($\mathbf{\Delta S}$ and $\mathbf{\Delta F}$) are given in Table 6.20. There is not a clear tendency of displacements regarding gender, but it seems that PD participants produced larger displacements compared to HC participants on average. Whether these results might be associated with HD is a question that requires also further study, in the sense that a large weight magnitude means that small sweeps in formants are associated with large displacements in the reference point, otherwise, small weight magnitudes mean that small displacements in the reference point may produce large sweeps in formants. In this case, it may be hypothesized that if the effective oral cavity is reduced by HD, small changes in its cross-section could produce a substantial change in the formants.

The down-sampling procedure, as mentioned in Table App. 5 (Appendix I.8), has the added benefit of making the methodology presented in this work compatible with telephonic recordings not necessarily reliant on high-quality data.

This is possible due to the characteristics of the first and second formant ranges being below 3 kHz (Huang, Acero, and Hon, 2001), therefore, the bandwidth of the telephonic channel would not become an issue, as it is restricted to 4 kHz (sampling frequency of 8 kHz), allowing for enough spectral resolution. With this reflection in mind, a future line of study would be to explore the characterization of PD dysarthria by kinematic projection models on data collected remotely, such as the database taken within the project MonParLoc[13] (Palacios et al., 2020). It contains recordings produced by the eight diadochokinetic exercises mentioned in subsection Table App. 5 (Appendix I.8), including data from 45 PD participants of both genders with the collaboration of PD associations of Spain and Portugal, containing 696 valid utterances (by males) and 637 (by females) for diadochokinetic analysis. This platform is to be adapted to monitor patients with respiratory diseases, including covid-19, as this technology allows contact-free testing.

As a general comment derived from the overall perspective of this study, it must be highlighted that time realignment is a more relevant procedure than iteration refinement to reduce the estimation error, although a combination of both techniques could improve the estimation accuracy. This issue is being left for further study.

[13] https://monparloc.github.io

7.4 Study limitations due to methodological issues

Three different but tightly interconnected experimental scenarios have been presented, analysed, and discussed in chapters 5, 6, and 7 on the modelling of glottal signals and biomechanics, and mandibular articulation, aiming to characterise voice from a neuromotor point of view to provide deeper insight to PD HD description. These studies cover partial aspects of the problems related to the main objective; therefore, many limitations are to be expected due to the restrictions and assumptions made to focus on specific research questions. These limitations are briefly summarized in what follows.

Regarding the study supported by the estimation of the GFAD discussed in subsection 7.1, the summarized limitation factors found during the experimental execution to be taken into account are the following:

- The size of the database used. Although the number of phonation samples may seem reasonable (144 gender- and case-balanced, consisting of 6 groups of 24 subjects each) this amount should be enlarged to give more statistical relevance to results, because the specificities of each group (gender, and affection/nonaffection condition) reduce the number of samples per dataset to a very specific limit of 24, which is not considered large enough under inter-participant requirements.
- Age-matching between PD and HC population samples must be improved. Although both PD and HC distributions cannot be considered far apart by hypothesis testing, a better matching in the female datasets is to be considered.

- The low performance of hierarchical clustering regarding geometrical three-distance classification is to be reviewed. Other clustering methods are to be considered in future work (Ahmadlou and Adeli, 2010; Rafiei and Adeli, 2017). The main restriction in this sense is the low number of samples available per dataset, which is a sensitive issue when dealing with advanced machine learning methodologies.

- Possible information content alteration due to feature selection and potential over-fitting regarding feature selection methods used is to be investigated. This question, together with preserving explainability, interpretability, and clinical acceptance, are the most concerning issues.

- The relevance of different linguistic origins of the databases should be taken into account. Although the effect of this difference is considered negligible in this study as far as open vowel phonation is concerned, it should not be the case when running speech features are considered.

- The different instrumentation used in recording PD and HC subjects on one side and NS on the other is another question to be taken into account. The high quality of instrumentation and recording procedures, the use of downsampling, and the inverse filtering methodology ensured full compatibility in the present study, but this question is wide open to be taken into account in future studies as well.

Regarding the study supported by the biomechanical characterization of the glottal NMA in subsection 7.2, the summarized limitation factors found during the experimental execution in the study involving the VFBS to be taken into account are the following:

- The not exhaustive character of the experimental protocol might be a main drawback of the study because the size of the sample studied is a limitation to the findings observed. Nevertheless, it offers the possibility of conducting a more ambitious, extensive, and exhaustive study including the combination of EEG recordings and other traits derived from articulatory movements to include EEG, sEMG, and audio recordings. In this sense, recent advancements in brain connectivity combining EEG, MEG, fMRI, and NIRS characterization by graph theory (Ahmadlou et al., 2014; delEtoile and Adeli, 2017; Yaqub et al., 2022; Graña and Silva, 2021) and probabilistic neural networks (Hirschauer, Adeli, and Buford, 2015) could offer new insights for future studies.
- The application of this methodology to synchronized mixed EEG-audio databases (Verwoert et al., 2022) should offer new insights into speech production comprehension and eventually might allow further disclosure of brain functionality and physiological responses in PD.

Regarding the study based on the APM discussed in subsection 7.3, the summarized limitation factors found during the experimental execution to be taken into account are the following:

- The low number of participants included, which does not allow the generalization of results, has become a common problem found in the other studies as well, but in this case, it becomes a more critical issue, because the recording protocol is much more complex due to multi-trait signal acquisition. Some extra difficulties arise from the data acquisition procedures, which demand direct physical contact with participants. The number of participants per experiment is therefore reduced because the collection of multi-trait signals with standard off-the-shelf instrumentation is cumbersome and exhausting for the participants. Besides, the impact of the covid-19 pandemic has been devastating, especially during the phase of data gathering for the studies conducted (during 2020 and 2021), limiting the direct contact with volunteers especially sensitive to viral infection. This fact had an impact on the significance of the results produced.
- The intrinsic non-linear behaviour of the model needs further study, and its time variance evidenced by the correlation modelling reported needs a specific modelling effort out of the limits of the present study. The dependence of time alignment shifts on formant estimation is also an important issue. An effort in this sense is made to establish reliable relationships between formant bandwidths and delay estimations.

- In a joint recording of sEMG, EEG, 3DAcc, and speech, strict synchronisation is mandatory. Time lags may result in substantial model estimation accuracy reduction. This problem was dealt with by strict adaptive matching optimisation in Gómez-Rodellar et al. (2021b) regarding sEMG, 3DAcc, and speech. As the availability of perfectly synchronised multi-trait signals is a critical limitation, special recording platforms granting this strict requirement are needed, and in this sense, specific actions are planned to create them as portable compact equipment for easy and comfortable database recording, with possible application in the clinical assessment of neuropathological speech disorders.

- Besides, data acquisition is complicated by the difficulty found in participants perceiving and correctly implementing data acquisition protocols. This issue may become a source of variability affecting the robustness of the methodology and deserves a specific treatment in itself.

- Another problem is signal quality in sEMG and 3DAcc. The design of a new fixture integrating all the required sensors is necessary and envisioned in a future research line foreseeing the design of an autonomous unit for data gathering and clinical assessment.

- The inversion chain affecting articulation signals is more complex than the laryngeal one because many more muscles are involved, depending on different neural pathways (pharyngeal, velar, lingual, hypoglossal, orofacial).

CHAPTER 8

8 Conclusions

8.1 Contributions, findings, and insights

Given the polyhedric character of the three main studies included in the present manuscript, the nature of the conclusions to be summarized is also connected with each of the previously introduced studies. Regarding the study based on the GFAD in subsection 7.1, intended to assess the validity of glottal flow amplitude distributions in differentiating PD from HC and normative phonations based on MI contents, the following conclusions can be outlined:

- Phonation instability is present in the amplitude distributions of the glottal flow as far as HC and PD datasets are concerned. The NS dataset showed distributions that were bimodal in both amplitude extremes, contrary to distributions from aged HC and PD patients, which tend to show modes in the centre of the distribution amplitude.
- A clear separation between PD and NS distributions should be possible under relevant statistics using bisector and hierarchical clustering.
- On the contrary, it has been observed that the behaviour of glottal flow amplitude distributions from PD relative to HC distributions is not different enough to permit a clear distinction between both groups in terms of statistical relevance using bisector or hierarchical clustering.

- Nevertheless, good separation may be granted by feature selection and SVM methods (accuracies of 94.8% for the male subset and 92.8 for the female subset), although it must be taken into account that the resource to feature selection may hide contents relevant under IT, and produce data dependency, also affecting the interpretability of results.

- A clear separation between the older population subsets (PD and HC) relative to the mid-age normative subjects (NS) regarding phonation has been observed. This fact indicates that phonation alterations may appear and aggravate with age affecting both HC and PD subjects, in this last case independently, and concurrently with neuromotor deterioration.

- The difficulty in separating age-matched HC and PD phonation during the ON state requires further explanation. The effects of medication and vocal fold tissue deterioration due to ageing are plausible assumptions that would require further investigation.

- SVMs were employed to distinguish between the distributions from PD and HC speech, these two were the ones that exhibited the most significant overlap, with normative groups being omitted from the analysis. The SVM was supplied with the histograms with a hundred bins as features. The RELIEF algorithm was utilized to identify a subset of superior features, which in this context, are the histogram bins that account for the most difference between histograms. This methodology could potentially pinpoint regions of the histogram where the distributions diverge the most, warranting a more detailed examination of these areas. By scrutinizing the selected channels and employing analytical techniques such as Principal Component Analysis,

it could potentially lead to the development of a novel exploratory method for techniques based on Information Theory.

- Classification performance has been sacrificed in exchange for explainability.

Regarding the study on the use of the VFBS to explore the possibilities of predicting the interactions on the EEG-related ϑ-γ frequency bands of the NMA from the phonation acoustical signal discussed in subsection 7.2, the conclusions to be summarized are the following:

- Although the size of the sample studied is a limiting factor to the findings observed, and these results are tentative given we have limited participants which need to be verified in longer and larger trials, it may be asserted that the results are well aligned with ongoing studies in the field, especially in the use of log-likelihood ratios to assess functional improvements in phonation after rTMS. This finding is based on the possibility of using VFBS to serve as a correlate to monitor disruption events in vocal emission attributable to PD consequences. Consequently, visualizing EEG-related frequency bands could help in understanding some of the phenomena underlying vocal emission disruptions.

- The positive effects of rTMS are evident in the results, although observations on phonation instability behaviour from active and sham cases in pre-stimulus and post-stimulus recordings offered mixed results, concluding that this nature of findings needs to be further explored.

- There is a clear promise in these tentative findings grounded on previous work that can demonstrate differences in functional behaviour based purely on acoustic signal analysis, which delves deeper and provides insights over and above what can be assessed by classical acoustic analysis, and allows for the assessment of a closer-to-brain functionality and predict physiological responses in PD. However, it may be appreciated that there may be limitations and confounding factors when examining this methodology in detail, such as emotional or comorbid conditions that might alter the sensitivity of functional assessment, and their effects should be taken into account in future studies.

The study based on the APM discussed in subsection 7.3 has been conceived to provide further insights into the acoustic-to-kinematic model of the jaw-tongue articulation joint, based on preliminary approaches. In summary, the key findings derived from it can be summarized as follows:

- The feasibility and applicability of an acoustic-to-kinematic model to predict jaw-tongue joint kinematics from acoustic dynamics expressed in formants have been examined in depth, with special emphasis on weight estimation procedures.
- A weight estimation refinement method based on an iterative gradient algorithm has been explored. It has been found that a reduction in the estimation error functions is always possible at a reasonable number of iteration steps, although its benefit in terms of error reduction is not uniform, depending on specific participant data.

- A complementary correlation optimization study based on signal realignment has been also proposed, and a method to predict the relative time displacements to be included eventually in the acoustic-to-kinematic projection model has also been defined. Time delays from the male and female datasets used in the study have been estimated and discussed.
- A comparative study on the common characteristics of the estimated projection weights has also been carried on. An average gender-independent model has been established based on the dataset available, valid for both the HC and PD datasets.
- As a summarizing reflection, although many questions remain open and will require a deeper study in future work, it is essential to move along the progress on these methodologies allowing a remote monitoring of different diseases using convenient and cost-effective technology.

8.2 Future lines

The methodology presented in the study of phonation using the amplitude distributions of the glottal signals in the time domain may be applied in different fields of PD phonation monitoring, especially when frequent longitudinal patient evaluations have to be carried on, as in estimating the performance of binaural stimulations (Gálvez et al., 2018; Gálvez et al., 2019) and speech therapies (Ramig, Fox, and Sapir, 2008) in the potential improvement of patients' neuromotor function. This possibility could be further capitalized by end-to-end machine learning approaches.

The studies presented and discussed regarding the use of phonation and articulation mechanistic models in the characterization of PD speech are to be considered the first steps towards a more ambitious objective of describing central and peripheral brain activity from speech acoustics by inverse model methods. In this sense, a first contribution has been already published (Gómez-Rodellar et al. 2023), although this research line is not mature enough to be considered already consolidated, therefore, it has not been included as a consistent body of knowledge in the present manuscript. The small number of participants available for the study at the time that the research was accomplished, and the lack of ground truth support availing the results were the main limitations to consider it more than a pilot study.

The validation of the sEMG-EEG cortico-muscular coupling projection models requires correlation matching techniques, based on wavelet transforms, or transfer entropy. Joint concurrently recorded multi-trait signals involving EEG, sEMG, 3DAcc, and speech are needed. A possible approach to overcome these difficulties might be afforded by using third-party databases where speech and direct brain activity are found (Verwoert et al., 2022), and applying advanced methods based on MI. This is one of the most promising future lines of study to advance in this rather promising research field, but there is no known database including the four traits. In this sense, new actions are to be taken as a continuation of the research undertaken in the present PhD thesis proposal, by formulating a project proposal, pursuing the design and implementation of an autonomous unit for multi-trait strict synchronous recording, with immediate application in projection model research, and clinical assessment of neuropathological speech studies.

The inversion chain affecting glottal signals offers new challenges, as determining if the agonist-antagonist of the cricothyroid and thyroarytenoid joints is coordinated from a

bulbar circuit, from an extrapyramidal ganglion, or the independent poles in the laryngeal control cortex.

The implementation of these methodologies can allow to assess and quantify the effects of rehabilitiative therapies, by providing objective markers that allow to quantify, and more importantly, compare temporally the progression and effectiveness of a tiven treatment protocol, rehabilitation and medication effects, as well as providing PwP, caregivers and support staff an observable metric that allows to establish a progression and a sense of assurance where data is sometimes difficult to visualize.

REFERENCES

Abdi, H. and Molin, P., (2007) Lilliefors/Van Soest's test of normality, Encyclopedia of Measurement and Statistics, N. Salkind (Ed.). Thousand Oaks, CA (2007).

Abitbol, J., Abitbol, P., and Abitbol, B. (1999) Sex hormones and the female voice. J. Voice 13, 424–446. https://doi.org/10.1016/S0892-1997(99)80048-4.

Ackermann, H., and Ziegler, W. (1991). Articulatory deficits in parkinsonian dysarthria: an acoustic analysis. Journal of Neurology, Neurosurgery & Psychiatry, 54(12), 1093-1098. http://dx.doi.org/10.1136/jnnp.54.12.1093.

Aguilera, M., Douchamps, V., Battaglia, D., and Goutagny, R. (2022). How Many Gammas? Redefining Hippocampal Theta-Gamma Dynamic During Spatial Learning. Frontiers in behavioral neuroscience, 16, 811278. http://doi.org/10.3389/fnbeh.2022.811278.

Ahmadlou, M., and Adeli, H. (2010) Enhanced Probabilistic Neural Network with Local Decision Circles: A Robust Classifier, Integrated Computer-Aided Engineering 17(3) 197-210. https://doi.org/10.3233/ICA-2010-0345.

Ahmadlou, M., Adeli, A., Bajo, R., and Adeli, H. (2014). Complexity of functional connectivity networks in mild cognitive impairment subjects during a working memory task. Clinical Neurophysiology, 125(4), 694-702. https://doi.org/10.1016/j.clinph.2013.08.033.

Ahmadlou, M., and Adeli, H. (2017) Complexity of weighted graph: A new technique to investigate structural complexity of brain activities with applications to aging and autism. Neuroscience Letters, 650 103-108. https://doi.org/10.1016/j.neulet.2017.04.009.

Alku, P. (1992). Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. Speech communication, 11(2-3), 109-118. https://doi.org/10.1016/0167-6393(92)90005-R.

Alku, P., Murtola, T., Malinen, J., Kuortti, J., Story, B., Airaksinen, M., Salmi, M., Vilkman, E., Geneid, A., (2019) OPENGLOT-An open environment for the evaluation of glottal inverse filtering, Speech Communication 107 38-47, https://doi.org/10.1016/j.specom.2019.01.005.

Álvarez-Marquina, A., Gómez, A., Palacios-Alonso, D., Mekyska, J., Tsanas, A., Gómez, P., and Martínez, R. (2020). Parkinson's Disease Glottal Flow Characterization: Phonation Features vs Amplitude Distributions. In BIOSIGNALS (pp. 359-368). https://doi.org/10.5220/0009189403590368.

Álvarez-Marquina, A., Gómez-Rodellar, A., Gómez-Vilda, P., Palacios-Alonso, D., and Díaz-Pérez, F. (2022, May). Identification of Parkinson's Disease from Speech Using CNNs and Formant Measures. In *International Work-Conference on the Interplay Between Natural and Artificial Computation* (pp. 332-342). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-06242-1_33.

Arora, S., Baghai-Ravary, L. and Tsanas, A. (2019) Developing a large scale population screening tool for the assessment of Parkinson's disease using telephone-quality speech, Journal of the Acoustical Society of America; 145 (5), 2871-2884. https://doi.org/10.1121/1.5100272.

Arora, S. and Tsanas, A. (2021). Assessing Parkinson's disease at scale using telephone-recorded speech: insights from the Parkinson's voice initiative, Diagnostics, Vol. 11(1) e1892. https://doi.org/10.3390/diagnostics11101892.

Arora, S., Lo, C., Hu, M., and Tsanas, A. (2021). Smartphone speech testing for symptom assessment in rapid eye movement sleep behavior disorder and Parkinson's disease, IEEE Access, 9, 44813-44824. https://doi.org/10.1109/ACCESS.2021.3057715.

Atal, B. S and Hanauer, S. L. (1971) Speech Analysis and Synthesis by Linear Prediction of the Speech Wave, J. Acoust. Soc. Am. 50 (2) 637-655. https://doi.org/10.1121/1.1912679.

Baken, R. J., and Orlikoff, R. F. (2000) Clinical Measurement of Speech and Voice, 2nd ed. San Diego, CA, USA: Singular Thomson Learning, 2000.

Barios, J. A., Ezquerro, S., Bertomeu-Motos, A., Catalan, J. M., Sanchez-Aparicio, J. M., Donis-Barber, L., Fernández, E., and Garcia-Aracil, N. (2021). Movement-related EEG oscillations of contralesional hemisphere discloses compensation mechanisms of severely affected motor chronic stroke patients. International journal of neural systems, 31(12), 2150053. https://doi.org/10.1142/S0129065721500532.

Barzilai, J. and Borwein, J. M. (1988) Two-point Step Size Gradient Methods, IMA Journal of Numerical Analysis; 8: 141-148. https://doi.org/10.1093/imanum/8.1.141.

Belalcázar-Bolaños, E. A., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Orozco-Arroyave, J. R. (2015). Nonlinear glottal flow features in Parkinson's disease detection, Proc. 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA), Bogota, Colombia 1–6. https://doi.org/10.1109/STSIVA.2015.7330420.

Berry, D. A. (2001). Mechanisms of modal and nonmodal phonation. Journal of Phonetics, 29(4), 431-450. https://doi.org/10.1006/jpho.2001.0148.

Bhat, S., Acharya, U. R., Dadmehr, N., and Adeli, H. (2018). Parkinson's Disease: cause factors, measurable indicators, and early diagnosis, Computers in Biology and Medicine, 102 234-241. https://doi.org/10.1016/j.compbiomed.2018.09.008.

Blitzer, A., Crumley, R. L., Dailey, S. H., Ford, C. N., Floeter, M. K., Hillel, A. D., Hoffman, H. T., Ludlow, C. L., Merati, A., Munin, M. C., Robinson, L. R., Rosen, C., Saxon, K. G., Sulica, L., Thibeault, S. L., Titze, I. R., Woo, P., and Woodson, G. E. (2009). Recommendations of the Neurolaryngology Study Group on laryngeal electromyography. Otolaryngology—Head and Neck Surgery, 140(6), 782-793. https://doi.org/10.1016/j.otohns.2009.01.026.

Blom, K., Emmelot-Vonk, M. H., and Koek, H. D. L. (2013). The influence of vascular risk factors on cognitive decline in patients with dementia: a systematic review. *Maturitas*, 76(2), 113-117. https://doi.org/10.1016/j.maturitas.2013.06.011.

Bloem, B. R., Okun, M. S. and Klein, C. (2021). Parkinson's disease. The Lancet, 397(10291), (2021) 284-2303. https://doi.org/10.1016/S0140-6736(21)00218-X.

Boller, F., and Forbes, M. M. (1998). History of dementia and dementia in history: an overview. *Journal of the neurological sciences*, 158(2), 125-133. https://doi.org/10.1016/S0022-510X(98)00128-2.

Braak, H., Del Tredici, K., Rüb, U., A.I de Vos, R., Jansen Steur, E. N. H., Braak, E., Staging of brain pathology related to sporadic Parkinson's disease, Neurobiology of Aging, Volume 24, Issue 2, 2003, Pages 197-211, ISSN 0197-4580, https://doi.org/10.1016/S0197-4580(02)00065-9.

Brabenec, L., Mekyska, J., Galaz, Z., and Rektorová, I. (2017). Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation, J. Neural Transm. 124(3) 303–334. https://doi.org/10.1007/s00702-017-1676-0.

Brabenec, L., Klobusiakova, P., Simko, P., Kostalova, M., Mekyska, J., and Rektorova, I. (2021). Non-invasive brain stimulation for speech in Parkinson's disease: A randomized controlled trial. *Brain Stimulation*, *14*(3), 571-578. https://doi.org/10.1016/j.brs.2021.03.010.

Brambilla, C., Pirovano, I., Mira, R. M., Rizzo, G., Scano, A., and Mastropietro, A. (2021). Combined use of EMG and EEG techniques for neuromotor assessment in rehabilitative applications: a systematic review. Sensors, 21(21), 7014. https://doi.org/10.3390/s21217014.

Brittain, J. S., and Brown, P. (2014). Oscillations and the basal ganglia: motor control and beyond. Neuroimage, 85, 637-647. https://doi.org/10.1016/j.neuroimage.2013.05.084.

Brown, S., Laird, A. R., Pfordresher, P. Q., Thelen, S. M., Turkeltabu, P., Liotti, M. (2009). The somatotopy of speech: Phonation and articulation in the human motor cortex. Brain and Cognition 70 31-41. https://doi.org/10.1016/j.bandc.2008.12.006.

Brückl, M., Ghio, A., and Viallet, F. (2018) Measurement of Tremor in the Voices of Speakers with Parkinson's Disease. Procedia Computer Science, 128, 47-54. https://doi.org/10.1016/j.procs.2018.03.007.

Burk, B. R., and Watts, C. R. (2019). The effect of Parkinson disease tremor phenotype on cepstral peak prominence and transglottal airflow in vowels and speech. Journal of Voice, 33(4), 580-e11. https://doi.org/10.1016/j.jvoice.2018.01.016.

Caiola, M., and Holmes, H. M. (2019). Model and Analysis for the Onset of Parkinsonian Firing Patterns in a Simplified Basal Ganglia, International Journal of Neural Systems, 29(1), 1850021 (21 pages). https://doi.org/10.1142/S0129065718500211.

Caligiuri, M. P. (1989). The influence of speaking rate on articulatory hypokinesia in Parkinsonian dysarthria. Brain and Language, 36(3), 493-502. https://doi.org/10.1016/0093-934X(89)90080-1.

Cavanaugh, J. E., and Neath, A. A. (2019) The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. WIREs Comput Stat, 11:3 e 1460. https://doi.org/10.1002/wics.1460.

Chang, C., and Lin, C. J. (2011) LIBSVM: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology, 2(3) Art. 27. https://doi.org/10.1145/1961189.1961199.

Chen, G., Kreiman, J., Shue, Y. L., and Alwan, A. (2011). Acoustic correlates of glottal gaps. In Twelfth Annual Conference of the International Speech Communication Association. https://doi.org/10.1017/S002510031400019X.

Chiang, J., Wang, Z. J., and McKeown, M. J. (2012) A multiblock PLS model of cortico-cortical and corticomuscular interactions in Parkinson's disease, Neuroimage, 63 1498-1509. https://doi.org/10.1016/j.neuroimage.2012.08.023.

Cirillo, D.; Catuara-Solarz, S.; Morey, C.; Guney, E.; Subirats, L.; Mellino, S.; Gigante, A.; Valencia, A.; Rementeria, M.J.; Chadha, A.S.; et al. (2020) Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ Digit. Med. 3, 81. https://doi.org/10.1038/s41746-020-0288-5.

Colamarino, E., De Seta, V., Masciullo, M., Cincotti, F., Mattia, D., Pichiorri, F., and Toppi, J. (2021). Corticomuscular and intermuscular coupling in simple hand movements to enable a hybrid brain–computer interface. International Journal of Neural Systems, 31(11), 2150052. https://doi.org/10.1142/S0129065721500520.

Cover, T. M., Thomas, J. A. (2012). Elements of Information Theory, John Wiley and Sons, Hoboken, NJ, USA.

Christine, C. W., and Aminoff, M. J. (2004). Clinical differentiation of parkinsonian syndromes: prognostic and therapeutic relevance. The American journal of medicine, 117(6), 412-419. https://doi.org/10.1016/j.amjmed.2004.03.032.

Darbin, O., and Montgomery, E.B. (2022) Challenges for future theories of Parkinson pathophysiology. Neuroscience Research 177 1-7. https://doi.org/10.1016/j.neures.2021.11.010.

Dauer, W., Przedborski, S. (2003). Parkinson's disease: Mechanisms and models. Neuron, 39(6) 889–909. https://doi.org/10.1016/S0896-6273(03)00568-3.

Davis, P. J., Zhang, S. P., Winkworth, and Bandler, R. (1996). Neural Control of Vocalization: Respiratory and Emotional Influences, Journal of Voice 10(1) 23-38. https://doi.org/10.1016/S0892-1997(96)80016-6.

De Lau, L. M., and Breteler, M. M. (2006). Epidemiology of Parkinson's disease. The Lancet Neurology, 5(6), 525-535. https://doi.org/10.1016/S1474-4422(06)70471-9.

delEtoile, J., and Adeli, H. (2017) Graph Theory and Brain Connectivity in Alzheimer's Disease. The Neuroscientist, 23:6 616-626. https://doi.org/10.1177/1073858417702621.

Deller, J. R., Proakis, J. G. and Hansen, J. H. L. (1993). Discrete-Time Processing of Speech Signals. Macmillan, NewYork.

Demonet, J. F., Thierry, G., Cardebat. D. (2005). Renewal of the Neurophysiology of Language: Functional Neuroimaging. Physiol. Rev. 85, 49-95, https://doi:10.1152/physrev.00049.2003.

Dietrich, M., Andreatta, R. D., Jiang, Y., and Stemple, J. C. (2020). Limbic and cortical control of phonation for speech in response to a public speech preparation stressor. Brain Imaging and Behavior 14, 1696-1713. https://doi.org/10.1007/s11682-019-00102-x.

Dorsey, E. R., Sherer, T., Okun, M. S., Bloem, B. R. (2018a) The Emerging Evidence of the Parkinson Pandemic. Journal of Parkinson's Disease 8 S3-S8. https://doi.org/10.3233/JPD-181474.

Dorsey, E.R., Elbaz, A., Nichols, E., Abd-Allah, F., Abdelalim, A., Adsuar, J.C., Ansha, M.G., Brayne, C., Choi, J.Y., Collado-Mateo, D., Dahodwala, N., Do, H. P., Edessa, D., Endres, M., Fereshtehnejad, S. M., Foreman, K. J., Gankpe, F. G., Gupta, R., Hamidi, S., Hankey, G. J., Hay, S. I., Hegazy, M. I., Hibstu, D. T., Kasaeian, A., Khader, Y., Khalil, I., Khang, Y. H., Kim, Y. J., Kokubo, Y., Logroscino, G., Massano, J., Mohamed Ibrahim, N. M., Mohammed, M. A., Mohammadi, A., Moradi-Lakeh, M., Naghavi, M., Thanh Nguyen, B. T., Nirayo, Y. L., Ogbo, F. A., Ojo Owolabi, M. O., Pereira, D. M., Postma, M. J., Qorbani, M., Rahman, M. A., Roba, K. T., Safari, H., Safiri, S., Satpathy, M., Sawhney, M., Shafieesabet, A., Shiferaw, M. S., Smith, M., Szoeke, C. E. I., Tabarés, R., Truong, N. T., Ukwaja, K. N., Venketasubramanian, N., Villafaina, S., Weldegwergs, K. G., Westerman, R., Wijeratne, T., Winkler, A. S., Xuan, B. T., Yonemoto, N., Feigin, F. L., Vos, T., Murray, C. J. L. (2018b) Global, regional, and national burden of Parkinson's disease, 1990-2016: A systematic analysis for the Global Burden of Disease Study. Lancet Neurol. 17, 939–953. https://doi.org/10.1016/S1474-4422(18)30295-3.

Dromey, C., Jang, G. O., Hollis, K. (2013). Assessing correlations between lingual movements and formants. Speech Communication 55:2 315-328, https://doi:10.1016/j.specom.2012.09.001.

Drugman, T., Bozkurt, B., and Dutoit, T. (2012). A comparative study of glottal source estimation techniques. Computer Speech & Language, 26(1), 20-34. https://doi.org/10.1016/j.csl.2011.03.003.

Dugger, B. N., and Dickson, D. W. (2017). Pathology of neurodegenerative diseases. *Cold Spring Harbor perspectives in biology*, 9(7), a028035. https://doi.org/10.1101/cshperspect.a028035.

Duffy, J. R. (2013). Motor Speech Disorders: Substrates, Differential Diagnosis, and Management, 3rd Ed., Elsevier, River Lane, St. Louis, Missouri, USA.

Durbin, J. (1960). The fitting of time-series models. Revue de l'Institut International de Statistique, 233-244. https://www.jstor.org/stable/1401322.

Dworkin, J. P. and R. J. Meleca (1997) Vocal Pathologies: Diagnosis, Treatment and Case Studies, Singular Pub. Group, San Francisco, 1997.

Edelman, A., and Kostlan, E. (1995). How many zeros of a random polynomial are real?. Bulletin of the American Mathematical Society, 32(1), 1-37. https://doi.org/10.1090/S0273-0979-1995-00571-9.

Emamzadeh, F. N., and Surguchov, A. (2018). Parkinson's disease: biomarkers, treatment, and risk factors. *Frontiers in neuroscience*, 12, 612. https://doi.org/10.3389/fnins.2018.00612.

Endres, D. M., and Schindelin, J. E. (2003). A New Metric for Probability Distributions, IEEE Trans. on Information Theory 49(7) (2003) 1858-1860. https://doi.org/10.1109/TIT.2003.813506.

Epperson, J. F. (2013). An introduction to numerical methods and analysis. John Wiley & Sons.

Fant, G. (1960) Acoustic Theory of Speech Production, Mouton, The Hague. The Netherlands.

Fant, G. (1981) The source filter concept in voice production. STL-QPSR 1, 21–37.

Fant, G., Liljencrants, J., and Lin, Q. (1985) A four-parameter model of glottal flow, STL-QPRS 26(4) 1-13.

Feng, N., Hu, F., Wang, H., and Zhou, B. (2021). Motor intention decoding from the upper limb by graph convolutional network based on functional connectivity. International Journal of Neural Systems, 31(12), 2150047. https://doi.org/10.1142/S0129065721500477.

Filipović, S. R., Rothwell, J. C., van de Warrenburg, B. P., and Bhatia, K. (2009). Repetitive transcranial magnetic stimulation for levodopa-induced dyskinesias in Parkinson's disease. Movement disorders: official journal of the Movement Disorder Society, 24(2), 246-253. https://doi.org/10.1002/mds.22348.

Forrest, K., Weismer, G., and Turner, G. S. (1989). Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. The Journal of the Acoustical Society of America, 85(6), 2608-2622. https://doi.org/10.1121/1.397755.

Forrest, K., and Weismer, G. (1995). Dynamic aspects of lower lip movement in Parkinsonian and neurologically normal geriatric speakers' production of stress. Journal of Speech, Language, and Hearing Research, 38(2), 260-272. https://doi.org/10.1044/jshr.3802.260.

Fu, Q., and Murphy, P. (2006) Robust Glottal Source Estimation Based on Joint Source-Filter Model Optimization, IEEE Trans. on Audio, Speech, and Lang. Proc. 14(2) 492-501. https://doi.org/10.1109/TSA.2005.857807.

Gálvez, G., Recuero, M., Canuet, L., and Del Pozo, F. (2018) Short-term effects of Binaural Beats on EEG power, functional connectivity, cognition, gait and anxiety in Parkinson's Disease, International Journal of Neural Systems 28(5) 1750055 (16 pages). https://doi.org/10.1142/S0129065717500551.

Gálvez, G., Gómez, A., Palacios, D., De Arcas, G., Gómez, P. (2019) Temporal Reversion of Phonation Instability in Parkinson's Disease by Neuroacoustical Stimulation, Proc. MAVEBA 2019, C. Manfredi (Ed.), Florence University Press 21-24.

Gao, Y., Ren, L., Li, R., and Zhang, Y. (2018) Electroencephalogram–electromyography coupling analysis in stroke based on symbolic transfer entropy. Frontiers in neurology, 8, 716. https://doi.org/10.3389/fneur.2017.00716.

García, B., Fernández, A., Martínez, A., Alcaraz, R., and Novais, P. (2022). Evaluation of Brain Functional Connectivity from Electroencephalographic Signals Under Different Emotional States. International Journal of Neural Systems, 32(10), 2250026. https://doi.org/10.1142/S0129065722500265.

García, J. L., and Díaz, J. C. (2014). Finding the number of roots of a polynomial in a plane region using the winding number. Computers & Mathematics with Applications, 67(3), 555-568. https://doi.org/10.1016/j.camwa.2013.11.013.

Georgiou, T. T., and Lindquist, A. (2003). Kullback-Leibler approximation of spectral density functions. IEEE Transactions on Information Theory, 49(11), 2910-2917. https://doi.org/10.1109/TIT.2003.819324.

Gillivan-Murphy, P. (2013) Voice tremor in Parkinson's disease (PD): identification, characterisation and relationship with speech, voice and disease variables (Doctoral dissertation, Newcastle University). https://theses.ncl.ac.uk/jspui/bitstream/10443/2170/1/Gillivan-Murphy%2013.pdf.

Gillivan-Murphy, P., Miller, N., and Carding, P. (2018). Voice tremor in Parkinson's disease: an acoustic study. Journal of Voice, 33(4), 526-535. https://doi.org/10.1016/j.jvoice.2017.12.010.

Goberman, A. M. and Coelho, C. (2002). Acoustic analysis of Parkinsonian speech I: Speech characteristics and L-Dopa therapy. Neurorehabilitation 17, 237-246, https://doi.org/10.3233/NRE-2002-17310.

Goberman, A., Coelho, C., and Robb, M. (2002). Phonatory characteristics of parkinsonian speech before and after morning medication: the ON and OFF states. Journal of communication disorders, 35(3), 217-239. https://doi.org/10.1016/S0021-9924(01)00072-7.

Godino-Llorente, J. I., Shattuck-Hufnagel, S., Choi, J. Y., Moro, L., Gómez, J. A. (2017) Towards the identification of Idiopathic Parkinson's Disease from the speech. New articulatory kinematic biomarkers. PLOS One, December 14, 2017 (35 pages), https://doi.org/10.1371/journal.pone.0189583.

Goetz, C. G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G. T., Counsell, C., Giladi, N., Holloway, R. G., Moore, C. G., Wenning, G. K., Yahr, M. D., Seidl, L. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: Status and recommendations The Movement Disorder Society Task Force on rating scales for Parkinson's disease, Movement Disorders 19 (9) (2004) 1020–1028. https://doi.org/10.1002/mds.20213.

Goetz, C. G., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stebbins, G. T., Stern, M. B., Tilley, B. C., Dodel, R., Dubois, B., Holloway, R., Jankovic, J., Kulisevsky, J., Lang, A. E., Lees, A., Leurgans, S., LeWitt, P. A., Nyenhuis, D., Olanow, C. W., Rascol, O., Schrag, A., Teresi, J. A., Van Hilten, J. J., LaPelle, N. (2007) Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Process, format, and clinimetric testing plan. Mov Disord 22:41–47. https://doi.org/10.1002/mds.21198.

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M. B., Dodel, R., Dubois, B., Holloway, R., Jankovic, J., Kulisevsky, J., Lang. A. E., Lees, A., Leurgans, S., LeWitt, P. A., Nyenhuis, D., Olanow, C. W., Rascol, O., Schrag, A., Teresi, J. A., van Hilten, J. J., LaPelle, N., Movement Disorder Society URTF (2008) Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Movement disorders: official journal of the Movement Disorder Society, 23(15), 2129-2170. https://doi.org/10.1002/mds.22340.

Gómez-Rodellar, A., De Arcas, G., Gómez, P., Álvarez-Marquina, A. and López, J. M. (2018) Estimating Facial Neuromotor Activity from sEMG and Accelerometry for Speech Articulation", Proc. of the 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA) 1-6, Rome https://doi.org/10.1109/MeMeA.2018.8438744.

Gómez-Rodellar, A., Tsanas, A., Gómez, P., Palacios, D., Álvarez-Marquina, A., Martínez. R. (2019a) A Neuromechanical Model of Jaw-Tongue Articulation in Parkinson's Disease Speech. Proc. of MAVEBA 19; 25-28, Firenze University Press, December 17-19.

Gómez-Rodellar, A., Palacios, D., Mekyska, J., Álvarez-Marquina, A., and Gómez, P. (2019b) Comparing Parkinson's Disease Dysarthria and Aging Speech using Articulation Kinematics, Proc. 12th International Joint Conference on Biomedical Engineering Systems and Technologies, F. Putze, A. Fred and H. Gamboa (Eds.), SCITEPRESS, Lisbon, Portugal 52-61. https://doi.org/10.5220/0007355700520061.

Gómez-Rodellar, A., Palacios, D., Ferrández-Vicente, J. M., Mekyska, J., Álvarez-Marquina, A., and Gómez, P. (2019c) Evaluating Instability on Phonation in Parkinson's Disease and Aging Speech, Lecture Notes on Computer Science, 11487(2) 340-351. https://doi.org/10.1007/978-3-030-19651-6_33.

Gómez-Rodellar, A., Palacios, D., Ferrández-Vicente, J. M., Mekyska, J., Álvarez-Marquina, A., and Gómez, P. (2020a). A methodology to differentiate Parkinson's disease and aging speech based on glottal flow acoustic analysis. International Journal of Neural Systems, 30(10), 2050058. https://doi.org/10.1142/S0129065720500586.

Gómez-Rodellar, A., Tsanas, A., Gómez-Vilda, P., Álvarez-Marquina, A., and Palacios-Alonso, D. (2020b). Individual Mandibular Motor Actions Estimated from Speech Articulation. In LREC 2020 Language Resources and Evaluation Conference 11-16 May 2020 (p. 74).

Gómez-Rodellar, A., Gómez-Vilda, P., Palacios, D., Rodellar, V., Nieto, V., Álvarez-Marquina, A., and Tsanas, A. (2021a). A Neuromotor to Acoustical Jaw-Tongue Projection Model With Application in Parkinson's Disease Hypokinetic Dysarthria. Frontiers in Human Neuroscience, 15, 622825. https://doi.org/10.3389/fnhum.2021.622825.

Gómez-Rodellar, A., Tsanas, A., Gómez, P., Palacios, D., Rodellar, V. and Álvarez-Marquina, A. (2021b). Acoustic to Kinematic Projection in Parkinson's Disease Dysarthria. Biomedical Signal Processing and Control 66 102422. https://doi.org/10.1016/j.bspc.2021.102422.

Gómez-Rodellar, A. and Tsanas, A. (2021c). F0 Estimation in Irregular Vocal Emissions using Ridge Detection Methods. Claudia Manfredi (Ed.), Models and Analysis of Vocal Emissions for Biomedical Applications: 12th International Workshop, December, 14-16, 2021, Firenze University Press (www.fupress.com), pp. 43-46. https://doi.org/10.36253/978-88-5518-449-6.

Gómez-Rodellar, A., Mekyska, J., Brabenec, L., Simko, P., Rektorová, I., Gómez, P., and Tsanas, A. (2021d). Longitudinal Effect of Repetitive Transcranial Magnetic Stimulation on Phonation in a Patient with Parkinson's Disease: A Case Study. Claudia Manfredi (Ed.), Models and Analysis of Vocal Emissions for Biomedical Applications: 12th International Workshop, December, 14-16, 2021, Firenze University Press (www.fupress.com), pp. 157-160.. https://doi.org/10.36253/978-88-5518-449-6.

Gómez-Rodellar, A., Gómez-Vilda, P., Ferrández-Vicente, J., and Tsanas, A. (2022a). Characterizing Masseter Surface Electromyography on EEG-Related Frequency Bands in Parkinson's Disease Neuromotor Dysarthria. In International Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 219-228). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-06242-1_22.

Gómez-Rodellar, A., Mekyska, J., Gómez-Vilda, P., Brabenec, L., Simko, P., and Rektorová, I. (2022b). Evaluation of TMS Effects on the Phonation of Parkinson's Disease Patients. In International Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 199-208). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-06242-1_20.

Gómez-Rodellar, A., Mekyska, J., Gómez-Vilda, P., Brabenec, L., Šimko, P., and Rektorová, I. (2023). A Pilot Study on the Functional Stability of Phonation in EEG Bands After Repetitive Transcranial Magnetic Stimulation in Parkinson's Disease. International Journal of Neural Systems, 2350028. https://doi.org/10.1142/S0129065723500284.

Gómez-Vilda, P., Fernández, R., Nieto, A., Díaz, F., Fernández, F. J., Rodellar-Biarge, V., Álvarez-Marquina, A., and Martínez, R. (2007). Evaluation of voice pathology based on the estimation of vocal fold biomechanical parameters. Journal of Voice, 21(4), 450-476. https://doi.org/10.1016/j.jvoice.2006.01.008.

Gómez-Vilda, P., Fernández, R., Rodellar-Biarge, V., Nieto, V., Álvarez-Marquina, A., Mazaira, L. M., Martínez, R., Godino, J. I. Glottal Source biometrical signature for voice pathology detection, Speech Communication 51(9) (2009) 759-781. https://doi.org/10.1016/j.specom.2008.09.005.

Gómez-Vilda, P., Mazaira, L. M., Martínez, R., Álvarez-Marquina, A., Hierro, J. A., and Nieto, R. (2012). Distance metric in forensic voice evidence evaluation using dysphonia-relevant features. In Proceedings of the VI Meeting of Biometric Recognition of Persons, pp. 169-178.

Gómez-Vilda, P., Ferrández-Vicente, J. M., and Rodellar-Biarge, V. (2013). Simulating the phonological auditory cortex from vowel representation spaces to categories. Neurocomputing, 114, 63-75. https://doi.org/10.1016/j.neucom.2012.07.036.

Gómez-Vilda, P., Mekyska, J., Ferrández-Vicente, J. M., Palacios, D., Gómez-Rodellar, A., Rodellar-Biarge, V., Galaz, Z., Smekal, Z., Eliasova, I., Kostalova, M., Rektorová, I. (2017a) Parkinson Disease Detection form Speech Articulation Neuromechanics. Frontiers in Neuroinformatics 11, 1-17, https://doi.org/10.3389/fninf.2017.00056.

Gómez-Vilda, P., Mekyska, J., Gómez-Rodellar, A., Palacios, D., Rodellar-Biarge, V. and Álvarez-Marquina, A., (2017b) Articulation Dynamics in Parkinson Dysarthria, Proc. of MAVEBA 17; 81-84 Firenze University Press, December 13-15.

Gómez-Vilda, P., Gómez-Rodellar, A., Ferrández-Vicente, J. M., Mekyska, J., Palacios, D., Rodellar, V., Álvarez-Marquina, A., Smekal, Z., Eliasova, I., Kostalova, M., Rektorová, I. (2019a) Neuromechanical Modelling of Articulatory Movements from Surface Electromyography and Speech Formants. International Journal of Neural Systems, 29:2, 1850039, https://doi.org/10.1142/S0129065718500399.

Gómez-Vilda, P., Mekyska, J., Gómez-Rodellar, A., Palacios, D., Rodellar-Biarge, V., Álvarez-Marquina, A. (2019b) Characterization of Parkinson's disease dysarthria in terms of speech articulation kinematics. Biomedical Signal Processing and Control, 52, 312-320, https://doi.org/10.1016/j.bspc.2019.04.029.

Gómez-Vilda, P., Galaz, Z., Mekyska, J., Ferrández-Vicente, J. M., Gómez-Rodellar, A., Palacios, D., Smekal, Z., Eliasova, I., Kostalova, M., Rektorová, I. (2019c) Vowel Articulation Dynamic Stability Related to Parkinson's Disease Rating Features: Male Dataset, Int. Journal of Neural Systems 28(2), 1850037 (13pages). https://doi.org/10.1142/S0129065718500375.

Gómez-Vilda, P., Gómez-Rodellar, A., Palacios-Alonso, D., and Tsanas, A. (2021). Performance of monosyllabic vs multisyllabic diadochokinetic exercises in evaluating Parkinson's disease hypokinetic dysarthria from fluency distributions. In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies—BIOSIGNALS* (pp. 114-123). https://www.researchgate.net/publication/370701277.

Gómez-Vilda, P., Gómez-Rodellar, A., Palacios-Alonso, D., Álvarez-Marquina, A., and Tsanas, A. (2022). Characterization of Hypokinetic Dysarthria by a CNN Based on Auditory Receptive Fields. In *International Work-Conference on the Interplay Between Natural and Artificial Computation* (pp. 343-352). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-06242-1_34.

Gómez-Vilda, P., Mekyska, J., Brabenec, L., Šimko, P., Rektorová, I., Gómez-Rodellar, A., and Rodellar-Biarge, V. (2023a). Description of PD Phonation in Terms of EEG-Related Frequency Bands. In Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2023) - Volume 4: BIOSIGNALS, pages 226-233. https://doi.org/10.5220/0011669100003414.

Gómez-Vilda, P., Gómez-Rodellar, A., Palacios, D., and Tsanas, A. (2023). Evaluating the Performance of Diadochokinetic Tests in Characterizing Parkinson's Disease Hypokinetic Dysarthria. In Biomedical Engineering Systems and Technologies: 14th International Joint Conference, BIOSTEC 2021, Virtual Event, February 11–13, 2021, Revised Selected Papers (pp. 102-119). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-20664-1_6.

Graña, M., and Silva., M. (2021) Impact of Machine Learning Pipeline Choices in Autism Prediction From Functional Connectivity Data. International Journal on Neural Systems 31:4 2150009. https://doi.org/10.1142/S012906572150009X.

Greenberg, S., Ainsworth, W. A., Popper, A. N., and Fay, R. R. (2004). Speech Processing in the Auditory System. Springer, New York.

Griffin, J. P. (2008). Changing life expectancy throughout history. *Journal of the Royal Society of Medicine*, 101(12), 577-577, https://doi.org/10.1258/jrsm.2008.08k037

Hammer, G. P., Windisch, G., Prodinger, P. M., Anderhuber, F., and Friedrich, G. (2010) The Cricothyroid Joint – Functional Aspects with regard to Different Types of its Structure, J. Voice, 24 (2) 140-145. https://doi.org/10.1016/j.jvoice.2008.07.001.

Hannam, A. G., Stavness, I., Lloyd, J. E., and Fels, S. (2008) A dynamic model of jaw and hyoid biomechanics during chewing, J. Biomechanics; 41(5): 1069-1076. https://doi.org/10.1016/j.jbiomech.2007.12.001.

Hanratty, J., Deegan, C., Walsh, M., Kirkpatrick, B. (2016). Analysis of glottal source parameters in Parkinsonian speech, Proc. of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL 3666–3669. https://doi.org/10.1109/EMBC.2016.7591523.

Harel. B., Cannizzaro, M. and Snyder, P. J. (2004a). Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study. Brain and Cognition 56, 24-29, https://doi.org/10.1016/j.bandc.2004.05.002.

Harel, B. T., Cannizzaro, M. S., Cohen, H., Reilly, N. and Snyder, P. J. (2004b). Acoustic characteristics of Parkinsonian speech: a potential biomarker of early disease progression and treatment. Journal of Neurolinguistics 17, 439-453, https://doi.org/10.1016/j.jneuroling.2004.06.001.

Hardiman, O., Doherty, C. P., Elamin, M., and Bede, P. (2011). *Neurodegenerative disorders* (pp. 115-117). London, UK:: Springer. https://doi.org/10.1007/978-3-319-23309-3.

Henderson, M. X., Trojanowski, J. Q., Lee, V. M. α-Synuclein pathology in Parkinson's disease and related α-synucleinopathies. Neurosci Lett. 2019 Sep 14;709:134316. doi: 10.1016/j.neulet.2019.134316. Epub 2019 Jun 3. PMID: 31170426; PMCID: PMC7014913.

Hertrich, I., and Ackermann, H. (1995). Gender-specific vocal dysfunctions in Parkinson's disease: electroglottographic and acoustic analyses. Annals of Otology, Rhinology & Laryngology, 104(3), 197-202. https://doi.org/10.1177/000348949510400304.

Hidalgo, I., Gómez, P., and Garayzábal, E. (2017) Biomechanical Description of Phonation in Children Affected by Williams Syndrome, Journal of Voice 32(4) 515.e15-e28. https://doi.org/10.1016/j.jvoice.2017.07.002.

Hirschauer, T., Adeli, H., and Buford, T. (2015). Computer-Aided Diagnosis of Parkinson's Disease using an Enhanced Probabilistic Neural Network, Journal of Medical Systems, 39(179) (12 pages). https://doi.org/10.1007/s10916-015-0353-9.

Hlavnička, J., Tykalová, T., Ulmanová, O., Dušek, P., Horáková, D., Růžička, E., Klempir, J., Rusz, J. (2020) Characterizing vocal tremor in progressive neurological diseases via automated acoustic analyses. Clinical Neurophysiology, 131:5 1155-1165. https://doi.org/10.1016/j.clinph.2020.02.005.

Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., and Gates, S. (1998). Speech impairment in a large sample of patients with Parkinson's disease. Behavioural neurology, 11(3), 131-137.

Ho, A. K., Bradshaw, J. L., Iansek, R. (2008) For Better or Worse: The Effect of Levodopa on Speech in Parkinson's Disease, Mov. Disorders 23(4) 574-580. https://doi.org/10.1002/mds.21899.

Holmes, R. J., M. Oates, J., J. Phyland, D., and J. Hughes, A. (2000). Voice characteristics in the progression of Parkinson's disease. International Journal of Language & Communication Disorders, 35(3), 407-418. https://doi.org/10.1080/136828200410654.

Hou, Y., Dan, X., Babbar, M., Wei, Y., Hasselbalch, S. G., Croteau, D. L., and Bohr, V. A. (2019). Ageing as a risk factor for neurodegenerative disease. *Nature Reviews Neurology*, 15(10), 565-581, https://doi.org/10.1038/s41582-019-0244-7.

Huang, X., Acero, A., and Hon, H. W. (2001). Spoken language processing: A guide to theory, algorithm, and system development. Prentice hall PTR.

Ibarra-Lecue, I., Haegens, S., and Harris, A. Z. (2022) Breaking Down a Rhythm: Dissecting the Mechanisms Underlying Task-Related Neural Oscillations, Front. Neural Circuits, 16 846905. https://doi.org/10.3389/fncir.2022.846905.

Illes, J. (1989). Neurolinguistic features of spontaneous language production dissociate three forms of neurodegenerative disease: Alzheimer's, Huntington's, and Parkinson's. Brain and language, 37(4), 628-642. https://doi.org/10.1016/0093-934X(89)90116-8.

Inamoto, Y.; Saitoh, E.; Okada, S.; Kagaya, H.; Shibata, S.; Baba, M.; Onogi, K.; Hashimoto, S.; Katada, K.; Wattanapan, P.; et al. (2015) Anatomy of the larynx and pharynx: Effects of age, gender and height revealed by multidetector computed tomography. J. Oral Rehabil., 42, 670–677. https://doi.org/10.1111/joor.12298.

IPA (2005), International Phonetic Association: https://www.internationalphoneticalphabet.org/ipa-charts/ipa-symbols-chart-complete/ (retrieved 2021/01/18).

Itakura, F., and Saito, S. (1970): A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies. Electronics and Communications in Japan 53A, 36-43.

Itakura,F., and Saito,S. (1968). Analysis synthesis telephony based on the maximum likelihood method. In Proc. 6th of the International Congress on Acoustics (pp. C17–C–20). LosAlamitos, CA: IEEE.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2017) An Introduction to Statistical Learning; Springer: Berlin/Heidelberg, Germany, 8th Ed. https://doi.org/10.1007/978-1-0716-1418-1.

Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis, J. Neurol. Neurosurg. Psychiatry 79(4) 368–376. https://doi.org/10.1136/jnnp.2007.131045.

Jiménez, F. J., Gamboa, J., Nieto, A., Guerrero, J., Orti-Pareja, M., Molina, J. A., García-Albea, E., and Cobeta, I. (1997). Acoustic voice analysis in untreated patients with Parkinson's disease. Parkinsonism & Related Disorders, 3(2), 111-116. https://doi.org/10.1016/S1353-8020(97)00007-2.

Jürgens, U. (2002) Neural pathways underlying vocal control. Neurosci. and Behav. Rev. 26, 235-258, https://doi.org/:10.1016/S0149-7634(01)00068-9.

Jürgens, U. (2009) The Neural Control of Vocalization in Mammals: A Review, Journal of Voice 23(1) 1-10. https://doi.org/10.1016/j.jvoice.2007.07.005.

Kadiri, S. R., Kethireddy, R., and Alku, P. (2020). Parkinson's Disease Detection from Speech Using Single Frequency Filtering Cepstral Coefficients. In Interspeech (pp. 4971-4975). https://doi.org/10.21437/Interspeech.2020-3197.

Kandel, E., et al. (2013) Principles of Neural Science, New York, USA: McGraw- Hill.

Karakullukcu, N., and Yilmaz, B. (2022) Detection of Movement Intention in EEG-Based Brain-Computer Interfaces Using Fourier-Based Synchrosqueezing Transform. International Journal on Neural Systems, 32:1. https://doi.org/10.1142/S0129065721500593.

Kegl, J., Cohen, H., and Poizner, H. (1999). Articulatory consequences of Parkinson's disease: Perspectives from two modalities. Brain and cognition, 40(2), 355-386. https://doi.org/10.1006/brcg.1998.1086.

Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., and Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. American Journal of Speech-Language Pathology 18, 124–132. https://doi.org/10.1044/1058-0360(2008/08-0017).

Kent, R. D., and Kim, Y. J. (2003). Toward an acoustic typology of motor speech disorders. Clinical linguistics & phonetics, 17(6), 427-445. https://doi.org/10.1080/0269920031000086248.

Koc, T. Ciloglu, T. (2016) Nonlinear Interactive source-filter models for speech, Com. Speech and Lang. 36 365-394. https://doi.org/10.1016/j.csl.2014.12.002.

Lanciego, J. L., Luquin, N., and Obeso, J. A. (2012). Functional neuroanatomy of the basal ganglia. Cold Spring Harbor perspectives in medicine, 2(12), a009621. https://doi.org/10.1101%2Fcshperspect.a009621.

Le Dorze, G., Ryalls, J., Brassard, C., Boulanger, N., & Ratté, D. (1998). A comparison of the prosodic characteristics of the speech of people with Parkinson's disease and Friedreich's ataxia with neurologically normal speakers. Folia Phoniatrica et Logopaedica, 50(1), 1-9. https://doi.org/10.1159/000021444.

Leodori, G., Fabbrini, A., De Bartolo, M. I., Costanzo, M., Asci, F., Palma, V., Belvisi, D., Conte, A., and Berardelli, A. (2021). Cortical mechanisms underlying variability in intermittent theta-burst stimulation-induced plasticity: a TMS-EEG study. Clinical Neurophysiology, 132(10), 2519-2531. https://doi.org/10.1016/j.clinph.2021.06.02.

Levinson, N. (1946) The Wiener RMS (root mean square) error criterion in filter design and prediction, Journal of Mathematics and Physics, 25(1-4), 261-278. https://doi.org/10.1002/sapm1946251261.

Liberski, P. P. and Ironside, J. W., Prion Diseases, Editor(s): Michael J. Zigmond, Lewis P. Rowland, Joseph T. Coyle, Neurobiology of Brain Disorders, Academic Press, 2015, Pages 356-374, ISBN 9780123982704, https://doi.org/10.1016/B978-0-12-398270-4.00023-9.

Lin, J. (1991) Divergence Measures Based on the Shannon Entropy, IEEE Trans. on Information Theory 37(1) 145-151. https://doi.org/10.1109/18.61115.

Liotti, M., Ramig, L. O., Vogel, D., New, P., Cook, C. I., Ingham, R. J., Ingham, J. C., and Fox., P. T. (2003). Hypophonia in Parkinson's disease: Neural correlates of voice treatment revealed by PET, Neurology 60 432-440. https://doi.org/10.1212/WNL.60.3.432.

Loh, H. W., Ooi, C. P., Palmer, E., Barua, P. D., Dogan, S., Tuncer, T., Baygin, M., and Acharya, U. R. (2021). GaborPDNet: Gabor transformation and deep neural network for Parkinson's disease detection using EEG signals. Electronics, 10(14), 1740. https://doi.org/10.3390/electronics10141740.

Longemann, J. A., Fisher, H. B., Boshes, B., and Blonsky, E. R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. Journal of Speech and hearing Disorders, 43(1), 47-57. https://doi.org/10.1044/jshd.4301.47.

Louis, E. D., Klatka, L. A., Liu, Y., and Fahn, S. (1997). Comparison of extrapyramidal features in 31 pathologically confirmed cases of diffuse Lewy body disease and 34 pathologically confirmed cases of Parkinson's disease. Neurology, 48(2), 376-380. https://doi.org/10.1212/WNL.48.2.376.

Ludlow, C. L., Connor, N. P., and Bassich, C. J. (1987). Speech timing in Parkinson's and Huntington's disease. Brain and language, 32(2), 195-214. https://doi.org/10.1016/0093-934X(87)90124-6.

Ludlow, C. L. (2005). Central nervous system control of the laryngeal muscles in humans. Respiratory physiology & neurobiology, 147(2-3), 205-222. https://doi.org/10.1016/j.resp.2005.04.015.

Ludlow, C. L. (2015). Central nervous system control of voice and swallowing. J Clin Neurophysiol. 2015 August; 32(4): 294–303. https://doi.org/10.1097/WNP.0000000000000186.

Madruga, M., Campos-Roca, Y., and Pérez, C. J. (2023). Addressing smartphone mismatch in Parkinson's disease detection aid systems based on speech. Biomedical Signal Processing and Control, 80, 104281. https://doi.org/10.1016/j.bspc.2022.104281.

Makhoul, J. (1975). Linear prediction: A tutorial review. Proceedings of the IEEE, 63(4), 561-580. https://doi.org/10.1109/PROC.1975.9792.

Manríquez, R., Peterson, S. D., Prado, P., Orio, P., Galindo, G. E., Zañartu, M. (2019) Neurophysiological Muscle Activation Scheme for Controlling Vocal Fold Models, IEEE Trans. on Neural Syst. and Rehab. Eng. 27:1 1043-1052, https://doi.org/10.1109/TNSRE.2019.2906030.

Martin, A., Doddington, G., Kamm, T., Ordowski, M. and Przybocki, M. (1997). The DET curve in assessment of detection task performance. NIST, Gaithersburg, MD. https://apps.dtic.mil/sti/pdfs/ADA530509.

Mayo Clinic (2023), Parkinson's test (a-Synuclein seed amplification assay), https://www.mayoclinic.org/tests-procedures/parkinsons-testing-alpha-synuclein-seed-amplification/about/pac-20556191. (Retrieved May 6, 2024).

Meghraoui, D., Boudraa, B., Merazi-Meksen, T., Gómez-Vilda, P. (2021) A novel pre-processing technique in pathologic voice detection: Application to Parkinson's disease phonation, Biomedical Signal Processing and Control, 68 (2021), Art. 102604. https://doi.org/10.1016/j.bspc.2021.102604.

Mekyska, J., Janousova, E., Gómez, P., Smekal, Z., Rektorová, I., Eliasova, I., Kostalova, M., Mrackova, M., Alonso, J. B., Faúndez, M., López-de-Ipiña, K. (2015). Robust and complex approach of pathological speech signal analysis, Neurocomputing 167 94-111. https://doi.org/10.1016/j.neucom.2015.02.085.

Mekyska, J., Galaz, Z., Kiska, T., Zvoncak, V., Mucha, J., Smekal, Z., Eliasova, I., Kostalova, M., Mrackova, M., Fiedorova, D., Faúndez, M., Solé, J., Gómez, P., and Rektorova, I. (2018). Quantitative analysis of relationship between hypokinetic dysarthria and the freezing of gait in Parkinson's disease. Cognitive computation, 10, 1006-1018. https://doi.org/10.1007/s12559-018-9575-8.

Midi, I., Dogan, M., Koseoglu, M., Can, G., Sehitoglu, M. A., and Gunal, D. I. (2008) Voice abnormalities and their relation with motor dysfunction in Parkinson's Disease, Acta Neurol. Scand. 117 26-34. https://.doi.org/10.1111/j.1600-0404.2007.00965.x.

Moro-Velazquez, L., Gomez-Garcia, J. A., Arias-Londoño, J. D., Dehak, N., and Godino-Llorente, J. I. (2021). Advances in Parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects. Biomedical Signal Processing and Control, 66, 102418. https://doi.org/10.1016/j.bspc.2021.102418.

Mu, J., Chaudhuri, K. R., Bielza, C., de Pedro-Cuesta J., Larrañaga, P., and Martínez, P. (2017) Parkinson's disease subtypes identified from cluster analysis of motor and non-motor symptoms. Front Aging Neurosci. 9(9):1–10. https://doi.org/10.3389/fnagi.2017.00301.

Naylor, P. A., Kounoudes, A., Gudnason, J., and Brookes, M. (2007). Estimation of Glottal Closure Instants in Voiced Speech Using the DYPSA Algorithm, IEEE Transactions on Audio, Speech, and Language Processing, 15-1 34-43. https://doi:10.1109/TASL.2006.876878.

Novotný, M., Dušek, P., Daly, I., Ružička, E., and Rusz, J. (2020) Glottal Source Analysis of Voice Deficits in Newly Diagnosed Drug-naïve Patients with Parkinson's Disease: Correlation Between Acoustic Speech Characteristics and Non-Speech Motor Performance, Biomedical Signal Processing and Control, 57 101818. https://doi.org/10.1016/j.bspc.2019.101818.

Numssen, O., Zier, A. L., Thielscher, A., Hartwigsen, G., Knösche, T. R., and Weise, K. (2021). Efficient high-resolution TMS mapping of the human motor cortex by nonlinear regression. NeuroImage, 245, 118654. https://doi.org/10.1016/j.neuroimage.2021.118654.

Orozco-Arroyave, J.R., Belalcázar, E.A., Arias, J.D., Vargas, J.F., Skodda, S., Rusz, J., Daqrouq, K., Honig, F., Noth, E. (2015) Characterization Methods for the Detection of Multiple Voice Disorders: Neurological, Functional, and Laryngeal Diseases. IEEE J. Biomed. Health Inf. 19, 1820–1828. https://doi.org/10.1109/JBHI.2015.2467375.

Orozco-Arroyave, J. R., Hönig, F., Arias, J. D., Vargas, F., Daqrouq, K., Skodda, S., Rusz, J., Nöth, E., (2016) Automatic detection of Parkinson's Disease in running speech spoken in three different languages, J. Acoust. Soc. Am. 139 481-500. https://doi.org/10.1121/1.4939739.

Ozbolt, A. S., Moro-Velazquez, L., Lina, I., Butala, A. A., and Dehak, N. (2022). Things to consider when automatically detecting parkinson's disease using the phonation of sustained vowels: Analysis of methodological issues. Applied Sciences, 12(3), 991. https://doi.org/10.3390/app12030991.

Palacios, D., Meléndez, G., López, A., Lázaro, C., Gómez, A. and Gómez, P. (2020) MonParLoc: A speech-based system for Parkinson's disease analysis and monitoring. IEEE Access, Early Access, https://.doi.org/10.1109/ACCESS.2020.3031646.

Papavramidou, N. (2018). The ancient history of dementia. *Neurological Sciences*, *39*, 2011-2016. https://doi.org/10.1007/s10072-018-3501-4.

Papoulis, A. (1991) Probability, random variables, and stochastic processes. New York: McGraw-Hill.

Parkinson, J. (1817) "An Essay on the Shaking Palsy". J. Neuropsychiatry Clin. Neurosci. 14:2 (2002) 223-236. (Re-edited in Neuropsychiatry Classics from the 1817 monograph, by Sherwood, Neely and Jones).

Perez, K. S., Ramig, L. O., Smith, M. E., and Dromey, C. (1996). The Parkinson larynx: tremor and videostroboscopic findings. Journal of Voice, 10(4), 354-361. https://doi.org/10.1016/S0892-1997(96)80027-0.

Pickering, M., and Jones, J. F. (2002) The diaphragm: two physiological muscles in one. J Anat.; 201(4):305-12. http://doi.org/10.1046/j.1469-7580.2002.00095.x.

Pinho, P., Monteiro, L., Soares, M. F. D. P., Tourinho, L., Melo, A., and Nóbrega, A. C. (2018). Impact of levodopa treatment in the voice pattern of Parkinson's disease patients: a systematic review and meta-analysis. In CoDAS (Vol. 30, No. 5). Sociedade Brasileira de Fonoaudiologia, https://doi.org/10.1590/2317-1782/20182017200.

Postuma, R., Lang, A. E., Gagnon, J. F., Pelletier, A., Montplaisir, J. I. (2012) How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder, Brain, 35 (Pt6) 1860–1870. https://doi.org/10.1016/j.eswa.2012.01.102.

Radomski, T. (2005). Manuel García (1805–1906): A bicentenary reflection, Australian Voice. 11: 25–41.

Rafiei. M. H., and Adeli, H. (2017). A New Neural Dynamic Classification Algorithm, IEEE Transactions on Neural Networks and Learning Systems, 28(12) 3074-3083. https://doi.org/10.1109/TNNLS.2017.2682102.

Rajput, A. and Noyes, E. (2024), Merck Manual: Parkinsonism, URL: https://www.merckmanuals.com/home/brain,-spinal-cord,-and-nerve-disorders/movement-disorders/parkinsonism (retrieved May 6, 2024).

Ramig, L. A., and Ringel, R. L. (1983). Effects of physiological aging on selected acoustic characteristics of voice. Journal of Speech, Language, and Hearing Research, 26(1), 22-30. https://doi.org/10.1044/jshr.2601.22.

Ramig, L. O., Fox, C., Sapir, S. (2008). Speech treatment for Parkinson's disease, Expert Rev. Neurother., 8(2) 297–309. https://doi.org/10.1586/14737175.8.2.297.

Rektorová, I., Mikl, M., Barrett, J. Marecek, R., Rektor, I., Paus, T. (2012). Functional neuroanatomy of vocalization in patients with Parkinson's disease, J. Neu. Sci. 313 7-12. https://doi.org/10.1093/cercor/bhi06.

Ricciardi, L., Ebreo, M., Graziosi, A., Barbuto, M., Sorbera, C., Morgante, L., and Morgante, F. (2016). Speech and gait in Parkinson's disease: When rhythm matters. Park. Relat. Disord., vol. 32, pp. 42–47. https://doi.org/10.1016/j.parkreldis.2016.08.013.

Ringel, R. L., and Chodzko-Zajko, W. J. (1987). Vocal indices of biological age. Journal of Voice, 1(1), 31-37.https://doi.org/10.1016/S0892-1997(87)80021-8.

Robbins, J. A., Logemann, J. A., and Kirshner, H. S. (1986). Swallowing and speech production in Parkinson's disease. Annals of neurology, 19(3), 283-287. https://doi.org/10.1002/ana.410190310.

Robnik-Šikonja, M., and Kononenko, I. (2003) Theoretical and empirical analysis of ReliefF and RReliefF. Machine Learning 53(1–2) 23–69.

Rödel, R. M. W., Olthoff, A., Tergau, F., Simonyan, K., Kraemer, D., Markus, H., and Kruse, E. (2004) Human Cortical Motor Representation of the Larynx as Assessed by Transcranial Magnetic Stimulation (TMS). The Laryngoscope 114 918-922. https://doi.org/10.1097/00005537-200405000-00026.

Rose, G. (2001). Sick individuals and sick populations. International journal of epidemiology, 30(3), 427-432. https://doi.org/10.1093/ije/30.3.427.

Rosenbrock, H. H. (1960) An automatic method for finding the greatest or least value of a function, The Computer Journal; 3 (3): 175–184, https://doi.org/doi:10.1093/comjnl/3.3.175.

Rothenberg, M. (1973) A new inverse-filtering technique for deriving the glottal airflow waveform during voicing, J. Acoust. Soc. Am. 53(6) 1632-1645. https://doi.org/10.1121/1.1913513.

Rusz, J., Cmelja, R., Tykalova, T., Ruzickova, H., Klempir, J. Majerova, V., Picausova, J., Roth, J. and Ruzicka, E. (2013) Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task, Journal of the Acoustical Society of America; 134: 2171–2181. https://doi.org/10.1121/1.4816541.

Salenius, S., Avikainen, S., Kaakkola, S., Hari, R., Brown, P. (2002) Defective cortical drive to muscle in Parkinson's disease and its improvement with levodopa. Brain 125-3, 491–500. https://doi.org/10.1093/brain/awf042.

Salicrú, M., Morales, D., Menéndez, M. L., and Pardo, L. (1994) On the Applications of Divergence Type Measures in Testing Statistical Hypotheses, J. of Multivar. Anal. 51 (2) 372-391. https://doi.org/10.1006/jmva.1994.1068.

Sanguinetti, V., Laboissière, R. and Payan, Y. (1997). A control model of human tongue movements in speech. Biol. Cybern. 77 11-22, https://doi.org/10.1007/s004220050362.

Sapir, S. (2014) Multiple Factors Are Involved in the Dysarthria Associated With Parkinson's Disease: A Review With Implications for Clinical Practice and Research. Journal of Speech, Language and Hearing Research, 57, 1330-1343, https://doi.org/10.1044/2014_JSLHR-S-13-0039.

Sapir, S., Ramig, L. O., Spielman, J. L., Fox, C. (2010). Formant Centralization Ratio: A Proposal for a New Acoustic Measure of Dysarthric Speech, Journal of Speech, Language and Hearing Research, 53(1) 114-125. https://doi.org/10.1044/1092-4388(2009/08-0184.

Saravanamuttu, J., Radhu, N., Udupa, K., Baarbé, J., Gunraj, C., Chen, R. (2021) Impaired motor cortical facilitatory-inhibitory circuit interaction in Parkinson's disease, Clinical Neurophysiology, 132-10 2685-2692. https://doi.org/10.1016/j.clinph.2021.05.032.

Schulz, G. M., Varga, M., Jeffires, K., Ludlow, C. L., and Braun A. R. (2005), Functional Neuroanatomy of Human Vocalization: An H215O PET Study, Cerebral Cortex 15 1835-1847. https://doi.org/10.1093/cercor/bhi061.

Sharma, M., Patel, S., and Acharya, R. (2020) Automated detection of abnormal EEG signals using localized wavelet filter banks. Pattern Recognition Letters 133 188-194. https://doi.org/10.1016/j.patrec.2020.03.009.

Shi, Z., Li, Y., Lin H., and Zhu, F., (2022) Coupling Analysis of EEG and EMG Signals Based on Transfer Entropy after Consistent Empirical Fourier Decomposition, 8th Int. Conf. on Control, Automation and Robotics (2022) 436-441. https://doi.org/10.1109/ICCAR55106.2022.9782665.

Šimek, M., and Rusz, J. (2021). Validation of cepstral peak prominence in assessing early voice changes of Parkinson's disease: Effect of speaking task and ambient noise. The Journal of the Acoustical Society of America, 150(6), 4522-4533. https://doi.org/10.1121/10.0009063.

Simon, D. K., Tanner, C. M., and Brundin, P. (2020). Parkinson disease epidemiology, pathology, genetics, and pathophysiology. *Clinics in geriatric medicine*, 36(1), 1-12. https://doi.org/10.1016/j.cger.2019.08.002.

Skaper, D., Facci, L., Zusso, M. and Giusti, P., An Inflammation-Centric View of Neurological Disease: Beyond the Neuron, Frontiers in Cellular Neuroscience; (2018), https://doi.org/10.3389/fncel.2018.00072.

Skodda, S., Visser, W. and Schlegel, U. (2011) Vowel Articulation in Parkinson's Disease, Journal of Voice 25 (4): 467-472. https://doi.org/10.1016/j.jvoice.2010.01.009.

Skodda, S., Grönheit, W., and Schlegel, U. (2012) Impairment of Vowel Articulation as a Possible Marker of Disease Progression in Parkinson's Disease, PLoS ONE 7 (2) e32132. https://doi.org/10.1371/journal.pone.0032132.

Skodda, S., Grönheit, W., Mancinelli, N., and Schlegel, U. (2013) Progression of Voice and Speech Impairment in the Course of Parkinson's Disease: A Longitudinal Study, Parkinson's Disease; Article ID 389195. https://doi.org/10.1155/2013/389195.

Solomon, E. A., Kragel, J. E., Sperling, M. R., Sharan, A., Worrell, G., Kucewicz, M., Inman, C. S., Lega, B., Davis, K. A., Stein, J. M., Jobst, B. C., Zaghloul, K. A., Sheth, S. A., Rizzuto, D. S., and Kahana, M. J. (2017). Widespread theta synchrony and high-frequency desynchronization underlies enhanced cognition. Nature communications, 8(1), 1704. 2017). https://doi.org/10.1038/s41467-017-01763-2.

Story, B. H., and Titze, I. R. (1995). Voice simulation with a body-cover model of the vocal folds. The Journal of the Acoustical Society of America, 97(2), 1249-1260. https://doi.org/10.1121/1.412234.

Svensson, P., Henningson, C., and Karlsson, S. (1993). Speech motor control in Parkinson's disease: a comparison between a clinical assessment protocol and a quantitative analysis of mandibular movements. Folia Phoniatrica et Logopaedica, 45(4), 157-164.https://doi.org/10.1159/000266243.

Taroni, F., Aitken, C., Garbolino, P. and Biedermann, A. (2006). Bayesian Networks and Probabilistic Inference in Forensic Science, Wiley.

Taroni, F., Bozza, S., Biedermann, A., Garbolino, P., and Aitken, C. (2010) Data Analysis in Forensic Science: A Bayesian Decision Perspective; John Wiley and Sons: Hoboken, NJ, USA.

Tennenholtz, G.; Zahavy, T.; Mannor, S. (2018) Train on validation: Squeezing the data lemon. arXiv 2018. arXiv:1802.05846. https://doi.org/10.48550/arXiv.1802.05846.

Titze, I. R. (1994a). Principles of Voice Production. Prentice-Hall, Inc., Hoboken, New Jersey, USA.

Titze, I. R. (1994b) Summary Statement. Workshop on Acoustic Voice Analysis, National Center for Voice and Speech.

Titze, I. R., and Palaparthi, A. (2016). Sensitivity of source–filter interaction to specific vocal tract shapes. IEEE/ACM transactions on audio, speech, and language processing, 24(12), 2507-2515. https://doi.org/10.1109/TASLP.2016.2616543.

Titze, I. R., and Story, B. H. (2002). Rules for controlling low-dimensional vocal fold models with muscle activation. The Journal of the Acoustical Society of America, 112(3), 1064-1076. https://doi.org/10.1121/1.1496080.

Töpfer, F., and Wiesing, U. (2005a). The medical theory of Richard Koch I: theory of science and ethics. Medicine, Health Care and Philosophy, 8, 207-219. https://doi.org/10.1007/s11019-004-7445-5.

Töpfer, F., and Wiesing, U. (2005b). The medical theory of Richard Koch II: natural philosophy and history. Medicine, Health Care and Philosophy, 8, 323–334. https://doi.org/10.1007/s11019-004-7446-4.

Tsanas, A., Little, M. A., McSharry, P. E., and Ramig, L. O. (2010a) New nonlinear markers and insights into speech signal degradation for effective tracking of Parkinson's disease symptom severity, in Proc. Int. Symp. Nonlinear Theory Appl. (NOLTA), Sep. 2010, pp. 457460.

Tsanas, A., Little, M. A., McSharry, P. E., and Ramig, L. O. (2010b) Enhanced classical dysphonia measures and sparse regression for telemonitoring of Parkinson's disease progression. In: 2010 IEEE Int. Conf. Acoust. Speech Signal Process., no. March, pp. 594–597, https://doi.org/10.1109/ICASSP.2010. 5495554.

Tsanas, A., Little, M. A., McSharry, P. E., and Ramig, L. O. (2010c) Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE Trans Biomed Eng. 57(4):884–93. https:// doi. org/ 10. 1109/ TBME. 2009. 20360 00.

Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J., and Ramig, L. O. (2010d) Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease, IEEE Trans. on Biomed. Eng. 59 1264-1271.

Tsanas, A., Little, M. A. McSharry, P. E., Ramig L.O. (2011). Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. Journal of the Royal Society Interface 8, 842-855, https://doi.org/10.1098/rsif.2010.0456.

Tsanas, A. (2012). Accurate telemonitoring of Parkinson's disease symptom severity using nonlinear speech signal processing and statistical machine learning, Ph.D. thesis, Oxford Centre for Industrial and Applied Mathematics, University of Oxford, Oxford, UK.

Tsanas, A., Little, M. A., McSharry, P. E., Scanlon, B. K., and Papapetropoulos, S. (2012). Statistical analysis and mapping of the Unified Parkinson's Disease Rating Scale to Hoehn and Yahr staging. Parkinsonism and related disorders, 18(5), 697-699. https://doi.org/10.1016/j.parkreldis.2012.01.011.

Tsanas, A. (2013) Acoustic analysis toolkit for biomedical speech signal processing: concepts and algorithms, 8th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), pp. 37-40, Florence, Italy.

Tsanas, A., and Gómez, P. (2013) Novel robust decision support tool assisting early diagnosis of pathological voices using acoustic analysis of sustained vowels. In Proceedings of the Multidisciplinary Conference of Users of Voice, Speech and Singing (JVHC 13), Las Palmas de Gran, Canaria, Spain, 27–28.

Tsanas, A., Little, M. A., Fox, C., Ramig, L. O. (2014) Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease. IEEE Trans Neural Syst Rehabil Eng. 22(1):181–90. https:// doi. org/ 10. 1109/ TNSRE. 2013. 22935 75.

Tsanas A. (2019) New insights into Parkinson's disease through statistical analysis of standard clinical scales quantifying symptom severity. In: 41st IEEE Engineering in Medicine and Biology Conference 3412–3415, https:// doi. org/ 10. 1109/ EMBC. 2019. 88565 59.

Tsanas, A., and Arora, S. (2019) Biomedical speech signal insights from a large scale cohort across seven countries: the Parkinson's voice initiative study. In: Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), 2019, pp 45–48.

Tsanas, A., and Arora, S. (2020) Large-scale clustering of people diagnosed with Parkinson's disease using acoustic analysis of sustained vowels: Findings in the Parkinson's voice initiative study. In: BIOSIGNALS 2020—13th International Conference on Bio-Inspired Systems and Signal Processing, Proceedings; Part of 13th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2020, pp. 369–376, https:// doi. org/ 10. 5220/ 00093 61203 690376.

Tsanas, A., and Arora, S. (2021) Assessing Parkinson's disease speech signal generalization of clustering results across three countries: findings in the Parkinson's voice initiative study. In: BIOSIGNALS 2021—14th International Conference on Bio-Inspired Systems and Signal Processing; Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021, pp. 124–131, https:// doi. org/ 10. 5220/00103 83001 240131.

Tsanas, A., Little, M. A., Ramig, L. O. (2021). Remote assessment of Parkinson's disease symptom severity using the simulated cellular mobile telephone network, IEEE Access, Vol. 9, pp. 11024-11036, 2021. https://doi.org/10.1109/ACCESS.2021.3050524.

Tsanas, A., Arora, S. (2022). Data-driven subtyping of Parkinson's using acoustic analysis of sustained vowels and cluster analysis: findings in the Parkinson's voice initiative study, Springer Nature Computer Science, Vol. 3:232. https://doi.org/10.1007/s42979-022-01123-y.

Turcano, P., Chen, J.J., Bureau, B.L. Savica, R., Early ophthalmologic features of Parkinson's disease: a review of preceding clinical and diagnostic markers. J Neurol 266, 2103–2111 (2019). https://doi.org/10.1007/s00415-018-9051-0

Tykalová, T., Rusz, J., Švihlík, J., Bancone, S., Spezia, A., and Pellecchia, M. T. (2020). Speech disorder and vocal tremor in postural instability/gait difficulty and tremor dominant subtypes of Parkinson's disease. Journal of Neural Transmission, 127(9), 1295-1304. https://doi.org/10.1016/j.specom.2019.04.003.

Tysnes, O. B., and Storstein, A. (2017). Epidemiology of Parkinson's disease, J. Neural Transm, 124-8, 901–905. https://doi.org/10.1007/s00702-017-1686-y.

Verwoert, M., Ottenhoff, M. C., Goulis, S., Colon, A. J., Wagner, L., Tousseyn, S., van Dijk, J. P., Kubben, P. L., and Herff, C. (2022). Dataset of speech production in intracranial electroencephalography. Scientific data, 9(1), 434. https://doi.org/10.1038/s41597-022-01542-9.

Vicente, R., Wibral, M., Lindner, M., and Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. Journal of computational neuroscience, 30, 45-67. https://doi.org/10.1007/s10827-010-0262-3.

Wakita, H. (1973). Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms. IEEE Transactions on Audio and Electroacoustics, 21(5), 417-427. https://doi.org/10.1109/TAU.1973.1162506.

Weismer, G., and Wildermuth, J. (1998). Fomant Trajectory Characteristics in Persons with Parkinson, Cerebella, and Upper Motor Neuron Disease. https://web.archive.org/web/20170811011111id_/http://www.icacommission.org/Proceedings/ICA1998Seattle/pdfs/vol_2/1267_1.pdf.

Whitfield, J. A., and Goberman, A. M. (2014) Articulatory-acoustic vowel space: Application to clear speech in individuals with Parkinson's disease", Journal of Communication Disorders; 51: 19-28 https://doi.org/10.1016/j.jcomdis.2014.06.005.

Wiener, N. (1964). The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Prediction. https://doi.org/10.7551/mitpress/2946.001.0001.

World Health Organization (2021). GHE: Life expectancy and healthy life expectancy. The Global Health Observatory [Internet]. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-life-expectancy-and-healthy-life-expectancy (accessed Sep. 04, 2023).

World Health Organization (2023) Parkinson disease. https://www.who.int/news-room/fact-sheets/detail/parkinson-disease (visited 2023.03.16).

Yaqub, M. A., Hong, K. S., Zafar, A., and Kim, C. S. (2022). Control of transcranial direct current stimulation duration by assessing functional connectivity of near-infrared spectroscopy signals. International Journal of Neural Systems, 32(01), 2150050. https://doi.org/10.1142/S0129065721500507.

Yunusova, Y., Weismer, G. G., Westbury, J. R. and Lindstrom, M. J. (2008). Articulatory Movements During Vowels in Speakers with Dysarthria and Healthy Controls. Journal of Speech, Language and Hearing Research 51 596-611. https://doi.org/10.1044/1092-4388(2008/043).

Yunusova, Y., Weismer, G. G. and Lindstrom, M. J. (2011). Classifications of Vocalic Segments From Articulatory Kinematics: Healthy Controls and Speakers With Dysarthria. Journal of Speech, Language and Hearing Research 54 1302-1311, https://doi.org/10.1044/1092-4388(2011/09-0193).

Yuvaraj, R., Murugappan, M., Acharya, U. R., Adeli, H., Ibrahim, N. M., Mesquita, E. (2016). Brain Functional Connectivity Patterns for Emotional State Classification in Parkinson's Disease Patients Without Dementia, Behavioural Brain Research, 298 248-260. https://doi.org/10.1016/j.bbr.2015.10.036.

Zhang, T., Lin, L., Tian, J., Xue, Z., and Guo, X. (2023a). Voice feature description of Parkinson's disease based on co-occurrence direction attribute topology. Engineering Applications of Artificial Intelligence, 122, 106097. https://doi.org/10.1016/j.engappai.2023.106097.

Zhang, T., Lin, L., and Xue, Z. (2023b). A voice feature extraction method based on fractional attribute topology for Parkinson's disease detection. Expert Systems with Applications, 219, 119650. https://doi.org/10.1016/j.eswa.2023.119650.

Zwirner, P., Murry, T., and Woodson, G. E. (1991). Phonatory function of neurologically impaired patients. Journal of communication disorders, 24(4), 287-300. https://doi.org/10.1016/0021-9924(91)90004-3.

APPENDICES

Appendix I Databases used in the present study

Data resources are a fundamental component of any comprehensive research effort. This appendix is dedicated to offering an in-depth description of the datasets utilized throughout the duration of this doctoral study. These datasets employed are categorized into two groups: those that were self-recorded and those that were externally provided. Subsequent sections will provide comprehensive descriptions of each dataset type. The pandemic imposed significant constraints on data collection efforts between 2020 and 2022. While data gathering did resume later, the delay was substantial, and the influx of new data arrived too late to exert a meaningful influence on the outcome of the PhD project. During this period, most research lines had already transitioned towards the utilization of externally provided datasets. This is an essential point to highlight because, during the initial planning of the PhD project, there was a greater emphasis on self-recorded datasets. The primary objective was to create a distinctive database for the investigation of articulation and phonation, which would have been a significant contribution to the field. However, due to the delays caused by the pandemic, only a partial recording of this dataset was completed. Consequently, the study's conclusions lacked the necessary statistical power to support further exploration of the developed concepts. As a result, to advance the PhD project, efforts were made to gain access to other databases. This shift in approach constrained the outcomes to align with the specifications and standards set by other research groups, given the limitations in controlling recording conditions and tasks.

Appendix I.1 PARCZ dataset

One of the speech databases used in the study (PARCZ) was collected at St. Anne's University Hospital in Brno (Czech Republic), containing recordings produced by PD patients of both genders. The database also includes demographic and clinical information from each patient. This database contains also speech recordings and demographic information from age-matched HC subjects. Each participant signed an informed consent form that was approved by the ethics committee of St. Anne's University Hospital.

The recordings included in PARCZ were acquired in a quiet isolated room (about 70 m3) under environmental noise lower than 30 dB SPL (measured with an NTI Acoustilyzer AL1). The recordings used a large capsule cardioid microphone M-AUDIO Nova (https://www.m-audio.com/products/view/nova) mounted on a boom arm RODE PSA1 fixed at a distance of approximately 20 cm from the speaker's mouth. These signals were taken at a sampling rate of 48 kHz (16-bit resolution) on an M-AUDIO Fast Track Pro platform.

Appendix I.2 PARCZ dataset participant description

The following table gives the demographic and clinical description of the participants' subset selected from PARCZ for the experimental framework described in subsection 5.1.1.1. The codes P1xxx-a refer to files containing emissions of a sustained vowel [a:] from female PD participants. Similarly, the code P2xxx-a refers to similar utterances from male PD participants. Codes of type K1xxx-a and K2xxx-a refer to female and male participants of the HC group.

Table App. 1 PARCZ PD patient and HC subject set demographical data.
Gen: Gender, Con: Condition, PD: PD patient subject; HC: healthy control subject; UIII: Evaluation according to UPDRS-III scale, LED: Levodopa Equivalent Dosage.

Code	Gen	Age	Con	UIII	LED	Code	Gen	Age	Con	UIII	LED	Code	Gen	Age	Con	UIII	LED
P1006-a	F	59	PD	24	875	P1027-a	F	65	PD	8	740	P1058-a	F	71	PD	20	464
P1007-a	F	76	PD	55	1185	P1031-a	F	59	PD	10	918	P1064-a	F	60	PD	11	660
P1008-a	F	78	PD	23	1444	P1033-a	F	73	PD	14	650	P1066-a	F	68	PD	26	1230
P1020-a	F	64	PD	8	160	P1040-a	F	70	PD	32	1115	P1068-a	F	73	PD	11	1124
P1021-a	F	65	PD	5	600	P1041-a	F	72	PD	31	990	P1071-a	F	70	PD	35	1320
P1022-a	F	72	PD	6	800	P1051-a	F	62	PD	13	300	P1073-a	F	64	PD	30	2102
P1025-a	F	64	PD	8	1033	P1052-a	F	49	PD	33	700	P1076-a	F	72	PD	5	460
P1026-a	F	76	PD	12	540	P1053-a	F	56	PD	19	1305	P1103-a	F	60	PD	27	518
P2005-a	M	46	PD	25	2135	P2024-a	M	77	PD	10	1173	P2039-a	M	64	PD	37	1058
P2009-a	M	66	PD	14	150	P2028-a	M	72	PD	9	160	P2043-a	M	64	PD	29	1800
P2010-a	M	66	PD	39	931	P2029-a	M	68	PD	5	535	P2044-a	M	54	PD	37	1318
P2012-a	M	71	PD	35	2186	P2030-a	M	70	PD	8	767	P2045-a	M	77	PD	49	580
P2017-a	M	63	PD	19	625	P2032-a	M	80	PD	22	759	P2046-a	M	52	PD	26	1630
P2018-a	M	63	PD	32	750	P2034-a	M	74	PD	15	870	P2047-a	M	58	PD	37	600
P2019-a	M	73	PD	12	785	P2037-a	M	86	PD	36	1185	P2049-a	M	62	PD	31	1139
P2023-a	M	73	PD	13	610	P2038-a	M	71	PD	55	1330	P2055-a	M	67	PD	20	700
K1003-a	F	63	HC	-		K1019-a	F	64	HC	-		K1029-a	F	**57**	HC	-	
K1004-a	F	65	HC	-		K1020-a	F	49	HC	-		K1030-a	F	69	HC	-	
K1005-a	F	59	HC	-		K1021-a	F	49	HC	-		K1031-a	F	87	HC	-	
K1006-a	F	64	HC	-		K1022-a	F	70	HC	-		K1036-a	F	65	HC	-	
K1007-a	F	59	HC	-		K1023-a	F	56	HC	-		K1040-a	F	74	HC	-	
K1012-a	F	**67**	HC	-		K1024-a	F	55	HC	-		K1048-a	F	65	HC	-	
K1017-a	F	61	HC	-		K1025-a	F	61	HC	-		K1051-a	F	61	HC	-	
K1018-a	F	45	HC	-		K1026-a	F	63	HC	-		K1053-a	F	78	HC	-	
K2001-a	M	59	HC	-		K2015-a	M	76	HC	-		K2041-a	M	79	HC	-	
K2002-a	M	68	HC	-		K2016-a	M	65	HC	-		K2042-a	M	66	HC	-	
K2008-a	M	70	HC	-		K2027-a	M	74	HC	-		K2043-a	M	65	HC	-	
K2009-a	M	68	HC	-		K2032-a	M	63	HC	-		K2044-a	M	56	HC	-	
K2010-a	M	83	HC	-		K2033-a	M	59	HC	-		K2045-a	M	76	HC	-	
K2011-a	M	55	HC	-		K2034-a	M	63	HC	-		K2046-a	M	49	HC	-	
K2013-a	M	54	HC	-		K2035-a	M	60	HC	-		K2047-a	M	56	HC	-	
K2014-a	M	62	HC	-		K2038-a	M	59	HC	-		K2049-a	M	78	HC	-	

Appendix I.3 rTMS dataset

This dataset has been recruited by the Applied Neuroscience Research Group, CEITEC, Masaryk University, Brno, Czech Republic, from PwP participants showing mild to moderate HD directly related to PD, all of them right-handed, and native speakers of Czech, following a program of repetitive transcranial magnetic stimulation (rTMS), according to the protocol described in Appendix I.5. All were on stable dopaminergic medication for the duration of the whole study. The patients were tested in the ON state (they had received the adequate dosage of dopaminergic medication according to their respective prescriptions two hours before the evaluations were conducted). The study considered only those cases having produced one pre-stimulus recording, and four post-stimulus recordings spaced in time. Eighteen cases fulfilling this condition were selected from the 33 participants included in the original database. All other cases completed less than four post-stimulus examinations and therefore were not considered for this first study. Half the participants received an active stimulation, and half submitted to a sham stimulation (same duration protocol, but no active transcranial stimulation). All participants were informed of the nature of the research and gave their written consent. The trial was registered in clinicaltrials.gov (Number NCT04203615).

Appendix I.4 rTMS dataset participant description

The study included participants recorded at pre-stimulus and after four post-stimulus sessions spaced in time. This reduced the number of cases to 18 out of the total number of 33 subjects included in the original database. After random selection, half the participants received an active stimulation, and half were submitted to a sham stimulation (identical protocol and recording conditions, but no active stimulus being applied). The stimulation description of the participants is given in Table App. 2.

The cohort distributions are broadly similar in terms of UPDRS grade (females: 16.6±4.1; males: 12.3±3.9) and age (females: 74.6±3.0; males: 69.7±8.4).

Table App. 2 Participants' demographic and clinical data from rTMS.

A: active stimulation; S: sham stimulation; F: Female; M: Male; Y: years. UPDRS-III: Unified Parkinson Disease Rating Scale, section III (motor section).

PwP code (pre)	Active/Sham	Gender	Age (Y)	UPDRS-III
0100	A	F	71	10
0800	A	M	58	9
1100	A	M	73	14
1200	A	M	72	21
1400	A	M	64	10
1600	S	F	79	20
1700	S	M	70	16
1800	S	M	61	9
1900	S	M	77	8
2000	A	F	76	28
2200	S	M	66	13
2300	S	M	55	7
2400	S	M	72	10
2500	S	M	81	14
2600	S	F	73	16
2700	A	M	77	14
2800	A	M	80	15
2900	A	F	74	17

Appendix I.5 rTMS dataset protocol description

The stimulation protocol and speech recording conditions are described in detail in Brabenec et al. (2021), of which a summary is offered here. Participants were subject to rTMS (DuoMAG™ XT-100, Deymed Diagnostic) in ten stimulation sessions over two weeks at CEITEC, Masaryk University. Each stimulation session took 40 min. to complete, during which an eight-shaped coil applied pulses at a frequency of 1 Hz, 100% intensity of the pre-estimated resting motor threshold (1800 pulses per stimulation session) over the right posterior superior temporal gyrus (STG, MNI coordinates X = 40, Y = -38, Z=14). The same coil was used in quite a similar fixture for sham stimulation sessions, producing the same sounds as in the active case described before, but no magnetic field was applied.

The stimulation thresholds and settings were established in a preliminary study; the interested reader will be referred for further clarification by Brabenec et al. (2021).

Each participant went through a baseline assessment (pre-stimulus evaluation at the first session: T0) before being submitted to ten stimulation sessions (stimulation process) within two weeks; a follow-up evaluation session two weeks after stimulation (post-stimulus at T1); additional follow-up evaluations around six weeks (post-stimulus at T2), and around ten weeks (post-stimulus at T3). The 18 participants of the subset included in the study submitted also to a fourth post-treatment evaluation session around fourteen weeks after the stimulation process (post-stimulus at T4). The evaluation dates are listed in Table App. 3 (Appendix I.5). Each participant in the study was randomly assigned to active or sham stimulation. A perceptual assessment was conducted by a speech therapist rating speech performance, faciokinesis, phonorespiration, and phonetic competence at each evaluation step. Audio recordings of the following utterances from each participant were taken before (pre-stimulus) and after (post-stimulus) the stimulation process: one free-topic monologue; one short neutral reading in Czech; 1 short emission of vowels [a:], [i:], [u:] (of ∼1.5 s); one long emission of a sustained [a:] (of around 15 s); 1 long emission of a diadochokinetic exercise consisting in the repetition of [pataka] (lasting >10 s), and one single emission of ten different selected tri-syllabic words in Czech.

A large capsule cardioid microphone M-AUDIO Nova mounted to a boom arm RODE PSA1 at a distance of approximately 20 cm from the patient's mouth was used for audio recordings. Acoustic signals were digitized by the M-AUDIO Fast Track Pro audio interface with fs = 48 kHz sampling frequency and 16-bit resolution. The details of recording data and time between stimuli are given in Table App. 3.

Table App. 3 Intervals between pre-stimulus and post-stimulus evaluations in days.

Code	Interval	Dates	Time Lap	Weight	Code	Interval	Dates	Time Lap	Weight
0100	T0	4.9.2017	0.00	0.00	2000	T0	27.9.2018	0.00	0.00
0101	T1	15.9.2017	11.00	0.11	2001	T1	12.10.2018	15.00	0.15
0102	T2	16.10.2017	42.00	0.42	2002	T2	9.11.2018	43.00	0.43
0103	T3	13.11.2017	70.00	0.70	2003	T3	7.12.2018	71.00	0.72
0104	T4	13.12.2017	100.00	1.00	2004	T4	4.1.2019	99.00	1.00
0800	T0	8.2.2018	0.00	0.00	2200	T0	9.11.2018	0.00	0.00
0801	T1	23.2.2018	15.00	0.14	2201	T1	23.11.2018	14.00	0.14
0802	T2	6.3.2018	26.00	0.24	2202	T2	18.12.2018	39.00	0.40
0803	T3	27.4.2018	78.00	0.72	2203	T3	18.1.2019	70.00	0.71
0804	T4	28.5.2018	109.00	1.00	2204	T4	15.2.2019	98.00	1.00
1100	T0	6.4.2018	0.00	0.00	2300	T0	12.11.2018	0.00	0.00
1101	T1	20.4.2018	14.00	0.13	2301	T1	4.12.2018	22.00	0.20
1102	T2	21.5.2018	45.00	0.41	2302	T2	4.1.2019	53.00	0.47
1103	T3	22.6.2018	77.00	0.71	2303	T3	1.2.2019	81.00	0.72
1104	T4	24.7.2018	109.00	1.00	2304	T4	4.3.2019	112.00	1.00
1200	T0	4.6.2018	0.00	0.00	2400	T0	12.11.2018	0.00	0.00
1201	T1	15.6.2018	11.00	0.11	2401	T1	4.12.2018	22.00	0.20
1202	T2	18.7.2018	44.00	0.44	2402	T2	4.1.2019	53.00	0.47
1203	T3	15.8.2018	72.00	0.72	2403	T3	1.2.2019	81.00	0.72
1204	T4	12.9.2018	100.00	1.00	2404	T4	4.3.2019	112.00	1.00
1400	T0	30.8.2018	0.00	0.00	2500	T0	3.12.2018	0.00	0.00
1401	T1	14.9.2018	15.00	0.15	2501	T1	14.12.2018	11.00	0.12
1402	T2	16.10.2018	47.00	0.47	2502	T2	10.1.2019	38.00	0.40
1403	T3	12.11.2018	74.00	0.75	2503	T3	8.2.2019	67.00	0.71
1404	T4	7.12.2018	99.00	1.00	2504	T4	8.3.2019	95.00	1.00
1600	T0	6.9.2018	0.00	0.00	2600	T0	4.2.2019	0.00	0.00
1601	T1	21.9.2018	15.00	0.15	2601	T1	15.2.2019	11.00	0.12
1602	T2	23.10.2018	47.00	0.47	2602	T2	18.3.2019	42.00	0.44
1603	T3	19.11.2018	74.00	0.75	2603	T3	12.4.2019	67.00	0.71
1604	T4	14.12.2018	99.00	1.00	2604	T4	10.5.2019	95.00	1.00
1700	T0	7.9.2018	0.00	0.00	2700	T0	18.2.2019	0.00	0.00
1701	T1	1.10.2018	24.00	0.20	2701	T1	1.3.2019	11.00	0.12
1702	T2	31.10.2018	54.00	0.45	2702	T2	1.4.2019	42.00	0.44
1703	T3	27.11.2018	81.00	0.68	2703	T3	3.5.2019	74.00	0.78
1704	T4	4.1.2019	119.00	1.00	2704	T4	24.5.2019	95.00	1.00
1800	T0	8.10.2018	0.00	0.00	2800	T0	11.3.2019	0.00	0.00
1801	T1	19.10.2018	11.00	0.12	2801	T1	25.3.2019	14.00	0.15
1802	T2	15.11.2018	38.00	0.40	2802	T2	18.4.2019	38.00	0.41
1803	T3	14.12.2018	67.00	0.71	2803	T3	17.5.2019	67.00	0.72
1804	T4	11.1.2019	95.00	1.00	2804	T4	12.6.2019	93.00	1.00
1900	T0	3.10.2018	0.00	0.00	2900	T0	11.3.2019	0.00	0.00
1901	T1	12.10.2018	9.00	0.09	2901	T1	25.3.2019	14.00	0.15
1902	T2	12.11.2018	40.00	0.42	2902	T2	18.4.2019	38.00	0.41
1903	T3	7.12.2018	65.00	0.68	2903	T3	17.5.2019	67.00	0.72
1904	T4	7.1.2019	96.00	1.00	2904	T4	12.6.2019	93.00	1.00

Appendix I.6 HUGMM dataset

Another database containing sustained vowel recordings from normative subjects was collected at Hospital Universitario Gregorio Marañón of Madrid (HUGMM). The inclusion criteria in this database were not having suffered any organic, neurological, or psychological dysfunction affecting phonation or known addictive habits. Participants had to pass a laryngoscopy inspection to disregard laryngeal problems, and they were required to sign an informed consent form that was approved by the local ethics committee of Universidad Politécnica de Madrid. The normative database (HUGMM) was recorded in a quiet isolated room of about 50 m3 using a Cardioid Sennheiser ME4 "clip-on" microphone at a fixed distance of 20 cm from the participant's mouth. The sampling rate was 44,100 Hz (16-bit resolution) on a MOTU Traveller Firewire audio recording platform, USB-connected to a portable computer.

Appendix I.7 HUGMM dataset description

The following table gives the demographic description of the participants' subset selected from HUGMM for the experimental framework described in subsection 5.1.1.1. In this case, NS participants have been labelled as N10xx (MNS: Male Normative Subject) and N11xx (FNS: Female Normative Subject). No clinical information is supplied as these participants were selected under strict inclusion criteria excluding any type of laryngeal, neurological, or psychological disorder.

Table App. 4 Normative subject set demographic data from HUGMM (NS).

Code	Gend	Age	Cond	Code	Gend	Age	Cond	Code	Gend	Age	Cond
N1005-a	M	21	N	N1035-a	M	35	N	N1043-a	M	39	N
N1018-a	M	29	N	N1036-a	M	42	N	N1044-a	M	45	N
N1027-a	M	34	N	N1037-a	M	50	N	N1045-a	M	48	N
N1028-a	M	28	N	N1038-a	M	41	N	N1046-a	M	60	N
N1030-a	M	25	N	N1039-a	M	58	N	N1047-a	M	36	N
N1032-a	M	48	N	N1040-a	M	47	N	N1048-a	M	44	N
N1033-a	M	42	N	N1041-a	M	53	N	N1049-a	M	62	N
N1034-a	M	30	N	N1042-a	M	42	N	N1050-a	M	56	N
N1105-a	F	43	N	N1126-a	F	40	N	N1139-a	F	42	N
N1108-a	F	22	N	N1127-a	F	59	N	N1142-a	F	27	N
N1112-a	F	20	N	N1128-a	F	52	N	N1143-a	F	38	N
N1116-a	F	45	N	N1130-a	F	35	N	N1144-a	F	46	N
N1117-a	F	25	N	N1132-a	F	30	N	N1145-a	F	29	N
N1120-a	F	33	N	N1134-a	F	28	N	N1146-a	F	24	N
N1121-a	F	57	N	N1137-a	F	24	N	N1147-a	F	55	N
N1125-a	F	38	N	N1138-a	F	33	N	N1149-a	F	50	N

Appendix I.8 APARKAM dataset

This dataset includes speech, accelerometry, and sEMG data from eight Spanish native speakers (four males and four females, stage 2 on H&Y scale) who were recruited from a PD patient association in the metropolitan area of Madrid (Asociación de Pacientes de Parkinson de Alcorcón y Móstoles, APARKAM), as well as from eight HC age-paired volunteers (four males and four females) participating in the study. The sEMG on the masseter was recorded, as well as the acceleration on the chin, simultaneously with the speech signal during the utterance of specific diadochokinetic exercises, as shown in Figure 5.1. The selection of the masseter as the target muscle obeys to the following reasons: it is a powerful muscle developing a strong sEMG when contracting, it is accessible (beneath the caudal section of the cheek), it may modify strongly the oral cavity when contracting or relaxing leaving a clear acoustic signature in formants, and its biomechanical activity is well understood. The equipment used allows the simultaneous and synchronous recording of masetter sEMG, 3DAcc, and speech, as illustrated in Figure 5.1. The demographic and clinical information relative to the participant subsets is given in Table App. 5.

Table App. 5 Biometrical Description of the participants included in the study.

Label	Age	Gender	H&Y	State	Label	Age	Gender	H&Y	State
CM1	69	M	-	-	CF1	66	F	-	-
CM2	70	M	-	-	CF2	62	F	-	-
CM3	61	M	-	-	CF3	65	F	-	-
CM4	68	M	-	-	CF4	65	F	-	-
PM1	73	M	2	on	**PF1**	69	F	2	on
PM2	71	M	2	on	PF2	73	F	2	on
PM3	73	M	2	on	PF3	71	F	2	on
PM4	69	M	2	on	PF4	70	F	2	on

The study was approved by the Ethical Committee of UPM (MonParLoc, 18/06/2018). The voluntary participants were informed about the experiments to be conducted, and the protection of their personal data, and signed an informed consent. The methodology was strictly aligned with the Declaration of Helsinki.

Appendix I.9 APARKAM dataset recording protocol

The equipment used allows the simultaneous and synchronous recording of masetter sEMG, 3DAcc, and speech, as illustrated in Figure 5.1. The facial sEMG is recorded by the two attachments of the masseter complex to the jaw and skull, and the 3DAcc signals are obtained from an accelerometer attached to the chin. These signals were digitised and collected with a Biopac MP150 EMG100 at 2 kHz and 16 bits. A Sennheiser cardioid wireless microphone (ew320 g2) on a MOTU Traveler MK1 sound card was used to record speech at 40 kHz with 32-bit resolution. The speech was later down-sampled to 8 kHz for the analysis of the first and second formants since the ranges for both formants are below 4 kHz (Huang, Acero, and Hon, 2001). The formant estimation is based on adaptive lattice filters (Deller, Proakis, and Hansen, 1993) producing a formant pair every 2 ms. Consequently, sEMG and 3D accelerometer signals were down-sampled to 500 Hz to match this time resolution. Recordings were carried out following a protocol that comprises the sustained vowels [*a*: *e*: *i*: *o*: *u*:], the fast repetition of the syllables [*pa*], [*ta*] and [*ka*], the three connected syllables [*pataka*] and [*pakata*] and the diphthong […a→i→a→i...].

Appendix II GFAD Three-way comparison results

Table App. 6 JSDs between each PD, HC, and NS distribution to NS and PD averages.

MPD	avMNS	avMPD	MHC	avMNS	avMPD	MNS	avMNS	avMPD
P2005-a	0.2872	0.1870	K2001-a	0.2662	0.2377	N1005-a	0.1157	0.2587
P2009-a	0.1358	0.1214	K2002-a	0.1768	0.1097	N1018-a	0.0956	0.2541
P2010-a	0.1117	0.1844	K2008-a	0.3299	0.2228	N1027-a	0.0976	0.2173
P2012-a	0.2534	0.1146	K2009-a	0.0852	0.1806	N1028-a	0.1377	0.2447
P2017-a	0.2150	0.1323	K2010-a	0.3860	0.2579	N1030-a	0.0929	0.2548
P2018-a	0.2234	0.0643	K2011-a	0.3160	0.2299	N1032-a	0.0848	0.2644
P2019-a	0.3012	0.1843	K2013-a	0.1627	0.1091	N1033-a	0.1093	0.2618
P2023-a	0.3666	0.3302	K2014-a	0.2621	0.2221	N1034-a	0.0859	0.2517
P2024-a	0.2986	0.1933	K2015-a	0.2036	0.2393	N1035-a	0.0765	0.2204
P2028-a	0.1557	0.1900	K2016-a	0.3367	0.2324	N1036-a	0.0886	0.1946
P2029-a	0.3669	0.2392	K2027-a	0.2467	0.1341	N1037-a	0.1635	0.1833
P2030-a	0.3369	0.2196	K2032-a	0.2429	0.1521	N1038-a	0.0876	0.1826
P2032-a	0.3068	0.1846	K2033-a	0.2219	0.1740	N1039-a	0.1001	0.1740
P2034-a	0.2801	0.1451	K2034-a	0.2612	0.2210	N1040-a	0.1799	0.1727
P2037-a	0.1678	0.1483	K2035-a	0.3989	0.3345	N1041-a	0.1485	0.2018
P2038-a	0.2086	0.1152	K2038-a	0.2376	0.1788	N1042-a	0.0815	0.2561
P2039-a	0.2157	0.1204	K2041-a	0.4253	0.3498	N1043-a	0.1201	0.2172
P2043-a	0.2721	0.1567	K2042-a	0.4252	0.3561	N1044-a	0.1400	0.1920
P2044-a	0.2805	0.1324	K2043-a	0.2110	0.1044	N1045-a	0.1443	0.1655
P2045-a	0.0975	0.1493	K2044-a	0.3432	0.2087	N1046-a	0.1055	0.2435
P2046-a	0.2553	0.1183	K2045-a	0.2528	0.1558	N1047-a	0.1283	0.2612
P2047-a	0.1589	0.1117	K2046-a	0.2861	0.1603	N1048-a	0.0901	0.2623
P2049-a	0.2979	0.1910	K2047-a	0.2961	0.2169	N1049-a	0.0798	0.2483
P2055-a	0.1316	0.1934	K2049-a	0.2716	0.2505	N1050-a	0.1051	0.1421
FPD	**avFNS**	**avFPD**	**FHC**	**avFNS**	**avFPD**	**FNS**	**avFNS**	**avFPD**
P1006-a	0.1969	0.1024	K1003-a	0.3966	0.2427	N1105-a	0.1034	0.2799
P1007-a	0.2317	0.1871	K1004-a	0.3324	0.1916	N1108-a	0.0973	0.2856
P1008-a	0.2658	0.0634	K1005-a	0.3468	0.1253	N1112-a	0.0997	0.2503
P1020-a	0.2301	0.1597	K1006-a	0.3829	0.2005	N1116-a	0.1849	0.3292
P1021-a	0.3542	0.1658	K1007-a	0.4057	0.2387	N1117-a	0.1198	0.2062
P1022-a	0.1780	0.1826	K1012-a	0.4341	0.2950	N1120-a	0.1129	0.3124
P1025-a	0.2137	0.0854	K1017-a	0.3346	0.1270	N1121-a	0.1245	0.3045
P1026-a	0.1776	0.2065	K1018-a	0.4568	0.3463	N1125-a	0.1449	0.1980
P1027-a	0.2666	0.1914	K1019-a	0.2502	0.1835	N1126-a	0.1170	0.3173
P1031-a	0.2214	0.1280	K1020-a	0.3424	0.1255	N1127-a	0.1137	0.2148
P1033-a	0.2452	0.2797	K1021-a	0.3784	0.1883	N1128-a	0.1362	0.3198
P1040-a	0.2244	0.1804	K1022-a	0.2998	0.3100	N1130-a	0.1257	0.3136
P1041-a	0.4473	0.2631	K1023-a	0.3294	0.1859	N1132-a	0.0910	0.2829
P1051-a	0.3196	0.2132	K1024-a	0.3106	0.1180	N1134-a	0.1152	0.2746
P1052-a	0.4562	0.3663	K1025-a	0.2389	0.1000	N1137-a	0.1383	0.2697
P1053-a	0.2967	0.1482	K1026-a	0.2590	0.2334	N1138-a	0.1126	0.2717
P1058-a	0.3871	0.2048	K1029-a	0.1944	0.1269	N1139-a	0.2215	0.3777
P1064-a	0.3525	0.1559	K1030-a	0.1778	0.1568	N1142-a	0.1548	0.3355
P1066-a	0.3895	0.1883	K1031-a	0.3961	0.1824	N1143-a	0.1975	0.3234
P1068-a	0.3116	0.1189	K1036-a	0.3523	0.1757	N1144-a	0.1836	0.3424
P1071-a	0.3763	0.2234	K1040-a	0.3323	0.1986	N1145-a	0.1511	0.1727
P1073-a	0.2245	0.1495	K1048-a	0.3926	0.2609	N1146-a	0.1680	0.3132
P1076-a	0.3594	0.2119	K1051-a	0.3011	0.1843	N1147-a	0.1595	0.2410
P1103-a	0.3507	0.1222	K1053-a	0.3279	0.1815	N1149-a	0.2073	0.2149

Appendix III Different configurations of the ONPT

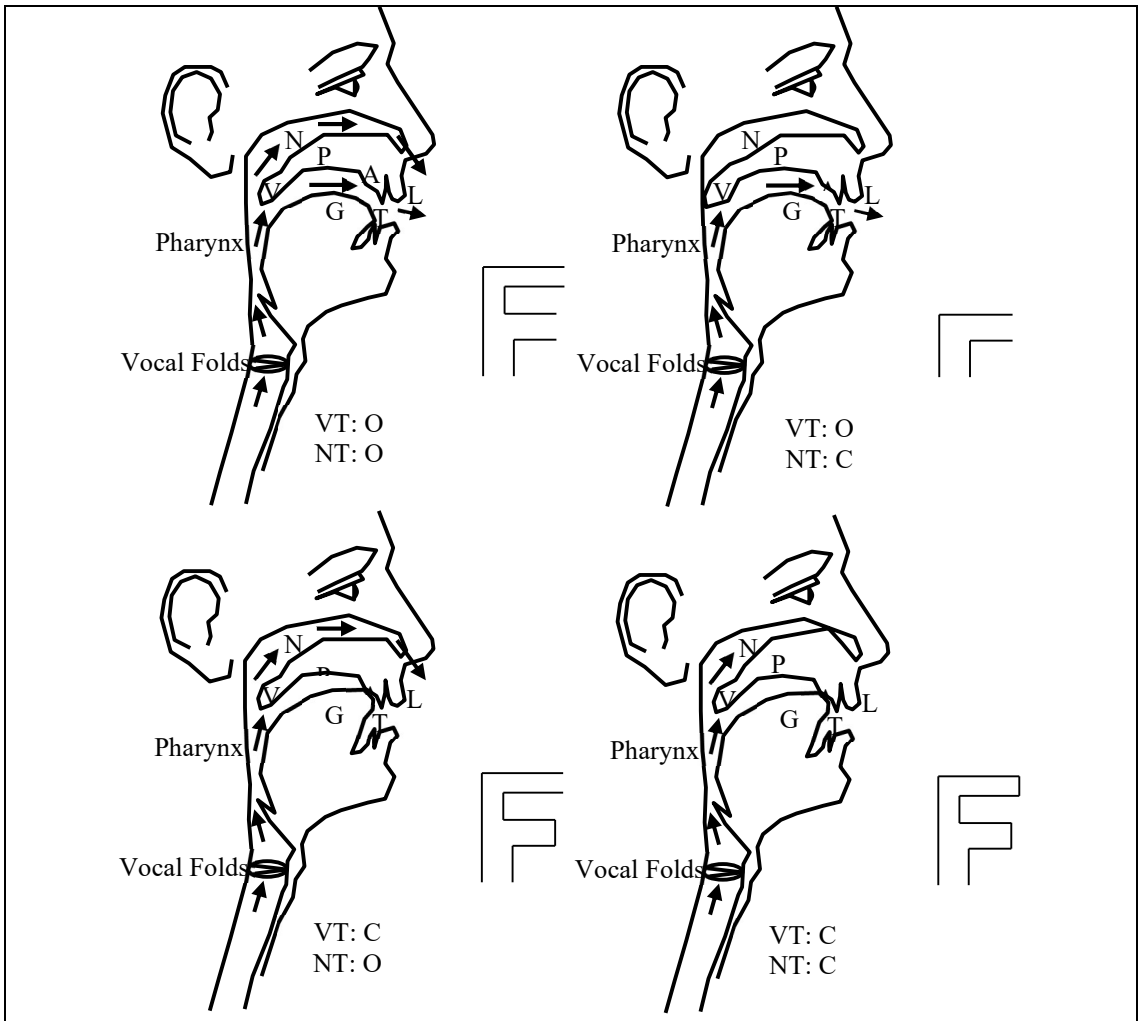

Figure App. 1 Different configurations of the ONPT.
Description: a) VT: O, NT: O, vocal tract open, nasal tract open. Antiresonances provoked by the open nasal tract at one-fourth of wavelength correspond to a pole-zero system, compatible with nasalised vowels. b) VT: O, NT: C, vocal tract open, nasal tract closed. Idealised one-path vocal tract responding to an all-pole system, compatible with pure vocal vowels. c) VT: C, NT: O, vocal tract closed, nasal tract open; depending on the oral closure (labial, dental, or velar) the system may behave as a nasal tract with vocal antiresonances (pole-zero system), or as a pure nasal tract (all-pole nasal system). d) VT: C, NT: PC, vocal tract closed, nasal tract partly closed; this configuration is compatible with a hyponasalised phonation.