MODEL-BASED TRAJECTORY ESTIMATION OF LARYNGEAL BIOMECHANICS FOR CHARACTERIZING CENTRAL AND AUTONOMIC DYSREGULATION IN VOICE PRODUCTION

BY

Rocío Belén Ortega Vargas

A dissertation submitted

in partial fulfillment of the requirements

for the degree

Master of Science in Electronic Engineering

Universidad Técnica Federico Santa María

Valparaíso, Chile

January 2026

CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☐ Memoria o trabajo de título ☒ Tesis de Postgrado

Título del trabajo: Model-based trajectory estimation of laryngeal biomechanics for characterizing central and autonomic dysregulation in voice production.

Nombre del candidato(a): Rocío Belén Ortega Vargas

Carrera / Grado: Magíster en Ciencias de la Ingeniería Electrónica

Campus: Campus Casa Central **Departamento:** Departamento de Electrónica

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, **Matías Zañartu**, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por **(marcar una opción)**:

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

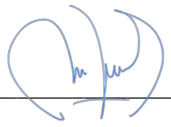
_____

_____

_____

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: jueves, 12 de marzo de 2026 **Firma:** 

Estudiante o Candidato(a):

Fecha: jueves, 12 de marzo de 2026 **Firma:** 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

"Model-based trajectory estimation of laryngeal biomechanics for characterizing central and autonomic dysregulation in voice production," a dissertation prepared by Rocío Belén Ortega Vargas in partial fulfillment of the requirements for the degree, Master of Science, has been approved and accepted by the following:

Matías Zañartu
Thesis Advisor

Alejandro J. Weinstein
Chair of the Examining Committee

Date

Committee in charge:

Dr. Matias Zañartu

Dr. Alejandro J. Weinstein, Chair

Dr. Kimberly L. Dahl

Dr. Defne Abur

DEDICATION

I would like to dedicate this thesis to my partner and my beloved dogs. Thank you for your unwavering support and for providing the emotional comfort and containment I needed throughout this journey. You have been my anchor.

To my family and friends, for their constant encouragement, love, and for being present at every stage of my life.

Finally, to the entire team at USM, VPLab, and AC3E. Thank you for providing the academic environment that made this research possible and for contributing significantly to my professional growth.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Professor Matías Zañartu. His mentorship, vast expertise, and scientific rigor have been the cornerstone of this Master's thesis. I am sincerely thankful for the trust he placed in me to conduct this research and for his constant guidance, which has significantly shaped my development as a researcher during this program.

I also extend my appreciation to the members of my thesis committee: Dr. Alejandro Weinstein, Dr. Kimberly Dahl, and Dr. Defne Abur, for their time, constructive feedback, and valuable insights, which contributed to refining the quality of this work.

A special acknowledgement goes to the team at the VPLab. I am particularly grateful to Jesús Parra, Emiro Ibarra, Nicolás Quinteros, Juan Pablo Cortés, Josué Martinez, and Christian Castro. Beyond being exceptional colleagues, you have become close friends and a vital support network. Your encouragement, camaraderie, and our shared experiences were essential in helping me navigate and complete this process.

Finally, I gratefully acknowledge the financial support that made this research possible, provided by the NIH, National Institute on Deafness and Other Communication Disorders, grant P50 DC015446; ANID grants FONDECYT 1230828, BASAL CIA250006, and National Master's Scholarship 22231668.

ABSTRACT

MODEL-BASED TRAJECTORY ESTIMATION OF LARYNGEAL BIOMECHANICS FOR CHARACTERIZING CENTRAL AND AUTONOMIC DYSREGULATION IN VOICE PRODUCTION

Rocío Belén Ortega Vargas

Master of Science in Electronic Engineering

Universidad Técnica Federico Santa María

Valparaíso, Chile, 2026

Dr. Matías Zañartu Salas, Chair

This thesis focuses on the development and application of a model-based trajectory estimation framework for laryngeal motor control, with the goal of inferring latent biomechanical variables from acoustic signals to improve the assessment of neurological and functional voice disorders. The motivation lies in the need to bridge the gap between observable acoustic outputs and the underlying physiological processes, allowing for a deeper understanding of vocal motor impairments associated with both central neurodegeneration and autonomic dysregulation.

The first stage of this research addresses the methodological configuration and operational characterization of the regression–based forward mapping used to approximate the laryngeal plant. Building upon established simulations of the Triangular Body–Cover Model, this work systematically evaluates distinct combinations of laryngeal control parameters and acoustic outputs to identify the specific input–output interface that maximizes estimation accuracy and numerical stability. This selection process establishes the robust parametric foundation required for the reliable inverse estimation of control trajectories in subsequent analyses.

The second stage applies this configured framework to the characterization of central motor impairment in Parkinson's disease. By reconstructing continuous biomechanical trajectories from sustained vowel phonations, the study reveals that PD–related motor deficits manifest as reduced stability and altered temporal coordination of laryngeal motor commands compared to healthy speakers. These biomechanically informed features demonstrate superior discriminative power compared to surface–level acoustic descriptors, highlighting their potential as objective markers for the automated assessment of hypokinesia and rigidity in voice production.

Extending the framework to functional dysregulations, the third stage examines the autonomic modulation of vocal motor control under elevated cognitive load. Analyzing continuous speech produced during a Stroop task, the research

quantifies how cognitive demand reshapes laryngeal muscle activation and aerodynamic control. The results indicate that individuals with vocal hyperfunction exhibit larger and less adaptive modulations of biomechanical control variables compared to healthy controls, suggesting that autonomic arousal acts upon an already constrained motor system, amplifying maladaptive compensatory strategies.

This thesis advances quantitative voice assessment through three complementary contributions. First, from a methodological standpoint, it validates the feasibility of using regression–based forward mappings to infer continuous laryngeal motor trajectories from standard, non–invasive audio. This bridges the gap between theoretical biomechanical modeling and clinical signal processing. Second, in the domain of neurological impairment, it operationalizes biomechanical stability as a specific marker of central neurodegeneration in PD, distinguishing it from healthy phonation. Third, regarding functional adaptation, it provides a physiological account of the cost of cognitive load in vocal hyperfunction, revealing that while healthy systems adapt, hyperfunctional systems exhibit distinct, maladaptive aerodynamic and muscular adjustments. Collectively, this work shifts the analytical focus from descriptive acoustics toward physics–based modeling of human motor control.

Contents

List of Tables

List of Figures

Chapter 1
Introduction

1.1 Motivation

Neurological integrity is fundamental to human motor performance in daily life [2], yet it is frequently compromised by dysfunctions occurring at two distinct physiological levels: central motor processing [3] and autonomic regulation [4]. Central neurodegenerative disorders, such as Parkinson's Disease (PD), represent one of the leading neurological conditions worldwide, predominantly affecting the elderly population by impairing motor planning and execution capabilities [5, 6, 7]. Conversely, autonomic nervous system (ANS) dysregulation, often manifesting as chronic physiological arousal or maladaptive stress responses, affects a broad demographic spectrum, contributing to occupational health issues and functional disorders [8, 9]. Both axes of dysregulation are significant challenges to public health, not only due to their prevalence but also because they severely limit an individual's ability to interact with their environment and perform daily tasks efficiently.

These conditions operate through divergent pathophysiological mechanisms that systemically alter bodily function. Central dysregulation, exemplified by PD, is driven by the progressive loss of dopaminergic neurons in the basal ganglia, resulting in a systemic breakdown of motor modulation characterized by rigidity, bradykinesia, and tremor [10]. In contrast, autonomic dysregulation involves an imbalance in the sympathetic-parasympathetic homeostasis [11]. In this state, the body may remain in a persistent state of heightened physiological arousal ("fight or flight"), leading to increased muscle tone [12], altered cardiovascular variability [13, 14], and a reduced capacity to adapt to environmental or cognitive stressors [15].

The diagnosis of central neurodegeneration is primarily based on clinical observation and subjective rating scales, frequently supported by advanced neuroimaging techniques or invasive procedures like lumbar punctures to analyze cerebrospinal fluid biomarkers [16, 17]. Similarly, the evaluation of autonomic function typically relies on the measurement of physiological markers such as heart rate variability, electrodermal activity, or cortisol levels, which usually require controlled laboratory conditions [18]. These methods, while effective for diagnosis, are impractical for long-term monitoring, creating a critical need for accessible, non-invasive markers that can reflect these internal states.

In this context, the human voice emerges as a powerful candidate, as phonation constitutes a highly complex motor task that depends on precise neuromuscular

control and is therefore vulnerable to central neurological degeneration, while at the same time serving as a primary channel for emotional and physiological expression that is strongly modulated by autonomic arousal [19, 20]. In PD, central motor deficits manifest directly in phonation, where rigidity and bradykinesia impair access to motor planning and induce stiffness in the laryngeal muscles [21]. Consequently, these deficits produce a characteristic dysregulation across phonatory and prosodic domains, manifesting as hypophonia (reduced loudness), dysphonia, and mono-pitch or mono-loudness (reduction in dynamic range) [22, 23].

Conversely, autonomic dysregulation impacts vocal motor control through a distinct pathway driven by systemic physiological arousal, often triggered by elevated cognitive load or psychosocial stress [13, 14]. This process alters vocal behavior, leading to modifications in pitch, loudness, and phonation stability [24, 25, 26]. These acoustic adjustments are driven by the modulation of laryngeal function, specifically through increased muscle activation [27, 28], which serves as a compensatory mechanism to maintain intelligibility and communicative under cognitive demand [29]. Crucially, such activation profiles mirror the characteristic patterns observed in subjects with vocal hyperfunction (VH), suggesting that hyperfunctional behaviors may represent a maladaptive compensatory strategies [30].

To utilize the voice as a clinical marker, research has traditionally relied on acoustic features extracted from the output signal, such as jitter, shimmer, cep-

stral peak prominence (CPP), mel-frequency cepstral coefficients, fundamental frequency (f_o), and sound pressure level (SPL) [22, 31, 32]. While these features correlate with both PD and autonomic alterations, these parameters are not the underlying physiological variables that define phonation motor control function function, such as laryngeal muscle activity and subglottal pressure (P_S). Although indirect approaches exist to estimate these aerodynamic [33] and muscular parameters [34], obtaining "ground truth" data typically relies on invasive procedures such as tracheal punctures or electromyography (EMG) [35, 36].

To address this limitation, numerical computational models of phonation offer a powerful non-invasive alternative for estimating latent physiological variables. By leveraging acoustic signals, recent frameworks have successfully employed machine learning to infer vocal folds (VFs) physiology [37]. Furthermore, deep learning architectures have been developed to estimate specific biomechanical parameters, such as P_S and laryngeal muscle activity, primarily utilizing neck–surface vibration signals [38, 39, 40]. More recently, inverse mapping strategies have been applied to ambulatory acoustic data to infer physiological parameters in hyperfunctional speakers [41]. Building on these foundations, Parra et al. [1] recently expanded this approach by integrating a laryngeal control model capable of mapping biomechanical trajectories directly from acoustic measures, such as f_o and SPL. This framework combines a physiologically grounded plant based on the Triangular Body-Cover Model (TBCM) [42], which captures the layered viscoelastic

dynamics of the VFs, with a feedback controller employing an inverse-Jacobian algorithm [43]. This dual structure allows for the iterative adjustment of motor inputs, effectively linking aerodynamic drive and laryngeal coordination within a unified motor-control framework.

Building on this modeling paradigm, the present work adopts a model–based trajectory estimation approach to move beyond acoustic descriptors toward representations of laryngeal motor behavior. Specifically, acoustic features are used to drive a biomechanical–control model that estimates time-varying laryngeal muscle activation and P_S, allowing vocal output to be interpreted in terms of underlying motor control strategies. This framework provides a means to examine how central motor impairment and autonomic modulation differentially shape phonatory behavior, while maintaining a non–invasive and physiologically interpretable methodology.

In summary, this thesis seeks to determine whether laryngeal motor control patterns estimated from acoustic signals provide sensitive markers of neurological impairment and cognitive loading. To this end, the study focuses on two key objectives: first, to evaluate whether these patterns differ systematically between speakers with PD and healthy controls (HC), determining if they improve group discrimination relative to acoustic features alone; and second, to assess the sensitivity of these motor control patterns to increased cognitive load in speakers with vocal hyperfunction compared to HC.

1.2 Aims and hypotheses

1.2.1 Overall aim

The overall aim of this thesis is to determine whether laryngeal motor control patterns estimated from acoustic speech features provide sensitive and interpretable markers of altered vocal motor control associated with neurological impairment and cognitive loading.

1.2.2 Specific aims

1. To determine whether estimated laryngeal motor control patterns derived from acoustic speech features differ systematically between speakers with Parkinson's disease and healthy controls, and whether these patterns improve group discrimination relative to acoustic features alone.

2. To determine whether estimated laryngeal motor control patterns are sensitive to increased cognitive load in speakers with vocal hyperfunction, compared to healthy speakers.

1.2.3 Hypotheses

- **H1:** Estimated laryngeal motor control patterns derived from acoustic speech features differ systematically between speakers with Parkinson dis-

ease and healthy controls, and provide statistically improved discrimination between groups compared to acoustic features alone.

PD arises from progressive dopaminergic degeneration within basal ganglia circuits, leading to impairments in motor planning, reduced movement amplitude, rigidity, and deficits in temporal coordination [5, 10]. From a motor control perspective, these alterations reflect a reduced capacity to generate stable, precisely timed motor commands and to flexibly coordinate multiple effectors. Such central motor constraints extend to voice production [6], where they manifest as hypophonia, reduced prosodic variability, and phonatory instability [7, 23]. Most prior work has characterized these vocal deficits using acoustic features such as jitter, shimmer, and spectral measures, which capture surface-level variability in the acoustic signal [22, 31]. While informative, these descriptors do not directly index the latent motor control variables that govern phonation. In contrast, physiologically informed voice production models explicitly link neural control signals to biomechanical execution, enabling the estimation of laryngeal muscle activations, P_S, and their temporal coordination. Given that PD-related motor impairment constrains the ability to initiate, sustain, and finely modulate laryngeal gestures, such biomechanical control variables are expected to exhibit reduced stability and altered coordination compared to HC. Estimating these variables through a model–based trajectory framework therefore

provides a principled means of revealing latent motor control constraints imposed by central neurodegeneration, beyond what can be inferred from acoustic variability alone.

- **H2:** Under increased cognitive load, speakers with vocal hyperfunction exhibit statistically larger and more consistent changes in estimated laryngeal motor control patterns than healthy speakers, reflecting a heightened and maladaptive autonomic response in the laryngeal mechanism.

Cognitive load and psychosocial stress reliably elicit autonomic activation that biases the laryngeal system toward increased intrinsic muscle engagement, elevated glottal stiffness, and augmented aerodynamic demands during phonation [27, 44, 28]. From a motor control perspective, these autonomically mediated adjustments reflect a reorganization of laryngeal control strategies under increased cognitive or emotional demand, resulting in systematic changes in muscle activation patterns and P_S regulation. In healthy speakers, such adjustments are generally transient and adaptive, allowing the vocal motor system to maintain acoustic stability and intelligibility despite increased task demands [29]. The system retains sufficient flexibility to modulate laryngeal muscle activation and aerodynamic control without entering maladaptive regimes. In contrast, individuals with VH exhibit chronically elevated or imbalanced intrinsic laryngeal muscles (ILMs) activation,

reflecting a persistent alteration of vocal motor control [30, 45]. This baseline state effectively constrains the available motor control space. Under conditions of elevated cognitive load, autonomic modulation is therefore expected to operate on an already constrained motor system, amplifying laryngeal muscle coactivation and aerodynamic adjustments, and further limiting the flexibility of phonatory control. As a consequence, changes in estimated biomechanical control variables (ILMs and P_S) are expected to be larger and less adaptive than those observed in healthy speakers, consistent with increased autonomic influence on vocal motor behavior.

1.3 Methodology overview

The methodology of this thesis is explicitly grounded in the laryngeal motor control framework recently proposed by Parra et al. [1]. This computational architecture employs a closed–loop feedback controller that operates in acoustic task space, utilizing an inverse–Jacobian formulation to update motor commands based on the error between target and realized acoustic outputs. Within this framework, the acoustic outputs are generated through a forward mapping that approximates the laryngeal plant using simulations of the TBCM and machine–learning regression. Accordingly, the methodology is structured into three main stages, corresponding to the three core chapters of the thesis.

First, the framework is customized and optimized at the level of the forward mapping. Since the original design allows for the independent manipulation of five ILMs in addition to P_S, but the specific constraints of the available experimental data may not support the estimation of the full high–dimensional motor space, this stage focuses on adapting the model to the specific needs of this study. This process involves the generation of specific synthetic datasets to train the regression models and the systematic evaluation of different input–output configurations. Specifically, distinct subsets of control variables and various acoustic feature vectors are tested to determine the configuration that yields the best trade–off between estimation accuracy, numerical stability, and physiological plausibility. This step establishes the parametric and structural basis required for reliable inverse trajectory estimation in subsequent analyses.

Second, in alignment with the first specific aim, the validated framework is applied to the study of central motor impairment in PD using sustained vowel phonation. Spanish voice recordings from the PC-GITA [46] and NeuroVoz [47] databases are preprocessed to isolate stable voiced segments suitable for numerical modeling. The model–based estimator is then used to reconstruct continuous trajectories of ILMs activations and P_S from measured acoustic targets (f_o and SPL). These biomechanical trajectories are analyzed alongside conventional acoustic features to assess their capacity to characterize PD–related motor deficits. Statistical group comparisons and supervised classification analyses, utilizing a Support Vec-

tor Machine (SVM) with stratified group k-fold cross-validation, are conducted to evaluate the discriminative power of the estimated biomechanical variables as markers of PD.

Third, in alignment with the second specific aim, the same trajectory estimation framework is applied to continuous speech produced under varying cognitive load in healthy speakers and individuals with VH. Speech recordings obtained during a color–word Stroop task are preprocessed to remove silences and outliers, and segmented at the word level into temporal intervals. The model is used to estimate time–varying laryngeal control trajectories during continuous speech, enabling the examination of how autonomic arousal associated with cognitive demand modulates ILMs activation and aerodynamic control. Linear mixed-effects (LME) models are employed to quantify the main and interaction effects of group, task condition (low cognitive vs. high cognitive demand), and temporal segment on both acoustic and biomechanical variables, with post hoc analyses conducted to probe significant effects and contrasts.

Together, these three methodological stages establish a unified, physiologically interpretable framework for estimating and analyzing vocal motor control trajectories, enabling a direct comparison between motor rigidity arising from PD and functional constraints associated with cognitive load.

1.4 Significance and contributions

Understanding the motor control mechanisms underlying neurological and autonomic dysregulations remains a central challenge in contemporary speech science. Despite the widespread use of non–invasive acoustic recordings, a persistent gap exists between observable vocal outputs and the latent physiological processes that generate them. This thesis addresses this gap by advancing a model–based framework capable of estimating continuous biomechanical control trajectories from acoustic signals, thereby linking measurable vocal behavior to physiologically meaningful motor variables in both clinical and functional contexts.

Building on this validated framework, a first major contribution is the evaluation of estimated biomechanical control variables as physiologically interpretable markers of PD. By reconstructing continuous trajectories of ILMs activation and P_S from sustained vowel phonation, this work enables the characterization of PD-related hypokinesia and rigidity directly within the motor control space. Unlike traditional acoustic descriptors, which capture surface–level signal variability, the proposed approach provides insight into how central neurodegeneration constrains the stability, coordination, and temporal organization of laryngeal motor commands. These findings support the use of model–derived biomechanical features as complementary markers for automated classification and objective assessment of motor dysfunction in PD.

A second contribution of this thesis is the application of the same trajectory estimation framework to the study of vocal motor control modulation under cognitive load. By analyzing continuous speech produced during a Stroop task, this work quantifies how increased cognitive demand associated with autonomic arousal reshapes laryngeal muscle activation and aerodynamic control in real time. The comparison between healthy speakers and individuals with VH provides a physiologically grounded account of how compensatory control strategies that are adaptive in healthy systems may become amplified and potentially maladaptive in the presence of chronic motor constraints. This analysis reveals subtle increases in laryngeal muscle coactivation and P_S that are often masked in conventional acoustic analyses, offering new insight into the physiological cost of cognitive load in vocally hyperfunctional speech.

In conclusion, this thesis advances quantitative voice assessment by shifting the analytical focus from descriptive acoustic features toward physics–based, physiologically interpretable modeling of vocal motor control. The unified inverse-mapping framework presented here serves as a non-invasive and transferable tool for estimating biomechanical motor control variables, enabling a principled distinction between motor rigidity arising from PD and functional constraints associated with cognitive load. As such, this work contributes both methodological rigor and conceptual clarity to the study of voice as a window into human motor control.

1.5 Publications

This thesis is supported by a set of publications.

1.5.1 Journals papers in review

1. **Ortega R.B.**, Parra J.A., Quinteros N.F., Ibarra E.J., Stepp C.E., and Zañartu M. *"Effects of cognitive load on vocal motor control: A model-based analysis in individuals with hyperfunctional and healthy voices"*. Submitted for publication at JSLHR Journal of Speech, Language, and Hearing Research.

1.5.2 Conferences

1. **Ortega R.B.**, Parra J.A., Quinteros N.F., Ibarra E.J., Stepp C.E., and Zañartu M. *"Effects of cognitive load on vocal motor control: Changes in subglottal pressure and laryngeal muscle activation"*. 16th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research, AQL 2025, June, 24–27, 2025. Groningen, Netherlands.

2. **Ortega R.B.**, Quinteros N.F., and Zañartu M. *"Método para detectar carga cognitiva elevada a través de un modelo de control motor laringeo"*. 80° Congreso Chileno de Otorrinolaringología, ORL 2024, November, 20–23,

2024. Valdivia, Chile.

1.6 Document structure

The chapters in this thesis are organized as follows. Section chapter 2 establishes the theoretical foundations of the study. It introduces the biomechanical principles of phonation, including the body–cover theory of the VFs and the source–filter framework. This chapter further reviews the hierarchical organization of neural control of voice production, encompassing cortical planning, sensorimotor feedback, and autonomic modulation. Particular emphasis is placed on distinguishing central motor dysregulation in Parkinson's disease from autonomic imbalance associated with vocal hyperfunction. Finally, computational modeling of phonation is introduced as a non-invasive methodological approach for estimating latent laryngeal control variables. Section chapter 3 presents the implementation and systematic configuration of the model-based trajectory estimation framework. This chapter focuses on the forward mapping component of the model, examining how biomechanical variables derived from the TBCM relate to different acoustic representations of phonation. Correlation analyses and performance metrics are used to evaluate alternative input–output configurations and polynomial model complexities, guiding the selection of the most robust and numerically stable forward mappings for subsequent inverse trajectory estimation.

Section chapter 4 applies the validated framework to the characterization of central motor impairment in PD. Using sustained vowel phonation from Spanish-language speech databases, this chapter describes the reconstruction of continuous trajectories of laryngeal muscle activation and P_S from acoustic targets. The estimated biomechanical variables are analyzed alongside conventional acoustic features to assess their discriminative capacity as objective markers of disease–related motor dysfunction, using statistical group comparisons and supervised classification approaches. Section chapter 5 examines autonomic modulation of vocal motor control under conditions of increased cognitive load. The framework is applied to continuous speech recordings obtained during a color–word Stroop task in healthy speakers and individuals with VH. This chapter analyzes how mental demand shapes laryngeal control strategies in real time and tests the hypothesis of reduced regulatory flexibility and exaggerated biomechanical responses in vocally hyperfunctional speech through mixed-effects modeling and post hoc analyses. Finally, Section chapter 6 summarizes the main findings and contributions of the thesis. It integrates the results obtained across both central and autonomic axes of motor control, emphasizing the transition from descriptive acoustic analysis to physics-based, physiologically interpretable modeling. The chapter concludes by outlining limitations and proposing future research directions in quantitative voice assessment and vocal motor control.

Chapter 2
Background

This chapter establishes the theoretical and methodological framework necessary to understand the interplay between physiological mechanisms, neurological control, and the computational modeling of voice production. The review begins by defining the fundamental biomechanics of phonation (Section 2.1), with a specific focus on the body–cover theory and the source–filter model as the physical basis of speech. Subsequently, it explores the hierarchical organization of neural control, detailing how cortical planning, sensorimotor feedback, and autonomic modulation interact to regulate laryngeal function.

Building upon these foundations, the chapter examines the mechanisms of neurological dysregulation (Section 2.2), contrasting the central motor rigidity characteristic of Parkinson's Disease with the autonomic compensatory patterns observed in Vocal Hyperfunction. This section also reviews current clinical assessment methods and the evidence linking these systemic alterations to specific vocal changes, thereby validating the potential of the voice as a non–invasive clinical marker. Finally, the chapter introduces computational modeling (Section 2.3)

as a methodological solution to estimate biomechanical motor control variables directly from acoustic signals.

2.1 Mechanisms of voice production

Vocie production results from the coordinated interaction among the respiratory, laryngeal, and supralaryngeal systems, functioning under the hierarchical control of the nervous system. These systems operate interdependently to generate a dynamic process in which pulmonary airflow, the viscoelastic properties of the VFs, and the resonances of the vocal tract are integrated to produce acoustic outputs [48]. From a neurophysiological perspective, speech production is not merely an acoustic or aerodynamic phenomenon; rather, it represents a complex motor control process that integrates physiological, biomechanical, and neural mechanisms.

2.1.1 Physiology and biomechanics

A fundamental framework for describing the physiology and biomechanics of voice, considering it as a process where aerodynamic energy is transformed into acoustic energy, is the source–filter theory [49]. This model proposes that voice production involves two distinct stages: the generation of a sound source via the modulation of glottal airflow by the vibrating VFs, and the subsequent acoustic

filtering performed by the vocal tract configuration.

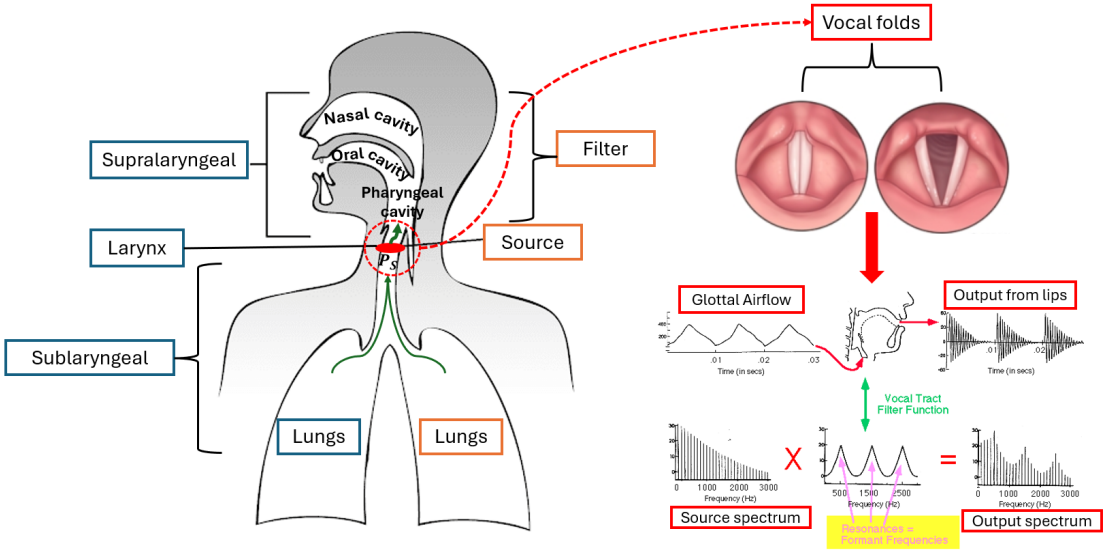


Figure 2.1: Schematic overview of the main components involved in voice production, highlighting the source–filter relationship (orange boxes) and its interaction with the underlying physiological subsystems (blue boxes). Adapted from MIT OpenCourseWare [50].

Figure 2.1 illustrates the relationship between the source–filter model and the physiological anatomy. Through this process, three primary components interact to constitute the phonatory mechanism: (1) the VFs, (2) the glottal airflow dynamics, and (3) the vocal tract resonances. These components are described in detail below.

Vocal folds

The vocal folds, located within the larynx (see Figure 2.1), are the primary oscillators responsible for modulating expiratory airflow to generate the periodic vibration essential for voice. These structures define the glottis, the space through which airflow is modulated during phonation, and serve as the central anatomical reference in phonation. Anatomically, the VFs exhibit a "V" configuration [51], and their structural composition is best described by the body–cover theory proposed by Hirano [48]. According to this model, the VFs consist of multiple layers with distinct mechanical properties (as shown Figure 2.2): a superficial epithelial layer; the lamina propria, subdivided into superficial, intermediate, and deep layers; and an internal muscular layer corresponding to the thyroarytenoid (TA) muscle.

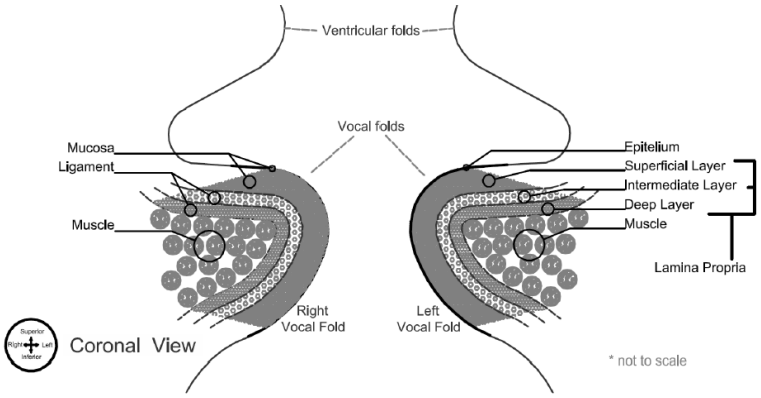


Figure 2.2: VFs layered structure. Reprinted from Titze [52]

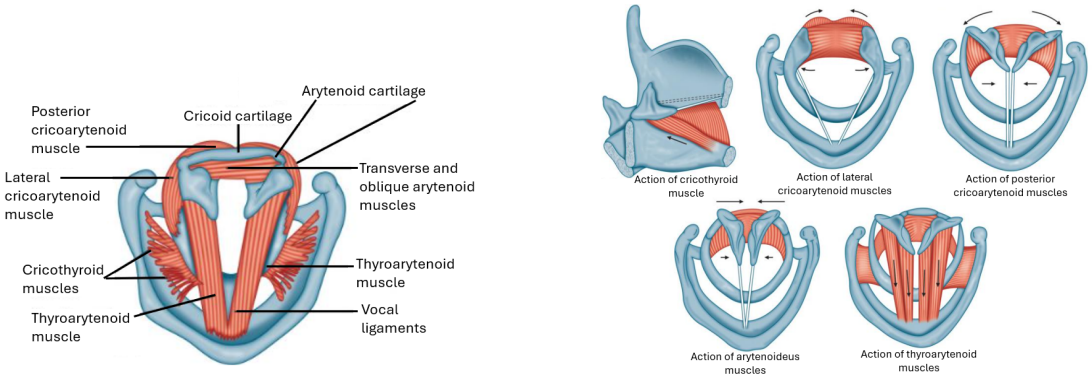
The epithelium forms a thin, lubricated surface that facilitates contact dur-

ing glottal closure. Beneath it, the superficial layer of the lamina propria, also known as Reinke's space, contains a loose matrix favorable for vibration. The intermediate and deep layers are denser, dominated by elastic and collagen fibers respectively, providing tensile strength. Together, these layers form the "body", while the epithelium and superficial layer constitute the "cover", which is more flexible and responsible for the mucosal wave propagation. This organization confers nonlinear viscoelastic properties to the VFs, enabling stable and adaptable oscillation across a wide range of frequencies and intensities [53, 54].

The dynamic control of these properties relies on the coordinated action of the ILMs, which adjust the position, shape, and tension of the VFs (Figure 2.3). The lateral cricoarytenoid (LCA) and interarytenoid (IA) muscles generate glottal adduction (closure). In contrast, the posterior cricoarytenoid (PCA) muscle serves as the sole abductor. The cricothyroid (CT) muscle elongates the VFs, increasing longitudinal tension to raise the f_o. Conversely, the TA muscle shortens and thickens the VFs, modulating both pitch and loudness. The antagonistic interaction between the CT and TA muscles is the primary mechanism for regulating VFs stiffness and length, determining oscillation frequency and phonatory quality [55].

The extrinsic laryngeal muscles, or strap muscles of the neck, control the vertical movements of the larynx by altering the angles and distances between cartilages through their elastic arrangement. These adjustments modify the resting length of the ILMs, indirectly influencing fine phonatory control and overall voice

quality [56].



(a) (b)

Figure 2.3: Anatomical and functional representation of the ILMs. Panel (a) shows the position of each ILMs. Panel (b) illustrates the principal actions of the ILMs involved in phonation. Adapted from Sataloff [55].

Glottal airflow

To obtain glottal airflow, the lungs first release subglottic airflow, which travels through the lower region of the larynx toward its midsection, where the glottis and VFs are located. Since the VFs are adducted (or closed), the P_S must exceed the mechanical resistance of the VFs to force abduction (or opening) of the VFs in the convergent phase. Subsequently, the acceleration of airflow causes a reduction in glottal pressure (P_g) due to the Bernoulli effect, leading to VFs closure during the divergent phase, thereby generating a cycle of self-sustained oscillations. This repeated oscillation modulates the airflow exhaled by the lungs, as illustrated in

Figure 2.4 and constitutes the glottal airflow or sound source [57].

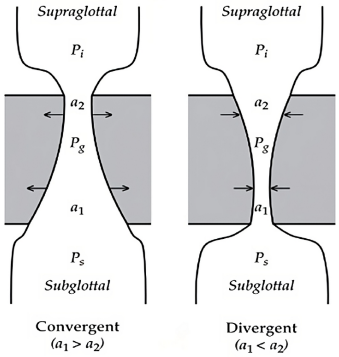


Figure 2.4: Diagram of the glottal cycle showing the convergent and divergent phases of VFs oscillation. Reprinted from Story [57].

Mathematically, the glottal airflow can be approximated as a function of the glottal area (a_g) and the pressure differences. Equation 2.1 describes this relationship, where P_i is the supraglottal pressure, A^* is the effective area between the upper and lower portions of the VFs, k_t is an empirical transglottal pressure coefficient, ρ is the air density, and c is the speed of sound:

$$u_g = \frac{a_g c}{k_t}\left(-\frac{a_g c}{A^*} + \left[\left(\frac{a_g}{A^*}\right)^2 + \frac{4k_t}{c^2\rho}(P_S - P_i)\right]^{\frac{1}{2}}\right) \tag{2.1}$$

Vocal tract

The vocal tract is composed of the pharyngeal, oral, and nasal cavities, functioning as the resonating system that transforms the glottal airflow generated by VFs oscillation into articulated sound. The spatial configuration of these struc-

tures is dynamically determined by the movements of the articulators, specifically the tongue, lips, jaw, soft palate (velum), and pharyngeal walls, as shown in Figure 2.1. This geometric modulation alters the acoustic features of the sound, defining the quality, timbre, and intelligibility of the voice. From the perspective of the source–filter theory [49], the vocal tract behaves as an acoustic resonator that filters the glottal sound source, producing resonances known as formants, whose frequency and bandwidth depend directly on the instantaneous geometry of the articulators (see Figure 2.1).

2.1.2 Neural motor control

Speech relies on a highly distributed neural control system that integrates cortical planning, subcortical modulation, and motor execution at the brainstem level [19, 58, 6]. This network regulates not only the precise movements of the larynx [59, 20] but also its coordination with respiration, prosody, and emotional expression [60]. As a whole, the voice emerges as a manifestation of hierarchical motor control that allows continuous adjustment of vocal production to communicative and contextual demands [19, 61].

Figure 2.5 illustrates the cortical architecture supporting this process. The diagram highlights the functional connectivity between anterior regions, responsible for the planning and generation of motor commands, and posterior regions involved in the processing of auditory and somatosensory feedback. This cortical

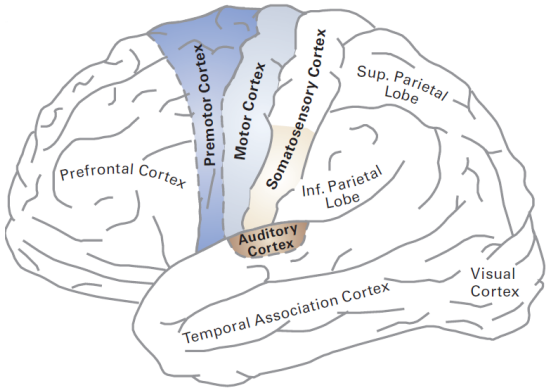


Figure 2.5: Schematic representation of the primary cortical regions involved in speech motor control. Reprinted from Guenther [19].

network constitutes the structural basis for the neural motor control integration mechanisms described in the following sections.

Cortical and subcortical control

The speech process begins in the cortical areas responsible for planning and programming motor actions. Broca's area, located in the left inferior frontal gyrus (see Figure 2.5), participates in organizing phonological patterns and converting linguistic representations into motor commands [62]. These signals are integrated in the supplementary and premotor cortices, which coordinate the temporal sequencing and initiation of speech [63]. At a more direct level, the primary motor cortex of the precentral gyrus contains somatotopic representations of the laryngeal, articulatory, and respiratory muscles, generating the motor impulses that control phonation [64].

Cortical activity is modulated by subcortical loops involving the basal ganglia, thalamus, and cerebellum (Figure 2.6a). The basal ganglia play a critical role in selecting and inhibiting appropriate motor programs, preventing redundant or inefficient movements during speech [65]. The cerebellum, in turn, ensures temporal synchronization and precision of articulatory and phonatory gestures [66], being essential for rhythmic stability and speech fluency [67].

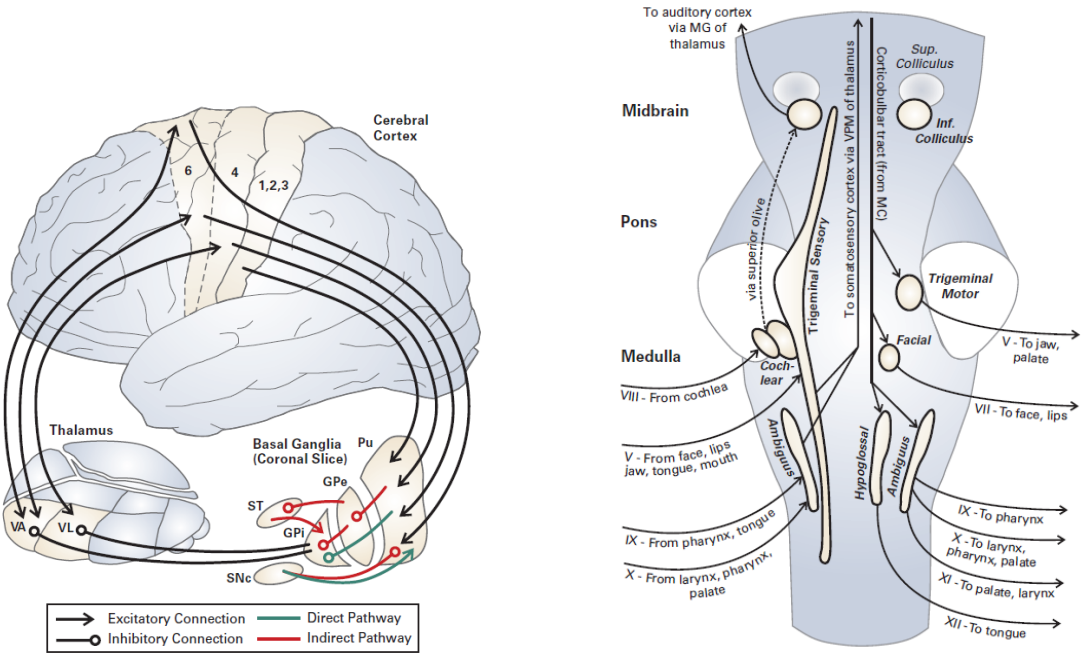
Motor execution and brainstem pathways

Once planned, descending motor commands travel from the motor cortex through the corticobulbar tract, establishing synapses in the motor nuclei of the brainstem [68]. As illustrated in Figure 2.6a, this transmission allows for the integration of descending cortical signals with reflex loops prior to muscle activation.

In the medulla oblongata, the nucleus ambiguus gives rise to the laryngeal branches of the vagus nerve (cranial nerve X), as detailed in the anatomical view of Figure 2.6b. Among these, the recurrent laryngeal nerve innervates most ILMs, whereas the external branch of the superior laryngeal nerve innervates the CT muscle [69]. This organization enables independent control of VFs tension, adduction, and length, thereby modulating the f_o and loudness.

Articulatory muscles are innervated by the trigeminal (V), facial (VII), glossopharyngeal (IX), vagus (X), and hypoglossal (XII) nerves, which are also coordinated at the brainstem level. These nuclei receive bilateral cortical projections,

ensuring symmetrical movements during speech production. Thus, vocal motor execution operates as an integrated system that combines high–resolution cortical commands with automatic postural and respiratory control mechanisms [19, 70].



(a) (b)

Figure 2.6: Motor execution pathways. (a) Schematic representation of the descending motor pathways, illustrating how cortical commands are refined by the cerebellum before execution. (b) Anatomical location of the cranial nerve nuclei in the brainstem involved in speech and voice production. Adapted from Guenther [19].

Sensorimotor integration

The precision of phonation depends on auditory and somatosensory feedback mechanisms that enable real–time adjustments in vocal output. Auditory feedback is processed in the superior temporal gyrus, where the perceived acoustic features are compared with internal predictions generated by the motor system [19, 71]. Discrepancies between the two signals trigger rapid corrective adjustments in voice production [72].

Somatosensory feedback, including tactile and proprioceptive information from the vocal tract and larynx, is transmitted to the thalamus and projected to the somatosensory cortex, contributing to fine control of articulation and phonation. These circuits form a closed–loop control system, supported by specific transcerebellar loops as illustrated in Figure 2.7. Within this architecture, efference copies of motor commands are compared with sensory outcomes, ensuring vocal stability in the presence of external or internal variations such as cognitive load, physical effort, or emotional state [73].

2.1.3 Autonomic modulation

Beyond the voluntary control of speech, the phonatory system is subject to continuous modulation by the ANS, which regulates the baseline physiological state of the vocal mechanism. This modulation is mediated by limbic structures,

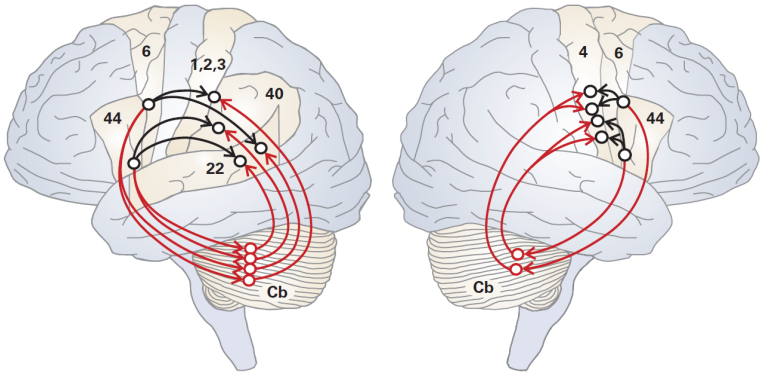


Figure 2.7: Schematic representation of the transcerebellar pathways involving distinct hemispheric roles. The left panel depicts the left-hemisphere loop connecting premotor cortical areas with auditory (Brodmann area 22) and somatosensory regions (Brodmann areas 1, 2, 3, 40), which is responsible for generating sensory targets. The right panel illustrates the right–hemisphere loop connecting premotor areas to the bilateral primary motor cortex, responsible for issuing corrective motor commands. Cb: Cerebellum. Reprinted from Guenther [19].

including the anterior cingulate cortex, amygdala, and periaqueductal gray, which project directly to the laryngeal motor nuclei in the brainstem, operating in parallel with the corticobulbar motor pathways [19, 20].

Under homeostatic conditions, the interplay between the sympathetic and parasympathetic branches of the ANS maintains optimal laryngeal muscle tone. However, physiological or psychological arousal, triggered by cognitive load or emotional stress, shifts this balance toward sympathetic dominance. Research indicates that sympathetic activation leads to a generalized increase in the tension of

both intrinsic and extrinsic laryngeal muscles, altering the viscoelastic properties of the VFs, whereas parasympathetic predominance tends to favor more relaxed and stable phonation [74, 75, 27].

Within this framework, vocal production emerges from the coordinated interaction of neural motor control, physiological regulation, and biomechanical processes. Central motor planning and subcortical modulation shape laryngeal motor commands, which are executed at the brainstem level and translated into biomechanical actions through respiratory drive, laryngeal muscle activation, and tissue dynamics, all under continuous autonomic modulation. Because these mechanisms operate jointly to determine phonatory behavior, changes in their functional state are systematically reflected in the acoustic structure and temporal dynamics of the voice. Consequently, vocal output can be exploited as an indirect yet sensitive marker of neurophysiological regulation, capturing alterations arising from both central motor dysfunction and autonomic imbalance. The following section examines how dysregulation across these neural, physiological, and biomechanical control levels has been associated with characteristic alterations in voice production.

2.2 Neurological dysregulation and voice markers

Neurological health relies on the precise regulation of motor and autonomic systems to maintain homeostasis and facilitate interaction with the environment. When these regulatory mechanisms fail, whether due to structural neurodegeneration or functional dysregulation, the consequences manifest across multiple physiological systems, significantly impacting the individual's daily life [76]. In the context of this thesis, we distinguish between two primary axes of dysregulation: one of central motor origin, exemplified by PD, and another of autonomic origin, associated with stress responses and VH. Understanding these conditions primarily as systemic disorders is essential before analyzing their specific impact on voice production.

2.2.1 Central motor control dysregulation and Parkinson's disease

Parkinson's Disease is the second most prevalent neurodegenerative disorder worldwide, affecting approximately 2–3% of individuals over the age of 65 [16]. Beyond its clinical characterization as a movement disorder, PD fundamentally represents a disorder of motor control, arising from progressive dysfunction

within cortico–subcortical circuits responsible for the planning, modulation, and execution of voluntary movement. Central to this pathology is the degeneration of dopaminergic neurons in the substantia nigra pars compacta, which disrupts the balance of excitatory and inhibitory signaling within the basal ganglia–thalamocortical loops that regulate movement amplitude, timing, and coordination [77, 10]. As a consequence, motor output becomes globally hypokinetic and rigid, manifesting clinically as bradykinesia, rigidity, and resting tremor [64].

The clinical assessment of PD motor symptoms relies primarily on observational rating scales, most notably the Movement Disorder Society–Unified Parkinson's Disease Rating Scale (MDS–UPDRS) [78]. While widely adopted, these assessments are inherently subjective, episodic, and limited in their capacity to capture subtle or continuous changes in motor function. Objective confirmation may be supported by neuroimaging techniques such as positron emission tomography or electroencephalography (EEG), but these approaches are costly, invasive, or impractical for longitudinal monitoring [16]. This diagnostic landscape highlights the need for accessible, non–invasive markers capable of reflecting the integrity of central motor control in PD.

Crucially, the dopaminergic dysfunction that compromises limb motor control also affects the neural circuits governing speech production. Because the ILMs are regulated by the same basal ganglia–cortical pathways involved in general motor control, vocal impairment in PD is not merely a secondary consequence

of the disease but often an early and pervasive manifestation of central motor dysregulation. Deficits in motor scaling and coordination disrupt the fine control of VFs activation, adduction, and timing, leading to reduced phonatory precision and stability [21]. Clinically, these impairments present as hypophonia, dysphonia, and reduced articulatory range, producing the characteristic mono–pitch and mono–loudness patterns commonly observed in individuals with PD [79].

Beyond impairments in motor execution, PD also compromises the sensorimotor integration mechanisms that support vocal self–regulation. In neurologically healthy speakers, speech production relies on the continuous interaction between motor commands and sensory feedback, allowing the system to anticipate and correct deviations in vocal output through auditory and somatosensory monitoring. This auditory–motor integration enables dynamic adjustments of loudness and pitch, as evidenced by adaptive behaviors like the Lombard effect [80, 81]. In PD, degeneration within cortico–subcortical networks disrupts this integration, leading to attenuated perception of self–generated vocal changes and delayed or reduced compensatory responses.

Empirical evidence supports this disruption. Individuals with PD exhibit diminished sensitivity to auditory feedback perturbations, showing reduced ability to modulate loudness and [82] pitch [83] in response to external or internal demands [84]. Neurophysiological studies using EEG have demonstrated delayed and attenuated N100 and P200 responses following pitch or loudness perturba-

tions, indicating impaired auditory error detection [85, 86] and slower corrective motor adjustments [87, 88]. These deficits extend beyond phonatory control to affect emotional prosody, as modulation of pitch and loudness is essential for affective expression. Functional neuroimaging studies have linked altered activity in structures, such as the subthalamic nucleus and amygdala to impairments in the recognition and expression of vocal emotions, underscoring the close coupling between motor and affective circuits in PD [89, 90, 91].

The neurophysiological and biomechanical consequences of central motor dysregulation in PD are ultimately reflected in the acoustic structure of the voice. Increased laryngeal rigidity and micro–instabilities in VFs vibration give rise to cycle–to–cycle perturbations, which are commonly quantified using acoustic features such as jitter, shimmer, and Harmonics–to–Noise Ratio (HNR) [92, 93]. Similarly, reduced articulatory range resulting from hypokinesia manifests acoustically as vowel centralization, captured by metrics such as the Vowel Articulation Index and the Triangular Vowel Space Area, which consistently show reduced articulatory distinctiveness in speakers with PD [94].

While these acoustic features have demonstrated sensitivity to Parkinsonian speech, they primarily reflect the downstream consequences of central motor impairment rather than the mechanisms that generate them. As such, vocal alterations in PD should be interpreted as expressions of disrupted motor scaling, timing, and sensorimotor integration within basal ganglia–cortical networks, rather

than as isolated acoustic anomalies.

2.2.2 Autonomic dysregulation and vocal hyperfunction conditions

Autonomic dysregulation refers to a persistent alteration in the balance and adaptability of the ANS [95], resulting in impaired coordination between sympathetic and parasympathetic activity [96, 97]. Under healthy conditions, the ANS dynamically regulates involuntary physiological processes [11], including cardiovascular function, respiration [98, 99], thermoregulation, and muscle tone in response to internal demands and environmental challenges [100]. Autonomic dysregulation emerges when this adaptive regulation is compromised, leading to sustained states of arousal that are disproportionate to situational demands [11].

The causes of autonomic dysregulation are multifactorial and include chronic psychological stress [101, 102], maladaptive emotional regulation, traumatic experiences [103, 104], prolonged cognitive overload [97, 105], and dysfunctional cortical–subcortical integration [106, 107]. Neurophysiologically, these conditions are associated with altered communication between limbic structures, prefrontal cortical regions, brainstem nuclei, and peripheral autonomic pathways, resulting in reduced autonomic flexibility and impaired recovery to baseline states following stress exposure [101, 108]. Rather than reflecting a transient response, autonomic

dysregulation is characterized by a persistent pattern of physiological regulation [109, 104].

Systemically, autonomic dysregulation manifests across multiple bodily systems, independent of vocal behavior. Common physiological correlates include elevated resting heart rate, reduced heart rate variability (HRV), irregular respiratory patterns, altered electrodermal activity, increased baseline muscle tone, gastrointestinal disturbances, and disrupted sleep–wake cycles [96, 110, 111]. Clinically, these alterations are often reported as palpitations, dyspnea, excessive sweating, chronic fatigue, or heightened somatic awareness. Importantly, such manifestations indicate a generalized state of autonomic imbalance rather than localized dysfunction [112, 95].

Assessment of autonomic dysregulation therefore relies on multimodal physiological evaluation. Clinically, autonomic function is commonly examined through cardiovascular reflex tests, respiratory measures, and standardized questionnaires targeting autonomic symptoms. In experimental and research settings, objective evaluation increasingly relies on wearable and laboratory–based sensors, including heart rate and HRV, respiratory rate and variability, electrodermal activity, skin temperature, and, in some cases, EEG to index central autonomic regulation [112, 113, 111]. These measures provide quantitative indices of sympathetic–parasympathetic balance and autonomic reactivity during rest and task engagement, allowing the characterization of both baseline arousal and stress responsiv-

ity.

Within this broader physiological framework, phonation represents a motor behavior that is particularly sensitive to autonomic modulation [114]. Because speech production requires fine coordination between respiratory drive, laryngeal muscle activation, and articulatory timing, changes in autonomic state directly influence vocal motor control [98]. Sympathetic hyperactivation increases overall muscle tone, elevates intrinsic and extrinsic laryngeal muscle activation, and biases the phonatory system toward increased VFs stiffness and P_S [27, 115]. Conversely, reduced autonomic activation may result in insufficient muscular engagement and unstable airflow control. In both cases, autonomic imbalance interferes with the sensorimotor mechanisms that normally ensure efficient and adaptable vocal production.

Cognitive load and psychosocial stress are among the most potent triggers of autonomic activation during speech. Under conditions of elevated cognitive or emotional demand, sympathetic dominance induces coordinated changes across respiratory, cardiovascular, and muscular systems, which are consistently reflected in altered phonatory behavior [116, 117, 118]. Acoustic studies have demonstrated systematic changes in f_o, loudness, and voice quality during cognitively demanding tasks, supporting the notion that vocal output acts as a downstream expression of autonomic state [44, 28].

VH can be understood within this context as a chronic, maladaptive expres-

sion of autonomic dysregulation [119, 120, 30]. Rather than representing a purely local laryngeal phenomenon, VH appears to reflect a persistent state of elevated autonomic arousal in which sympathetic activation remains chronically upregulated [121]. Individuals with VH typically exhibit increased baseline muscle tension and reduced capacity to downregulate physiological arousal, rendering the phonatory system continuously biased toward excessive muscular engagement [119]. This sustained autonomic arousal compromises motor efficiency and limits the system's ability to flexibly adapt to changing communicative demands.

Clinically, VH is characterized by excessive or inefficient recruitment of laryngeal musculature during phonation, leading to strained, effortful, or poorly coordinated glottal closure patterns [122, 30]. It encompasses both phonotraumatic VH, in which prolonged mechanical stress produces structural tissue changes, and non-phonotraumatic VH, where dysphonia arises from dysfunctional motor coordination without overt tissue pathology [123, 30]. Across both forms, the persistence of autonomic arousal contributes to a self-reinforcing cycle of physiological stress, muscular tension [120], and vocal fatigue [124].

Taken together, these findings support the interpretation of VH as a manifestation of sustained autonomic arousal, in which elevated sympathetic activity biases the phonatory system toward excessive muscular tension and reduced regulatory flexibility. Within this framework, vocal alterations reflect a systemic dysregulation of autonomic control rather than a purely local laryngeal disorder.

Across both central motor and autonomic axes, the evidence reviewed in this section highlights a common principle: voice production constitutes a highly sensitive motor output shaped by the interaction of neural planning [19, 6], subcortical modulation [125], autonomic regulation [107, 114], and peripheral biomechanics [52, 126]. Because these systems converge on the control of respiration, laryngeal muscle activation, and VFs dynamics, alterations in their functional integrity are systematically expressed in vocal behavior. From this perspective, the voice does not merely reflect communicative intent but encodes information about the underlying neurophysiological state of the speaker.

This sensitivity has motivated extensive research on vocal acoustic features as non–invasive markers of neurological dysregulation [31]. Measures such as f_o variability, intensity modulation, spectral perturbation indices, and articulatory space metrics have demonstrated robust associations with both PD and conditions involving heightened autonomic arousal [117, 121, 29]. These approaches offer clear practical advantages, including ease of acquisition, low cost, and suitability for longitudinal monitoring. However, acoustic features primarily characterize the outcome of vocal production rather than the internal mechanisms that generate it. Similar acoustic patterns may arise from distinct neuromotor or biomechanical sources, limiting their interpretability when the goal is to understand the nature of the underlying dysregulation [57].

Importantly, the core alterations in both central motor disorders and auto-

nomic dysfunctions are hypothesized to reside at an internal level of control–namely, in the regulation of laryngeal muscle activation, VFs stiffness, P_S, and sensorimotor coordination [30, 27, 83]. Direct measurement of these variables in humans is technically challenging and often invasive, requiring procedures such as EMG, endoscopic assessments, or P_S estimation, which are impractical for large-scale or repeated evaluations [33, 127]. This methodological constraint has historically limited the ability to directly link neural dysregulation to its biomechanical manifestations in voice.

In this context, physiologically grounded mathematical models of phonation provide a compelling alternative. By integrating knowledge of VFs biomechanics, motor control principles, and acoustic generation, such models enable the estimation of latent internal parameters from observable vocal signals [128, 38, 39]. Rather than treating the voice solely as an acoustic phenomenon, this approach conceptualizes it as the measurable output of an underlying dynamical system governed by neuromotor and autonomic regulation. Consequently, model–based estimation frameworks offer a principled pathway to bridge the gap between neural dysfunction and vocal acoustics outputs, allowing internal motor and biomechanical states to be inferred non–invasively.

This perspective reframes the voice as a functional marker of neurological dysregulation whose diagnostic and explanatory power can be substantially enhanced when acoustic observations are interpreted through physiologically informed mod-

els. The following section builds upon this rationale by introducing the modeling framework adopted in this thesis and detailing how internal vocal motor parameters can be estimated to characterize central and autonomic dysregulation in voice production.

2.3 Computational modeling of voice motor control

Computational models constitute a central theoretical and methodological tool for the quantitative study of voice production, as they provide an explicit formalization of the relationships between neural control, biomechanical execution, and acoustic output. Their development responds to a fundamental limitation in human voice research: the difficulty of directly accessing internal motor and physiological variables without invasive procedures. By simulating the phonatory system under controlled conditions, computational models enable the systematic exploration of mechanisms that are otherwise inaccessible in vivo, while preserving physiological interpretability.

Within this framework, voice production is conceptualized as the observable output of an underlying dynamical system governed by motor planning, sensorimotor control, and biomechanical constraints. Computational models have been instrumental in elucidating how neural commands are transformed into laryngeal

muscle activation patterns, how these activations shape VFs dynamics, and how biomechanical states ultimately give rise to acoustic signals [52, 129]. Importantly, these models provide a common quantitative space in which physiological and pathological conditions can be compared, offering a reproducible grounded basis for interpreting clinical and experimental observations.

Beyond forward simulation, computational modeling has enabled the investigation of neurological dysregulation in speech by linking alterations in control strategies to their biomechanical and acoustic consequences. Model–based approaches have been used to study impairments in sensorimotor integration, predictive control, and feedback regulation, revealing how disruptions at the neural level propagate through the phonatory system [19, 61, 130]. Collectively, these advances have contributed to a deeper understanding of the hierarchical organization of voice motor control and its tight coupling with laryngeal biomechanics, an understanding that is essential for interpreting voice as a marker of neurophysiological dysregulation.

2.3.1 Motor control theory overview

Within the framework of control theory, phonation is conceived as a dynamic process that requires the precise synchronization of multiple muscles with complex biomechanical properties and nonlinear relationships between their physiological variables and the resulting acoustic signal. To describe this process, motor con-

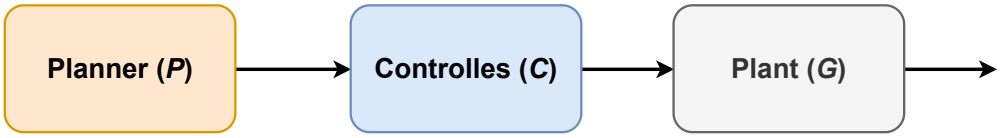


Figure 2.8: Schematic representation of a hierarchical motor control architecture applied to phonation. The planner P defines task-level goals in *task space*, the controller C transforms these goals into motor commands, and the plant G represents the biomechanical and physiological system that executes the commands.

trol models are typically structured hierarchically into three main components: planner (P), controller (C), and plant (G) [130] (see Figure 2.8).

The *Planner* generates a reference vector or motor target, represented as

$$r = P(a, t) \tag{2.2}$$

where a contains high-level information (e.g., the motor program associated with a phonetic, prosodic, or cognitive sequence) and t represents time. This vector defines the motor goal in the *task space*, such as a formant trajectory, a specific articulatory configuration, or an acoustic target associated with the vocal task [19].

The *Controller* transforms this target into motor commands in the *mobility space*, that is, into muscle activation signals or articulatory positions that allow achieving the desired goal [61]. In the absence of sensory feedback, the controller operates in feedforward mode according to

$$(u, \dot{u}) = C(r) \tag{2.3}$$

where u represents motor commands (e.g., ILMs activation) and $\dot{u}$ their temporal derivatives. When sensory feedback is included, the controller adjusts the commands based on the error between the desired state r and the estimated state of the plant $\hat{x}$ [130]:

$$e = r - \hat{x}, (u, \dot{u}) = C(e) \tag{2.4}$$

The estimated state $\hat{x}$ is obtained by combining the system's internal predictions $\tilde{x}$ with sensory signals from the auditory (y_{aud}) and somatosensory (y_{somat}) channels, such that

$$\hat{x} = f(\tilde{x}, y_{aud}, y_{somat}) \tag{2.5}$$

In the context of this thesis, these variables may correspond to P_S or ILMs activation, and the glottal configuration associated with the vibratory pattern.

To implement this control law mathematically, one of the most widely used approaches in speech motor control is the **Jacobian–based formulation** [71, 131]. Since the mapping between the high-dimensional *mobility space* (muscles) and the low-dimensional *task space* (acoustics) is nonlinear and redundant (i.e., multiple muscle patterns can produce the same sound), the controller relies on the

Jacobian matrix $J(u)$ to linearize this relationship around the current state. The Jacobian defines how small changes in motor commands $\dot{u}$ map to changes in the system's output $\dot{x}$:

$$\dot{x} \approx J(u)\dot{u} \tag{2.6}$$

To correct an error e, the controller must solve the inverse problem: finding the motor change $\dot{u}$ that eliminates the discrepancy. Because $J(u)$ is often not square (due to motor redundancy), this is typically solved using the pseudoinverse $J(u)^{-1}$ [132, 133], allowing the system to resolve the redundancy by selecting the motor solution with the minimum norm (smallest effort):

$$\dot{u} = J(u)^{-1}e \tag{2.7}$$

This formulation is central in the framework used in this thesis, as it provides a robust mechanism to transform sensory errors into precise motor corrections.

The *Plant*, in turn, represents the biomechanical and physiological system that executes the motor commands and generates the resulting acoustic signal [130]. Formally, it can be expressed as

$$(u, u, \dot{x}, \dot{x}) = G(u, \dot{u}) \tag{2.8}$$

In voice motor control, this may include the dynamics of the VFs, laryngeal muscle activation, P_S, and the acoustic response of the vocal tract. In summary,

the *mobility space* describes all possible configurations of the plant, while the *task space* reflects the functional state relevant to the task, such as formant value or lip aperture.

Within this hierarchical structure, motor control models can be implemented following different principles:

Feedforward control, generates motor commands solely from the plan, $(u, \dot{u}) = C(r)$, without sensory feedback. This approach allows fast and smooth movements but is vulnerable to cumulative errors or unexpected perturbations. For example, in speech, the tongue trajectory for producing /i/ might follow a pre-established path without correcting transient deviations [61].

Feedback control, computes motor commands based on the error $e = r - \hat{x}$, such that $(u, \dot{u}) = C(e)$, allowing deviations between the reference and the estimated plant state to be corrected. This scheme is robust to external perturbations, such as a temporary vocal tract obstruction, although its performance is limited by delays and noise in the sensory signal [71].

Model predictive control (MPC), uses an internal model of the plant $\tilde{x}$ to anticipate the effects of motor commands before sensory feedback is received, allowing real–time adjustments despite delays inherent to sensory pathways. For instance, the system may predict the effect of an increase in P_S or the f_o and adjust VFs activation before an acoustic deviation occurs [134].

Although these control architectures can operate independently, they are often combined into hybrid architectures to achieve the speed of feedforward control, the error correction of feedback control, and the anticipatory capabilities of internal models [130, 71]. This integration typically relies on an observer (e.g., a Kalman filter or Smith predictor) to merge sensory signals and state estimates [61]. Such an architecture enables the phonatory system to transform abstract representations in *task space* into motor commands in *mobility space*, balancing speed, accuracy, and adaptability to internal or external perturbations.

Taken together, this hierarchical and multi–level architecture provides a physiologically plausible framework to model the planning, execution, and correction of phonation, offering a basis for analyzing how planning, motor control, and biomechanics are integrated to produce adaptive and stable acoustic signals [130].

Within phonation, this control architecture operates over variables such as P_S, ILMs activation, and VFs configuration, whose coordinated regulation is required to achieve stable and flexible vocal output. Crucially, disturbances at any level of this hierarchy can lead to measurable changes in vocal behavior, making the phonatory system particularly sensitive to neurological and autonomic dysregulation.

2.3.2 Voice motor control models

Computational models of voice motor control are grounded in general principles of motor control theory, formalizing the interaction between motor planning, sensorimotor regulation, and physical execution through hierarchical architectures. These architectures typically decompose voice production into three functional components: a planner that specifies task–level goals, a controller that transforms these goals into motor commands, and a plant that executes the commands to generate articulatory or biomechanical movements and their resulting acoustic output.

Articulatory–based control architectures

Among the most influential neurocomputational models of speech and voice motor control are Directions Into Velocities of Articulators (DIVA) [19], State Feedback Control (SFC) [61], and Feedback–Aware Control of Tasks in Speech (FACTS) [134]. Despite differences in formal implementation, these models share a common control–theoretic perspective in which speech movements emerge from the interaction between internal predictions, sensory feedback, and an articulatory plant representing the vocal tract.

In DIVA, the plant corresponds to a kinematic–acoustic model of the supralaryngeal articulators, implemented via the Maeda articulatory synthesizer [135].

Motor commands generated by the controller drive this articulatory plant, while auditory and somatosensory feedback signals are compared against task-level targets defined by the planner. Through a learning process analogous to infant babbling, DIVA establishes internal forward and inverse mappings between motor commands and their sensory consequences, enabling stable feedforward control after training. A distinctive strength of DIVA lies in its explicit neuroanatomical grounding, which allows its components to be associated with cortical and subcortical structures observed in neuroimaging studies.

The SFC model adopts a similar planner–controller–plant decomposition but reframes speech production as a state estimation problem. The plant again represents the articulatory system, while the controller relies on an internal forward model to predict the plant state. Delayed and noisy sensory feedback is optimally integrated with these predictions through a Kalman–filter like observer [61]. This formulation provides a principled account of how voice motor control remains stable and accurate in the presence of sensory uncertainty and temporal delays.

FACTS extends this architecture by explicitly combining state estimation with task–based dynamical control [134, 136]. In FACTS, the plant is defined in articulatory space using the Configurable Articulatory Synthesizer [137], while speech gestures are modeled as dynamical systems in *task space*. An internal observer continuously estimates the articulatory state, allowing real–time corrective control based on sensory feedback.

Despite their conceptual and neurophysiological sophistication, DIVA, SFC, and FACTS share an important limitation for the study of voice production: phonation is treated implicitly within a source–filter paradigm, and the laryngeal system is not modeled as an explicit biomechanical plant. Consequently, these architectures are limited in their ability to account for voice–specific phenomena arising from variations in VFs stiffness, ILMs activation, or P_S, variables that are critically involved in both central motor disorders and autonomic regulation.

Biomechanical laryngeal plants

To overcome these limitations, LaDIVA was introduced as an extension of the DIVA framework [43], incorporating a physiologically grounded laryngeal plant based on the extended Body–Cover Model (ext–BCM) [138]. Whereas the classical Body–Cover Model (BCM) [129, 139] represents the vocal folds using a layered mass–spring system that captures body–cover interactions, the ext–BCM further incorporates explicit modeling of ILMs activations and their effects on tissue stiffness, and aerodynamic coupling.

Within LaDIVA, the controller generates motor commands in terms of P_S and ILMs activations, which are executed by the laryngeal plant to produce acoustic outputs such as f_o and SPL. This integration allows the model to directly link neural control variables to biomechanical execution and acoustic outcome, representing a substantial advance over purely articulatory plants.

However, this physiological realism comes at a cost. Mass–spring biomechanical models such as the BCM and ext–BCM require numerical integration of nonlinear differential equations, resulting in a high computational burden. In LaDIVA, this limitation is partially mitigated by replacing direct biomechanical simulation with precomputed lookup table (LUT) that approximate the forward mapping of the laryngeal plant. While LUT–based approaches reduce computation time, they introduce limitations in numerical smoothness, scalability, adaptability to individual anatomy, and suitability for continuous control or real–time estimation.

Machine learning approaches to laryngeal modeling

Motivated by the computational constraints of explicit biomechanical simulation, a growing body of work has explored machine learning (ML) techniques to approximate nonlinear mappings in voice production. ML–based approaches have been successfully applied to estimate laryngeal source parameters from acoustic and aerodynamic features, including P_S [140, 141], VFs biomechanical variables [142, 143, 37], and ILMs activations [38, 39, 41]. These methods offer substantial advantages in computational efficiency, robustness, and adaptability to inter–subject variability.

ML–based laryngeal models span a broad methodological spectrum. Feature–driven regression approaches rely on physiologically interpretable acoustic or aerodynamic descriptors to estimate control parameters with relatively low computa-

tional cost and improved transparency [38, 37]. In contrast, signal–based deep learning models operate directly on raw waveforms or trajectories [143]. Although these models can achieve high predictive accuracy, they typically require large training datasets, incur higher computational cost, and offer limited physiological interpretability.

Recent work has shown that ML regressors can also function as forward models of the laryngeal plant, mapping motor commands ($x = [P_S, a_{CT}, a_{TA}]$) directly to acoustic outputs ($y = [f_o, SPL]$). When embedded within closed–loop feedback controllers formulations, such learned plants enable dynamic trajectory tracking while remaining compatible with established motor control architectures.

Data–driven laryngeal plants within control

architectures

Within this context, Parra et al. [1] propose an extension of the LaDIVA framework in which LUT–based representations of the laryngeal plant are replaced by data–driven regression models, while explicitly preserving the planner–controller–plant architecture, as illustrated Figure 2.9. In this formulation, the plant is implemented as a learned forward mapping that predicts acoustic output from laryngeal motor commands, whereas the controller retains a task–space feedback structure grounded in classical control theory.

Rather than numerically integrating the underlying biomechanical equations

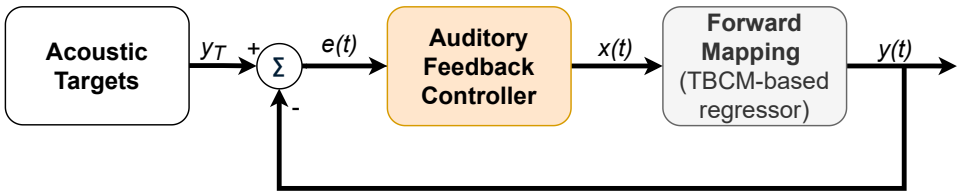


Figure 2.9: Block diagram of the closed-loop vocal control framework. The architecture integrates an auditory feedback controller and a data–driven forward plant, represented here by a TBCM-based regressor, which maps motor commands $x(t)$ to acoustic outputs $y(t)$. The controller utilizes the error $e(t)$ between acoustic targets y_T and current feedback to generate corrective motor updates.

at runtime, the forward mapping is approximated using regression models trained on physiologically grounded simulations. This approach preserves the functional role of the biomechanical plant while enabling efficient computation, numerical smoothness, and continuous control.

Feedback control is achieved by locally linearizing the learned forward mapping around the current operating point. Small deviations between current ($y(t)$) and target acoustic ($y_T(t)$) trajectories are translated into corrective updates in motor command space (Eq. 2.9) using the inverse Jacobian of the forward model (Eq. 2.10).

$$x(t + \Delta t) = x(t) + \alpha J^{-1}(x, t)\big(y(t) - y_T(t)\big) \tag{2.9}$$

$$J^{-1}(x,t) = J^T(x,t)\big(J(x,t)J^T(x,t) + \gamma^2 I\big)^{-1} \tag{2.10}$$

To ensure numerical robustness, the Jacobian inverse (J^{-1}) is computed using a damped least-squares formulation, which stabilizes control updates in regions of redundancy or ill-conditioning [132]. Update gain (α), damping factor (γ), numerical-derivative step size, and sampling frequency define the temporal behavior and stability of the closed-loop system.

A further distinguishing feature of this framework is the choice of biomechanical reference model. In addition to regressors trained on ext–BCM data, the Triangular Body–Cover Model (TBCM) [42] is incorporated as a physiologically grounded plant. Compared to BCM–based formulations, the TBCM provides improved anatomical realism, a richer description of ILMs stress–strain relationships, and enhanced capacity to capture asymmetric and dynamic phonatory behaviors.

Parra et al. [1] systematically evaluated multiple regression strategies, including Random Forests, Neural Networks, and Polynomial Regression. While Random Forests achieved the highest predictive accuracy, Neural Networks and Polynomial Regression produced smoother control signals and lower computational cost, making them better suited for closed–loop control. These results demonstrate that data–driven laryngeal plants can be embedded within physiologically grounded control architectures without sacrificing stability or interpretability.

By integrating learned biomechanical plants within an explicit control frame-

work, this approach bridges the gap between physiological realism and computational tractability. Because it builds upon the TBCM, which has shown strong agreement with experimental and in vivo data [38, 40, 41], it provides a validated and physiologically constrained basis for modeling laryngeal motor control. In the present thesis, this framework is adopted to infer latent vocal motor states from observable acoustic trajectories, offering a principled link between neural regulation and vocal expression.

2.4 Chapter conclusions

This chapter has established the physiological, neurological, and methodological foundations necessary to conceptualize voice production as the dynamic outcome of interacting central motor, biomechanical, and autonomic processes. By reviewing the core mechanisms of phonation, from the layered structure of the VFs and the role of intrinsic and extrinsic laryngeal muscles to the source–filter framework, voice production was framed as a tightly regulated motor behavior governed by nonlinear biomechanical constraints and hierarchical neural control.

Building upon this biomechanical basis, the chapter highlighted how vocal behavior reflects the integrity of both central motor planning and autonomic regulation. In particular, contrasting neurological conditions such as PD and functional disorders such as VH illustrated that distinct patterns of dysregulation, central

motor rigidity on one hand and sustained autonomic arousal on the other, can converge on similar phonatory subsystems. This convergence supports the view of the voice as a sensitive, non–invasive marker of underlying neurophysiological states, while also underscoring the interpretative limitations of purely acoustic descriptions.

Although extensive research has demonstrated associations between acoustic features and neurological or autonomic dysfunction, this review emphasized that acoustic measures primarily characterize the observable outcome of phonation rather than the internal control variables that generate it. Direct assessment of key physiological parameters, such as ILMs activation or P_S, remains technically challenging and often invasive, limiting their applicability in longitudinal or large-scale studies. As a result, a critical methodological gap persists between observable vocal signals and the latent motor and biomechanical mechanisms they encode.

In this context, computational modeling was introduced as a principled solution to bridge this gap. Physiologically grounded models of phonation provide a framework for linking neural control, muscle activation, and VF biomechanics to acoustic output within a unified dynamical system. By enabling the estimation of biomechanical control variables from acoustic data, these models shift voice analysis from purely descriptive acoustics toward an interpretation grounded in underlying control processes.

The following chapters build upon this framework by introducing and imple-

menting computational model of laryngeal motor control to infer internal vocal motor states from observed acoustic trajectories. These models are then applied to characterize patterns of central and autonomic dysregulation in voice production, with the goal of advancing both the theoretical understanding of phonatory control and the development of non-invasive tools for clinical and experimental research.

Chapter 3

Methodological configuration of the forward mapping for laryngeal biomechanics estimation

This chapter presents the methodological foundations for characterization and selection of a regression–based forward mapping that links laryngeal biomechanical variables to acoustic features derived from microphone recordings. The approach is strictly grounded in the biomechanical modeling framework developed by Parra et al. [1] and does not introduce a novel motor control architecture or modify the underlying physics of the TBCM. Instead, the contribution of this chapter lies in establishing an operational and empirically optimized configuration of the forward model inputs and outputs to ensure its suitability for the specific signals analyzed in this thesis.

Specifically, this chapter focuses on a systematic evaluation of the input–output interface of the forward mapping. Given that the Parra et al. [1] framework allows for the independent manipulation of five ILMs and P_S, and considering that multiple acoustic representations of phonation can be derived from the model

simulations, this stage addresses two complementary methodological tasks. First, correlation analyses are performed among candidate acoustic features to assess their information content and redundancy, guiding the selection of a compact and physiologically meaningful acoustic feature space. Second, multiple configurations of the forward mapping are evaluated by testing different subsets of control variables, acoustic outputs, and polynomial regression degrees. These configurations are quantitatively compared using estimation accuracy and numerical stability metrics to identify the most robust approximation for the specific experimental needs of this work.

By grounding the selection of acoustic features and biomechanical control variables in a systematic performance evaluation rather than heuristic assumptions, this chapter establishes the parametric basis for the forward component of the estimator. This optimized configuration provides the necessary foundation for the implementation of the model–based trajectory estimation in subsequent chapters, ensuring consistency, interpretability, and numerical reliability for the analysis of sustained phonation in PD and continuous speech under cognitive load.

3.1 Methodology scope within the laryngeal control architecture

This chapter operates within the model–based laryngeal control architecture established by Parra et al. [1], the structural organization of which was detailed in Chapter 2 (refer to Figure 2.9). While the complete Parra et al. [1] framework supports closed–loop motor control and inverse trajectory estimation via a planner–controller–plant hierarchy, the methodological focus of this chapter is strictly restricted to the configuration and performance evaluation of the forward biomechanical mapping corresponding to the Plant component.

As described in the reference framework, the architecture relies on acoustic targets ($y_T(t)$) defined in an auditory feature space to serve as planning variables. Deviations between desired and produced acoustic outputs ($e(t) = y_T(t) - y(t)$) are mapped onto motor command updates ($x(t)$) via a feedback controller using an inverse Jacobian formulation (Eq. 2.10). However, the implementation of these closed–loop control mechanisms is not the subject of the present chapter. Instead, the objective is to formalize and empirically characterize the forward relationship between laryngeal motor commands and acoustic outputs, ensuring that the Plant (TBCM) component is optimally configured for the specific signals analyzed in this thesis.

Crucially, although this chapter does not explicitly address the auditory planner component of the control architecture, the configuration of the forward mapping performed here implicitly defines the operational limits of the planner. Specifically, the selection of acoustic outputs $y(t)$ for the regression–based approximation of the biomechanical plant effectively delimits the acoustic *task space* that can be utilized as target trajectories $y_T(t)$ in the feedback control scheme. In this sense, the methodological decisions regarding feature selection addressed in this chapter directly constrain the formulation of the auditory planner employed in subsequent chapters.

Within this methodological scope, two sets of configuration choices are central to adapting the Parra et al. [1] framework to the experimental data. The first concerns the definition of the acoustic output space, which involves the selection and combination of auditory features that are both physiologically meaningful and robustly extractable from the available microphone signals. Complementing this, the second choice addresses the specification of the accessible *mobility space*, involving the determination of the specific subsets of ILMs activations and P_S to be utilized within the TBCM.

The following sections systematically examine how these choices affect the accuracy, numerical stability, and interpretability of the regression–based forward mapping. This empirical evaluation establishes the forward model configuration that serves as the necessary parametric foundation for the biomechanical trajec-

tory inference and closed–loop analyses developed in the remainder of this work.

3.1.1 Definition of the acoustic output space

The definition of the acoustic output space constitutes a critical design step for ensuring a meaningful and physiologically interpretable mapping between laryngeal motor commands ($x(t)$) and acoustic outputs ($y(t)$). In the present framework, candidate acoustic variables must satisfy two fundamental constraints: they must be reliably extractable from microphone recordings, and they must be compatible with the acoustic output representations generated by TBCM simulations used to train the forward mappings. While these variables may later serve as auditory targets within a feedback control scheme, in this chapter they are treated exclusively as outputs of the biomechanical forward model.

TBCM simulations provide a controlled environment in which a wide range of phonatory features can be computed directly from known motor configurations [42]. These include glottal airflow–related measures such as open quotient, normalized open quotient, speed quotient, AC flow, and maximum flow declination rate, as well as spectral tilt measures, such as the amplitude difference between the first and second harmonics of the voice spectrum (H_1–H_2) and acoustic features such as f_o, SPL, and cepstral peak prominence (CPP).

In simulation, these features are well defined and directly accessible: f_o, CPP, H_1–H_2, and all glottal airflow–related measures are computed from the simulated

glottal volume velocity signal, whereas SPL is derived from the radiated pressure waveform. This enables systematic exploration of how underlying laryngeal motor variables shape phonatory output under controlled biomechanical conditions.

However, when working with real speech recordings captured via microphones, the availability of auditory features is substantially more constrained. Although glottal flow–based measures can, in principle, be estimated via inverse filtering using techniques such as Iterative Adaptive Inverse Filtering [144], Quasi–Closed Phase analysis [145], or Impedance–Based Inverse Filtering [138], their reliable extraction, particularly in continuous speech, requires additional assumptions regarding vocal tract filtering, radiation effects, and dynamic source–filter interactions. Extending inverse filtering approaches from sustained vowels to running speech therefore introduces substantial methodological challenges and would constitute a distinct line of investigation beyond the scope of the present work.

Accordingly, this study focuses on four acoustic features that can be robustly extracted from microphone signals across tasks and recording conditions: f_o, SPL, CPP, and corrected H_1–H_2. These features are widely used in clinical and experimental voice research [146], exhibit sensitivity to laryngeal configuration and aerodynamic–myoelastic control variables, and remain applicable in both sustained phonation and connected speech contexts.

Crucially, the suitability of these auditory targets is not assumed a priori. Instead, their informativeness and redundancy with respect to laryngeal motor vari-

ables are systematically evaluated in the following sections through correlation analyses and forward–mapping performance metrics. This empirical evaluation provides a principled basis for constructing low–dimensional, physiologically interpretable auditory target spaces that support stable and robust biomechanical inference.

3.1.2 Regression–based forward mapping approximation of the biomechanical plant

To enable the estimation of acoustic output from specified laryngeal motor commands, regression–based forward mappings were implemented using synthetic data generated with the TBCM. These forward architectures approximate the nonlinear transformation from laryngeal motor control variables to acoustic features, providing a computationally efficient surrogate of the biomechanical plant suitable for large–scale evaluation and subsequent control–oriented analyses.

All simulations were conducted using the TBCM configured to represent male laryngeal anatomy during sustained /ae/ vowel production. Phonation was driven by coordinated activation of the five ILMs, a_{LCA}, a_{IA}, a_{PCA}, a_{CT}, and a_{TA}, together with P_S. This configuration defines a physiologically grounded phonatory reference space in which variations in acoustic output can be directly interpreted in terms of underlying biomechanical control mechanisms.

Two simulation datasets were generated to evaluate alternative assumptions regarding laryngeal posture and motor space accessibility. In the first dataset, a common adductory configuration was enforced by fixing $a_{LCA} = a_{IA} = 0.6$ and $a_{PCA} = 0$, following the recommendations of Alzamendi et al. [42]. Under this configuration, P_S was varied from 400 to 2000 Pa in increments of 50 Pa, while a_{CT} and a_{TA} were systematically varied from 0 to 1 in steps of 0.025. This dataset isolates the effects of longitudinal tension and aerodynamic drive on acoustic output under a stable and modal adductory posture.

In the second dataset, adductory constraints were partially relaxed to allow broader exploration of laryngeal configurations. Specifically, a_{IA} was fixed at 0.6, a_{PCA} was varied from 0 to 0.5, and a_{LCA}, a_{CT}, and a_{TA} were systematically varied from 0 to 1 in steps of 0.025, with P_S spanning the same range as in the fixed–posture dataset. This configuration enables examination of how increased variability in adductory control influences motor–acoustic relationships and the robustness of forward mappings.

For both datasets, simulations were filtered to retain only stable and physiologically meaningful phonatory conditions. Specifically, retained outputs were constrained to $0 < \text{SPL} < 100$ dB, $\text{CPP} > 0$ dB, $75 < f_o < 400$ Hz, and $H_1 - H_2 > 0$ dB. In addition, only configurations with complete glottal closure ($a_{PCA} = 0$) were retained for subsequent regression analyses. These criteria were applied uniformly across datasets to exclude non–phonatory, unstable, or

acoustically implausible configurations. The resulting number of simulations and parameter ranges retained after filtering are summarized in Table 3.1.

The decision not to systematically vary a_{IA} and a_{PCA} in the regression stage reflects their functional role within the TBCM. In this modeling framework, these muscles primarily act as posture–defining elements, with variations often inducing abrupt transitions between phonatory regimes rather than supporting smooth, continuous modulation of acoustic output. Treating them as quasi–discrete parameters allows the analysis to focus on muscles such as CT, TA, and LCA, which provide graded control over VFs tension and adduction within the phonatory regime of interest.

The implementation of the forward mappings relies on polynomial regression with an eighth–degree polynomial. This specific design choice is grounded in the comparative analysis conducted by Parra et al. [1], who evaluated the performance of three regression architectures, Random Forests (RF), Neural Networks (NN), and Polynomial Regression (PR), for approximating the TBCM behavior.

Parra et al. [1]'s findings indicated that while RF achieved the lowest static prediction errors (f_o error $\approx$ 4 Hz), it exhibited squared patterns in the feature space and oscillatory instability during dynamic trajectory tracking due to the discrete partitioning nature of tree–based models. In contrast, PR produced smoother response surfaces consistent with the physiological underlying data. Although PR presented slightly higher errors (f_o error $\approx$ 6 Hz) compared to RF, it

Table 3.1: Motor and acoustic parameter ranges for two TBCM simulation datasets: one with fixed adductory posture ($a_{LCA} = 0.6$) and one with variable a_{LCA}. Columns labeled "Raw" indicate the full simulated motor space prior to physiological screening, whereas columns labeled "Filter" report the subset of simulations retained after applying physiologically plausible constraints on the acoustic outputs.

	$a_{LCA} = 0.6$		Variable a_{LCA}	
	Raw	Filter	Raw	Filter
$N_{simulation}$	80.751	37.014	4.818.366	40.907
a_{CT} (a.u.)	[0.025 : 1]	[0.025 : 1]	[0 : 1]	[0 : 1]
a_{TA} (a.u.)	[0.025 : 1]	[0.025 : 1]	[0 : 1]	[0 : 1]
a_{LCA} (a.u.)	[0.4 : 0.6]	0.6	[0.2 : 1]	[0.2 : 0.8]
a_{PCA} (a.u.)	0	0	[0 : 0.5]	0
P_S (P_a)	[428 : 2164]	[448 : 2164]	[226 : 2169]	[231 : 2165]
f_o (HZ)	[80 : 384]	[80 : 381]	[80 : 920]	[80 : 400]
SPL (dB)	[43 : 104]	[50 : 100]	[−1518 : 140]	[0.04 : 100]
CPP (dB)	[−15 : 45]	[0 : 45]	[−21 : 54]	[0 : 53]
$H_1 - H_2$ (dB)	[0 : 32]	[0 : 29]	[0 : 110]	[0 : 60]

maintained high goodness-of-fit ($R^2 > 0.94$) while ensuring mathematical differentiability. This smoothness property is critical for the stability of the inverse–Jacobian controller used in this framework, which requires continuous gradients to

update motor commands without introducing numerical artifacts. Consequently, an eighth–degree polynomial was adopted as the optimal trade–off between expressive power and control–relevant stability.

Prior to regression training, all motor input and acoustic output variables were normalized using min–max scaling to ensure comparable numerical ranges and to prevent dominance of variables with larger magnitudes during model fitting. Each filtered dataset was then randomly partitioned into independent training and testing subsets, using 80% of the samples for model estimation and the remaining 20% for out–of–sample evaluation. This split was performed at the level of individual simulations, ensuring that the test set contained motor–acoustic configurations not encountered during training.

Model performance was quantified using three complementary metrics: mean absolute error (MAE), which captures average prediction accuracy; root mean square error (RMSE), which penalizes larger deviations and is sensitive to outliers; and the coefficient of determination (R^2), which reflects the proportion of variance in the acoustic outputs explained by the regression model. Together, these metrics provide a comprehensive basis for comparing alternative forward–mapping configurations and identifying the models best suited for subsequent biomechanical trajectory inference.

Table 3.2: Regression–based forward mapping configurations evaluated in this study, differing in the dimensionality of the laryngeal motor input space and the predicted acoustic output space, and applied to TBCM datasets with fixed and variable adductory posture.

TBCM dataset	**Model**	**Motor inputs**	**Acoustic outputs**
$a_{LCA} = 0.6$	Model A	a_{CT} and a_{TA}	f_o
	Model B	a_{CT}, a_{TA} and P_S	f_o and SPL
	Model C	a_{CT}, a_{TA} and P_S	CPP and SPL
	Model D	a_{CT}, a_{TA} and P_S	$H_1 - H_2$ and SPL
Variable a_{LCA}	Model E	a_{CT}, a_{TA}, a_{LCA} and P_S	f_o and SPL
	Model F	a_{CT}, a_{TA}, a_{LCA} and P_S	f_o, SPL and $H_1 - H_2$
	Model G	a_{CT}, a_{TA}, a_{LCA} and P_S	f_o, SPL and CPP

3.2 Evaluation of forward mapping configurations

This section presents the quantitative evaluation of the different forward mapping configurations parameterized in Section 3.1. The results are intended to assess how distinct definitions of the acoustic *task space* and the motor command dimensionality influence the identifiability, redundancy, and predictive accuracy

of the regression–based approximations.

It is important to emphasize that the analyses reported here are driven by operational criteria rather than physiological exploration. The primary objective is to guide the selection of a specific forward model configuration that offers the optimal trade–off between estimation accuracy and numerical conditioning, ensuring its suitability for the inverse control tasks performed in subsequent chapters.

The section is organized into two parts: first, a correlation analysis of motor–auditory relationships to characterize the underlying structure and potential collinearities of the simulated datasets; and second, a comparative performance evaluation of the candidate forward mappings to identify the most robust configuration.

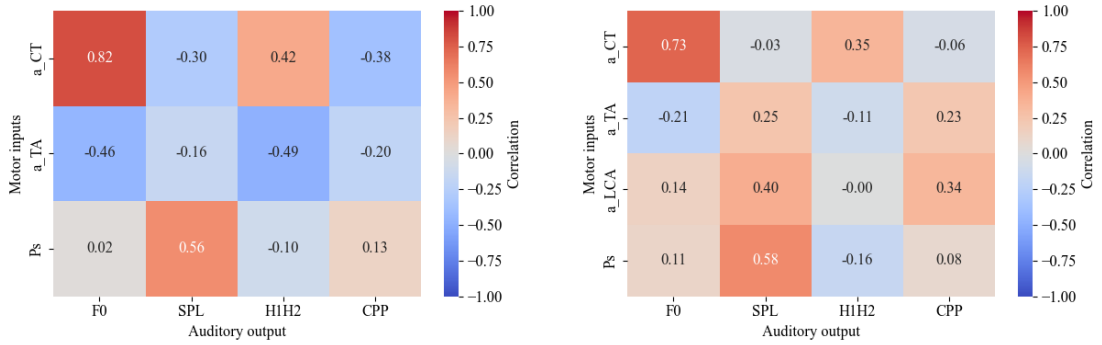
3.2.1 Motor–auditory correlation structure

To evaluate the sign and strength of the relationships between the laryngeal control space and the resulting acoustic output, a Pearson correlation analysis was performed on the synthetic datasets generated using the TBCM. Correlation heatmaps were generated for two distinct simulation scenarios: one with a fixed LCA activation ($a_{LCA} = 0.6$) and another with variable a_{LCA}, as illustrated in Figure 3.1.

In the fixed a_{LCA} condition (Figure 3.1a), the system exhibited strong and clearly defined physiological dependencies. f_o showed a robust positive correla-

tion with a_{CT} and a moderate negative correlation with a_{TA}, in line with known antagonistic roles of these muscles in regulating VF tension and pitch. Similarly, a high positive correlation was observed between P_S and SPL, confirming that aerodynamic drive remains the primary regulator of vocal loudness in the model. The spectral tilt measure, $H_1 - H_2$, exhibited a correlation pattern similar to that of f_o, suggesting a coupling between pitch raising and changes in the glottal flow harmonic structure. For the CPP, the highest correlation was observed with a_{CT}, although its magnitude was significantly lower compared to other acoustic features, likely due to the high temporal stability of the sustained phonation simulated. Because CPP is particularly sensitive to cycle–to–cycle irregularities and aperiodic components in the acoustic signal, the absence of temporal perturbations in the simulated phonation limits its variability and, consequently, its correlation strength.

In contrast, when a_{LCA} was allowed to vary (Figure 3.1b), the correlation structure became more complex and diffuse. The individual correlations between f_o and the tensor muscles (a_{CT}, a_{TA}) were substantially reduced, indicating that changes in glottal adduction, driven by a_{LCA}, interact nonlinearly with the tensioning mechanism. In this scenario, CPP and SPL showed increased correlations with a_{TA} and a_{LCA}, suggesting that these acoustic parameters are sensitive indicators of the adductory patterns and glottal configuration. While P_S remained a significant predictor of SPL, its influence was partially redistributed across the



(a) Fixed $a_{LCA} = 0.6$

(b) Variable a_{LCA}

Figure 3.1: Correlation matrix between motor control inputs ($a_{CT}, a_{TA}, a_{LCA}, P_S$) and acoustic features (f_o, SPL, $H_1 - H_2$, CPP) in the TBCM datasets.

increased degrees of freedom of the laryngeal plant.

These results have significant implications for the proposed modeling framework. In the fixed a_{LCA} scenario, a subset of acoustic features (f_o, SPL, and $H_1 - H_2$) appears sufficient to uniquely characterize the motor state. However, under the variable a_{LCA} condition, the inclusion of CPP becomes essential to capture the variance introduced by adductory adjustments. These findings emphasize that a multi–feature auditory *task space* is essential for supporting sufficiently informative mappings between acoustic features and biomechanical states in subsequent control and inference stages. By integrating spectral and cepstral features alongside primary acoustic targets, the framework has the potential to facilitate the disentanglement variables associated with central and autonomic dysregulation, such as laryngeal stiffness and compensatory tension. However,

while a richer *task space* improves mapping accuracy, the overall efficiency and stability of the tracking process depend on the selection of a forward model that balances complexity and performance.

3.2.2 Forward mapping performance across model configurations

Based on the outcomes of the correlation analysis, Models C and F were excluded from the forward mapping evaluation due to their limited additional informativeness and reduced robustness observed in preliminary analyses. The remaining models were trained and tested using two datasets corresponding to fixed and variable a_{LCA} configurations. For the fixed–a_{LCA} condition, the dataset comprised 29,611 training samples (80%) and 7,403 test samples (20%). The variable–a_{LCA} dataset included 97,406 training samples (80%) and 24,352 test samples (20%). Model performance was quantified using MAE, MSE, RMSE, and the coefficient R^2. While all metrics are reported in Table 3.3, the analysis emphasizes MAE and R^2, as these metrics offer a more interpretable characterization of prediction accuracy and explained variance across model configurations.

Under the fixed a_{LCA} condition, Model B achieved the highest predictive accuracy among the evaluated configurations. f_o was estimated with an R^2 of 0.986 and an MAE of 1.039 Hz, while SPL prediction yielded an R^2 of 0.923 with an

MAE of 1.662 dB. Model A, which relied on a reduced set of motor inputs and predicted only f_o, exhibited lower accuracy, with an R^2 of 0.958 and larger error values. Model D showed comparable SPL performance to Model B, while predictions of $H_1 - H_2$ presented higher errors and reduced explained variance.

For the variable a_{LCA} dataset, overall predictive accuracy decreased across all models. Notably, the increase in absolute error (e.g., MAE values above 11 Hz for f_o) should not be interpreted as a limitation of the regression framework itself, but rather as a direct consequence of the increased dimensionality and degrees of freedom introduced by allowing adductory control to vary. Models E and G yielded comparable performance for f_o estimation, with R^2 values of 0.929 and MAE values above 11 Hz. SPL prediction accuracy was also reduced, with R^2 values ranging between 0.851 and 0.865. When additional acoustic outputs were included, as in Model G, the prediction of CPP resulted in an R^2 of 0.739, indicating lower agreement between predicted and simulated values relative to f_o and SPL.

These performance metrics are consistent with the benchmarks reported by Parra et al. [1] for the TBCM forward mapping. In their evaluation of polynomial regressors, they observed MAE of approximately 6 Hz for f_o and 2 dB for SPL. The results obtained here for the fixed–a_{LCA} condition ($MAE_{f_o} \approx 1.04$ Hz; $MAE_{SPL} \approx 1.66$ dB) surpass these baselines, likely due to the reduced dimensionality of the constrained motor space. Conversely, the increased errors observed in

the variable–a_{LCA} condition reflect the inherent challenge of mapping a higher–dimensional motor space (5 active ILMs) to the same acoustic outputs, confirming the trade–off between physiological completeness and estimation accuracy highlighted in the reference framework.

Across both datasets, models trained under the fixed–a_{LCA} condition consistently outperformed those trained with variable a_{LCA} in terms of absolute error and explained variance. The numerical results reported in Table 3.3 characterize how forward mapping accuracy varies across model configurations and acoustic outputs, and provide the quantitative basis for the comparative analysis developed in the subsequent discussion section.

Taken together, the correlation patterns and forward mapping performance results delineate a constrained set of viable model configurations for biomechanical inference from acoustic features. While multiple architectures achieve acceptable predictive accuracy under specific assumptions, the results highlight clear differences in robustness, interpretability, and sensitivity to biomechanical variability. These findings motivate a focused discussion of the design trade–offs underlying the selected model configuration, which is developed in the following section.

Table 3.3: Performance metrics of regression–based forward mapping models evaluated under fixed ($a_{LCA} = 0.6$) and variable a_{LCA} conditions. Results are reported on independent test sets (N_{test} = 7,403 and 24,352, respectively) using MAE, MSE, RMSE, and coefficient of determination (R^2).

TBCM dataset	**Model**	**Output**	MAE	MSE	RMSE	R^2
$a_{LCA} = 0.6$	Model A	f_o	11.680	216.162	14.702	0.958
	Model B	f_o	1.039	2.187	1.479	0.986
		SPL	1.662	6.476	2.545	0.923
	Model D	$H_1 - H_2$	6.102	71.534	8.458	0.878
		SPL	1.660	6.196	2.489	0.926
Variable a_{LCA}	Model E	f_o	11.465	401.070	20.027	0.929
		SPL	3.726	41.474	6.440	0.851
	Model G	f_o	11.326	396.815	19.920	0.929
		SPL	3.610	36.150	6.013	0.865
		CPP	3.737	23.915	4.890	0.739

3.3 Discussion

The results presented in this chapter support the selection of Model B trained on the fixed–a_{LCA} dataset as the reference configuration for subsequent analyses.

This decision is grounded in both technical performance criteria and physiological considerations related to laryngeal control during stable phonation [1].

From a modeling perspective, Model B achieved the highest predictive accuracy across the primary acoustic targets, with near–ceiling performance for f_o and robust SPL estimation. Importantly, this accuracy was obtained using a relatively compact motor input space, involving a_{CT}, a_{TA}, and P_S. This configuration yielded stable regression behavior and smooth input–output mappings, properties that are essential for integration within inverse Jacobian–based feedback control schemes, where numerical stability and continuity of the forward model directly affect convergence and tracking performance [43, 1].

Physiologically, fixing a_{LCA} at a moderate activation level establishes a consistent adductory posture that approximates typical modal phonation conditions. Under this constraint, variations in pitch and loudness are predominantly governed by longitudinal tension (mediated by CT and TA activation) and aerodynamic drive, respectively, which aligns with well–established principles of laryngeal biomechanics [129]. The strong and interpretable correlations observed in this configuration suggest that the resulting acoustic space provides sufficient observability of the underlying motor state without introducing excessive redundancy.

In contrast, allowing a_{LCA} to vary introduced additional degrees of freedom associated with glottal adduction and medial compression. While this expanded motor space increased the generality of the model, it also led to reduced for-

ward mapping accuracy and more diffuse motor–auditory relationships, reflecting the increasingly ill–posed nature of the mapping, in which multiple adductory–tensional configurations give rise to acoustically similar outputs. This outcome arises from the nonlinear interaction between adduction and tension mechanisms, whereby changes in VFs contact patterns modulate spectral and cepstral features in a manner that is not uniquely attributable to a single motor variable [42, 147].

The inclusion of additional acoustic outputs, such as CPP, partially mitigated this limitation by providing sensitivity to changes in glottal closure and phonatory stability [148, 149]. However, because the simulated sustained phonation relies on deterministic lumped–mass dynamics that lack the natural temporal perturbations and cycle–to–cycle irregularities inherent to human tissue [52, 42], CPP does not contribute sufficient independent information to fully compensate for the increased biomechanical complexity introduced by variable adduction. As a result, the fixed–a_{LCA} configuration offers a more favorable balance between physiological relevance, computational efficiency, and inferential reliability.

Importantly, the selection of Model B does not imply that adductory control is irrelevant for voice production. Rather, it reflects a deliberate modeling choice aligned with the goals of this thesis: to establish a robust and interpretable framework for inferring laryngeal motor control from acoustic features in continuous speech and sustained vowels. By constraining the adductory posture, the model isolates motor dimensions that are known to be sensitive to both PD and

cognitive demand at the level of their acoustic manifestations, providing a stable reference space for subsequent investigations.

3.4 Chapter Conclusions

This chapter detailed the methodological configuration and empirical characterization of the regression–based forward mapping used to approximate the laryngeal plant within the TBCM simulation environment. Building upon the computational architecture established by Parra et al. [1], this study focused on identifying the specific input–output parametrization that maximizes estimation accuracy and numerical stability for the experimental tasks of this thesis.

The correlation analyses demonstrated that constraining the laryngeal adduction parameter (a_{LCA}) yields a significantly more identifiable and structured relationship between motor inputs and acoustic outcomes. In contrast, allowing a_{LCA} to vary introduces a degree of dimensionality and redundancy that challenges the robustness of the regression mapping given the available acoustic feature set. These observations provided a principled, data–driven basis for restricting the motor control space to ensure reliable inference.

Performance evaluations confirmed that model configurations trained on the fixed–a_{LCA} dataset consistently achieved higher predictive accuracy. Among the candidate architectures, Model B was identified as the optimal configuration, of-

fering a favorable trade–off between physiological interpretability and low estimation error for f_o and SPL. Consequently, this specific configuration was selected as the reference forward mapping to be embedded in the trajectory estimation framework.

In summary, these results establish the operational boundaries for the biomechanical analyses that follow. By defining a robust and numerically stable forward relationship ($x(t) \rightarrow y(t)$), this chapter implicitly delimits the acoustic *task space* available for the planner in the subsequent closed–loop control formulations. This validated configuration provides the necessary parametric foundation to extend the Parra et al. [1] framework toward the specific clinical and functional applications addressed in the remainder of this work.

Chapter 4

Estimating laryngeal biomechanical trajectories for Parkinson's disease classification

This chapter investigates whether biomechanically informed descriptors of laryngeal motor control provide added value for the characterization and classification of PD from speech. Building on the methodological framework developed in previous chapters (Chapter 3), the focus here is on applying a model–based trajectory estimation framework to sustained phonation tasks in order to derive physiologically interpretable control variables underlying phonation.

Specifically, continuous trajectories of P_S and ILMs activations (a_{CT} and a_{TA}) are estimated from microphone recordings by integrating the regression–based forward mapping selected in the previous chapter (Model B) within a laryngeal motor control framework. In this formulation, the forward model approximates the biomechanical plant that maps laryngeal motor commands onto acoustic outputs (f_o and SPL), while motor trajectory estimation is achieved through a feedback control loop based on the inverse Jacobian.

From the estimated trajectories, statistical descriptors are extracted to characterize temporal variability, coordination, and stability of laryngeal control. These biomechanically informed features are analyzed alongside conventional acoustic features computed directly from the speech signal.

The chapter evaluates and compares the discriminative power of acoustic and biomechanical feature sets through statistical analysis and supervised classification experiments. The goal is to determine whether features grounded in a biomechanical model of phonation provide complementary information for PD detection beyond that captured by standard acoustic representations.

The analysis is conducted using two independent Spanish speech databases, allowing assessment of the robustness and generalizability of the proposed features across datasets. In alignment with the first specific objective of this thesis, this chapter operationalizes the model–based trajectory estimation framework to extract laryngeal control descriptors that may reflect altered motor regulation in PD. In line with Hypothesis 1, it is expected that these biomechanically informed features capture aspects of motor coordination and stability that are not fully reflected by conventional acoustic measures, thereby improving classification performance.

4.1 Voice–based characterization of Parkinson's disease

PD is a progressive neurodegenerative disorder characterized by impairments in motor planning, execution, and sensorimotor integration, which manifest prominently in speech production. Alterations in phonation, articulation, and prosody are among the earliest and most prevalent symptoms, motivating extensive research into speech–based markaers as non–invasive indicators of the disease [10, 150]. Voice analysis is particularly attractive in this context due to its low acquisition cost, minimal invasiveness, and suitability for longitudinal monitoring.

4.1.1 Acoustic feature–based approaches

The majority of speech–based studies in PD have relied on acoustic features extracted directly from the speech signal. These typically include measures related to f_o, intensity, temporal organization, spectral envelope, and perturbation metrics such as jitter, shimmer, and HNR [79, 93]. Such features have been shown to capture salient perceptual correlates of PD speech, including hypophonia, monopitch, increased temporal variability, and vocal instability [151].

Building on these representations, a wide range of ML techniques have been applied to PD detection and severity estimation. Traditional classifiers, includ-

ing support vector machines (SVM), decision trees, random forests [152], and gradient boosting models [153, 154], have demonstrated good discriminative performance when trained on carefully selected acoustic feature sets . More recently, deep learning approaches based on convolutional [155, 156] and recurrent neural networks [157] have achieved high classification accuracy by learning hierarchical representations from time–frequency transforms such as spectrograms or short–time Fourier transforms .

While these approaches consistently confirm that speech carries diagnostically relevant information, their reliance on surface–level acoustic descriptors limits physiological interpretability. Acoustic features provide only indirect access to the neuromotor mechanisms underlying speech production and offer limited insight into how central motor impairments associated with PD translate into observable vocal symptoms.

4.1.2 Model–based and motor control–oriented approaches

To overcome these limitations, a smaller body of work has adopted model–based approaches grounded in theories of speech motor control. Under the hypothesis that Parkinsonian speech deficits arise from disruptions in motor planning, coordination, and feedback regulation, these studies aim to infer latent control

variables associated with speech production rather than relying exclusively on acoustic outcomes.

Neurocomputational frameworks, such as the DIVA model, have been widely employed to estimate articulatory and control parameters from speech data, serving as valuable features for disease characterization and severity prediction [71, 158]. Building on this foundation, recent extensions have incorporated physiologically motivated voice production models to enable the direct estimation of ILMs activations. Specifically, Ciccarelli et al. [159] and Quatieri et al. [160] integrated the extended body–cover model [138] within a SimpleDIVA framework [161]. Their work demonstrated that the inferred activations of the CT (a_{CT}) and TA (a_{TA}) muscles provide discriminative features for the classification of Parkinson's Disease.

These studies provide evidence that model–derived motor variables may capture disease–related alterations that are not fully reflected in conventional acoustic descriptors. However, existing implementations are often restricted to simplified speech tasks (e.g., sustained phonation), static or segment–level representations, or relatively coarse temporal resolutions. Moreover, the application of biomechanically informed models to connected speech remains limited, particularly in the context of estimating continuous motor trajectories that reflect dynamic laryngeal control.

Within this research landscape, the present chapter extends prior model–based approaches by introducing a laryngeal motor control framework that enables the estimation of continuous biomechanical trajectories during sustained phonation tasks. While previous studies have demonstrated the feasibility of inferring ILMs activations, most notably a_{CT} and a_{TA} from acoustic observations, these efforts have largely relied on simplified control formulations, restricted acoustic targets, and implicit assumptions regarding aerodynamic drive, often focusing exclusively on f_o [159].

In contrast, the framework adopted here explicitly incorporates P_S as an estimated control variable and leverages a regression–based forward mapping that jointly relates laryngeal motor commands to both f_o and SPL. By embedding this forward model within a feedback control architecture based on the inverse Jacobian, the proposed approach supports the estimation of dynamically evolving laryngeal control trajectories from microphone recordings of sustained vowels, providing a physiologically grounded description of phonatory control under stable task conditions.

By moving beyond static or purely acoustic representations and by explicitly modeling the interaction between ILMs and aerodynamic inputs, this chapter positions biomechanically informed control variables as a complementary feature space for PD characterization. This controlled phonatory setting establishes a necessary validation step for the proposed framework and serves as a foundation

for its subsequent extension to connected and continuous speech in the following chapter.

4.2 Biomechanical and acoustic analysis pipeline for Parkinson's disease

Figure 4.1 provides an overview of the analysis pipeline implemented in this chapter, which integrates acoustic analysis and biomechanical modeling to characterize phonatory behavior in individuals with PD and HC. The workflow begins with the acquisition and preprocessing of sustained vowels obtained from two Spanish–language speech databases (PC–GITA and Neurovoz), ensuring clean and reliable voiced segments suitable for analysis. From these curated segments, acoustic features were extracted to capture phonatory stability, spectral structure, cepstral organization, and articulatory configuration, with the specific feature sets tailored to the properties of sustained phonation tasks.

In parallel, the trajectories of f_o and intensity obtained from the preprocessed signals were used to estimate laryngeal biomechanical features via a physiologically based model, yielding continuous trajectories of muscle activations and P_S consistent with the observed acoustics.

Both acoustic and biomechanical features were evaluated statistically between group. This analysis served to identify which features exhibited the strongest

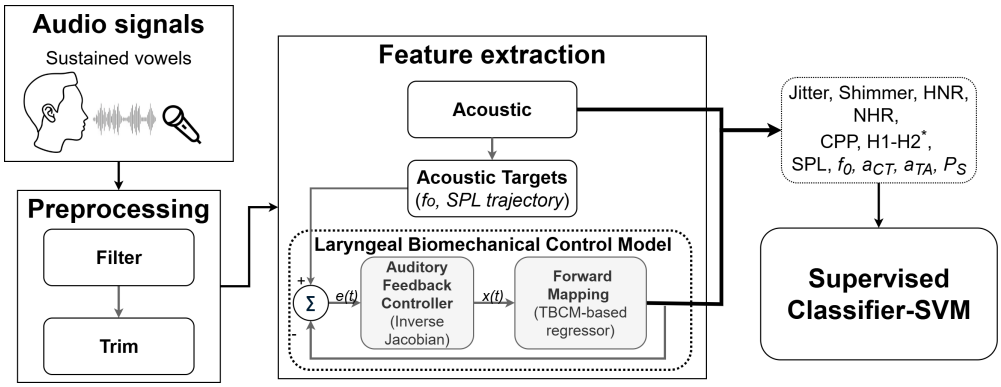


Figure 4.1: Overview of the methodological pipeline, including data acquisition, preprocessing, acoustic feature extraction, biomechanical estimation, statistical analysis, and ML classification.

group separations and were therefore most suitable for subsequent classification. Finally, the selected features were assessed within a supervised learning framework to determine their discriminative capacity for distinguishing PD from HC. Overall, this integrated pipeline provides a multilevel characterization of vocal function, combining signal processing, model–based estimation, and statistical learning.

4.2.1 Speech corpora for Parkinson's disease

Two independent speech corpora were employed in this study to investigate PD in Spanish–speaking populations: PC–GITA, comprising Colombian Spanish speakers [46], and NeuroVoz, comprising Castilian Spanish speakers [47]. Together, these datasets provide complementary linguistic, demographic, and clinical coverage, enabling a robust cross–corpus analysis of speech and voice characteris-

tics associated with PD within the proposed acoustic–biomechanical framework. A summary of the main demographic, clinical, and methodological characteristics of both corpora is provided in Table 4.1.

Although both corpora include a wide range of speech elicitation tasks, the present chapter focuses exclusively on sustained vowel phonations. This restriction ensures controlled phonatory conditions that are particularly suitable for the estimation and analysis of acoustic stability and laryngeal biomechanical trajectories, as developed in the preceding methodological chapters.

The PC–GITA corpus [46] contains speech recordings from 100 native speakers of Colombian Spanish, evenly distributed into 50 individuals with PD and 50 HC, with each group comprising 25 men and 25 women. Male participants with PD had a mean age of 62.2 ± 11.2 years, while female participants had a mean age of 60.1 ± 7.8 years. Among the HC group, men had a mean age of 61.2 ± 11.3 years, and women 60.7 ± 7.7 years, resulting in a dataset well balanced in terms of age and sex. Individuals with PD were diagnosed by expert neurologists and clinically assessed using the UPDRS and the Hoehn and Yahr (H&Y) scale; disease duration was also documented. All PD participants were recorded in the ON–medication state, defined as within three hours of morning dopaminergic intake.

Recordings were conducted under controlled acoustic conditions inside a sound–proof booth at Clínica Noel in Medellín, Colombia, using a Shure SM63L omnidirectional dynamic microphone and a professional M–Audio Fast Track C400

audio interface. All signals were acquired at a sampling rate of 44.1 kHz with 16–bit resolution. The recording protocol, approved by the corresponding institutional ethics committee and conducted in accordance with the Declaration of Helsinki, required written informed consent from all participants. While the full protocol included phonatory, articulatory, and prosodic tasks–such as sustained vowels, pitch–modulated vowels, diadochokinetic (DDK) sequences, word and sentence repetition, reading tasks, and spontaneous speech–only sustained phonation of the five Spanish vowels was considered for the analyses presented in this chapter.

The NeuroVoz corpus [47], although not numerically balanced across diagnostic groups, comprises recordings from 112 native speakers of Castilian Spanish, including 54 individuals with PD and 58 HC. The dataset exhibits an approximately balanced sex distribution (46 women, 61 men, and one participant with unreported sex). Participants in the PD group had a mean age of 71.13 ± 10.62 years, while those in the HC group had a mean age of 64.04 ± 10.26 years. PD participants were recruited from the Neurology Services of Hospital General Universitario Gregorio Marañón and Hospital Universitario de Fuenlabrada, where certified clinicians assessed disease severity using the UPDRS and H&Y scales. An otolaryngological examination was conducted to exclude organic vocal pathologies and to document symptoms such as vocal tremor, hypophonia, dysphagia, and sialorrhea. Most PD participants had a disease duration of less than ten years and were recorded in the ON–medication state, typically between two and five

hours after dopaminergic intake.

Recordings were carried out in a standardized quiet clinical room using an AKG C420 headset microphone connected to a phantom–powered preamplifier and a SoundBlaster Live! sound card, with a sampling rate of 44.1 kHz and a mean signal–to–noise ratio of 24.3 dB. The NeuroVoz protocol included sustained vowel phonation (three repetitions of each Spanish vowel), sentence repetition, a DDK task (/pa–ta–ka/), and spontaneous monologues elicited through picture description. Consistent with the scope of the present chapter, only sustained vowel phonations were retained for feature extraction and statistical analysis.

Together, PC–GITA and NeuroVoz constitute a heterogeneous and complementary collection of Spanish speech data encompassing two major dialectal varieties, standardized recording procedures, and rich clinical annotation. Their combined use enables a methodologically rigorous examination of acoustic and biomechanically inferred markers of PD across corpora. By restricting the analysis to sustained vowel phonation, this chapter emphasizes controlled phonatory behavior, providing a stable basis for the estimation and comparison of acoustic and laryngeal biomechanical descriptors.

Table 4.1: Summary of demographic, clinical, and methodological characteristics of the PC–GITA and NeuroVoz corpora. Only sustained vowel phonations were used in the present study.

	PC–GITA	**NeuroVoz**
Participants	100 (50 PD, 50 HC)	112 (54 PD, 58 HC)
Sex distribution	25 men and 25 women per group	46 women, 61 men, 1 not reported
Age distribution	PD men: 62.2 ± 11.2	PD: 71.13 ± 10.62
	PD women: 60.1 ± 7.8	HC: 64.04 ± 10.26
	HC men: 61.2 ± 11.3	
	HC women: 60.7 ± 7.7	
Language variety	Colombian Spanish	Castilian Spanish
Clinical assessment (PD)	UPDRS, H&Y; disease duration recorded	UPDRS, H&Y; neurological and ENT screening
Medication state (PD)	ON–state: within 3 h of morning medication	ON–state: 2–5 h after dopaminergic intake
Recording conditions	Sound–proof booth; Shure SM63L microphone; 44.1 kHz, 16–bit	Quiet clinical room; AKG C420 headset; 44.1 kHz
Speech material used	Sustained vowels	Sustained vowels

4.2.2 Database preprocessing

All audio recordings were preprocessed to ensure consistency across corpora and to isolate only the speech material relevant for the present analysis, which focuses exclusively on sustained vowel phonation. The preprocessing pipeline was designed to preserve phonatory stability and spectral integrity, both of which are critical for reliable acoustic analysis and biomechanical trajectory estimation.

As a first step, all signals were band–pass filtered using a 4th–order Infinite Impulse Response Butterworth filter with cutoff frequencies of 50 Hz and 4 kHz [162]. This configuration removes very low–frequency components, such as electrical hum and handling noise, and attenuates high–frequency content outside the informative speech band, while preserving the frequency region where most phonatory and articulatory information is concentrated. In particular, this range retains the f_o, its harmonic structure, and the lower formant region, which are essential for both conventional acoustic feature extraction and model–based estimation of laryngeal control variables. The Butterworth design was selected due to its maximally flat amplitude response in the passband, minimizing spectral distortion that could bias subsequent analyses.

Following filtering, the signals were trimmed to remove leading and trailing silent segments, retaining only the stable phonatory portion of each sustained vowel. This segmentation was performed through combined auditory and visual inspection of the waveform and spectrogram to ensure that only steady–state voiced material was preserved and that no relevant phonatory content was inadvertently removed. Although more automated segmentation strategies exist, manual inspection was preferred in this context to guarantee high–quality phonatory segments, which are particularly important for the estimation of continuous biomechanical trajectories under quasi–stationary vocal conditions.

The resulting preprocessed segments therefore consist of stable, sustained

phonation suitable for both acoustic feature extraction and biomechanical modeling, ensuring that variability in the estimated control trajectories reflects physiological differences rather than segmentation artifacts or transient vocal behaviors.

4.2.3 Acoustic feature extraction

Acoustic features were extracted from the preprocessed sustained vowels using `Parselmouth` (Python interface to Praat) [163] and custom analysis routines. The feature set was selected to characterize phonatory stability, spectral structure, cepstral organization, and articulatory–acoustic configuration, while remaining compatible with the requirements of the subsequent biomechanical inference framework.

Analysis window selection. Given the variability in vowel duration across recordings, analysis windows were adapted to ensure stable and comparable measurements. For signals shorter than 1 s, a central segment of 500 ms was analyzed. For signals between 1 and 4 s, a 1 s segment centered in the recording was used. For signals longer than 4 s, a central segment of 3 s was selected. This strategy ensured that all measurements were computed from the most stable portion of the phonation while maintaining consistency across corpora and experimental conditions.

Phonatory and spectral measures. f_o and perturbation measures were obtained from the glottal pulse train estimated using Praat's autocorrelation method [164], with a pitch floor of 75 Hz and a ceiling of 500 Hz. The analysis time step was set to 50 ms. Jitter and shimmer were computed using the set of perturbation metrics implemented in Praat, including local, RAP, and PPQ5 for jitter, and local, APQ3, and DDA for shimmer. Noise–related measures were derived from harmonicity analysis, yielding estimates of the HNR and noise–to–harmonic ratio (NHR).

Long–term intensity was computed using Praat's *Intensity* function with a Gaussian window of 50 ms. Since recordings were acquired without absolute calibration and using different microphone types across corpora, raw SPL values were not directly comparable. To mitigate this limitation, SPL trajectories were standardized to a reference mouth–to–microphone distance of 15 cm, consistent with the distance assumed in the simulations used to train the biomechanical model.

The correction followed the inverse distance law [165], which predicts an SPL decrease of approximately 6 dB for each doubling of distance under free–field conditions. Assuming a reference level of 70 dB ($L_{p@d_1}$) at 1 m (typical conversational speech, d_1), the corresponding SPL at 15 cm (d_2) is 86.48 dB ($L_{p@d_2}$), as shown in Eq. 4.1. A correction offset was computed as the difference between this theoretical reference and the mean SPL measured for each speaker, and this offset

was then applied uniformly to all SPL trajectories. This procedure was applied consistently across all databases.

$$
\begin{aligned}
L_{p@d_2} &= L_{p@d_1} - 20\log_{10}\left(\frac{d_2}{d_1}\right) \\
L_{p@d_{0.15m}} &= L_{p@d_{1m}} - 20\log_{10}\left(\frac{0.15m}{1m}\right) \\
L_{p@d_{0.15m}} &= 70\text{dB} - (-16.478) \\
L_{p@d_{0.15m}} &= 86.478\text{dB}
\end{aligned}
\tag{4.1}
$$

Formant and cepstral measures. Formant frequencies (F_1–F_4) were estimated via LPC analysis with pre–emphasis at 50 Hz and a 25 ms analysis window. Sex–specific formant ceilings were set to 5000 Hz for female speakers and 4500 Hz for male speakers, with LPC order determined automatically by Praat based on the analysis bandwidth. Based on the estimated formants, four vocal tract length (VTL) estimates were computed, one for each formant, following the methodology described in Pah et al. [162].

Cepstral organization was quantified using CPP, computed from a power cepstrogram with a quefrency step of 0.002 s. CPP was extracted using the CPPS algorithm with a time–averaging window of 0.001 s, a pitch range of 75–500 Hz, and a peak search window of 0.05 s. Spectral tilt was estimated using a corrected H_1–H_2 measure obtained via a custom implementation of the Iseli correction [166], which compensates for the influence of formant frequencies and bandwidths on

harmonic amplitudes.

Feature summarization and trajectory preparation. For all acoustic features, time–varying measurements computed within the selected analysis window were summarized to yield a single feature vector per recording. Unless otherwise stated, features were represented by their mean value over time. For formant–related measures (F_1–F_4), both the mean and standard deviation were retained to capture central tendency and articulatory variability. For f_o and SPL, the mean, median, and standard deviation were computed to account for potential asymmetries and variability in phonatory control.

In addition to these summary statistics, the time–varying trajectories of f_o and SPL were retained for subsequent biomechanical modeling. Temporal alignment across experimental vowels was achieved by time–normalizing each segment relative to its duration and resampling the trajectories to 100 frames per segment. This normalization ensured a consistent temporal representation across participants and conditions and is further exploited in the biomechanical trajectory estimation described in the following subsection.

4.2.4 Laryngeal biomechanical features estimation

The laryngeal biomechanical control model applied in this chapter builds directly on the framework established and systematically evaluated in Chapter chap-

ter 3. As demonstrated in that chapter, the model derived from the formulation of Parra et al. [1] provides a principled link between acoustic trajectories and underlying laryngeal motor control variables, and can be reliably used as an inference tool when appropriately configured.

Based on the comparative evaluation of alternative model configurations, the present analyses adopt the forward mapping corresponding to Model B, which exhibited the most robust and accurate performance under constrained adductory conditions. In this configuration, the biomechanical plant is approximated by an eighth–degree polynomial regression trained on simulations of the TBCM. The forward model implements a nonlinear mapping from a reduced motor command space (P_S, a_{CT}, and a_{TA}) to the primary acoustic outputs (f_o and SPL). This polynomial approximation provides an accurate yet interpretable representation of the motor–acoustic relationships, while maintaining low computational cost and numerical stability, consistent with the original framework proposed by Parra et al. [1].

The TBCM simulations used to construct the forward mapping were configured to represent a reference male laryngeal anatomy during sustained /a/ vowel production. As in Chapter chapter 3, a fixed adductory posture was enforced by setting $a_{LCA} = a_{IA} = 0.6$ and $a_{PCA} = 0$, while the remaining parameters were systematically varied. The P_S ranged from 400 to 2000 Pa in increments of 50 Pa, and muscle activations a_{CT} and a_{TA} varied from 0 to 1 in steps of 0.025. This

configuration defines a physiologically grounded phonatory reference space that is used consistently across all speakers, enabling the interpretation of observed acoustic trajectories in terms of relative laryngeal motor control dynamics.

For each simulated parameter configuration, acoustic and aerodynamic outputs were extracted from the final 50 ms of the simulation to avoid initialization transients [38]. f_o was computed from the glottal volume velocity signal, while SPL was derived from the radiated pressure assuming a mouth–to–microphone distance of 15 cm.

In the inverse control stage, deviations between observed acoustic trajectories $y(t) = [f_o(t), \mathrm{SPL}(t)]$ and their corresponding targets ($y_T(t)$) were mapped back into the motor command space using an inverse Jacobian approach [43]. The Jacobian matrix, which characterizes the local sensitivity of the acoustic outputs to changes in the motor variables $[a_{CT}, a_{TA}, P_S]$, was inverted using a damped least–squares pseudoinverse to enhance numerical stability [132]. The controller included two adjustable parameters: an update gain α and a damping factor γ, which were set to $\alpha = 0.3$ and $\gamma = 0.1$ for all analyses. Numerical derivatives were computed using a step size of 0.01, and the sampling frequency was fixed at $F_s = 1000$ Hz (see Eq. 2.10–2.9).

For each sustained–vowel recording, the control system tracked the corresponding acoustic trajectories of f_o and SPL and generated time–varying motor trajectories $x(t) = [a_{CT}(t), a_{TA}(t), P_S(t)]$. These trajectories are interpreted as biome-

chanical correlates of the observed phonatory patterns. In subsequent analyses, descriptive statistics were extracted from these trajectories to obtain a compact set of biomechanically informed features suitable for group–level statistical comparisons and supervised classification.

4.2.5 Statistical analysis

Statistical analyses were conducted to characterize the distribution of the extracted acoustic features and to evaluate group differences between individuals with PD and HC. All analyses were based exclusively on features derived from sustained vowel phonations, in accordance with the methodological scope of the present chapter.

For each recording, acoustic features were summarized as described in the previous section to yield a single feature vector per speaker. Perturbation–based measures (jitter and shimmer), harmonicity indices (HNR and NHR), cepstral measures (CPP), the corrected spectral tilt measure ($H_1 - H_2^*$), and the corresponding vocal tract length (VTL) were represented by a single representative value per vowel segment. Formant frequencies (F_1–F_4) were described using both their mean and standard deviation across the analysis window, capturing central tendency and articulatory variability. For f_o and SPL, mean, median, and standard deviation were computed to account for potential asymmetries and variability in phonatory control.

Normality of each feature distribution was assessed using the Anderson–Darling test [167]. Based on these results, group comparisons between PD and HC were performed using either an independent–samples t-test for features satisfying normality assumptions or the non–parametric Mann–Whitney U test otherwise. Statistical significance was defined as $p < 0.05$ for all hypothesis tests.

In addition to hypothesis testing, effect sizes were computed to quantify the magnitude of group differences independently of sample size. Cohen's d was used for parametric comparisons [168], while Cliff's δ was employed for non–parametric contrasts [169]. Effect sizes were interpreted according to established thresholds to facilitate the assessment of distributional separation between groups.

4.2.6 Classification model and performance metrics

To assess the discriminative capacity of the extracted features, a supervised classification analysis was performed using a Support Vector Machine (SVM) with a Gaussian Radial Basis Function (RBF) kernel [170]. The SVM constructs an optimal hyperplane to maximize the margin between classes, while the RBF kernel allows for the modeling of non–linear relationships within the feature space, a critical requirement for characterizing complex pathological speech patterns [162, 171]. For this analysis, the classifier hyperparameters were fixed at $C = 1$ and $\gamma = 0.1$.

To ensure a robust and unbiased estimate of model performance, evaluation

was conducted using a stratified group k–fold cross–validation scheme (with $k = 10$). Crucially, speaker identity was employed as the grouping variable. This constraint ensures that all samples from a specific subject are assigned exclusively to either the training or the test fold in any given iteration, thereby preventing data leakage that could occur if samples from the same speaker appeared in both sets simultaneously.

Furthermore, to strictly maintain the independence of the test sets, all data preprocessing steps were applied within each cross–validation loop. Specifically, missing values were imputed using the median, and features were standardized to zero mean and unit variance based solely on the statistics of the training partition. This pipeline ensures that the test data remains unseen during the model calibration phase.

To mitigate the risk of overfitting and the curse of dimensionality, two distinct feature selection strategies were implemented and compared prior to classification. The top ten features identified by each method were retained for the final models:

- **Relief–F Algorithm:** This similarity–based feature selection method estimates the quality of attributes by their ability to distinguish between instances that are near to each other. It is particularly effective at detecting feature dependencies and interactions that standard statistical filters might miss [172, 173].

- **Statistical Filter:** This approach prioritized features based on their statistical robustness across phonatory tasks [174, 175]. First, features showing statistically significant group differences ($p < 0.05$) were identified. From this pool, only those features that maintained significance across more than two vowels were retained. This consistency criterion ensures that the selected predictors represent generalized pathological alterations in voice production rather than vowel–specific articulatory effects.

Classification outcomes were evaluated using standard metrics applicable to diagnostic decision–support systems: accuracy, sensitivity, and specificity [176]. **Accuracy** quantifies the overall proportion of correctly classified samples (both PD and HC), providing a general measure of model correctness. **Sensitivity** (or recall) represents the true positive rate, reflecting the model's capacity to correctly identify individuals with PD, while **specificity** measures the true negative rate, indicating the ability to correctly identify HC subjects. Together, these metrics provide a comprehensive view of the classifier's performance, balancing the detection of pathology against the risk of false alarms.

4.3 Results

This section presents the findings of the acoustic and biomechanical analysis of sustained vowels in PD. The results are organized into three main parts: first,

a qualitative inspection of the estimated biomechanical trajectories; second, a statistical comparison of the extracted features between PD and HC groups across both databases; and finally, the evaluation of the classification performance using the proposed feature sets.

4.3.1 Acoustic and model–derived biomechanical trajectories

Figure 4.2 illustrates the time–varying profiles of the measured acoustic targets and the corresponding biomechanical control variables estimated by the inverse model for the PC–GITA corpus. Panel (a) presents representative trajectories from one HC subject and one individual with PD during sustained phonation of the vowel /a/, while Panel (b) summarizes the group–level behavior through averaged trajectories across all speakers in each group.

Figure 4.2a shows the tracking performance for individual trials. The upper panels display the acoustic targets $f_o(t)$ and SPL(t), where solid lines correspond to the measured signals and dotted lines to the model outputs. Across both subjects, the estimator closely reproduces the temporal evolution of the acoustic contours, indicating stable inverse tracking performance under both healthy and pathological conditions. Despite comparable overall task execution, clear qualitative differences are observed between groups. The PD subject exhibits a higher f_o

and a lower SPL relative to the HC subject. In addition, the PD trajectories show increased temporal irregularity, with rapid fluctuations and oscillatory behavior in both $f_o(t)$ and SPL(t), whereas the HC trajectories remain comparatively smooth and stable over time.

The lower panels of Figure 4.2a present the corresponding estimated biomechanical control variables. The PD subject exhibits substantially lower $P_S(t)$, with average values of approximately 800 Pa, compared to approximately 1100 Pa in the HC subject. This reduction in aerodynamic drive is consistent with the lower intensity observed at the acoustic level. With respect to laryngeal muscle activation, the PD subject shows lower a_{CT} and higher a_{TA} relative to the HC subject. Notably, the increased temporal variability observed in the acoustic signals is reflected in the estimated biomechanical trajectories, particularly in $P_S(t)$ and $a_{CT}(t)$, indicating that acoustic instability is accompanied by increased variability in the inferred motor control variables.

At the population level, Figure 4.2b presents the mean trajectories for the HC (blue) and PD (orange) groups in the PC–GITA cohort. The group–averaged acoustic trajectories show that the PD group maintains, on average, a slightly higher f_o throughout the phonation, while mean SPL is consistently lower and exhibits a more pronounced decay toward the end compared to the HC group.

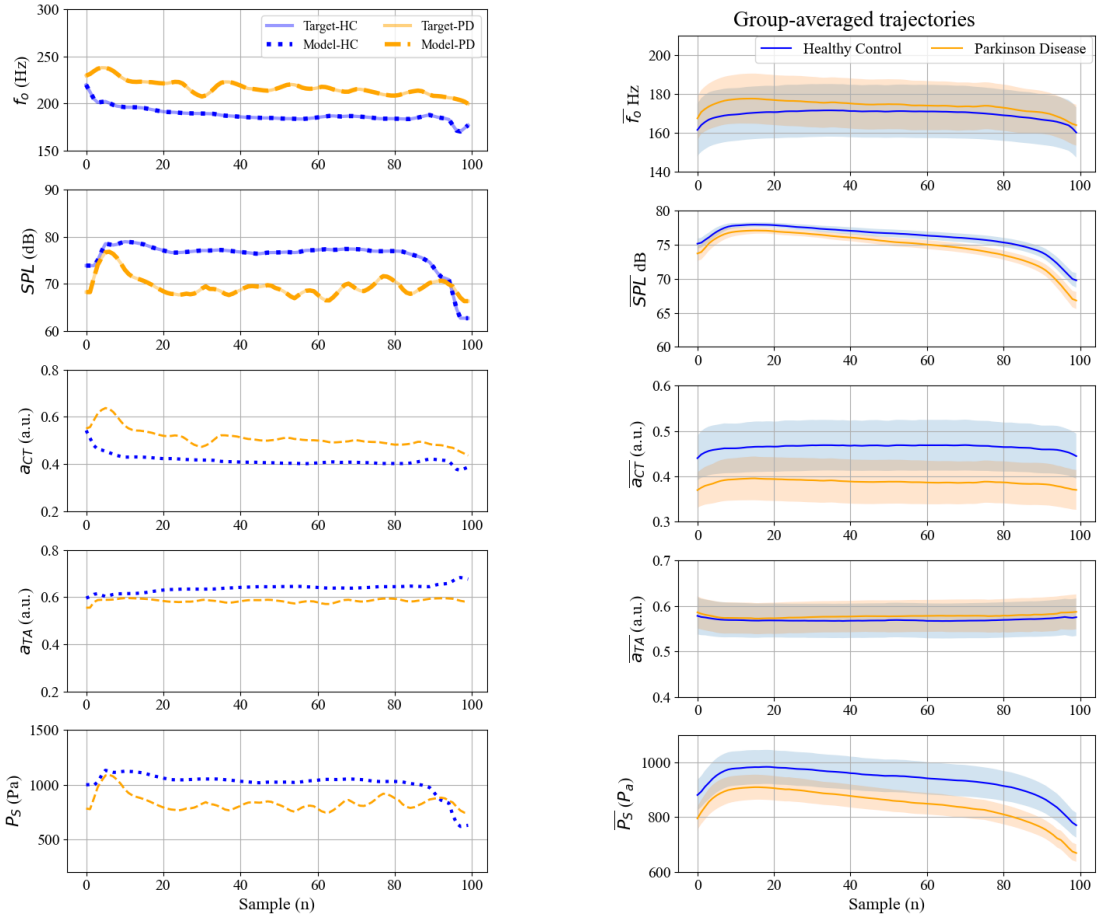
The corresponding group–averaged biomechanical trajectories further characterize these differences. The PD group exhibits lower $\overline{P_S}$ and reduced $\overline{a_{CT}}$ across

the duration of phonation. In contrast, the $\overline{a_{TA}}$ is only slightly higher in the PD group than in the HC group. This antagonistic pattern between a_{CT} and a_{TA} reflects systematic differences in the inferred laryngeal control strategies between groups. Finally, the shaded regions representing the 95% confidence intervals indicate minimal overlap between groups for SPL and P_S, suggesting that group differences in aerodynamic drive and vocal intensity are robust at the population level.

4.3.2 Statistical analysis of acoustic and biomechanical features

Since most feature distributions deviated from normality (Anderson–Darling test, $p < 0.05$), group comparisons were conducted using the non-parametric Mann–Whitney U test. Effect sizes were quantified using Cliff's delta (δ). To facilitate interpretation of the high–dimensional feature space, results are organized by feature domain and corpus.

First, a detailed analysis of the PC–GITA corpus is presented across articulatory, vocal quality, and acoustic–biomechanical domains. Second, a cross–corpus validation is performed using the NeuroVoz dataset, reporting only features showing statistically significant group differences ($p < 0.05$).



(a) (b)

Figure 4.2: Comparison of acoustic targets and estimated biomechanical trajectories for the sustained vowel /a/ (PC–GITA corpus). (a) Tracking performance for a representative Healthy Control (blue) and a subject with Parkinson's Disease (orange). (b) Mean trajectories ($\pm$ 95% confidence interval) for both groups.

PC–GITA corpus

Articulatory features. As summarized in Table 4.2, static articulatory estimates, including vowel–specific VTL measures, showed only sparse group differences and predominantly small effect sizes. This suggests that the average articulatory configuration remains largely preserved in PD speakers.

In contrast, dynamic articulatory stability was consistently reduced. Across nearly all vowels, PD speakers exhibited significantly higher variability in formant trajectories, particularly in F_1(SD) and F_2(SD), with small–to–medium effect sizes (e.g., F_1 SD for /a/: $p = 0.001$, $|\delta| = 0.403$; F_2 SD for /o/: $p < 0.001$, $|\delta| = 0.406$). Higher–order formants (F_3, F_4) showed fewer and weaker effects, indicating that instability primarily affects the lower spectral dimensions most directly linked to tongue height and advancement.

Table 4.2: Descriptive statistics of articulatory features for HC and PD groups in the PC–GITA corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	δ**-cliff**	**Effect size**
	a	37.386 ± 31.452	66.895 ± 54.032	0.001	-0.403	Medium
	e	19.379 ± 12.81	30.747 ± 28.34	0.011	-0.295	Small
F_1(SD)	i	16.855 ± 15.045	27.294 ± 26.683	0.004	-0.335	Medium
	o	28.573 ± 23.888	44.036 ± 36.501	0.013	-0.289	Small
	u	33.326 ± 33.925	45.067 ± 40.379	0.048	-0.23	Small
	a	37.386 ± 31.452	66.895 ± 54.032	0.001	-0.403	Medium
	e	19.379 ± 12.81	30.747 ± 28.34	0.011	-0.295	Small
F_1(SD)						

Table 4.2: Descriptive statistics of articulatory features for HC and PD groups in the PC–GITA corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	δ**-cliff**	**Effect size**
	i	16.855 ± 15.045	27.294 ± 26.683	0.004	-0.335	Medium
	o	28.573 ± 23.888	44.036 ± 36.501	0.013	-0.289	Small
	u	33.326 ± 33.925	45.067 ± 40.379	0.048	-0.23	Small
	a	56.131 ± 49.802	86.125 ± 82.346	0.03	-0.252	Small
	e	64.715 ± 56.152	88.948 ± 90.683	0.199	-0.15	-
F_2(SD)	i	70.948 ± 76.312	112.323 ± 113.489	0.007	-0.315	Small
	o	61.938 ± 98.452	127.291 ± 155.426	<0.001	-0.406	Medium
	u	147.402 ± 198.98	202.819 ± 250.68	0.143	-0.17	-
	a	100.829 ± 90.277	134.919 ± 96.557	0.03	-0.253	Small
	e	98.562 ± 82.468	97.633 ± 68.154	0.515	-0.076	-
F_3(SD)	i	99.665 ± 62.166	140.104 ± 105.241	0.091	-0.197	-
	o	78.592 ± 81.979	100.766 ± 82.092	0.051	-0.227	-
	u	108.768 ± 101.283	131.441 ± 125.394	0.493	-0.08	-
	a	172.975 ± 124.069	185.11 ± 102.046	0.243	-0.136	-
	e	249.799 ± 171.508	181.667 ± 124.112	0.079	0.204	-
F_4(SD)	i	182.839 ± 138.38	170.592 ± 129.454	0.588	0.063	-
	o	113.0 ± 110.593	143.471 ± 96.003	0.032	-0.249	Small
	u	142.391 ± 116.951	190.925 ± 149.547	0.155	-0.166	-
	a	11.016 ± 3.33	11.733 ± 1.726	0.005	-0.326	Small
	e	16.941 ± 2.346	18.027 ± 2.783	0.18	-0.156	-
VTL(F_1)	i	23.554 ± 4.764	24.14 ± 3.376	0.391	-0.1	-
	o	16.595 ± 2.235	17.772 ± 2.638	0.045	-0.233	Small
	u	21.284 ± 3.045	22.762 ± 4.056	0.134	-0.174	-
	a	19.794 ± 2.821	18.528 ± 2.185	0.035	0.246	Small
	e	12.667 ± 1.378	12.253 ± 1.7	0.066	0.214	-
VTL(F_2)	i	11.323 ± 1.107	11.31 ± 1.422	0.677	0.049	-

Table 4.2: Descriptive statistics of articulatory features for HC and PD groups in the PC–GITA corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	**δ-cliff**	**Effect size**
	o	28.893 ± 3.239	27.643 ± 3.603	0.088	0.198	-
	u	31.366 ± 5.475	29.319 ± 5.892	0.022	0.266	Small
	a	16.541 ± 1.511	15.806 ± 1.487	0.025	0.261	Small
	e	16.376 ± 1.356	15.915 ± 1.514	0.141	0.171	-
VTL(F_3)	i	15.363 ± 1.374	14.691 ± 1.443	0.018	0.274	Small
	o	16.407 ± 1.575	15.773 ± 1.7	0.028	0.256	Small
	u	16.235 ± 1.752	15.876 ± 1.545	0.115	0.183	-
	a	16.814 ± 1.374	16.511 ± 1.07	0.333	0.113	-
	e	16.706 ± 1.481	16.043 ± 1.615	0.042	0.236	Small
VTL(F_4)	i	16.368 ± 1.451	15.604 ± 1.609	0.009	0.304	Small
	o	16.776 ± 1.399	16.199 ± 1.378	0.036	0.244	Small
	u	16.497 ± 1.54	15.921 ± 1.481	0.046	0.232	Small

Vocal quality features. Table 4.3 reveals a robust degradation of vocal quality in the PD group. Measures related to periodicity and spectral noise, such as CPP and HNR, were consistently lower in PD speakers across several vowels, with medium effect sizes (e.g., CPP for /o/: $p = 0.001$, $|\delta| = 0.391$).

Concurrently, perturbation metrics exhibited systematic increases. Both frequency–based (Jitter) and amplitude–based (Shimmer) measures were significantly elevated, with some of the largest effect sizes observed for amplitude perturbation (e.g., $Shimmer_{local\_dB}$ for /a/: $|\delta| = 0.518$, Large). Together, these results indicate pronounced temporal instability in phonatory vibration rather than shifts

in average spectral shape.

Table 4.3: Descriptive statistics of vocal quality, perturbation, and spectral noise features for HC and PD groups in the PC–GITA corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	**δ-cliff**	**Effect size**
CPP	a	20.579 ± 2.047	18.732 ± 3.299	0.002	0.366	Medium
	e	20.36 ± 2.346	18.606 ± 3.107	0.002	0.356	Medium
	i	19.893 ± 2.634	18.074 ± 3.088	0.003	0.342	Medium
	o	20.023 ± 2.089	18.248 ± 2.83	0.001	0.391	Medium
	u	17.712 ± 2.335	16.745 ± 2.805	0.141	0.171	-
$H_1 - H_2$	a	6.25 ± 6.807	4.843 ± 8.394	0.754	0.037	-
	e	5.628 ± 7.658	6.53 ± 6.428	0.43	-0.092	-
	i	10.031 ± 6.388	8.334 ± 7.581	0.257	0.132	-
	o	4.806 ± 6.698	4.679 ± 7.51	0.593	-0.062	-
	u	9.183 ± 6.955	6.439 ± 9.129	0.29	0.123	-
HNR	a	21.465 ± 3.382	17.869 ± 4.671	<0.001	0.458	Medium
	e	21.677 ± 3.956	18.994 ± 4.294	0.001	0.394	Medium
	i	22.909 ± 4.577	21.348 ± 4.617	0.109	0.186	-
	o	26.22 ± 4.345	23.17 ± 5.137	0.001	0.377	Medium
	u	27.542 ± 4.964	24.888 ± 5.455	0.008	0.309	Small
$Jitter_{local}$	a	0.127 ± 0.102	0.322 ± 0.354	<0.001	-0.452	Medium
	e	0.152 ± 0.269	0.271 ± 0.294	0.002	-0.364	Medium
	i	0.134 ± 0.201	0.241 ± 0.308	0.006	-0.321	Small
	o	0.143 ± 0.149	0.293 ± 0.349	0.023	-0.264	Small
	u	0.165 ± 0.221	0.301 ± 0.374	0.007	-0.314	Small
$Jitter_{ppq5}$	a	0.03 ± 0.037	0.099 ± 0.132	0.001	-0.397	Medium
	e	0.044 ± 0.117	0.081 ± 0.103	0.004	-0.338	Medium
	i	0.032 ± 0.065	0.07 ± 0.114	0.021	-0.269	Small
	o	0.033 ± 0.05	0.087 ± 0.125	0.022	-0.265	Small
	u	0.04 ± 0.08	0.083 ± 0.132	0.008	-0.31	Small

Table 4.3: Descriptive statistics of vocal quality, perturbation, and spectral noise features for HC and PD groups in the PC–GITA corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	**δ-cliff**	**Effect size**
	a	0.011 ± 0.014	0.04 ± 0.057	0.001	-0.389	Medium
	e	0.016 ± 0.045	0.031 ± 0.041	0.004	-0.336	Medium
$Jitter_{RAP}$	i	0.013 ± 0.031	0.026 ± 0.046	0.024	-0.262	Small
	o	0.013 ± 0.022	0.034 ± 0.051	0.03	-0.252	Small
	u	0.015 ± 0.033	0.033 ± 0.054	0.009	-0.304	Small
	a	0.019 ± 0.019	0.057 ± 0.064	<0.001	-0.479	Large
	e	0.026 ± 0.068	0.042 ± 0.055	<0.001	-0.429	Medium
NHR	i	0.023 ± 0.07	0.03 ± 0.045	0.037	-0.242	Small
	o	0.016 ± 0.046	0.03 ± 0.043	0.003	-0.346	Medium
	u	0.017 ± 0.045	0.025 ± 0.039	0.024	-0.263	Small
	a	2.171 ± 0.948	3.198 ± 1.699	0.001	-0.383	Medium
	e	1.905 ± 1.361	2.713 ± 1.632	0.001	-0.38	Medium
$Shimmer_{apq3}$	i	1.793 ± 1.164	2.39 ± 1.832	0.103	-0.19	-
	o	1.781 ± 1.169	2.375 ± 1.58	0.048	-0.23	Small
	u	2.007 ± 1.427	2.346 ± 1.631	0.306	-0.119	-
	a	2.699 ± 1.14	4.098 ± 2.049	<0.001	-0.466	Medium
	e	2.365 ± 1.454	3.524 ± 1.949	<0.001	-0.425	Medium
$Shimmer_{apq5}$	i	2.311 ± 1.127	3.2 ± 1.944	0.022	-0.265	Small
	o	2.281 ± 1.241	3.093 ± 1.814	0.019	-0.274	Small
	u	2.559 ± 1.683	3.232 ± 1.884	0.048	-0.23	Small
	a	6.515 ± 2.843	9.595 ± 5.098	0.001	-0.383	Medium
	e	5.716 ± 4.084	8.139 ± 4.896	0.001	-0.38	Medium
$Shimmer_{DDA}$	i	5.379 ± 3.491	7.17 ± 5.496	0.103	-0.19	-
	o	5.343 ± 3.507	7.126 ± 4.739	0.048	-0.23	Small
	u	6.022 ± 4.282	7.038 ± 4.894	0.306	-0.119	-
	a	0.413 ± 0.161	0.632 ± 0.283	<0.001	-0.518	Large

$Shimmer_{local}$_db

Table 4.3: Descriptive statistics of vocal quality, perturbation, and spectral noise features for HC and PD groups in the PC–GITA corpus.

Measures	Vowel	HC Mean ± SD	PD Mean ± SD	p	δ-cliff	Effect size
	e	0.385 ± 0.267	0.575 ± 0.278	<0.001	-0.487	Large
	i	0.393 ± 0.269	0.555 ± 0.315	0.001	-0.377	Medium
	o	0.388 ± 0.219	0.537 ± 0.283	0.004	-0.338	Medium
	u	0.439 ± 0.284	0.583 ± 0.31	0.009	-0.302	Small

Acoustic targets and biomechanical features. Table 4.4 reports both the acoustic targets tracked by the control system (f_o, SPL) and the estimated biomechanical trajectories underlying their production (P_S, a_{CT} and a_{TA}).

Regarding acoustic targets, no significant group differences were observed in average pitch (f_o mean or median) across vowels, indicating that PD speakers generally achieved the intended tonal targets. Mean SPL showed vowel–dependent reductions, but these effects were not uniform. In contrast, the temporal stability of both targets was consistently reduced: the standard deviation of f_o and SPL was significantly higher in the PD group across most vowels (e.g., f_o SD for /a/: $p = 0.002$; SPL SD for /e/: $p = 0.001$).

This pattern was directly mirrored in the estimated biomechanical control signals. While central tendency measures (mean, median) of a_{TA}, a_{CT}, and P_S exhibited few group differences, their variability was markedly increased in PD speakers. In particular, P_S(SD) showed consistent medium effect sizes across all

vowels (e.g., /u/: $p < 0.001$, $|\delta| = 0.466$), accompanied by elevated variability in both a_{TA} and a_{CT}. These findings indicate that acoustic instability is associated with increased variability in the estimated aerodynamic and laryngeal motor commands.

Table 4.4: Descriptive statistics of the acoustic targets (f_o, SPL) and the estimated laryngeal biomechanical features (a_{TA}, a_{CT}, P_S) for HC and PD groups in the PC–GITA corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	δ**-cliff**	**Effect size**
	a	164.638 ± 45.562	166.462 ± 42.539	0.899	-0.015	-
	e	166.492 ± 46.836	170.416 ± 42.39	0.667	-0.05	-
f_o mean	i	174.736 ± 48.577	177.6 ± 45.564	0.855	-0.022	-
	o	170.476 ± 46.178	174.424 ± 44.695	0.893	-0.016	-
	u	176.697 ± 49.513	182.561 ± 47.527	0.667	-0.05	-
	a	163.344 ± 45.952	167.163 ± 43.942	0.749	-0.038	-
	e	165.125 ± 47.106	171.802 ± 43.684	0.493	-0.08	-
f_o median	i	173.227 ± 50.248	179.897 ± 45.965	0.515	-0.076	-
	o	168.912 ± 46.092	174.472 ± 45.2	0.791	-0.031	-
	u	175.551 ± 50.163	183.945 ± 49.267	0.528	-0.074	-
	a	3.735 ± 3.72	9.674 ± 10.923	0.002	-0.356	Medium
	e	4.157 ± 5.931	9.484 ± 10.696	0.003	-0.343	Medium
f_o(SD)	i	5.504 ± 9.196	8.856 ± 11.192	0.032	-0.25	Small
	o	4.999 ± 5.593	8.152 ± 8.491	0.058	-0.221	-
	u	5.898 ± 8.327	11.338 ± 12.613	0.009	-0.304	Small
	a	75.235 ± 2.089	74.005 ± 2.624	0.016	0.281	Small
	e	76.362 ± 2.005	74.642 ± 2.13	<0.001	0.46	Medium
SPL mean	i	76.255 ± 2.341	75.392 ± 2.328	0.1	0.191	-
	o	76.977 ± 2.234	75.717 ± 2.364	0.004	0.337	Medium

Table 4.4: Descriptive statistics of the acoustic targets (f_o, SPL) and the estimated laryngeal biomechanical features (a_{TA}, a_{CT}, P_S) for HC and PD groups in the PC–GITA corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	**δ-cliff**	**Effect size**
	u	77.572 ± 2.232	76.528 ± 2.241	0.01	0.299	Small
	a	75.305 ± 2.175	74.151 ± 2.648	0.025	0.261	Small
	e	76.432 ± 2.079	74.718 ± 2.181	<0.001	0.443	Medium
SPL median	i	76.266 ± 2.449	75.549 ± 2.504	0.199	0.15	-
	o	77.094 ± 2.251	75.897 ± 2.574	0.013	0.289	Small
	u	77.672 ± 2.324	76.674 ± 2.379	0.026	0.259	Small
	a	1.457 ± 0.671	1.835 ± 0.885	0.026	-0.259	Small
	e	1.471 ± 0.915	2.08 ± 1.181	0.001	-0.37	Medium
SPL(SD)	i	1.506 ± 0.882	2.226 ± 1.253	0.001	-0.372	Medium
	o	1.563 ± 1.045	2.265 ± 1.283	0.002	-0.36	Medium
	u	1.77 ± 1.076	2.437 ± 1.293	0.005	-0.326	Small
	a	0.546 ± 0.165	0.556 ± 0.161	0.97	-0.005	-
	e	0.56 ± 0.142	0.577 ± 0.15	0.672	-0.05	-
a_{TA} mean	i	0.564 ± 0.148	0.579 ± 0.145	0.463	-0.086	-
	o	0.583 ± 0.136	0.581 ± 0.156	0.785	-0.032	-
	u	0.585 ± 0.147	0.592 ± 0.144	0.702	-0.045	-
	a	0.36 ± 0.195	0.356 ± 0.181	0.959	-0.006	-
	e	0.368 ± 0.19	0.376 ± 0.161	0.588	-0.063	-
a_{TA} median	i	0.403 ± 0.21	0.401 ± 0.184	0.888	-0.017	-
	o	0.391 ± 0.183	0.389 ± 0.168	0.866	0.02	-
	u	0.421 ± 0.212	0.419 ± 0.175	0.97	-0.005	-
	a	0.004 ± 0.003	0.011 ± 0.013	0.002	-0.366	Medium
	e	0.006 ± 0.008	0.011 ± 0.011	0.023	-0.264	Small
a_{TA}(SD)	i	0.008 ± 0.015	0.011 ± 0.013	0.035	-0.246	Small
	o	0.007 ± 0.009	0.008 ± 0.008	0.097	-0.193	-

Table 4.4: Descriptive statistics of the acoustic targets (f_o, SPL) and the estimated laryngeal biomechanical features (a_{TA}, a_{CT}, P_S) for HC and PD groups in the PC–GITA corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	**δ-cliff**	**Effect size**
	u	0.01 ± 0.017	0.012 ± 0.012	0.056	-0.222	-
	a	0.409 ± 0.194	0.355 ± 0.178	0.099	0.192	-
	e	0.417 ± 0.189	0.375 ± 0.159	0.208	0.146	-
a_{CT} mean	i	0.452 ± 0.207	0.4 ± 0.183	0.155	0.166	-
	o	0.441 ± 0.184	0.391 ± 0.168	0.071	0.21	-
	u	0.471 ± 0.211	0.423 ± 0.175	0.178	0.157	-
	a	0.43 ± 0.195	0.356 ± 0.181	0.019	0.273	Small
	e	0.438 ± 0.19	0.376 ± 0.161	0.044	0.234	Small
a_{CT} median	i	0.473 ± 0.21	0.401 ± 0.184	0.04	0.238	Small
	o	0.461 ± 0.183	0.389 ± 0.168	0.013	0.288	Small
	u	0.491 ± 0.212	0.419 ± 0.175	0.042	0.236	Small
	a	0.008 ± 0.009	0.019 ± 0.021	0.001	-0.399	Medium
	e	0.009 ± 0.013	0.022 ± 0.026	0.001	-0.376	Medium
a_{CT}(SD)	i	0.013 ± 0.022	0.022 ± 0.028	0.035	-0.245	Small
	o	0.012 ± 0.016	0.016 ± 0.019	0.132	-0.175	-
	u	0.013 ± 0.021	0.025 ± 0.031	0.009	-0.302	Small
	a	865.184 ± 205.034	812.212 ± 176.014	0.103	0.19	-
	e	906.658 ± 223.032	833.801 ± 143.887	0.064	0.215	-
P_S mean	i	925.498 ± 207.659	867.123 ± 151.773	0.047	0.231	Small
	o	940.501 ± 222.337	878.586 ± 152.824	0.046	0.232	Small
	u	977.039 ± 215.429	928.349 ± 153.449	0.097	0.193	-
	a	874.682 ± 205.807	814.648 ± 178.411	0.091	0.197	-
	e	916.951 ± 223.71	834.615 ± 153.853	0.045	0.233	Small
P_S median	i	933.823 ± 209.929	871.185 ± 155.573	0.049	0.229	Small
	o	949.486 ± 223.781	878.825 ± 154.448	0.027	0.257	Small

Table 4.4: Descriptive statistics of the acoustic targets (f_o, SPL) and the estimated laryngeal biomechanical features (a_{TA}, a_{CT}, P_S) for HC and PD groups in the PC–GITA corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	**δ-cliff**	**Effect size**
	u	988.646 ± 218.006	930.265 ± 158.424	0.065	0.214	-
	a	30.059 ± 18.723	49.588 ± 32.368	0.002	-0.357	Medium
	e	31.265 ± 19.518	58.868 ± 40.985	<0.001	-0.449	Medium
P_S(SD)	i	33.343 ± 23.469	61.398 ± 42.54	<0.001	-0.42	Medium
	o	32.732 ± 21.855	52.591 ± 30.894	0.001	-0.403	Medium
	u	37.773 ± 26.055	65.688 ± 36.064	<0.001	-0.466	Medium

NeuroVoz corpus

The analysis was replicated on the NeuroVoz corpus to assess the robustness of the observed patterns. Given its exploratory nature, only features exhibiting statistically significant group differences ($p < 0.05$) are reported.

Significant acoustic features. As shown in Table 4.5, PD speakers again exhibited pronounced acoustic instability. Dynamic measures, particularly f_o(SD), showed large effect sizes across all vowels (e.g., /*a*/: $|\delta| = 0.722$, Large), exceeding those observed in PC–GITA. Variability in SPL and perturbation metrics was also consistently elevated.

Unlike PC–GITA, significant group differences were observed in average f_o (mean and median), suggesting corpus–specific differences in clinical severity, task

conditions, or recording protocols.

Table 4.5: Descriptive statistics of acoustic features showing statistically significant differences ($p < 0.05$) between HC and PD groups in the NeuroVoz corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	**δ-cliff**	**Effect size**
F_1(SD)	u	31.962 ± 71.574	11.486 ± 9.116	0.033	0.243	Small
F_2(SD)	u	199.199 ± 276.303	74.778 ± 158.751	0.024	0.257	Small
F_3(SD)	a	141.25 ± 179.801	172.661 ± 168.711	0.039	-0.23	Small
	u	162.38 ± 132.805	104.85 ± 104.892	0.022	0.26	Small
F_4(SD)	u	184.503 ± 135.891	126.697 ± 114.586	0.014	0.279	Small
VTL(F_1)	e	19.831 ± 3.31	18.465 ± 4.558	0.036	0.243	Small
	i	27.719 ± 4.101	25.522 ± 4.114	0.006	0.303	Small
VTL(F_2)	e	11.647 ± 1.074	12.306 ± 1.486	0.024	-0.262	Small
	i	10.286 ± 1.141	11.297 ± 2.73	0.003	-0.329	Small
	u	26.929 ± 9.054	33.141 ± 6.861	0.001	-0.391	Medium
VTL(F_3)	i	13.726 ± 1.418	14.692 ± 1.504	0.001	-0.368	Medium
	o	15.1 ± 1.499	16.153 ± 1.802	0.003	-0.331	Medium
	u	15.517 ± 1.762	16.452 ± 1.991	0.039	-0.235	Small
VTL(F_4)	o	15.597 ± 1.494	16.25 ± 1.629	0.043	-0.228	Small
	u	15.273 ± 1.692	16.258 ± 1.434	0.003	-0.338	Medium
CPP	u	14.38 ± 2.353	15.924 ± 2.065	0.001	-0.376	Medium
$Jitter_{ppq5}$	e	0.041 ± 0.07	0.034 ± 0.103	0.045	0.232	Small
$Jitter_{RAP}$	e	0.015 ± 0.031	0.014 ± 0.05	0.029	0.252	Small
$Shimmer_{apq3}$	u	3.39 ± 2.323	2.428 ± 1.498	0.033	0.243	Small
$Shimmer_{DDA}$	u	10.17 ± 6.968	7.284 ± 4.494	0.033	0.243	Small
f_o mean	a	143.552 ± 36.574	161.452 ± 39.314	0.021	-0.256	Small
	e	147.14 ± 38.755	171.848 ± 42.576	0.004	-0.338	Medium
f_o median	a	138.01 ± 37.685	162.942 ± 40.026	0.002	-0.347	Medium
	e	141.194 ± 40.843	172.003 ± 42.489	<0.001	-0.412	Medium
	i	149.912 ± 41.775	169.297 ± 46.524	0.041	-0.227	Small

Table 4.5: Descriptive statistics of acoustic features showing statistically significant differences ($p < 0.05$) between HC and PD groups in the NeuroVoz corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	**δ-cliff**	**Effect size**
	o	144.671 ± 38.132	165.613 ± 41.776	0.012	-0.281	Small
	u	150.638 ± 39.578	169.457 ± 43.388	0.041	-0.232	Small
f_o(SD)	a	2.241 ± 7.433	5.555 ± 10.756	<0.001	-0.722	Large
	e	3.926 ± 11.357	2.238 ± 3.659	<0.001	-0.623	Large
	i	2.62 ± 8.661	2.546 ± 3.47	<0.001	-0.63	Large
	o	1.783 ± 6.567	4.714 ± 9.85	<0.001	-0.707	Large
	u	1.648 ± 4.033	3.192 ± 6.788	<0.001	-0.561	Large
SPL median	e	74.729 ± 6.501	78.473 ± 7.824	0.019	-0.273	Small
	u	73.764 ± 6.09	76.258 ± 5.106	0.026	-0.254	Small
SPL(SD)	a	0.555 ± 0.305	0.933 ± 0.945	<0.001	-0.443	Medium
	e	0.719 ± 0.745	1.342 ± 2.112	0.009	-0.301	Small
	i	0.686 ± 0.666	0.799 ± 0.428	0.026	-0.247	Small
	o	0.84 ± 0.985	1.332 ± 1.693	0.003	-0.329	Small

Significant biomechanical features. Table 4.6 confirms that variability in the estimated motor commands is the most robust cross-corpus marker of PD. The standard deviation of P_S exhibited large effect sizes across all vowels (e.g., /a/: $|\delta| = 0.625$, Large), alongside significantly elevated variability in both a_{CT} and a_{TA}.

As in PC–GITA, central tendency measures showed fewer and weaker effects, reinforcing the dissociation between preserved average motor targets and impaired stability of motor execution.

Table 4.6: Descriptive statistics of biomechanical features showing statistically significant differences ($p < 0.05$) between HC and PD groups in the NeuroVoz corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	**δ-cliff**	**Effect size**
a_{CT} mean	a	0.256 ± 0.159	0.327 ± 0.169	0.034	-0.236	Small
	e	0.286 ± 0.159	0.367 ± 0.18	0.019	-0.271	Small
a_{CT} median	a	0.248 ± 0.16	0.329 ± 0.17	0.017	-0.265	Small
	e	0.274 ± 0.162	0.371 ± 0.18	0.005	-0.329	Small
	o	0.265 ± 0.166	0.344 ± 0.177	0.031	-0.242	Small
	u	0.294 ± 0.159	0.373 ± 0.192	0.039	-0.235	Small
a_{CT}(SD)	a	0.005 ± 0.015	0.011 ± 0.02	<0.001	-0.559	Large
	e	0.01 ± 0.027	0.015 ± 0.025	0.001	-0.368	Medium
	i	0.007 ± 0.017	0.006 ± 0.007	0.018	-0.261	Small
	o	0.006 ± 0.018	0.019 ± 0.044	<0.001	-0.401	Medium
	u	0.005 ± 0.007	0.012 ± 0.035	0.005	-0.322	Small
a_{TA} median	e	0.294 ± 0.162	0.371 ± 0.18	0.022	-0.265	Small
a_{TA}(SD)	a	0.003 ± 0.008	0.007 ± 0.011	<0.001	-0.494	Large
	e	0.005 ± 0.015	0.02 ± 0.053	<0.001	-0.422	Medium
	i	0.004 ± 0.009	0.004 ± 0.005	<0.001	-0.392	Medium
	o	0.005 ± 0.017	0.02 ± 0.066	0.001	-0.361	Medium
	u	0.002 ± 0.006	0.005 ± 0.008	<0.001	-0.458	Medium
P_S mean	e	777.004 ± 251.758	983.542 ± 338.499	0.002	-0.351	Medium
	u	784.815 ± 247.94	905.607 ± 264.6	0.025	-0.254	Small
P_S median	e	764.324 ± 260.223	985.747 ± 357.101	0.001	-0.375	Medium
	o	757.646 ± 261.062	901.699 ± 345.018	0.036	-0.236	Small
	u	779.99 ± 247.473	902.909 ± 265.267	0.024	-0.257	Small
P_S(SD)	a	8.499 ± 21.466	34.207 ± 40.861	<0.001	-0.625	Large
	e	20.169 ± 46.952	55.965 ± 99.108	<0.001	-0.592	Large
	i	14.451 ± 29.63	23.969 ± 15.776	<0.001	-0.505	Large
	o	13.433 ± 28.764	48.643 ± 75.758	<0.001	-0.536	Large

Table 4.6: Descriptive statistics of biomechanical features showing statistically significant differences ($p < 0.05$) between HC and PD groups in the NeuroVoz corpus.

Measures	**Vowel**	**HC Mean ± SD**	**PD Mean ± SD**	p	δ**-cliff**	**Effect size**
	u	10.606 ± 16.365	29.094 ± 27.456	<0.001	-0.54	Large

4.3.3 Discriminative analysis and classification performance

This section evaluates the discriminative potential of the proposed feature sets and their performance in automatically distinguishing between individuals with PD and HC. The analysis first examines feature ranking results to identify the most informative predictors and subsequently reports classification metrics obtained using a SVM classifier.

Feature ranking and selection

To identify the most informative markers of PD-related phonatory impairment, features were ranked using two complementary strategies: univariate statistical significance (ranking by p–value) and multivariate relevance estimation using the Relief–F algorithm. Table 4.7 summarizes the top–ranked features for the PC–GITA and NeuroVoz corpora, revealing distinct acoustic profiles alongside convergent biomechanical patterns.

Regarding the best acoustic features, the rankings highlight notable differ-

ences between the two databases. In the PC–GITA corpus, the most informative acoustic descriptors are predominantly related to short–term phonatory stability and glottal noise. Multiple variants of jitter and shimmer, together with spectral noise measures (NHR, HNR, CPP), consistently appear among the top–ranked features. This pattern indicates that fine-grained perturbation measures play a prominent role in distinguishing PD from HC in this cohort.

In contrast, the NeuroVoz ranking is dominated by measures capturing broader pitch and intensity variability. f_o(SD) emerges as the single most informative acoustic feature, accompanied by f_o median and SPL(SD). This divergence suggests that the acoustic manifestations of PD differ across corpora, potentially reflecting differences in clinical severity, demographic composition, or recording conditions.

In contrast to the acoustic features, the ranking of biomechanical descriptors reveals a high degree of consistency across both databases. Despite the heterogeneous acoustic patterns, the same estimated motor control variables are identified as the most informative predictors in PC–GITA and NeuroVoz. Specifically, variability measures of P_S, a_{CT}, and a_{TA}, consistently occupy the top positions.

This convergence indicates that distinct surface acoustic phenomena are mapped by the inverse model onto a shared physiological feature space characterized by increased motor variability. In addition, the presence of central tendency measures such as a_{CT} median and P_S median among the top–ranked biomechan-

ical features suggests that, beyond instability, shifts in the operating point of laryngeal control may also contribute to group discrimination.

The Relief–F analysis provided complementary insights, particularly for the PC–GITA corpus. As shown in Table 4.7, P_S mean and P_S median achieved the highest relevance scores (0.109 and 0.096, respectively), surpassing variability–based measures that dominated the univariate statistical ranking. Notably, these central tendency features did not exhibit statistically significant group differences in isolation (Section 4.3.2), suggesting that their discriminative contribution emerges through multivariate interactions captured by the Relief–F algorithm.

For the NeuroVoz corpus, the Relief–F ranking largely aligned with the univariate statistical results. Variability measures such as f_o(SD) (0.303) and P_S(SD) (0.169) retained the highest relevance scores, reinforcing the role of gross phonatory instability as the dominant discriminative factor in that dataset.

Table 4.7: Top-ranked features for the PC–GITA and NeuroVoz databases, selected based on statistical significance (Best Acoustic and Best Biomechanical sets) and Relief–F ranking (joint feature set). The asterisk (*) indicates that all computed variants of jitter and shimmer were considered.

PC-GITA				**NeuroVoz**			
Best acoustic	**Best biomechanical**	**Relief-F - score**		**Best acoustic**	**Best biomechanical**	**Relief-F - score**	
Shimmer*	P_S(SD)	P_S mean	0.109	f_o(SD)	P_S(SD)	f_o(SD)	0.303
Jitter*	a_{CT}(SD)	P_S median	0.096	f_o median	a_{TA}(SD)	P_S(SD)	0.169
SPL (SD)	a_{TA}(SD)	P_S(SD)	0.096	SPL (SD)	a_{CT}(SD)	a_{TA}(SD)	0.102

CPP	a_{CT} median	a_{CT} median	0.071	VTL(F_2-F_3-F_4)	P_S median	a_{CT} median	0.101
NHR	P_S median	F_1(SD)	0.065	f_o mean	a_{CT} median	a_{CT} mean	0.086
HNR	P_S mean	NHR	0.065	SPL median	P_S mean	SPL (SD)	0.083
f_o(SD)		$Jitter_{local}$	0.064		a_{CT} mean	a_{TA} mean	0.076
F_1-F_2(SD)		f_o median	0.054			a_{CT}(SD)	0.069
SPL median		a_{TA}(SD)	0.054			f_o median	0.059
VTL(F_1-F_3-F_4)		CPP	0.046			P_S median	0.053

Classification results

Table 4.8 reports the classification performance obtained independently for the PC–GITA and NeuroVoz corpora. In PC–GITA, models based exclusively on biomechanical features achieved the highest overall accuracy (87.6%), substantially outperforming acoustic–only configurations. Feature sets emphasizing variability consistently yielded more balanced sensitivity (85.60%) and specificity (89.60%) compared to those based on central tendency measures alone.

In NeuroVoz, classification performance was overall lower and more evenly distributed across feature sets. Combined feature configurations, including All Best and Top 10 Relief–F, achieved the highest accuracy (approximately 74%), indicating that integrating acoustic and biomechanical information is beneficial in this corpus.

Across both datasets, feature sets derived from Relief–F ranking exhibited robust and stable performance (PC-GITA: Accuracy:78.80% - Sensitivity:77.20% - Specificity:80.40%, NeuroVoz: Accuracy:74.21% - Sensitivity:68.90% - Speci-

ficity:80.20%), supporting the relevance of multivariate feature interactions for PD classification.

Table 4.8: Classification performance (Mean ± SD) for PC–GITA and NeuroVoz databases using SVM. Results are aggregated across all vowels.

PC-GITA			
Features sets	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
Acoustic	65.40 ± 8.95	65.60 ± 14.55	65.20 ± 20.40
Best Acoustic	62.80 ± 6.82	58.40 ± 12.67	67.20 ± 19.98
Biomechanical	**87.60 ± 6.97**	**85.60 ± 11.48**	**89.60 ± 10.15**
Best Biomechanical	73.60 ± 9.54	66.40 ± 16.22	80.80 ± 14.73
All	72.80 ± 6.88	75.20 ± 11.14	70.40 ± 14.99
All Best	67.60 ± 6.25	68.00 ± 13.86	67.20 ± 12.37
Top 10 Relief-F	**78.80 ± 9.04**	**77.20 ± 18.16**	**80.40 ± 15.01**
NeuroVoz			
Features sets	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
Acoustic	62.25 ± 5.51	55.53 ± 12.57	70.22 ± 11.44
Best Acoustic	67.40 ± 7.16	60.71 ± 15.64	74.92 ± 12.50
Biomechanical	67.95 ± 7.90	58.29 ± 12.12	78.71 ± 12.41
Best Biomechanical	67.80 ± 8.94	59.61 ± 12.75	75.54 ± 8.59
All	62.98 ± 6.75	56.83 ± 14.45	69.68 ± 13.07
All Best	**73.68 ± 6.65**	**68.71 ± 17.55**	**78.83 ± 9.12**
Top 10 Relief-F	**74.21 ± 6.70**	**68.90 ± 12.93**	**80.20 ± 8.14**

4.4 Discussion

The present study investigated PD–related alterations in voice production by integrating conventional acoustic analysis with biomechanically informed control

variables estimated through a laryngeal motor control model. Across two independent corpora, the results consistently revealed a differences between preserved central tendency measures and impaired temporal stability. While individuals with PD generally achieved average acoustic targets comparable to HC, they exhibited significantly increased variability in both acoustic outputs and inferred motor commands.

At the acoustic level, the most salient group differences were driven by measures capturing temporal variability rather than by average spectral or prosodic characteristics. In the PC–GITA corpus, short–term perturbation metrics such as jitter and shimmer, together with spectral noise measures including CPP and HNR, emerged as the most informative discriminative features (see Table 4.7). These descriptors are sensitive to cycle–to–cycle irregularities in VFs vibration and are commonly interpreted as indicators of reduced phonatory stability and impaired fine motor control of the larynx [177, 178].

In contrast, the NeuroVoz corpus was characterized by increased variability at a broader temporal scale. Measures such as f_o(SD) and SPL(SD) consistently ranked among the most informative features, indicating pronounced fluctuations in pitch and loudness over the duration of sustained phonation. Although classical descriptions of hypokinetic dysarthria emphasize reduced pitch variability and monotone speech in connected speech and prosodic tasks [179, 180], accumulating evidence shows that variability–based measures remain sensitive markers of vocal

impairment across tasks and temporal scales. In particular, systematic reviews report significant alterations in f_o variation and its standard deviation in PD compared to healthy speakers, reflecting impaired regulation of phonatory and prosodic control in both sustained and read speech [181].

Despite these differences in acoustic expression across corpora, both datasets converged on variability–based descriptors as the most informative features (see Table 4.2, 4.3 and 4.5). This observation is consistent with prior acoustic and physiological studies showing that Parkinsonian speech exhibits concurrent alterations in short–term perturbation measures (e.g., jitter, shimmer) and longer–term measures of pitch and intensity variability [182, 21]. Together, these findings indicate that PD–related phonatory impairment manifests across multiple temporal scales rather than being confined to a single acoustic dimension.

Crucially, the predominance of variability measures over central tendency metrics suggests that the observed acoustic differences are not primarily driven by systematic shifts in intended phonatory targets. Instead, they point toward an impaired ability to maintain stable vocal output over time. From a motor control perspective, such instability is consistent with increased execution noise or degraded regulation of ongoing motor commands, rather than with errors in target specification. This interpretation aligns with previous reports of heightened vocal variability in PD across sustained [183] and connected speech tasks [184, 185] and with broader theoretical accounts proposing that basal ganglia dysfunction com-

promises the suppression of motor variability and the stability of motor execution [10, 77].

While the acoustic profiles differed across corpora, the estimated biomechanical features revealed a notable convergence. In both PC–GITA and NeuroVoz, the most informative physiological descriptors were the standard deviations of P_S, a_{CT} and a_{TA} (see Table 4.4 and 4.6). These features consistently ranked among the top discriminative predictors, regardless of the dominant acoustic manifestation observed at the surface level ans shown in Table4.7.

This convergence suggests that distinct acoustic phenomena–ranging from fine–grained perturbations to broader pitch and intensity fluctuations, can be interpreted as surface–level expressions of a shared underlying biomechanical profile characterized by increased motor variability. Such a pattern is consistent with prior work describing PD as a disorder of impaired motor regulation rather than systematic target mis–specification, affecting both limb and speech motor systems [10, 77]. In the vocal domain, elevated acoustic variability has been linked to abnormal laryngeal muscle coordination and unstable aerodynamic drive, supporting the physiological plausibility of the present biomechanical findings [180].

Measures of central tendency in the biomechanical domain exhibited fewer and weaker group differences, mirroring the acoustic findings. This dissociation between preserved average control signals and elevated variability reinforces the interpretation that PD primarily affects the regulation and stability of motor ex-

ecution rather than the specification of phonatory targets [10, 77]. At the same time, the relevance of some mean or median biomechanical features in multivariate analyses suggests that compensatory shifts in operating point may coexist with increased instability, reflecting attempts to maintain phonatory goals in the presence of degraded motor regulation [182, 84].

From a methodological perspective, the classification results further reinforce the proposed interpretation of PD–related vocal impairment. The use of Relief–F feature ranking combined with compact feature sets is consistent with current trends in voice–based PD characterization, where increasing emphasis is placed on interpretable representations rather than on large feature spaces [153, 93]. Although numerous studies have demonstrated that classification accuracy can be improved by expanding the dimensionality of acoustic feature sets or by relying on deep learning architectures trained on time-–frequency representations [186, 155], such approaches often provide limited insight into the physiological mechanisms underlying the observed performance gains.

In contrast, the present framework prioritizes a reduced set of acoustically and biomechanically meaningful descriptors. Classification performance comparable to acoustic–only baselines was achieved using a limited number of features, while preserving direct interpretability in terms of phonatory control variables. Notably, incorporating biomechanical features yielded equal or improved discriminative performance relative to prior model–based approaches that inferred latent

motor variables from speech, particularly those relying on simplified representations or restricted acoustic targets [158, 159]. This balance between discriminative potential and explanatory value represents a key advantage over purely data–driven approaches, especially in a clinical context where understanding *why* a classifier succeeds is as important as its raw accuracy [151, 94].

Importantly, the Relief–F analysis revealed that certain central tendency measures of P_S (see table4.7), which did not emerge as significant in univariate group comparisons, contributed meaningfully to classification when combined with variability–based descriptors. This finding suggests that PD–related vocal impairment may involve an interaction between compensatory shifts in operating point and increased motor instability, a pattern that has been reported across multiple motor domains in PD [77, 84].

Taken together, these results position biomechanically informed features as a complementary and potentially more robust representation of PD–related vocal impairment. By grounding acoustic observations in motor control variables, the proposed framework bridges surface–level speech symptoms and underlying physiological mechanisms. This is particularly relevant in PD, where direct access to physical measurements of the speech production system, such as neuromuscular control or laryngeal biomechanics, is often limited or impractical in clinical settings. The present findings demonstrate that inverse biomechanical modeling can reveal such commonalities, providing a principled pathway toward interpretable,

mechanism–based voice markers.

However, several methodological limitations should be considered when interpreting the results of this chapter. The analysis was restricted to sustained vowel phonation, a controlled and quasi–stationary task chosen to ensure reliable biomechanical trajectory estimation. While this setting isolates fundamental aspects of phonatory control, it limits ecological validity and may not capture motor control deficits that emerge during connected or spontaneous speech. In addition, the biomechanical inference relies on a forward model trained on simulations of a reference male laryngeal configuration with fixed adductory posture. Although this model has demonstrated robustness, it does not explicitly account for inter–speaker anatomical variability, sex–related differences, or disease–related changes in tissue properties. Consequently, the estimated trajectories should be interpreted as relative indicators of motor control rather than absolute physiological measurements. Finally, the inverse control formulation assumes smooth compensatory adjustments in P_S, a_{CT}, and a_{TA} and does not explicitly model sensory delays, nonlinear feedback dynamics, or task–dependent control strategies, which may be altered in PD.

Future research should extend the proposed framework to more complex speech tasks, such as connected speech, which impose greater demands on motor planning, articulatory coordination, and feedback control. Methodological developments may include the incorporation of additional laryngeal muscle activations,

speaker–specific or anatomically scaled biomechanical models, and more explicit representations of sensory feedback and control dynamics. From a classification perspective, future studies should evaluate cross–database generalization to assess whether biomechanically informed features provide improved robustness under differents recording conditions and clinical settings.

4.5 Chapter Conclusions

This chapter examined PD–related alterations in voice production by integrating conventional acoustic analysis with biomechanically informed control variables estimated through a laryngeal motor control model. Across two independent speech corpora, the results consistently showed that PD–related vocal impairment is driven primarily by increased temporal variability rather than by systematic shifts in average phonatory targets. Although individuals with PD generally achieved mean acoustic outcomes comparable to HC, they exhibited reduced stability in both acoustic outputs and inferred motor commands. While the acoustic manifestations of this instability varied across corpora, ranging from short–term perturbations to broader fluctuations in pitch and intensity, the biomechanical estimates converged toward a shared physiological profile characterized by increased variability in P_S, a_{CT}, and a_{TA}. This convergence indicates that different acoustic outcomes can be interpreted as surface–level expressions of a common deficit

in motor coordination and stability affecting the regulation of aerodynamic drive and laryngeal muscle activations.

These findings are directly aligned with Hypothesis 1 of this thesis, which posited that biomechanically informed vocal control variables capture systematic alterations associated with central motor impairment in PD that are not fully reflected by conventional acoustic measures. In particular, the consistent prominence of variability–based biomechanical features and their relevance for classification support the view that biomechanical trajectory estimation provides access to physiologically meaningful dimensions of motor instability underlying PD–related voice impairment. By linking observable acoustic variability to coordinated fluctuations in motor control variables, the proposed framework offers a principled and interpretable approach for bridging speech acoustics and neural motor dysfunction in PD.

Chapter 5

Biomechanical correlates of autonomic arousal during cognitively demanding speech

Considering the specific aims of this thesis, this chapter examines how transient autonomic arousal associated with cognitive load (CL) is reflected in laryngeal motor control during continuous speech production. Building on the model–based trajectory estimation framework developed and validated in previous chapters, the present analysis extends its application from chronic neurological conditions to task–driven modulation of vocal motor strategies in neurologically healthy speakers.

CL was experimentally elicited using the Stroop color–word task, a well–established paradigm known to induce autonomic activation and increased mental demand. Speech produced under varying levels of CL was analyzed in both healthy controls and speakers with vocal hyperfunction, a population commonly associated with increased laryngeal muscle activation, elevated aerodynamic drive, and reduced adaptability in vocal motor coordination. In contrast to the PD case,

where sustained phonation was used to isolate phonatory deficits under controlled conditions, the present chapter focuses on connected speech in order to capture dynamic, context–dependent adjustments of vocal control under cognitive demand.

f_o and SPL trajectories were extracted from continuous speech and subsequently used to estimate time–varying laryngeal control variables, including P_S and ILMs activations, using the same physiologically grounded control framework adopted in Chapter chapter 4. By reusing an established forward mapping and inverse control formulation, this approach enables autonomic influences on vocal production to be interpreted in terms of underlying biomechanical adaptations, rather than solely through surface acoustic measures.

The estimated biomechanical trajectories were analyzed using LME models to assess the main effects of CL, group (HC vs. VH speakers), and their interaction. Importantly, these effects were evaluated in relation to the functional structure of the task, allowing the analysis to capture how autonomic and motor modulations unfold across different phases of speech production within the cognitively demanding context. Post–hoc analyses were conducted to further characterize condition–specific differences in laryngeal control strategies. Through this approach, the chapter investigates whether cognitively induced autonomic arousal is associated with increased amplitude, altered coordination, or reduced adaptability of laryngeal control trajectories in hyperfunctional speakers relative to HC, and whether such effects depend on the temporal context of the task.

By extending the proposed framework to continuous speech under cognitively demanding conditions, this chapter supports the second hypothesis of the thesis, demonstrating how biomechanically informed trajectory estimation can reveal latent adaptations in vocal motor control associated with autonomic modulation.

5.1 Cognitive load and vocal control

CL refers to the mental resources required to process information and sustain task performance under competing demands [187, 188]. In the context of speech production, increases in CL are particularly relevant because they engage regulatory mechanisms beyond voluntary motor planning, eliciting transient autonomic arousal that can influence multiple subsystems involved in voice production. In the present study, CL is operationalized as an experimental manipulation rather than a theoretical construct per se, serving as a controlled perturbation to the vocal motor system.

CL was induced using a color–word Stroop task [189], a well–established paradigm known to generate cognitive conflict and reliably increase activity within the ANS. Heightened autonomic arousal is associated with involuntary physiological adjustments, including changes in respiratory drive, cardiovascular [13, 14], and muscle tone regulation [115], which collectively support performance under increased mental demand [11]. In speech, these responses are expected to affect

both aerodynamic control, through modulation of P_S, and laryngeal motor output, through changes in ILMs activations.

Speech production constitutes a complex sensorimotor behavior requiring continuous coordination between neural, respiratory, and laryngeal subsystems [190, 191]. Accordingly, increased CL has been shown to induce systematic changes in vocal acoustics, including shifts in f_o [192], SPL, and measures of phonatory stability [118, 193]. These changes are often interpreted as compensatory adjustments that help maintain communicative effectiveness under cognitive stress [29]. However, because most prior studies rely on surface acoustic features, they provide limited insight into the underlying laryngeal control mechanisms that generate these effects.

Critically, the relevance of CL for the study of VH lies in the observation that the physiological responses elicited by transient autonomic arousal closely resemble the motor patterns that characterize VH at baseline. VH has been described as a chronic, maladaptive expression of autonomic dysregulation [119], in which sympathetic activation remains persistently elevated, biasing the phonatory system toward increased laryngeal muscle activation and heightened aerodynamic drive [120, 30]. Within this framework, CL does not act as an etiological factor but rather as an experimental probe that transiently shifts the vocal motor system toward a control regime that mirrors the one chronically observed in VH.

From this perspective, increases in cognitive demand are expected to elicit

adaptive, task–dependent laryngeal and respiratory adjustments in healthy speakers, whereas individuals with VH may exhibit amplified responses or a reduced capacity to modulate these adjustments across task conditions. Previous studies in healthy speakers suggest that cognitive stress is accompanied by increased ILMs activation and altered respiratory–laryngeal coordination [27, 28]. In contrast, the chronically elevated or imbalanced activation patterns characteristic of VH [45, 30] may limit the flexibility of these autonomically driven responses.

To address the limitations of purely acoustic analyses, the present chapter applies the model–based laryngeal motor control framework proposed by Parra et al. [1]. By combining a physiologically grounded plant based on the TBCM [42] with a feedback–aware inverse–Jacobian controller [43, 132], this approach enables the estimation of continuous biomechanical trajectories from acoustic signals during connected speech. The primary objective is to characterize how autonomic arousal induced by CL modulates latent laryngeal control variables in healthy controls and speakers with VH. In line with Hypothesis 2 of this thesis, it is expected that VH speakers will exhibit larger and less adaptable biomechanical responses to cognitive demand than healthy speakers, reflecting altered autonomic–motor coupling in vocal control.

5.2 Model-based trajectory analysis of speech under cognitive load

Figure 5.1 provides an overview of the analytical framework employed in this chapter to estimate laryngeal biomechanical trajectories during cognitively demanding speech. The speech recordings analyzed here were originally collected as part of the experimental protocol reported in Dahl and Stepp [121] and are reanalyzed to extend the original acoustic–based approach through the integration of model–based trajectory estimation.

As schematically illustrated in the Figure 5.1, speech produced during a color–word Stroop task was processed to isolate voiced sentence material suitable for trajectory analysis. From these segments, continuous acoustic trajectories of f_o and SPL were extracted and used as inputs to the laryngeal motor control framework. The model estimates the corresponding time-varying biomechanical control variables, including CT and TA muscle activations (a_{CT}, a_{TA}) and P_S, that are most consistent with the observed acoustic behavior.

Beyond trajectory estimation, the framework incorporates a temporally structured analysis aligned with the Stroop task. Each utterance was segmented into functionally defined intervals (Baseline, Effect, and Rest), allowing the examination of how CL modulates both acoustic and biomechanical trajectories over time.

Subsequent statistical analyses evaluate differences between HC and speakers with VH across task conditions and sentence segments, with the aim of characterizing group–specific patterns of autonomic modulation in vocal motor control.

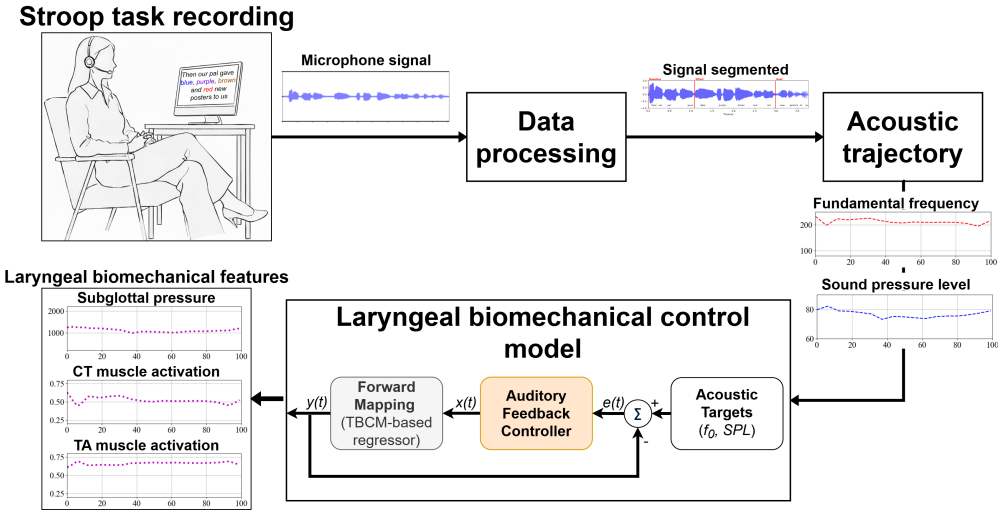


Figure 5.1: Schematic overview of the model–based trajectory analysis applied to speech produced during a Stroop task. Acoustic trajectories of f_o and SPL extracted from segmented sentence material are used to estimate time–varying laryngeal control variables via a physiologically grounded motor control model. The resulting acoustic and biomechanical trajectories are analyzed across task conditions and sentence segments to assess the effects of CL on vocal motor control.

5.2.1 Speech material and temporal structure of the Stroop task

The speech material analyzed in this chapter was drawn from a previously collected database reported by Dahl and Stepp [121], acquired with approval from the Boston University Institutional Review Board. The dataset is revisited here with the specific objective of enabling model–based estimation of laryngeal biomechanical trajectories under varying levels of CL. A concise description of the participants, recording conditions, and task structure is provided to contextualize the methodological decisions described below.

Participants and recording conditions The original dataset comprised recordings from 132 adult women, divided evenly between a HC group (n = 66) with no history of voice disorders and a VH group (n = 66). Participants in the HC group ranged in age from 18 to 68 years (31.8 ± 13.8 years, mean ± SD), while those in the VH group ranged from 18 to 66 years (31.8 ± 13.5 years). Within the VH group, diagnoses included muscle tension dysphonia (n = 41), VF nodules (n = 18), unspecified phonotraumatic disorders (n = 6), and VF polyp (n = 1). None of the participants reported visual, auditory, speech, language, or color–vision impairments.

Recordings were conducted either in a sound–treated booth at Boston Univer-

sity or in a quiet clinical room at Boston Medical Center. Speech signals were acquired at a sampling rate of 44.1 kHz with 16–bit resolution using a head–mounted microphone positioned approximately 7 cm from the mouth at a 45° angle. A condenser Shure SM35XLR microphone was used at Boston University, whereas a dynamic Shure WH20XLR microphone was used at Boston Medical Center. These recording conditions introduced unavoidable inter–session variability, which motivated several of the preprocessing steps described below.

Stroop task design and manipulation of cognitive load CL was elicited using a color–word Stroop paradigm [189], employing the full set of sentences described in Dahl and Stepp [28]. Participants were instructed to read aloud twelve sentences, each containing four color words printed in colored font (e.g., *"Then our pal gave* ***blue, purple, brown, and red*** *new posters to us"*). Crucially, participants were instructed to name the *font color* rather than the written word, thereby inducing cognitive conflict.

The first six sentences constituted the congruent condition, in which word meaning and font color matched, representing a lower–load baseline. The remaining six sentences formed the incongruent condition, in which word meaning and font color differed, reliably increasing CL and autonomic arousal. To minimize carryover effects and avoid contamination of baseline vocal measures by stress–related responses, the congruent condition was always presented prior to the incongruent

condition. This fixed ordering reflects a deliberate experimental choice to preserve a physiologically neutral reference state for subsequent comparisons.

Preprocessing overview and rationale Preprocessing was designed to ensure that the resulting speech signals were suitable for continuous trajectory analysis and biomechanical modeling, while minimizing confounds introduced by recording variability, task performance errors, and silence segments. As schematized in Figure 5.2, the preprocessing pipeline comprised three sequential stages: (a) screening and exclusion based on task performance and duration, (b) amplitude consistency control and outlier removal, and (c) silence removal with segmentation into task–relevant temporal intervals.

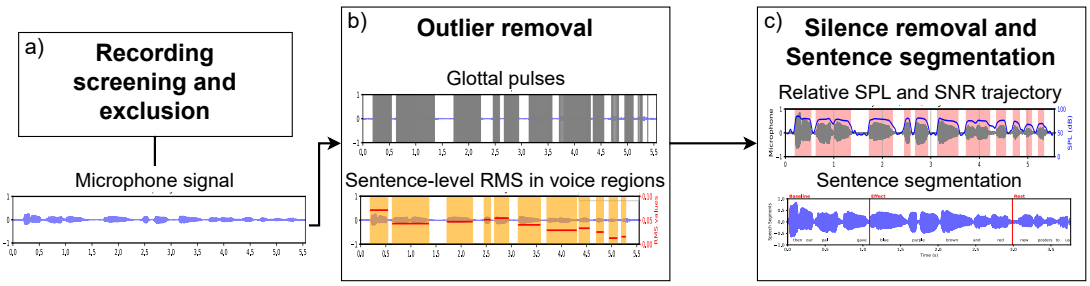


Figure 5.2: A schematic of the preprocessing pipeline for microphone recordings, including recording screening, outlier removal, and silence filtering with sentence segmentation, to generate continuous acoustic trajectories for model–based analysis.

a. Screening based on task performance and duration In the first stage, recordings were screened for abnormal duration patterns and task compliance.

Because CL is known to affect speech timing [118, 194], incongruent sentences whose duration exceeded that of their paired congruent sentence by more than 4 s were excluded, following the criteria established by Dahl and Stepp [121]. This resulted in the removal of 35 incongruent sentences along with their corresponding congruent counterparts. Additionally, 175 sentence pairs were excluded following perceptual inspection when color words were omitted, truncated, or mispronounced, as such productions do not faithfully reflect the Stroop task demands and could distort both acoustic and biomechanical trajectory estimation.

b. Amplitude consistency and RMS–based outlier removal The second stage (Fig. 5.2b) aimed to control for amplitude variability arising from differences in microphone type, placement, and recording conditions, while preserving task–related intensity changes associated with increased cognitive load. Voiced segments were first identified in `Praat` [163] using an autocorrelation–based glottal pulse detection algorithm. The signal was analyzed in consecutive 30 ms windows, which were classified as voiced if (i) the density of detected glottal pulses exceeded 10% of the theoretical maximum and (ii) the mean RMS amplitude exceeded 95% of the signal's amplitude range.

As illustrated in the *upper panel* of Fig. 5.2b, the acoustic waveform is shown together with the detected glottal pulses, indicated by vertical gray markers. This representation highlights the temporal correspondence between periodic excitation

and voiced regions identified by the detection algorithm.

Consecutive voiced windows meeting both criteria were concatenated into continuous voiced segments. RMS amplitude was computed for each segment, and segment–level RMS values were subsequently averaged to obtain a single sentence–level RMS estimate. As shown in the *lower panel* of Fig. 5.2b, the voiced segments are highlighted (yellow shading), while the corresponding RMS amplitude values are overlaid as a red contour.

This dual–criterion voicing strategy ensured robust detection of voiced regions even under conditions of low intensity or large amplitude variability, thereby reducing the risk of misclassifying weak phonation as silence. The resulting sentence–level RMS estimates served as the basis for identifying and excluding amplitude outliers while preserving physiologically meaningful intensity variations associated with task demands.

Outlier detection was performed using the distribution of sentence–level RMS values from the congruent condition only, which was treated as a baseline for comfortable phonation. This choice prevented the systematic removal of legitimate intensity increases associated with elevated CL in the incongruent condition. An interquartile range (IQR) criterion was applied, defined as:

$$IQR = Q_3 - Q_1,$$

$$\text{Lower bound} = Q_1 - 1.5 \times IQR, \tag{5.1}$$

$$\text{Upper bound} = Q_3 + 1.5 \times IQR.$$

Following this step, 1,020 recordings were retained, corresponding to 62 HC participants (30.8 ± 12.9 years) and 63 VH participants (32.3 ± 13.7 years). On average, four sentences per participant were available for analysis. The successive stages of data reduction are summarized in Table 5.1.

Table 5.1: Summary of the signal preprocessing and filtering pipeline applied to the speech recordings.

Stage	**Number of recordings**
Initial dataset	1584 (HC = 65, VHD = 65)
After remove > 4s	1514
After remove misarticulated sentences	1164
After remove outliers (IQR)	1020 (HC = 62, VHD = 63)

c. Silence removal and temporal segmentation of utterances The third stage (Fig. 5.2c) focused on generating continuous acoustic trajectories suitable for biomechanical modeling by removing silent intervals and segmenting each utterance into task–relevant temporal sections. Voiced regions were initially identified

using the glottal pulse detection procedure described above. To further refine the discrimination between voiced and silent intervals, an uncalibrated SPL contour was computed in `Praat` using 100 ms analysis windows. The minimum value of this contour was subtracted to obtain a relative signal–to–noise ratio (SNR) representation, and a threshold of –20 dB relative to the maximum SNR was applied. This conservative threshold was selected empirically to ensure robust performance across heterogeneous recording conditions, consistent with previous SNR–based voice detection approaches [195, 196].

As illustrated in the *upper panel* of Fig. 5.2c, the resulting relative SPL and SNR trajectories were used to identify voiced regions and guide the removal of silent intervals. Sections exceeding the SNR threshold were classified as voiced, whereas segments below the threshold were treated as silence and excluded from further analysis.

Following silence removal, the remaining voiced signal was concatenated to form a continuous utterance and subsequently segmented into three contiguous temporal intervals: **Baseline**, **Effect**, and **Rest**. **Baseline** corresponds to the portion of the sentence preceding the color word, **Effect** contains the uttered color word itself, and **Rest** corresponds to the post–color segment. Importantly, this segmentation was not intended to define the presence or absence of CL, as the Stroop manipulation applies to the entire utterance in incongruent trials. Instead, it constitutes a methodological extension introduced in the present study

to enable a fine–grained analysis of the temporal dynamics of speech production under sustained cognitive demand. To the authors' knowledge, such a temporally resolved analysis has not been previously reported for this dataset, nor combine

As shown in the *lower panel* of Fig. 5.2c, the resulting voiced waveform was divided into these three sections, providing temporally aligned segments for subsequent acoustic trajectory extraction and biomechanical modeling.

To enable consistent and automated segmentation across recordings, word–level time stamps were obtained using an automatic speech–to–text transcription algorithm (AssemblyAI). Given that the lexical content of each sentence was known *a priori*, these word–level alignments allowed reliable identification of the color word and precise delineation of the three temporal intervals across all utterances.

5.2.2 Acoustic and biomechanical trajectory estimation

This section describes the extraction of continuous acoustic trajectories from the preprocessed speech recordings and their subsequent mapping onto underlying laryngeal motor control variables using a biomechanical model. Importantly, the modeling framework and acoustic–to–motor inference strategy are identical to those established and validated in Chapter chapter 3 and applied in the Parkinson's disease analysis in Chapter chapter 4, thereby ensuring methodological consistency across chapters.

The specific contribution of the present chapter lies in extending this validated framework to the analysis of temporally resolved speech trajectories produced under sustained cl. This extension enables the characterization of dynamic laryngeal motor regulation within connected speech, rather than static or time–summarized phonatory configurations, while preserving full compatibility with the biomechanical reference space defined in previous analyses.

Acoustic trajectory extraction

From the preprocessed recordings, continuous trajectories of f_o and SPL were extracted using `Praat` [163]. Acoustic analysis was performed using 100 ms analysis windows with 50% overlap, a configuration that ensures the inclusion of multiple pitch periods per frame and yields stable f_o estimates in connected speech [164].

f_o was estimated using the autocorrelation method with a pitch range of 75–400 Hz. Octave errors were corrected, missing values were interpolated, and the resulting contours were smoothed using a five–sample moving average. SPL trajectories were computed using a minimum periodicity of 100 Hz and smoothed using the same procedure, ensuring temporal consistency between the acoustic dimensions.

Because the recordings were acquired without absolute SPL calibration and using two different headset microphones, raw SPL values were not directly com-

parable across recording sessions. As in the Parkinson's disease chapter, SPL trajectories were therefore expressed relative to a common acoustic reference compatible with the biomechanical simulations used to train the model. Specifically, all trajectories were standardized to an equivalent mouth–to–microphone distance of 15 cm, matching the reference distance assumed in the forward simulations.

This standardization was implemented using the free–field inverse distance law [165], which predicts an SPL decrease of approximately 6 dB for each doubling of distance. Under this assumption, a conversational speech level of 70 dB at 1 m corresponds to 86.48 dB at 15 cm (Eq. Equation 4.1). Rather than enforcing absolute SPL calibration, this value was used as a theoretical anchor to compute a constant offset equal to the difference between the expected SPL at 15 cm and the mean SPL observed in the congruent condition. This offset was then applied uniformly to all SPL trajectories.

Importantly, this procedure does not aim to recover absolute SPLs, but to ensure that the observed SPL trajectories occupy the same acoustic operating range as the biomechanical model, thereby enabling valid inverse inference while preserving relative intensity modulations associated with cognitive load.

Finally, to enable direct comparison across utterances and sentence segments, each acoustic trajectory was time–normalized within the Baseline, Effect, and Rest intervals and resampled to 100 frames per segment. This procedure yielded acoustically and temporally aligned trajectories suitable for model–based inference

and group–level analyses.

Biomechanical trayectory estimation

The laryngeal biomechanical control model employed in this chapter is identical to that introduced and validated in Chapter chapter 3. The model, derived from the TBCM formulation of Parra et al. [1], provides a physiologically grounded forward mapping between laryngeal motor commands and acoustic outputs, and has been shown to support stable inverse inference under constrained adductory conditions.

Based on the comparative evaluation presented previously, the analyses reported here adopt the forward mapping corresponding to Model B. In this configuration, the biomechanical plant is approximated by an eighth-degree polynomial regression trained on simulations of the TBCM. The model implements a nonlinear mapping from a reduced motor command space (P_S, a_{CT} and a_{TA})) to the primary acoustic outputs (f_o and SPL). This approximation preserves the structure of the underlying motor–acoustic relationships while enabling efficient and numerically stable inversion.

The simulations used to construct the forward mapping assumed male laryngeal anatomy during sustained /a/ vowel production. As in previous chapters, a fixed adductory posture was imposed ($a_{LCA} = a_{IA} = 0.6$, $a_{PCA} = 0$), defining a physiologically plausible phonatory reference space. Subglottal pressure ranged

from 400 to 2000 Pa, and a_{CT} and a_{TA} varied from 0 to 1. Acoustic outputs were extracted from the final 50 ms of each simulation to avoid transient effects [38], with SPL computed assuming a 15 cm mouth-to-microphone distance.

Inverse mapping was performed using a Jacobian-based control scheme [43], in which deviations between observed acoustic trajectories $y(t) = [f_o(t), \mathrm{SPL}(t)]$ and model predictions were mapped onto incremental updates in the motor command space $x(t) = [a_{CT}(t), a_{TA}(t), P_S(t)]$. The Jacobian was inverted using a damped least-squares pseudoinverse to improve numerical stability [132]. Controller parameters were fixed across all analyses ($\alpha = 0.3$, $\gamma = 0.1$), and numerical derivatives were computed with a step size of 0.01 at a sampling rate of 1000 Hz.

For each sentence segment (Baseline, Effect, and Rest), the controller tracked the corresponding acoustic trajectory and generated continuous motor trajectories. These trajectories are interpreted as model-consistent biomechanical correlates of the observed acoustic dynamics and constitute the primary variables used to assess the effects of cognitive load in subsequent analyses.

Biomechanical trajectory estimation

The laryngeal biomechanical control model employed in this chapter is identical to that introduced and validated in Chapter chapter 3. The model, derived from the TBCM formulation of Parra et al. [1], provides a physiologically grounded forward mapping between laryngeal motor commands and acoustic outputs and

has been shown to support stable inverse inference under constrained adductory conditions.

Based on the comparative evaluation presented previously, the analyses reported here adopt the forward mapping corresponding to Model B. In this configuration, the biomechanical plant is approximated by an eighth–degree polynomial regression trained on simulations of the TBCM. The model implements a nonlinear mapping from a reduced motor command space ($x(t) = [P_S, a_{CT}, a_{TA}]$) to the primary acoustic outputs ($y(t) = [f_o, SPL]$). This approximation preserves the structure of the underlying motor–acoustic relationships while enabling efficient and numerically stable inversion.

The simulations used to construct the forward mapping assumed a reference male laryngeal anatomy during sustained /*a*/ vowel production. As in previous chapters, a fixed adductory posture was imposed ($a_{LCA} = a_{IA} = 0.6$, $a_{PCA} = 0$), defining a physiologically plausible phonatory reference space. The P_S ranged from 400 to 2000 Pa, and a_{CT} and a_{TA} varied from 0 to 1. Acoustic outputs were extracted from the final 50 ms of each simulation to avoid transient effects [38], with SPL computed assuming a 15 cm mouth–to–microphone distance.

Inverse mapping was performed using a Jacobian–based control scheme [43], in which deviations between observed acoustic trajectories ($y_T(t)$) and model predictions ($y(t)$) were mapped onto incremental updates in the motor command space (see Eq: Equation 2.9). The Jacobian was inverted using a damped least–

squares pseudoinverse to improve numerical stability [132]. Controller parameters were fixed across all analyses ($\alpha = 0.3$, $\gamma = 0.1$), and numerical derivatives were computed with a step size of 0.01 at a sampling rate of 1000 Hz.

For each sentence segment (Baseline, Effect, and Rest), the controller tracked the corresponding acoustic trajectory and generated continuous motor trajectories.

5.2.3 Statistical analysis

To evaluate the effects of CL on both acoustic and biomechanical features, three complementary statistical analyzes were conducted. For this purpose, each participant's acoustic measures and model–derived biomechanical trajectories were averaged within each sentence segment to obtain representative mean values for the congruent and incongruent conditions. Thus, each participant contributed up to six averaged utterances per condition and segment, from which all variables (f_o, SPL, a_{CT}, a_{TA}, P_S) were derived, although the exact number varied slightly due to preprocessing exclusions.

The first analysis examined the fixed and interaction effects of Group (VH vs. HC), Condition (congruent vs. incongruent), and Segment (Baseline, Effect, Rest) on acoustic and biomechanical features, including all two–way and three–way interactions (Group × Condition, Group × Segment, Condition × Segment, and Group × Condition × Segment). Separate LME models were fitted for each

variable, with Participant included as a random intercept to account for within–subject variability. Models were estimated using restricted maximum likelihood. Statistical significance was evaluated using $F-$tests for main and interaction effects, with $p-$values reported for each factor and a significance criterion set at $\alpha = 0.05$. Effect sizes were quantified using partial eta squared (η_p^2) and interpreted following Witte and Witte [197]: small ($\eta_p^2 \approx 0.01$),medium ($\eta_p^2 \approx 0.06$), and large ($\eta_p^2 \approx 0.14$).

The second analysis examined in which sentence segments and groups the condition differences were most pronounced. *Post hoc* pairwise contrasts were derived from the estimated marginal means (EMMs) of the LME models, comparing congruent and incongruent conditions separately for VH and HC participants. Tukey's adjustment was applied to control for multiple comparisons. Given the relatively large number of observations per participant and the limited number of subjects, asymptotic degrees of freedom were used, resulting in the computation of $z-$ratios instead of $t-$ratios. Effect sizes were reported as Cohen's d and interpreted according to standard conventions (small: $d \leq 0.50$; medium: $0.50 \leq d \leq 0.79$; large: $d \geq 0.80$) [168].

Finally, a third analysis examined whether speakers with VH exhibited greater acoustic and biomechanical changes under CL compared with HC. For each participant, the difference between incongruent and congruent trials (Δ = Incongruent$-$ Congruent) was computed for every sentence and then averaged within each seg-

ment, yielding one Δ value per segment per participant. Independent–samples $t-$tests were conducted to compare these Δ values between VH and HC speakers. Bonferroni–corrected $p-$values were reported, together with Cohen's d to quantify the magnitude of group differences. Positive Δ values indicate greater increases in VH relative to HC under CL.

Together, these three complementary analyzes provided a comprehensive assessment of the influence of CL on both acoustic trajectories and underlying laryngeal biomechanics, from global effects and interaction patterns to localized segment-specific contrasts and between-group comparisons.

5.3 Results

The results are presented in two complementary parts. The first section provides a descriptive analysis of the measured acoustic trajectories and the corresponding biomechanical responses estimated by the laryngeal model, illustrating how these variables evolve across task segments, conditions, and groups. The second section describes the statistical analyzes performed to assess the influence of CL on both acoustic and biomechanical domains and to evaluate potential effects related to group, condition, and segment.

5.3.1 Acoustic and model–derived biomechanical trajectories

Figure 5.3 illustrates the trajectories of the measured acoustic signals, the corresponding acoustic trajectories reproduced by the model, and the biomechanical responses of the laryngeal motor control model that best fit the observed data. The figure shows one HC (left panels) and one VH participant (right panels) performing the congruent (red) and incongruent (blue) conditions during the effect segment of the Stroop task.

The upper panels display the trajectories of $f_o(t)$ and SPL(t), with dashed lines representing the measured acoustic data and dotted lines representing the corresponding model estimations. The estimated trajectories closely followed the temporal course and amplitude of the measured signals, accurately reproducing the local increases and decreases observed in SPL and f_o across the task segment. The lower panels present the model–derived biomechanical parameter trajectories of $P_S(t)$, $a_{CT}(t)$ and $a_{TA}(t)$, which showed overall stable patterns with small dynamic fluctuations synchronized with the acoustic variations. Although only one participant per group is shown, the model exhibited consistent qualitative behavior across the remaining participants and task segments.

Extending this analysis to the group level, Figure 5.4 summarizes the general behavior of the acoustic measurement trajectories and the biomechanical

trajectories obtained from the laryngeal model simulations, averaged by segment, condition, and group. In the figure, blue and orange lines represent HC and VH, respectively, with lighter tones corresponding to the congruent condition and darker tones to the incongruent one.

The upper two rows correspond to the mean group trajectories of the measured acoustic outcomes (denoted as $\overline{f_o}(t)$ and $\overline{SPL}(t)$). Both measures show a gradual decline from Baseline to Rest, with variations in magnitude across groups and conditions. For instance, in $\overline{f_o}(t)$, HC showed higher values than VH under the congruent condition at Baseline, whereas under the incongruent condition this relationship reversed. During the Effect and Rest phases, this pattern did not persist, with f_o values generally higher under CL for both groups. The $\overline{SPL}(t)$ remained consistently higher for the incongruent condition in both groups, although the separation between the congruent and incongruent conditions was more pronounced in VH.

The lower rows display the biomechanical trajectories derived from the laryngeal model ($\overline{P_S}(t)$, $\overline{a_{CT}}(t)$, and $\overline{a_{TA}}(t)$). A close correspondence can be observed between acoustic and biomechanical components: the $\overline{a_{CT}}(t)$ trajectory mirrors the temporal course of $\overline{f_o}(t)$, consistent with the physiological role of the CT muscle in regulating VFs tension [198, 199] and pitch [200, 201], while variations in $\overline{P_S}(t)$ parallel those in $\overline{SPL}(t)$, reflecting the aerodynamic contribution of P_S to vocal intensity [202]. In contrast, $\overline{a_{TA}}(t)$ exhibits a locally antagonistic pattern

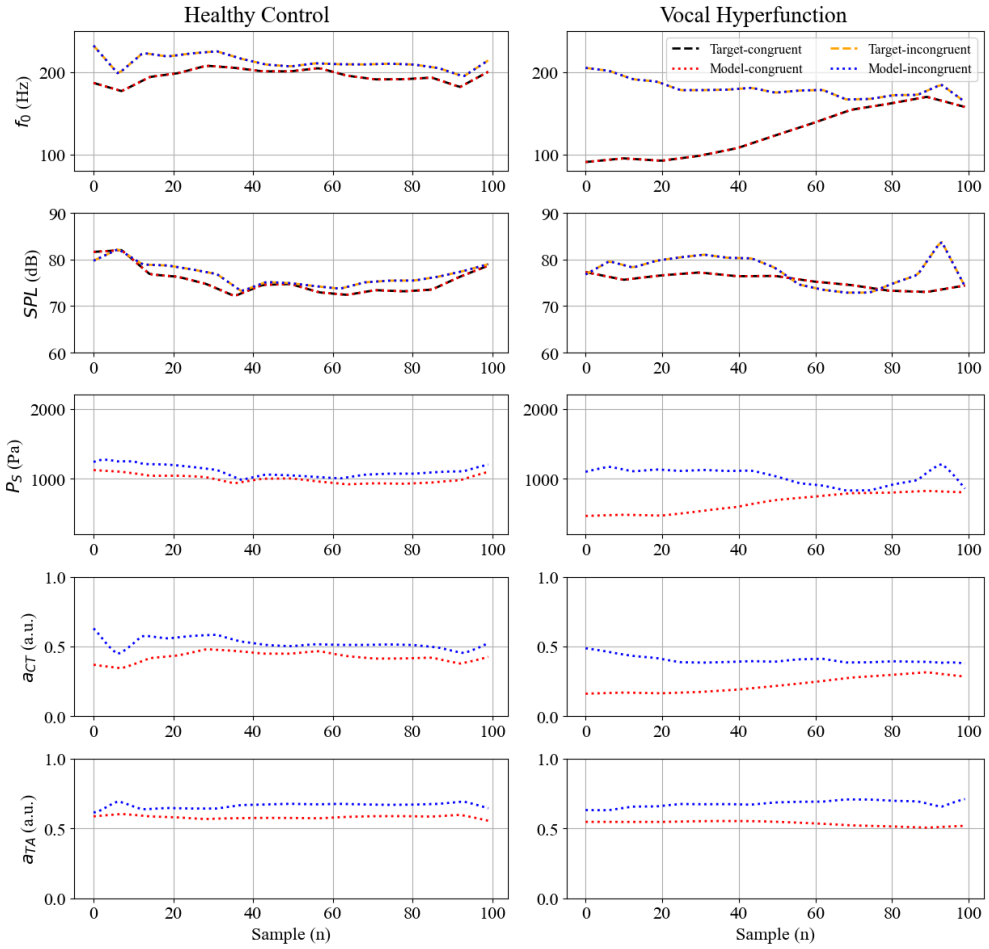


Figure 5.3: Trajectories of acoustic and biomechanical variables for one representative participant from each group (Healthy Control, left; Vocal Hyperfunctional, right) under the congruent (red) and incongruent (blue) conditions during the Effect segment of the Stroop task. The upper panels display the measured acoustic trajectories and those reproduced by the model ($f_o(t)$, SPL(t)), with dashed and dotted lines denoting measured and estimated signals, respectively. The lower panels show the corresponding biomechanical responses of the laryngeal control model ($P_S(t)$, $a_{CT}(t)$, $a_{TA}(t)$).

relative to $\overline{a_{CT}}(t)$: within each segment, its trajectory tends to vary in the opposite direction to $\overline{a_{CT}}(t)$, even though both muscles show an overall gradual decrease across the full utterance. This segment–wise opposition parallels the behavior observed in $\overline{f_o}(t)$ and reflects a progressive reduction in a_{CT}–a_{TA} co–activation as the task unfolds, consistent with the a_{TA}'s role in counteracting a_{CT}–driven elongation and tension of the VFs [203, 204].

The estimated trajectories closely followed the measured acoustic profiles across groups and segments, qualitatively reproducing the main acoustic variations with biomechanical responses consistent with the physiological principles of phonatory control reported in the literature. These results provide evidence that the model reproduces physiologically plausible control patterns, supporting its reliability for subsequent quantitative analyzes [205].

5.3.2 Statistical analysis of acoustic and biomechanical features

As described in Section 5.2.3, the statistical analysis of acoustic and biomechanical features was computed in three parts. Table 5.2 summarizes the results of the LME models, revealing distinct patterns across main effects and interactions. The main effect of Group was not significant for any acoustic or biomechanical measure ($p > .05$), indicating that, overall, the VH and HC groups did

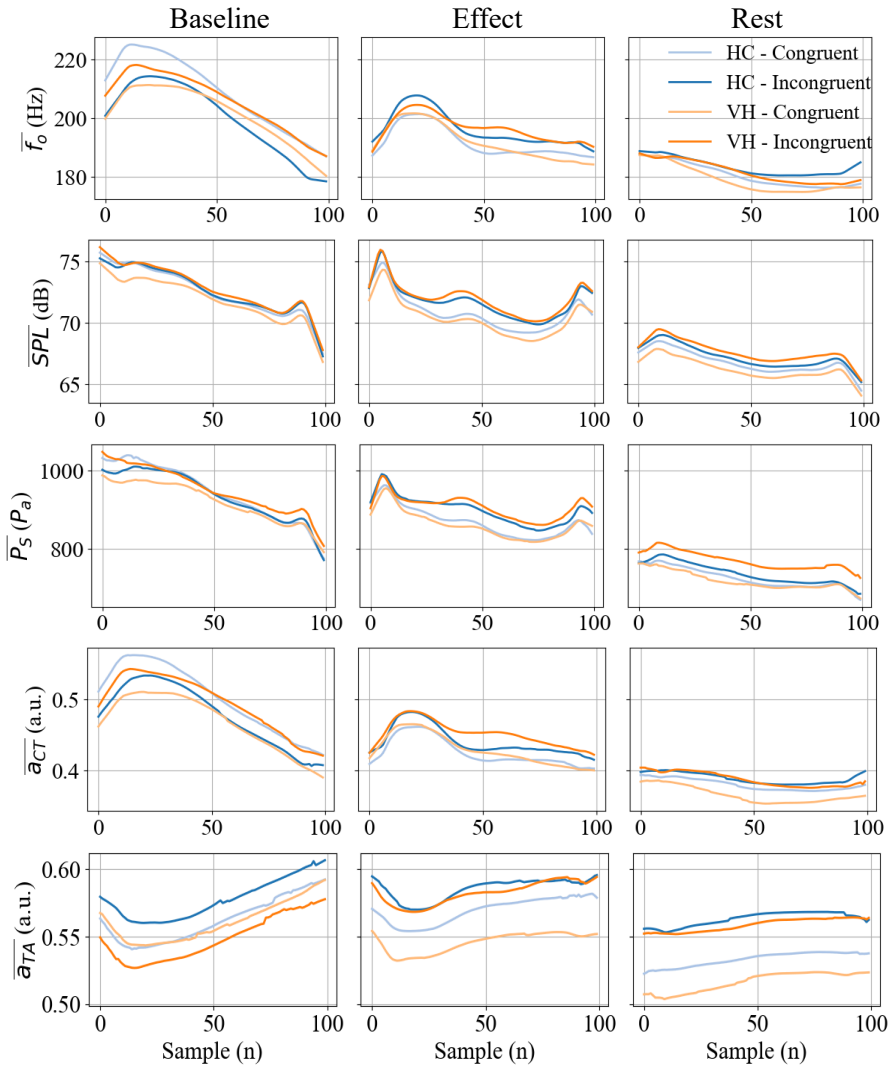


Figure 5.4: Average laryngeal motor control trajectories across segments and conditions. Each column corresponds to one segment, and each row represents either an acoustic measure (f_o or SPL) or a biomechanical parameter of the model ($a_{CT}(t)$, $a_{TA}(t)$, and $P_S(t)$). The colored lines within each panel represent the average trajectories of healthy controls and vocal hyperfunction participants under the congruent and incongruent conditions.

not differ statistically in their mean levels across tasks. In contrast, the main effect of Condition reached significance for all parameters ($p < .05$), with small effect sizes ($\eta_p^2 \approx .002 - .016$), suggesting consistent yet subtle differences between task conditions across measures. Both the Group and Condition effects are in line with previous findings by Dahl and Stepp [121]. The Segment factor also yielded robust effects for all variables ($p < .001$), with medium–to–large effect sizes ($\eta_p^2 = .149 - .622$), confirming systematic variability across speech segments and underscoring the importance of analyzing temporal structure within the Stroop task.

Table 5.2: Results of linear mixed–effects analyzes for acoustic and laryngeal biomechanical features.

Measure	**Term**	*df*	*F*	*p*	η_p^2	**Effect size**
	Group	1	0.007	0.935	–	–
	Condition	1	4.308	0.039	0.002	Small
	Segment	2	255.583	0.001	0.149	Medium
$f0_{Mean}$	Group × Condition	1	7.157	0.008	0.003	Small
	Group × Segment	2	0.405	0.668	–	–
	Condition × Segment	2	5.538	0.004	0.004	Small
	Group × Condition × Segment	2	6.999	0.001	0.005	Small
	Group	1	0.589	0.445	–	–
	Condition	1	173.734	0.001	0.057	Small
	Segment	2	2399.307	<0.001	0.622	Large
SPL_{Mean}	Group × Condition	1	41.015	0.001	0.014	Small
	Group × Segment	2	0.486	0.616	–	–
	Condition × Segment	2	9.220	0.001	0.007	Small

Table 5.2

Measure	Term	*df*	*F*	*p*	η_p^2	Effect size
	Group × Condition × Segment	2	0.222	0.802	–	–
	Group	1	0.011	0.920	–	–
	Condition	1	10.933	0.001	0.004	Small
	Segment	2	284.284	0.001	0.163	Medium
$a_{CT_{Mean}}$	Group × Condition	1	10.818	0.002	0.004	Small
	Group × Segment	2	2.038	0.131	–	–
	Condition × Segment	2	2.087	0.125	–	–
	Group × Condition × Segment	2	3.560	0.029	0.003	Small
	Group	1	0.267	0.607	–	–
	Condition	1	12.985	0.001	0.005	Small
	Segment	2	7.692	0.001	0.006	Small
$a_{TA_{Mean}}$	Group × Condition	1	0.022	0.883	–	–
	Group × Segment	2	0.054	0.949	–	–
	Condition × Segment	2	2.992	0.051	–	–
	Group × Condition × Segment	2	1.587	0.205	–	–
	Group	1	0.415	0.521	–	–
	Condition	1	44.804	0.001	0.016	Small
	Segment	2	857.323	0.001	0.370	Large
$P_{S_{Mean}}$	Group × Condition	1	15.116	0.001	0.006	Small
	Group × Segment	2	1.725	0.179	–	–
	Condition × Segment	2	3.754	0.024	0.003	Small
	Group × Condition × Segment	2	0.869	0.420	–	–

Across individual measures, f_o and SPL showed the largest effects for Condition and Segment, confirming that both factors consistently influenced the acoustic outcomes. In addition, small but significant interaction effects were observed, indicating that the influence of CL was not entirely uniform across groups or sentence

segments. The biomechanical parameters, such as a_{CT} and P_S, followed comparable patterns and closely paralleled those of f_o and SPL, respectively, reflecting the expected coupling between these biomechanical and acoustic domains. In contrast, a_{TA} showed smaller and less consistent variations, with significant main effects of Condition and Segment but no interactions.

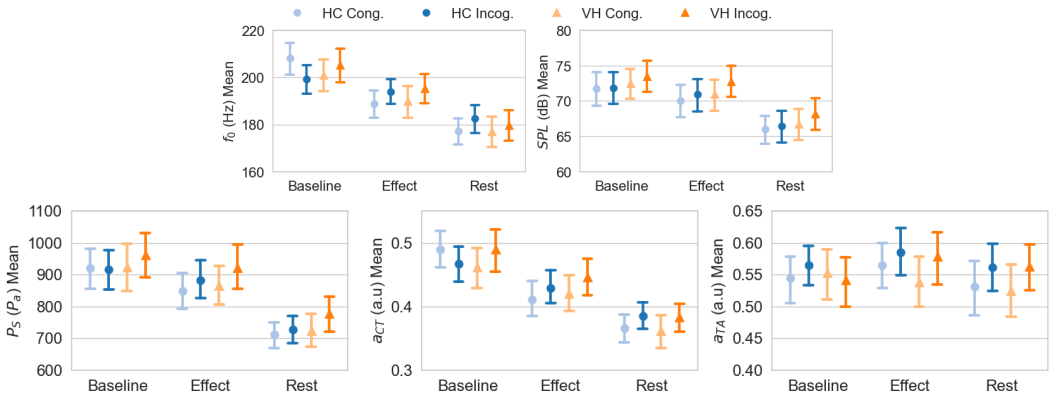


Figure 5.5: Estimated marginal means (EMMs) for acoustic and biomechanical parameters across sentence segments (Baseline, Effect, Rest) and experimental conditions (congruent, incongruent). The upper panel shows the acoustic outcomes (f_o, SPL), whereas the lower panel displays the model–derived biomechanical features (a_{CT} , P_S, a_{TA}). Both panels illustrate a general reduction in f_o, SPL, and P_S under the incongruent condition, particularly for speakers with vocal hyperfunction.

Overall, the LME results confirmed that the applied CL modulated both the acoustic outputs and their underlying biomechanical correlates, while group–related effects emerged primarily through interactions rather than global mean

differences. To visualize these patterns, Figure 5.5 provides a graphical complement to the LME results, displaying the EMMs for all acoustic and biomechanical parameters across segments, groups, and conditions. The figure highlights clear differences between the Baseline, Effect, and Rest segments across all measures, reflecting the strong Segment effects reported by the model. At the same time, subtle within–segment variations can be observed between the congruent and incongruent conditions, as well as between HC and VH speakers, supporting that the statistical differences observed between conditions depend primarily on group and Segment–related variability, as indicated by the LME analysis.

To further determine where these effects were concentrated, Table 5.3 summarizes the *post hoc* results, highlighting the sentence segments and groups in which condition–related differences were most evident. The results show that significant effects were present across all segments, although they were strongest and most consistent during the Effect phase—particularly among VH speakers. This pattern indicates that CL exerted its greatest influence during task execution, while smaller, but still noticeable, effects persisted during the Baseline and Rest phases.

Table 5.3: small Results of Post hoc contrasts (Congruent – Incongruent) derived from estimated marginal means of the linear mixed–effects model. analyzes were conducted on the interactions of interest (Condition × Segment and Group × Condition × Segment), allowing comparison of condition differences across groups (VH, HC) and segments (baseline, effect, rest).

Measure	**Segment**	**Group**	**Mean(SE)**	z	p	$d-cohen$	**Effect size**
$f_{0_{Mean}}$	Baseline	HC	8.762(2.058)	4.259	0.001	0.378	Small
		VH	-4.593(2.050)	-2.241	0.025	-0.198	Small
	Effect	HC	-4.422(2.058)	-2.149	0.032	-0.190	Small
		VH	-4.564(2.050)	-2.227	0.026	-0.196	Small
	Rest	HC	-2.830(2.058)	-1.375	0.169	–	–
		VH	-2.790(2.050)	-1.361	0.174	–	–
SPL_{Mean}	Baseline	HC	-0.102(0.168)	-0.612	0.540	–	–
		VH	-0.956(0.167)	-5.737	0.001	-0.507	Medium
	Effect	HC	-0.859(0.168)	-5.132	0.001	-0.455	Small
		VH	-1.632(0.167)	-9.792	0.001	-0.865	Large
	Rest	HC	-0.424(0.168)	-2.537	0.012	-0.225	Small
		VH	-1.418(0.167)	-8.504	0.001	-0.751	Medium
$a_{CT_{Mean}}$	Baseline	HC	0.024(0.009)	2.677	0.008	0.238	Small
		VH	-0.027(0.009)	-3.070	0.003	-0.271	Small
	Effect	HC	-0.015(0.009)	-1.692	0.091	–	–
		VH	-0.022(0.009)	-2.550	0.011	-0.225	Small
	Rest	HC	-0.008(0.009)	-1.005	0.315	–	–
		VH	-0.021(0.009)	-2.472	0.014	-0.218	Small
$a_{TA_{Mean}}$	Baseline	HC	-0.015(0.015)	-1.064	0.997	–	–
		VH	0.013(0.015)	0.844	1.000	–	–
	Effect	HC	-0.016(0.015)	-1.102	0.995	–	–
		VH	-0.036(0.015)	-2.536	0.317	–	–

Table 5.3

Measure	Segment	Group	Mean(SE)	z	p	$d-cohen$	Effect size
	Rest	HC	-0.029(0.015)	-2.057	0.654	–	–
		VH	-0.042(0.015)	-2.910	0.138	–	–
	Baseline	HC	7.285(10.089)	0.723	0.471	–	–
		VH	-31.463(10.050)	-3.130	0.002	-0.276	Small
$P_{S_{Mean}}$	Effect	HC	-30.324(10.089)	-3.005	0.003	-0.266	Small
		VH	-47.001(10.050)	-4.677	0.001	-0.413	Small
	Rest	HC	-11.558(10.089)	-1.145	0.252	–	–
		VH	-52.024(10.050)	-5.177	0.001	-0.457	Small

For the acoustic measures, f_o showed modest and variable effects. During Baseline, HC speakers produced lower f_o values under the incongruent condition ($z = 4.26, p = .001$), whereas VH speakers showed the opposite trend ($z = -2.24, p = .025$). In the Effect phase, both groups exhibited small but significant decreases in f_o ($p < .05$), with no differences during Rest. SPL, in contrast, showed consistent and robust increases under the incongruent condition, particularly in VH speakers, who exhibited the largest effects during the Effect segment ($p = .001$, $d = -0.87$). These increases were also significant during Baseline and Rest ($p = .001$), though with smaller magnitudes ($d = -0.51$ to -0.75). In HC speakers, SPL increased moderately in the Effect ($p = .001$, $d = -0.46$) and Rest ($p = .012$, $d = -0.23$) phases. Altogether, these results suggest that CL had a measurable influence on vocal intensity, most apparent during the active task segment and more noticeable in VH speakers.

Among the model–derived biomechanical parameters, a_{CT} and P_S once again exhibited clear parallels with their acoustic counterparts. Specifically, a_{CT} showed small but consistent condition related differences that mirrored the behavior of f_o: VH speakers displayed higher activation under the incongruent condition during the Effect and Rest phases ($p < .05$), whereas HC speakers showed a smaller opposite tendency during Baseline ($p = .008$). Similarly, P_S closely followed the SPL pattern, with strong and consistent increases under CL in VH speakers ($p \leq .003$ across all segments) and smaller effects in HC limited to the Effect phase ($p = .003$). In contrast, a_{TA} did not show significant condition effects in either group. This finding could suggest that a_{TA} was less influenced by CL. However, considering the laryngeal model used to derive a_{TA}, it might also reflect a not clear correlation with the acoustic features from which it was obtained [198, 206, 39].

The *post hoc* analysis suggests that the Effect segment elicited the strongest acoustic and biomechanical modulations, especially among VH speakers. The observed correspondence between SPL and P_S, and between f_o and a_{CT} , supports the view that the model–derived biomechanical features adequately represented the physiological mechanisms underlying the acoustic variations induced by CL.

To further examine whether the magnitude of these condition effects differed across segments and groups, Figure 5.6 illustrates the condition differences (expressed as Δ = Incongruent − Congruent) for both acoustic and biomechanical measures. For f_o and a_{CT} , the Δ magnitude for HC is lower than that for VH

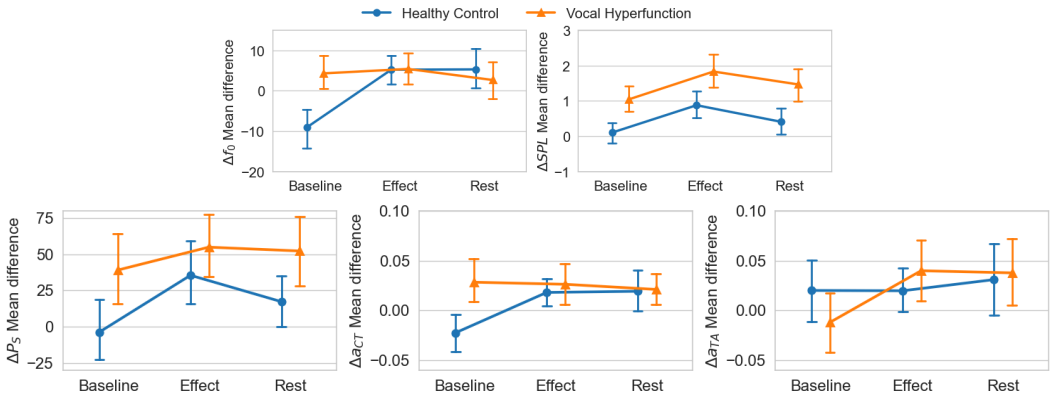


Figure 5.6: small Condition differences ($\Delta = Incongruent - Congruent$) for acoustic and biomechanical parameters across sentence segments (Baseline, Effect, Rest). The upper panel shows the mean Δ values for the acoustic outcomes (f_o, SPL), and the lower panel shows the corresponding differences for the model–derived biomechanical features (a_{CT} , P_S, a_{TA}). Positive Δ values indicate increases under CL, whereas negative values indicate reductions. Both panels reveal that the most pronounced effects occurred during the Effect segment, with VH speakers exhibiting larger increases in SPL and P_S, and smaller modulations in f_o and a_{CT} compared with HC.

during the Baseline segment. However, HC exhibits a pronounced increase from Baseline to Effect, reaching levels comparable to VH and maintaining similar values through the Rest phase. In contrast, VH shows a flatter pattern across segments, indicating less variation in their modulation by CL. SPL and P_S display greater magnitudes of change in VH compared with HC, both reaching their largest deviations during the Effect segment, consistent with the previously noted sensitivity of loudness–related measures to CL. Finally, as revealed in previous

analyzes, a_{TA} shows a less defined trend, with small and inconsistent variations across segments.

To reinforce this analysis, Table 5.4 presents the results of independent–samples $t-$tests comparing the condition differences (Δ = Incongruent − Congruent) between groups within each segment. Significant group differences were mainly observed during the Baseline segment, where VH exhibited larger Δ values than HC for most measures, including f_o ($p < .001$, $d = 0.73$), SPL ($p < .001$, $d = 0.70$), and a_{CT} ($p = .001$, $d = 0.62$). This indicates that, at the onset of the task, VH speakers showed stronger modulations associated with CL compared with HC participants. During the Effect segment, only SPL reached significance ($p = .009$, $d = 0.54$), suggesting a persistent but attenuated group difference as the task progressed. In the Rest phase, smaller yet consistent effects reappeared for SPL ($p = .002$, $d = 0.61$) and a tendency for P_S ($p = .061$), both showing larger Δ magnitudes in VH. In contrast, a_{TA} showed no significant group differences across segments.

The analysis of the magnitude differences between the congruent and incongruent conditions suggests that CL affected pitch, and loudness related measures differently, exhibiting distinct dynamics across segments. Group–related differences were most evident at the onset of the task across all parameters. However, as the task progressed, only measures associated with vocal intensity maintained consistent distinctions between the groups.

Table 5.4: Results of independent–samples t–tests comparing condition differences ($\Delta = Incongruent - Congruent$) between groups within each segment.

Measure	**Segment**	T	p	p_{corr}	$d - cohen$	
	Baseline	4.067	<0.001	<0.001	0.729	Medium
$\Delta f_{0_{Mean}}$	Effect	0.069	0.945	1.000	–	–
	Rest	-0.760	0.449	1.000	–	–
	Baseline	3.912	<0.001	<0.001	0.698	Medium
ΔSPL_{Mean}	Effect	3.031	0.003	0.009	0.541	Medium
	Rest	3.436	<0.001	0.002	0.614	Medium
	Baseline	3.458	0.001	0.002	0.617	Medium
$\Delta a_{CT_{Mean}}$	Effect	0.644	0.521	1.000	–	–
	Rest	0.136	0.892	1.000	–	–
	Baseline	-1.484	0.140	0.421	–	–
$\Delta a_{TA_{Mean}}$	Effect	1.029	0.306	0.917	–	–
	Rest	0.260	0.795	1.000	–	–
	Baseline	2.635	0.010	0.029	0.470	Small
$\Delta P_{S_{Mean}}$	Effect	1.217	0.226	0.677	–	–
	Rest	2.352	0.020	0.061	–	–

5.4 Discussion

The present study examined how CL influences vocal motor control by integrating acoustic measures with estimated biomechanical responses derived from a laryngeal motor control model. This approach was motivated by the established link between cognitive effort, autonomic activation, and vocal behavior, which suggests that changes in vocal output reflect underlying adjustments in laryn-

geal motor control [201]. By combining measured acoustic variables with model-derived biomechanical estimates, the study enables a physiologically interpretable analysis of how cognitive demands shape phonatory behavior. The findings align with previous acoustic evidence showing that CL modulates vocal intensity and laryngeal tension [26, 25, 194, 22, 24], and extend this evidence by demonstrating that such modulations are mirrored in estimated biomechanical control variables, with marked differences between HC and individuals with VH.

Segment-level analyses (see Table 5.4) revealed moderate group differences in modulation magnitude (incongruent–congruent) across both acoustic and model-derived biomechanical variables. In the baseline segment, an inverse pattern was observed between groups. Specifically, Figure 5.4 and 5.5 highlight distinct biomechanical strategies for managing cognitive demand. HC showed higher f_o and a_{CT} activation during congruent trials, whereas VH exhibited greater f_o and a_{CT} values under incongruent conditions. This pattern supports the notion that HC maintain a more flexible and efficient control strategy, relying on adaptive modulation of pitch and loudness to manage attentional demands. In contrast, VH appear to sustain elevated baseline laryngeal tension and reduced modulation capacity, consistent with the concept of accumulated cognitive and physiological load associated with chronic hyperfunction [207, 121, 74, 120].

Across subsequent segments, both groups exhibited increases in f_o and a_{CT} under incongruent conditions, accompanied by higher P_S and SPL values. These

effects are consistent with increased laryngeal muscle activation and aerodynamic drive under CL across participants. In line with previous work, increased cognitive demand has been associated with higher vocal intensity and with acoustic correlates of elevated laryngeal tension, including reductions in relative fundamental frequency, in both healthy and hyperfunctional speakers [28, 121]. In the present results, however, the magnitude of these load-related changes was greater in VH, suggesting a less adaptable control pattern (see Table 5.3).

At the physiological level, these patterns may reflect a redistribution of motor resources between the laryngeal and respiratory subsystems [192, 44, 208, 24]. In HC, such redistribution likely supports flexible pitch and loudness control through coordinated adjustments between subsystems. In contrast, in VH, the larger increases in SPL, model-derived P_S, and a_{CT} observed in Table 5.3 suggest the presence of overactivation or compensatory mechanisms [27, 209, 29, 210].

However, a_{TA} did not exhibit consistent modulation across conditions or groups (see Table 5.4). This pattern suggests that CL-related laryngeal activation does not manifest as a generalized co-contraction of intrinsic laryngeal muscles, but instead selectively engages control dimensions associated with pitch regulation and global output scaling. Although previous EMG studies have reported concurrent activation of multiple intrinsic and extrinsic laryngeal muscles under cognitive stress [27], the present model-based results indicate that not all activated muscles contribute equally to the functional modulation of acoustic output [199, 211].

Given the well-established association between a_{TA} and control of vocal intensity, the absence of systematic a_{TA} modulation suggests that arousal-related effects under cognitive load preferentially operate on more global or superficial control dimensions.

From a motor control perspective, these findings can be interpreted within neurocomputational frameworks of speech regulation that emphasize interactions between feedback-based corrections, and internal predictive models [190, 43]. In this context, cognitive control and vocal motor regulation rely on overlapping neural systems involved in attention, inhibition, and feedback monitoring [212, 213]. As CL increases, competition for these shared resources may lead to motor interference or dual-task cost effects mediated by physiological arousal mechanisms. Although autonomic responses were not directly measured in this study, the link between cognitive demand and autonomic arousal is well established [11]. Accordingly, the adaptive responses observed in HC suggest effective updating of internal models in response to increased cognitive demands, enabling flexible redistribution of control across subsystems. Conversely, the elevated model-derived P_S and reduced range of a_{CT} modulation observed in VH may represent an indirect manifestation of this interaction, whereby limited regulatory capacity amplifies arousal-driven effects on phonatory control, reinforcing habitual feedforward strategies and reducing sensitivity to contextual perturbations. This interpretation is consistent with previous reports of altered auditory–motor integration,

reduced capacity to update internal representations of vocal control, and atypical sensorimotor learning in individuals with VH [214, 215].

Taken together, these findings underscore the importance of considering cognitive and autonomic factors as integral components of vocal motor control in both healthy and hyperfunctional populations. The observed coupling between CL, autonomic arousal, and phonatory effort suggests that vocal behavior is highly context dependent, and that changes in task demands can elicit non-linear and subsystem-specific responses. From a clinical perspective, this has several implications. First, therapeutic tasks that increase cognitive complexity, such as dual-task paradigms, linguistic manipulation, or performance pressure may inadvertently alter vocal output by modulating respiratory drive and pitch-related control, even in the absence of overt changes in laryngeal posture. Such effects may be particularly pronounced in individuals with VH, who appear less able to adaptively redistribute effort across control dimensions.

Second, the results support therapeutic approaches that target not only vocal motor patterns, but also attentional and autonomic regulation. Interventions incorporating biofeedback, mindfulness-based strategies, or graded exposure to cognitively demanding tasks may help restore adaptive flexibility and prevent the consolidation of hyperfunctional control patterns [216, 217]. Importantly, these findings suggest that an optimal range of cognitive and autonomic activation may exist for vocal motor learning, beyond which performance and adaptability decline.

Finally, the integration of biomechanical modeling with acoustic analysis provides a promising avenue for objective, non-invasive assessment of vocal adaptability under cognitive stress. Model-based estimates of P_S and ILMs activation, combined with acoustic and autonomic markers, may enhance clinical characterization of VH and support personalized monitoring of treatment progress. More broadly, the sensitivity of vocal output to cognitive–motor interactions highlights the potential of the voice as a window into adaptive regulation, with relevance extending beyond functional voice disorders to neurological and cognitive conditions affecting speech motor control [26, 22, 218].

Although the present study provides new insights into the interaction between CL and vocal motor control, several limitations should be considered when interpreting the findings. First, the dataset was derived from a single experimental paradigm (the Stroop color-word task) with a fixed order of conditions. This design allowed clear contrasts between congruent and incongruent trials but also introduced potential carryover effects, as cognitive and vocal effort could accumulate across successive trials. Future studies should employ counterbalanced designs or include resting intervals to better isolate the effects of cognitive interference from those of task sequencing. Additionally, the sample size was relatively modest, and the groups were restricted to female speakers; thus, generalization to other populations, voice types, or communicative contexts should be approached with caution.

5.5 Chapter Conclusions

The findings presented in this chapter demonstrate the utility of a model–based trajectory estimation framework for characterizing the impact of CL on vocal motor control. By shifting the analytical focus from static acoustic descriptors to continuous biomechanical representations, the proposed approach enables inference of how changes in task–related arousal are reflected in the underlying organization of laryngeal motor commands. Within this framework, differences between healthy and hyperfunctional voices emerge not as isolated acoustic effects, but as distinct patterns of motor regulation across time.

The results indicate that, although both groups respond to increased cognitive demand, individuals with VH exhibit a reduced capacity for regulatory flexibility. In line with Hypothesis 2 of this thesis, the estimated trajectories for P_S and a_{CT}, together with limited and inconsistent modulation of a_{TA}, reveal exaggerated compensatory strategies in the VH group. These patterns suggest that task–related arousal acts upon an already constrained motor space, amplifying pre–existing imbalances in laryngeal motor variables rather than enabling adaptive redistribution of control.

This study further highlights the sensitivity of the TBCM combined with the inverse–Jacobian controller in capturing subtle, temporally structured physiological adjustments elicited by the Stroop task. The temporal coherence observed

between the inferred motor trajectories and the measured acoustic outputs supports the physiological plausibility of the framework, allowing task–related acoustic dynamics to be interpreted within a constrained biomechanical control space. While autonomic involvement is not directly measured and therefore cannot be asserted in causal terms, the resulting trajectory patterns are compatible with prior descriptions of voice modulation under increased cognitive and emotional demand.

At the same time, the presented estimates remain constrained by the simplified representation of the laryngeal plant and by the limited observability of certain intrinsic muscle activations, particularly a_{TA}. These limitations delineate clear avenues for future refinement of the biomechanical and control components of the model, while not detracting from its ability to reveal relative changes in motor coordination beyond the scope of traditional acoustic analyses.

In summary, by quantitatively linking task–related arousal with laryngeal motor coordination, this chapter provides a principled account of how cognitive processes modulate voice production dynamics. The findings support the view that VH reflects not only a local laryngeal dysfunction, but a broader reduction in the capacity to adapt vocal motor control to concurrent cognitive and emotional demands. This interpretation directly motivates the integrative framework developed in the final chapters of this thesis, where central, autonomic, and biomechanical mechanisms are considered jointly.

Chapter 6
Future work and conclusions

6.1 General conclusions

This thesis presented a unified framework for investigating vocal motor control by linking acoustic observations to physiologically interpretable biomechanical trajectories. In direct response to the overall aim of this work, the study successfully demonstrated that laryngeal motor control patterns estimated from acoustic speech features provide sensitive, interpretable, and discriminative markers of altered vocal function associated with neurological impairment and cognitive loading.

The realization of this aim relied on a foundational methodological contribution: the establishment and empirical validation of a robust forward biomechanical mapping for inverse estimation. Through a systematic evaluation of input–output configurations of the TBCM, this work demonstrated that a model with fixed adductory posture offers an optimal balance between physiological interpretability and numerical stability. This result provided the necessary reliability to address

the specific aims of the thesis across different populations and tasks.

Regarding the first specific aim, the analysis of PD phonation confirmed that estimated laryngeal motor control patterns differ systematically between speakers with PD and HC. These findings fully support Hypothesis 1. While average acoustic outputs were largely preserved in the PD group, the model–based analysis revealed a distinct biomechanical profile characterized by increased temporal instability in P_S and ILMs activations (a_{CT} and a_{TA}). Furthermore, as hypothesized, these biomechanically informed features significantly improved group discrimination compared to standard acoustic features alone. This validates the premise that the proposed framework captures latent motor regulation deficits, specifically, the inability to maintain stable motor execution, that are not fully resolved by surface–level acoustic descriptors.

Regarding the second specific aim, the investigation of CL effects on vocal function demonstrated that estimated motor patterns are indeed sensitive to increased cognitive demands, revealing distinct adaptation strategies between groups. The results support Hypothesis 2, showing that speakers with VH exhibit larger and more rigid changes in estimated laryngeal control compared to healthy speakers. While HC displayed flexible, task–dependent increases in laryngeal activation (interpretable as a normative adaptive response), individuals with VH responded with exaggerated and comparatively invariant control patterns. This confirms the hypothesis that autonomic modulation under cognitive load exacerbates maladap-

tive behaviors in an already constrained motor system, clarifying the distinction between adaptive physiological arousal and pathological vocal stiffness.

Taken together, the validation of both hypotheses confirms that model–based biomechanical trajectory estimation constitutes a powerful, non–invasive approach for investigating voice production. By shifting the analytical focus from acoustic outcomes to latent control variables, the proposed framework enables a principled differentiation between neurological instability and functional maladaptation, establishing a solid foundation for objective, physiology–informed approaches to clinical voice assessment and hypothesis–driven modeling of vocal motor control.

6.2 Future work

This thesis has focused on the characterization of vocal motor control through a physiologically inspired trajectory estimation framework. By bridging non–invasive acoustic signals with latent biomechanical control variables, the proposed approach provides a functional methodology for probing central motor impairments and autonomic influences on voice production. Given the complexity of speech motor control and its modulation by cognitive and physiological factors, several avenues for methodological refinement, physiological validation, and clinical translation naturally emerge from this work.

A first and critical direction for future research concerns the physiological vali-

dation of the estimated biomechanical trajectories. Although the forward model is grounded in the physics of the TBCM [42] and was systematically selected based on numerical stability (Chapter 3), the inferred control variables are derived from acoustic targets rather than from direct physiological measurements. Importantly, the TBCM and related modeling approaches have demonstrated strong agreement with experimental and in vivo observations across multiple studies [42, 156, 219], providing a physiologically constrained and experimentally supported approximation of laryngeal motor control.

Nevertheless, future work should explicitly validate the estimated P_S against aerodynamic measurements obtained through intraoral pressure interpolation or mask–based airflow recordings. Similarly, inferred ILMs activation (a_{CT} and a_{TA}) should be compared against EMG recordings, where available. While indirect estimation techniques for aerodynamic and muscular parameters have been proposed [33, 34], obtaining ground–truth measurements typically requires invasive procedures such as tracheal puncture or intramuscular EMG [36, 211, 35]. Systematic validation against such measurements would clarify whether the framework captures absolute physiological levels or primarily reflects relative control dynamics and coordination patterns, which may be sufficient for clinical characterization.

Building on the findings of Chapter 4, which focused on sustained phonation, a natural extension of this work is the application of the framework to continuous speech and diadochokinetic tasks. As demonstrated in Chapter 5, the proposed

methodology is capable of tracking dynamic biomechanical changes during connected speech under varying cognitive load. Extending this analysis to PD may reveal impairments in motor planning, articulatory–laryngeal coordination, and temporal regulation that remain hidden in quasi–stationary phonatory tasks.

From a clinical perspective, future studies should also assess the longitudinal sensitivity of biomechanical markers derived from the framework. Evaluating their responsiveness to dopaminergic medication states, disease progression, or behavioral voice therapies would help determine their potential as objective markers for monitoring disease evolution and treatment effects.

From a modeling standpoint, an important direction concerns the representation of laryngeal adduction. While fixing the adductory posture (a_{LCA}) enabled robust and stable trajectory estimation for modal phonation, this simplification limits the analysis of conditions involving glottal insufficiency, excessive medial compression, or task–dependent changes in vocal quality. Allowing controlled, time–varying adduction would provide a more realistic representation of connected speech and would be essential for capturing phenomena such as breathy or pressed voice, compensatory mechanisms, and context–dependent respiratory–laryngeal drive. This extension is particularly relevant in light of the cognitive load findings, where autonomic arousal was associated with changes in phonation and hyperfunctional behaviors, suggesting that adductory control may play a central role in linking cognitive state, autonomic regulation, and phonatory output.

A further promising avenue lies in the integration of the biomechanical plant into neurocomputational models of speech production. At present, the framework estimates peripheral motor control variables but does not explicitly model neural control processes. Embedding the TBCM–based estimator within architectures such as LaDIVA or FACTS would enable the simulation of altered sensory feedback gains, transmission delays, or increased motor noise associated with neurological disorders [43, 134]. Such integration would strengthen the mechanistic link between observed vocal instability, inferred biomechanical variability, and specific neural dysfunctions.

Finally, an important and necessary extension of the framework involves the development of subject–specific and sex–specific biomechanical models. The current implementation relies on a reference male laryngeal configuration, which favors numerical robustness but does not explicitly account for inter–speaker anatomical variability or well–documented sex–related differences in vocal fold geometry, tissue composition, and vibratory regimes. Future work could incorporate speaker–specific scaling of vocal fold length, mass, and stiffness based on anthropometric data or acoustic proxies, enabling more personalized trajectory estimation.

In particular, the development of forward models grounded in female laryngeal biology is essential for improving physiological interpretability and clinical relevance in mixed–sex populations. Female phonation is characterized by higher

fundamental frequencies, different body–cover ratios, and increased susceptibility to vocal hyperfunction and certain voice disorders [126, 220]. Incorporating sex–specific biomechanical parameterizations, or adaptive model selection strategies, would also directly connect with the findings of Chapter 5, where autonomic activation and vocal hyperfunction were shown to modulate phonatory behavior under cognitive load. Together, these developments would allow the framework to disentangle disease–related motor instability, cognitive–autonomic influences, and normative anatomical variability within a unified physiological modeling approach.

Bibliography

[1] Jesús A. Parra, Clara Sorolla, Nicolás F. Quinteros, Emiro J. Ibarra, Gabriel A. Alzamendi, Sean D. Peterson, Hasini R. Weerathunge, Frank H. Guenther, and Matías Zañartu. Modeling laryngeal biomechanical control of acoustic parameters using machine learning regressors. Manuscript submitted for publication, 2026.

[2] Yunzhi Wu, Yuqi Dong, Yunqi Tang, Weiran Wang, Yulong Bo, and Cui Zhang. Relationship between motor performance and cortical activity of older neurological disorder patients with dyskinesia using fnirs: a systematic review. *Frontiers in Physiology*, 14:1153469, 2023.

[3] Daniel S Peterson, Geetanjali Gera, Fay B Horak, and Brett W Fling. Corpus callosum structural integrity is associated with postural control improvement in persons with multiple sclerosis who have minimal disability. *Neurorehabilitation and neural repair*, 31(4):343–353, 2017.

[4] Gaetano Valenza, Francesco Di Ciò, Nicola Toschi, and Riccardo Barbieri. Sympathetic and parasympathetic central autonomic networks. *Imaging Neuroscience*, 2:1–17, 2024.

[5] Mahlon R DeLong and Thomas Wichmann. Circuits and circuit disorders of the basal ganglia. *Archives of neurology*, 64(1):20–24, 2007.

[6] Hermann Ackermann and Wolfram Ziegler. Brain mechanisms underlying speech motor control. *The handbook of phonetic sciences*, 2:202–250, 2010.

[7] Shimon Sapir. 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(4): 1330–1343, 2014.

[8] Julian F Thayer and Richard D Lane. A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of affective disorders*, 61 (3):201–216, 2000.

[9] Bruce S McEwen. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiological reviews*, 87(3):873–904, 2007.

[10] J Jankovic. Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(4):368–376, 2008. doi: 10.1136/jnnp.2007.131045.

[11] Mark F Bear, Barry W Connors, and Michael A Paradiso. *Neuroscience (Vol. 2)*. Lippincott Williams & Wilkins, 2007.

[12] D. Roman-Liu, I. Grabarek, P. Bartuzi, and W. Choromański. The influence of mental load on muscle tension. *Ergonomics*, 56(7):1125–1133, 2013. doi: 10.1080/00140139.2013.798429.

[13] Jukka Koskelo, Aleksi Lehmusaho, Tomi P. Laitinen, Juha E. K. Hartikainen, Taija M. M. Lahtinen, Tuomo K. Leino, and Kerttu Huttunen. Cardiac autonomic responses in relation to cognitive workload during simulated military flight. *Applied Ergonomics*, 121:104370, 2024. doi: 10.1016/j.apergo.2024.104370.

[14] Antonio Luque-Casado, José C. Perales, David Cárdenas, and Daniel Sanabria. Heart rate variability and cognitive processing: The autonomic response to task demands. *Biological Psychology*, 113:83–90, 2016. doi: 10.1016/j.biopsycho.2015.11.013.

[15] Xiaoni Wang, Binbin Liu, Lin Xie, Xiaolin Yu, Mengjun Li, and Jianbao Zhang. Cerebral and neural regulation of cardiovascular activity during mental stress. *BioMedical Engineering OnLine*, 15(2):160, 2016. doi: 10.1186/s12938-016-0255-1.

[16] Alexander J. Ehrenberg, Ayesha Khatun, Emma Coomans, Matthew J. Betts, Federica Capraro, Elisabeth H. Thijssen, Konstantin Senkevich, Tehmina Bharucha, Mehrsa Jafarpour, Peter N. E. Young, William Jagust, Stephen F. Carter, Tammaryn Lashley, Lea T. Grinberg, Joana B.

Pereira, Niklas Mattsson-Carlgren, Nicholas J. Ashton, Jörg Hanrieder, Henrik Zetterberg, Michael Schöll, and Ross W. Paterson. Relevance of biomarkers across different neurodegenerative diseases. *Alzheimer's Research & Therapy*, 12(1):56, 2020. doi: 10.1186/s13195-020-00601-w.

[17] Dorota Koníčková, Kateřina Menšíková, Lucie Tučková, Eva Hényková, Miroslav Strnad, David Friedecký, David Stejskal, Radoslav Matěj, and Petr Kaňovský. Biomarkers of Neurodegenerative Diseases: Biology, Taxonomy, Clinical Relevance, and Current Research Status. *Biomedicines*, 10(7):1760, 2022. doi: 10.3390/biomedicines10071760.

[18] Paul Ayres, Joy Yeonjoo Lee, Fred Paas, and Jeroen J. G. Van Merriënboer. The Validity of Physiological Measures to Identify Differences in Intrinsic Cognitive Load. *Frontiers in Psychology*, 12:702538, 2021. doi: 10.3389/fpsyg.2021.702538.

[19] Frank H. Guenther. *Neural Control of Speech*. The MIT Press, 2016. doi: 10.7551/mitpress/10471.001.0001.

[20] Uwe Jürgens. Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews*, 26(2):235–258, 2002.

[21] V Shibina and TM Thasleema. Acoustic signal based diagnosis of neurodegenerative parkinson's disease through machine learning approach: A

review. In *2022 IEEE 19th India Council International Conference (INDICON)*, pages 1–8. IEEE, 2022.

[22] Elizabeth Mahon and Margie E Lachman. Voice biomarkers as indicators of cognitive changes in middle and later adulthood. *Neurobiology of aging*, 119:22–35, 2022.

[23] Gemma Moya-Galé, Jennifer Spielman, Lorraine A Ramig, Luca Campanelli, and Youri Maryn. The acoustic voice quality index (avqi) in people with parkinson's disease before and after intensive voice and articulation therapies: Secondary outcome of a randomized controlled trial. *Journal of voice*, 38(6):1529–e7, 2024.

[24] Brittany L. Perrine and Ronald C. Scherer. Aerodynamic and acoustic voice measures before and after an acute public speaking stressor. *Journal of Speech, Language, and Hearing Research*, 63(10):3311–3325, 2020. doi: 10.1044/2020_JSLHR-19-00252. URL `https://pubs.asha.org/doi/abs/10.1044/2020_JSLHR-19-00252`.

[25] Hayley S Arnold, Megan K MacPherson, and Anne Smith. Autonomic correlates of speech versus nonspeech tasks in children and adults. *Journal of Speech, Language, and Hearing Research*, 57(4):1296–1307, 2014.

[26] Defne Abur, Megan K MacPherson, Adrianna C Shembel, and Cara E

Stepp. Acoustic measures of voice and physiologic measures of autonomic arousal during speech as a function of cognitive load in older adults. *Journal of Voice*, 37(2):194–202, 2023.

[27] Leah B. Helou, Wei Wang, Robin C. Ashmore, Clark A. Rosen, and Katherine Verdolini Abbott. Intrinsic laryngeal muscle activity in response to autonomic nervous system activation. *The Laryngoscope*, 123(11):2756–2765, 2013. doi: 10.1002/lary.24109.

[28] Kimberly L. Dahl and Cara E. Stepp. Changes in Relative Fundamental Frequency Under Increased Cognitive Load in Individuals with Healthy Voices. *Journal of Speech, Language, and Hearing Research*, 64(4):1189–1196, 2021. doi: 10.1044/2021_JSLHR-20-00134.

[29] Nicole E. Tomassi, Dea M. Turashvili, Alyssa Williams, Bridget Walsh, Emily P. Stephen, and Cara E. Stepp. Investigating cognitive load and autonomic arousal during voice production and vocal auditory-motor adaptation. *Journal of Speech, Language, and Hearing Research*, 68(4):1634–1653, 2025. doi: 10.1044/2024_JSLHR-24-00601. URL `https://pubs.asha.org/doi/abs/10.1044/2024_JSLHR-24-00601`.

[30] Robert E. Hillman, Cara E. Stepp, Jarrad H. Van Stan, Matías Zañartu, and Daryush D. Mehta. An updated theoretical framework for vocal hyper-

function. *American Journal of Speech-Language Pathology*, 29(4):2254–2260, 2020.

[31] Vikram Ramanarayanan, Adam C. Lammert, Hannah P. Rowe, Thomas F. Quatieri, and Jordan R. Green. Speech as a Biomarker: Opportunities, Interpretability, and Challenges. *Perspectives of the ASHA Special Interest Groups*, 7(1):276–283, 2022. doi: 10.1044/2021_PERSP-21-00174.

[32] John M. Tracy, Yasin Özkanca, David C. Atkins, and Reza Hosseini Ghomi. Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease. *Journal of Biomedical Informatics*, 104: 103362, 2020. doi: 10.1016/j.jbi.2019.103362.

[33] Martin Rothenberg. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *The Journal of the Acoustical Society of America*, 53(6):1632–1645, 1973.

[34] Ingo R. Titze and Johan Sundberg. Vocal intensity in speakers and singers. *The Journal of the Acoustical Society of America*, 91(5):2936–2946, 1992.

[35] Varun Varadarajan, Joel H. Blumin, and Jonathan M. Bock. State of the Art of Laryngeal Electromyography. *Current Otorhinolaryngology Reports*, 1(3):171–177, 2013.

[36] John M. Heinz and Kenneth N. Stevens. On the derivation of area functions

and acoustic spectra from cinéradiographic films of speech. *The Journal of the Acoustical Society of America*, 36(5):1037–1038, May 1964.

[37] Zhaoyan Zhang. Estimation of vocal fold physiology from voice acoustics using machine learning. *The Journal of the Acoustical Society of America*, 147(3):EL264–EL270, 2020.

[38] Emiro J. Ibarra, Jesús A. Parra, Gabriel A. Alzamendi, Juan P. Cortés, Víctor M. Espinoza, Daryush D. Mehta, Robert E. Hillman, and Matías Zañartu. Estimation of subglottal pressure, vocal fold collision pressure, and intrinsic laryngeal muscle activation from neck-surface vibration using a neural network framework and a voice production model. *Frontiers in Physiology*, 12, 2021.

[39] Joaquín Sepúlveda, Jesús A. Parra, Emiro J. Ibarra, Mauricio Araya, Patricio De La Cuadra, and Matías Zañartu. Estimation of physiological vocal features from neck surface acceleration signals using probabilistic bayesian neural networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 33:1576–1589, 2025. doi: 10.1109/TASLPRO.2025.3552938.

[40] Emiro J Ibarra, Julián D Arias-Londoño, Juan I Godino-Llorente, Daryush D Mehta, and Matías Zañartu. Subject-specific modeling by domain adaptation for the estimation of subglottal pressure from neck-surface

acceleration signals. *Biomedical Signal Processing and Control*, 106:107681, 2025.

[41] Jesús A. Parra, Emiro J. Ibarra, Carlos Calvache, Jarrad H. Van Stan, Robert E. Hillman, and Matías Zañartu. Estimating the pathophysiology of phonotraumatic vocal hyperfunction using ambulatory data and a computational model. *Journal of Speech, Language, and Hearing Research*, 68(3):949–962, 2025. doi: 10.1044/2024_JSLHR-24-00419. URL https://pubs.asha.org/doi/abs/10.1044/2024_JSLHR-24-00419.

[42] Gabriel A. Alzamendi, Sean D. Peterson, Byron D. Erath, Robert E. Hillman, and Matías Zañartu. Triangular body-cover model of the vocal folds with coordinated activation of the five intrinsic laryngeal muscles. *The Journal of the Acoustical Society of America*, 151(1):17–30, 2022. doi: 10.1121/10.0009169.

[43] Hasini R Weerathunge, Gabriel A Alzamendi, Gabriel J Cler, Frank H Guenther, Cara E Stepp, and Matías Zañartu. Ladiva: A neurocomputational model providing laryngeal motor control for speech acquisition and production. *PLOS Computational Biology*, 18(6):e1010159, 2022.

[44] Megan K MacPherson, Defne Abur, and Cara E Stepp. Acoustic measures of voice and physiologic measures of autonomic arousal during speech as a function of cognitive load. *Journal of Voice*, 31(4):504–e1, 2017.

[45] Jennifer Oates and Alison Winkworth. Current knowledge, controversies and future directions in hyperfunctional voice disorders. *International Journal of Speech-Language Pathology*, 10(4):267–277, 2008.

[46] Juan Rafael Orozco-Arroyave, Julián David Arias-Londoño, Jesús Francisco Vargas-Bonilla, María Claudia Gonzalez-Rátiva, and Elmar Nöth. New spanish speech corpus database for the analysis of people suffering from parkinson's disease. In *Lrec*, pages 342–347, 2014.

[47] Janaína Mendes-Laureano, Jorge A Gómez-García, Alejandro Guerrero-López, Elisa Luque-Buzo, Julián D Arias-Londoño, Francisco J Grandas-Pérez, and Juan I Godino-Llorente. Neurovoz: a castillian spanish corpus of parkinsonian speech. *Scientific Data*, 11(1):1367, 2024.

[48] Minoru Hirano, Yuki Kakita, Koichi Ohmaru, and Shigejiro Kurita. Structure and mechanical properties of the vocal fold. In *Speech and Language*, volume 7, pages 271–297. Elsevier, 1982.

[49] Gunnar Fant. *Acoustic Theory of Speech Production: With Calculations Based on X-Ray Studies of Russian Articulations*. Walter de Gruyter, 1971.

[50] Edward Flemming. Lecture 4: Source-filter model of speech production. MIT OpenCourseWare: 24.915 Linguistic Phonetics, 2015. URL `https://ocw.mit.edu/courses/24-915-linguistic-phonetics-fall-2015/`

resources/mit24_915f15_lec4/. Accessed: 2026-02-10. License: Creative Commons BY-NC-SA.

[51] Peter Birkholz, Bernd J Kröger, and Christiane Neuschaefer-Rube. Synthesis of breathy, normal, and pressed phonation using a two-mass model with a triangular glottis. In *Interspeech*, pages 2681–2684, 2011.

[52] Ingo R Titze and Daniel W Martin. *Principles of voice production*. Acoustical Society of America, 1998.

[53] Steven D Gray, Fariborz Alipour, Ingo R Titze, and Thomas Hale Hammond. Biomechanical and histologic observations of vocal fold fibrous proteins. *Annals of Otology, Rhinology & Laryngology*, 109(1):77–85, 2000.

[54] Marie D. Dubois, Martin Ch, and Jean M. Prades. Cellular, extracellular and histological analysis of the vocal folds: correlation between structure and function. *Revue de Laryngologie-Otologie-Rhinologie*, 128(5):267–271, 2007.

[55] Robert T Sataloff, Farhad Chowdhury, Joel E Portnoy, Mary J Hawkshaw, and Shruti Joglekar. *Surgical Techniques in Otolaryngology-Head & Neck Surgery: Laryngeal Surgery*. JP Medical Ltd, 2013.

[56] K Faaborg Andersen and Ahti Sonninen. The function of the extrinsic

laryngeal muscles at different pitch. *Acta Oto-Laryngologica*, 51(1-2):89–93, 1960.

[57] Brad H Story. Mechanisms of voice production. *The handbook of speech production*, pages 34–58, 2015.

[58] Uwe Jürgens. The neural control of vocalization in mammals: a review. *Journal of Voice*, 23(1):1–10, 2009.

[59] Christy L Ludlow. Central nervous system control of the laryngeal muscles in humans. *Respiratory physiology & neurobiology*, 147(2-3):205–222, 2005.

[60] Pamela J Davis, Shi Ping Zhang, Alison Winkworth, and Richard Bandler. Neural control of vocalization: Respiratory and emotional influences. *Journal of Voice*, 10(1):23–38, 1996.

[61] John F Houde and Srikantan S Nagarajan. Speech production as state feedback control. *Frontiers in human neuroscience*, 5:82, 2011.

[62] Peter Indefrey. The spatial and temporal signatures of word production components: a critical update. *Frontiers in psychology*, 2:255, 2011.

[63] F-Xavier Alario, Hanna Chainay, Stéphane Lehericy, and Laurent Cohen. The role of the supplementary motor area (sma) in word production. *Brain research*, 1076(1):129–143, 2006.

[64] Benjamin K Dichter, Jonathan D Breshears, Matthew K Leonard, and Edward F Chang. The control of vocal pitch in human laryngeal motor cortex. *Cell*, 174(1):21–31, 2018.

[65] Ann M Graybiel. The basal ganglia: learning new tricks and loving it. *Current opinion in neurobiology*, 15(6):638–644, 2005.

[66] Hermann Ackermann and Ingo Hertrich. The contribution of the cerebellum to speech processing. *Journal of Neurolinguistics*, 13(2-3):95–116, 2000.

[67] Joseph J Paton and Dean V Buonomano. The neural basis of timing: distributed mechanisms for diverse functions. *Neuron*, 98(4):687–705, 2018.

[68] Hanricus Gerardus Jacobus Maria Kuypers. Corticobulbar connexions to the pons and lower brain-stem in man: an anatomical study. *Brain*, 81(3): 364–388, 1958.

[69] Christy L Ludlow. Central nervous system control of voice and swallowing. *Journal of Clinical Neurophysiology*, 32(4):294–303, 2015.

[70] Susan Standring, Harold Ellis, J Healy, D Johnson, A Williams, P Collins, and C Wigley. Gray's anatomy: the anatomical basis of clinical practice. *American journal of neuroradiology*, 26(10):2703, 2005.

[71] Jason A. Tourville and Frank H. Guenther. The DIVA model: A neural the-

ory of speech acquisition and production. *Language and Cognitive Processes*, 26(7):952–981, 2011. doi: 10.1080/01690960903498424.

[72] John F Houde and Michael I Jordan. Sensorimotor adaptation in speech production. *Science*, 279(5354):1213–1216, 1998.

[73] Joseph S Perkell. Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of neurolinguistics*, 25(5):382–407, 2012.

[74] Maria Dietrich, Katherine Verdolini Abbott, Jackie Gartner-Schmidt, and Clark A Rosen. The frequency of perceived stress, anxiety, and depression in patients with common pathologies affecting voice. *Journal of voice*, 22 (4):472–488, 2008.

[75] Maria Dietrich, Richard D Andreatta, Yang Jiang, Ashwini Joshi, and Joseph C Stemple. Preliminary findings on the relation between the personality trait of stress reaction and the central neural control of human vocalization. *International journal of speech-language pathology*, 14(4):377–389, 2012.

[76] Rebecca C. Brown, Alan H. Lockwood, and Babasaheb R. Sonawane. Neurodegenerative Diseases: An Overview of Environmental Risk Fac-

tors. *Environmental Health Perspectives*, 113(9):1250–1256, 2005. doi: 10.1289/ehp.7567.

[77] Raúl Martínez-Fernández., Carmen Gasca-Salas C., Álvaro Sánchez-Ferro, and José Ángel Obeso. Actualización en la Enfermedad de Parkinson. *Revista Médica Clínica Las Condes*, 27(3):363–379, 2016. doi: 10.1016/j.rmclc.2016.06.010.

[78] Christopher G. Goetz, Barbara C. Tilley, Stephanie R. Shaftman, Glenn T. Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B. Stern, Richard Dodel, Bruno Dubois, Robert Holloway, Joseph Jankovic, Jaime Kulisevsky, Anthony E. Lang, Andrew Lees, Sue Leurgans, Peter A. LeWitt, David Nyenhuis, C. Warren Olanow, Olivier Rascol, Anette Schrag, Jeanne A. Teresi, Jacobus J. Van Hilten, and Nancy LaPelle. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders*, 23(15):2129–2170, 2008. doi: 10.1002/mds.22340.

[79] Andrew Ma, Kenneth K Lau, and Dominic Thyagarajan. Voice changes in Parkinson's disease: What are they telling us? *Journal of Clinical Neuroscience*, 72:1–7, 2020. doi: 10.1016/j.jocn.2019.12.029.

[80] Jeffery A Jones and Kevin G Munhall. Remapping auditory-motor representations in voice production. *Current biology*, 15(19):1768–1772, 2005.

[81] Christian Castro, Pavel Prado, Víctor M Espinoza, Alba Testart, Daphne Marfull, Rodrigo Manriquez, Cara E Stepp, Daryush D Mehta, Robert E Hillman, and Matías Zañartu. Lombard effect in individuals with nonphonotraumatic vocal hyperfunction: Impact on acoustic, aerodynamic, and vocal fold vibratory parameters. *Journal of Speech, Language, and Hearing Research*, 65(8):2881–2895, 2022.

[82] Yongxue Li, Mingdan Tan, Hao Fan, Emily Q Wang, Ling Chen, Jingting Li, Xi Chen, and Hanjun Liu. Neurobehavioral effects of lsvt® loud on auditory-vocal integration in parkinson's disease: A preliminary study. *Frontiers in neuroscience*, 15:624801, 2021.

[83] Defne Abur, Rosemary A Lester-Smith, Ayoub Daliri, Ashling A Lupiani, Frank H Guenther, and Cara E Stepp. Sensorimotor adaptation of voice fundamental frequency in parkinson's disease. *PloS one*, 13(1):e0191839, 2018.

[84] Christiane Arnold, Johannes Gehrig, Suzana Gispert, Carola Seifried, and Christian A Kell. Pathomechanisms and compensatory efforts related to parkinsonian speech. *NeuroImage: Clinical*, 4:82–97, 2014.

[85] Xi Chen, Xiaoxia Zhu, Emily Q Wang, Ling Chen, Weifeng Li, Zhaocong Chen, and Hanjun Liu. Sensorimotor control of vocal pitch production in parkinson's disease. *Brain research*, 1527:99–107, 2013.

[86] Henry Railo, Niklas Nokelainen, Saara Savolainen, and Valtteri Kaasinen. Deficits in monitoring self-produced speech in parkinson's disease. *Clinical Neurophysiology*, 131(9):2140–2147, 2020.

[87] Xiyan Huang, Xi Chen, Nan Yan, Jeffery A Jones, Emily Q Wang, Ling Chen, Zhiqiang Guo, Weifeng Li, Peng Liu, and Hanjun Liu. The impact of parkinson's disease on the cortical mechanisms that support auditory–motor integration for voice control. *Human brain mapping*, 37(12):4248–4261, 2016.

[88] Xiyan Huang, Hao Fan, Jingting Li, Jeffery A Jones, Emily Q Wang, Ling Chen, Xi Chen, and Hanjun Liu. External cueing facilitates auditory-motor integration for speech control in individuals with parkinson's disease. *Neurobiology of aging*, 76:96–105, 2019.

[89] Patricia Garrido-Vásquez, Marc D Pell, Silke Paulmann, Karl Strecker, Johannes Schwarz, and Sonja A Kotz. An erp study of vocal emotion processing in asymmetric parkinson's disease. *Social cognitive and affective neuroscience*, 8(8):918–927, 2013.

[90] Damien Benis, Claire Haegelen, Philippe Voruz, Jordan Pierce, Valérie

Milesi, Jean-François Houvenaghel, Marc Verin, Paul Sauleau, Didier Grandjean, and Julie Péron. Subthalamic nucleus oscillations during vocal emotion processing are dependent of the motor asymmetry of parkinson's disease. *Neuroimage*, 222:117215, 2020.

[91] Julie Péron, Sezen Cekic, Claire Haegelen, Paul Sauleau, Sona Patel, Dominique Drapier, Marc Vérin, and Didier Grandjean. Sensory contribution to vocal emotion deficit in parkinson's disease after subthalamic stimulation. *cortex*, 63:172–183, 2015.

[92] Patrizia Vizza, Giuseppe Tradigo, Domenico Mirarchi, Roberto Bruno Bossio, Nicola Lombardo, Gennarina Arabia, Aldo Quattrone, and Pierangelo Veltri. Methodologies of speech analysis for neurodegenerative diseases evaluation. *International Journal of Medical Informatics*, 122:45–54, 2019. doi: 10.1016/j.ijmedinf.2018.11.008.

[93] Benjamin G. Schultz, Venkata S. Aditya Tarigoppula, Gustavo Noffs, Sandra Rojas, Anneke Van Der Walt, David B. Grayden, and Adam P. Vogel. Automatic speech recognition in neurodegenerative disease. *International Journal of Speech Technology*, 24(3):771–779, 2021. doi: 10.1007/s10772-021-09836-w.

[94] Quoc Cuong Ngo, Mohammod Abdul Motin, Nemuel Daniel Pah, Peter Drotár, Peter Kempster, and Dinesh Kumar. Computerized analysis of

speech and voice for Parkinson's disease: A systematic review. *Computer Methods and Programs in Biomedicine*, 226:107133, 2022. doi: 10.1016/j.cmpb.2022.107133.

[95] Phillip A Low. Autonomic nervous system function. *Journal of Clinical Neurophysiology*, 10(1):14–27, 1993.

[96] Max J Hilz and Matthias Dütsch. Quantitative studies of autonomic function. *Muscle & Nerve: Official journal of the american association of electrodiagnostic medicine*, 33(1):6–20, 2006.

[97] Michael G Ziegler. Psychological stress and the autonomic nervous system. In *Primer on the autonomic nervous system*, pages 291–293. Elsevier, 2012.

[98] Edwin W Taylor, David Jordan, and John H Coote. Central control of the cardiovascular and respiratory systems and their interactions in vertebrates. *Physiological reviews*, 79(3):855–916, 1999.

[99] Richard Gordan, Judith K Gwathmey, and Lai-Hua Xie. Autonomic and endocrine control of cardiovascular function. *World journal of cardiology*, 7 (4):204, 2015.

[100] Robert W Hamill and Robert E Shapiro. Peripheral autonomic nervous system. In *Primer on the autonomic nervous system*, pages 20–28. Elsevier, 2004.

[101] Connor Liston, Bruce S McEwen, and BJ Casey. Psychosocial stress reversibly disrupts prefrontal processing and attentional control. *Proceedings of the National Academy of Sciences*, 106(3):912–917, 2009.

[102] Maria Dietrich and Katherine Verdolini Abbott. Vocal function in introverts and extraverts during a psychological stress reactivity protocol. *Journal of Speech, Language, and Hearing Research*, 55(5):1344–1361, 2012. doi: 10.1044/1092-4388(2012/11-0160).

[103] Harald G Wallbott and Klaus R Scherer. Stress specificities: Differential effects of coping style, gender, and type of stressor on autonomic arousal, facial expression, and subjective feeling. *Journal of Personality and Social Psychology*, 61(1):147, 1991.

[104] Brandon N Kyle and Daniel W McNeil. Autonomic arousal and experimentally induced pain: a critical review of the literature. *Pain Research and Management*, 19(3):159–167, 2014.

[105] Mara Mather and Matthew R Sutherland. Arousal-biased competition in perception and memory. *Perspectives on psychological science*, 6(2):114–133, 2011.

[106] Amy FT Arnsten. Stress signalling pathways that impair prefrontal cortex structure and function. *Nature reviews neuroscience*, 10(6):410–422, 2009.

[107] Hugo D Critchley. Neural mechanisms of autonomic, affective, and cognitive integration. *Journal of comparative neurology*, 493(1):154–166, 2005.

[108] Jessica K Alexander, Ashleigh Hillier, Ryan M Smith, Madalina E Tivarus, and David Q Beversdorf. Beta-adrenergic modulation of cognitive flexibility during stress. *Journal of cognitive neuroscience*, 19(3):468–478, 2007.

[109] William Lovallo. The cold pressor test and autonomic function: a review and integration. *Psychophysiology*, 12(3):268–282, 1975.

[110] Peter H Ellaway, Annapoorna Kuppuswamy, Alessia Nicotra, and Christopher J Mathias. Sweat production and the sympathetic skin response: Improving the clinical assessment of autonomic function. *Autonomic Neuroscience*, 155(1-2):109–114, 2010.

[111] Hugo D Critchley, Rebecca Elliott, Christopher J Mathias, and Raymond J Dolan. Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. *Journal of Neuroscience*, 20(8):3033–3040, 2000.

[112] David M Sletten, Guillermo A Suarez, Phillip A Low, Jay Mandrekar, and Wolfgang Singer. Compass 31: a refined and abbreviated composite autonomic symptom score. In *Mayo Clinic Proceedings*, volume 87, pages 1196–1201. Elsevier, 2012.

[113] Wolfram Boucsein. *Electrodermal activity*. Springer science & business media, 2012.

[114] Jennifer Kleinow and Anne Smith. Potential interactions among linguistic, autonomic, and motor factors in speech. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 48 (4):275–287, 2006.

[115] Leah B Helou, J Richard Jennings, Clark A Rosen, Wei Wang, and Katherine Verdolini Abbott. Intrinsic laryngeal muscle response to a public speech preparation stressor: Personality and autonomic predictors. *Journal of Speech, Language, and Hearing Research*, 63(9):2940–2951, 2020.

[116] Christopher Dromey and Erin Shim. The Effects of Divided Attention on Speech Motor, Verbal Fluency, and Manual Task Performance. *Journal of Speech, Language, and Hearing Research*, 51(5):1171–1182, 2008. doi: 10.1044/1092-4388(2008/06-0221).

[117] Jörg Peters, Marina Frank, and Tio Rohloff. Effects of Cognitive Load on Vocal Fold Vibratory Patterns in Bilingual Speakers of Low and High German. *Journal of Voice*, page S089219972300293X, 2023. doi: 10.1016/j.jvoice.2023.09.016.

[118] Scott E. Lively, David B. Pisoni, W. Van Summers, and Robert H. Bernacki.

Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *The Journal of the Acoustical Society of America*, 93(5):2962–2973, 05 1993. ISSN 0001-4966. doi: 10.1121/1.405815. URL https://doi.org/10.1121/1.405815.

[119] Lize Demmink-Geertman and Philippe Henri Dejonckere. Nonorganic habitual dysphonia and autonomic dysfunction. *Journal of Voice*, 16(4):549–559, 2002.

[120] Evelyne Van Houtte, Kristiane Van Lierde, and Sofie Claeys. Pathophysiology and treatment of muscle tension dysphonia: a review of the current knowledge. *Journal of Voice*, 25(2):202–207, 2011.

[121] Kimberly L. Dahl and Cara E. Stepp. Effects of Cognitive Stress on Voice Acoustics in Individuals with Hyperfunctional Voice Disorders. *American Journal of Speech-Language Pathology*, 32(1):264–274, 2023. doi: 10.1044/2022_AJSLP-22-00204.

[122] Robert E. Hillman, Eva B. Holmberg, Joseph S. Perkell, Michael Walsh, and Charles Vaughan. Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech, Language, and Hearing Research*, 32(2):373–392, 1989.

[123] Nelson Roy, Ray M. Merrill, Steven D. Gray, and Elaine M. Smith. Voice

disorders in the general population: Prevalence, risk factors, and occupational impact. *The Laryngoscope*, 115(11):1988–1995, 2005.

[124] Nancy Pearl Solomon. Vocal fatigue and its relation to vocal hyperfunction. *International Journal of Speech-Language Pathology*, 10(4):254–266, 2008.

[125] Kristina Simonyan and Barry Horwitz. Laryngeal Motor Cortex and Control of Speech in Humans. *The Neuroscientist*, 17(2):197–208, 2011. doi: 10.1177/1073858410386727.

[126] Ingo R. Titze. *The Myoelastic Aerodynamic Theory of Phonation*. National Center for Voice and Speech, 2006.

[127] Cara E. Stepp, Robert E. Hillman, and James T. Heaton. Modulation of neck intermuscular Beta coherence during voice and speech production. *Journal of Speech, Language, and Hearing Research*, 54(3):836–844, 2011.

[128] Pablo Gomez, Anne Schutzenberger, Marion Semmler, and Michael Dollinger. Laryngeal pressure estimation with a recurrent neural network. *IEEE Journal of Translational Engineering in Health and Medicine*, 7:1–11, 2019.

[129] Ingo R Titze and Brad H Story. Rules for controlling low-dimensional vocal fold models with muscle activation. *The Journal of the Acoustical Society of America*, 112(3):1064–1076, 2002.

[130] Benjamin Parrell, Adam C. Lammert, Gregory Ciccarelli, and Thomas F. Quatieri. Current models of speech motor control: A control-theoretic overview of architectures and properties. *The Journal of the Acoustical Society of America*, 145(3):1456–1481, 2019. doi: 10.1121/1.5092807.

[131] Lorenzo Sciavicco and Bruno Siciliano. *Modelling and control of robot manipulators*. Springer Science & Business Media, 2012.

[132] Arati S Deo and Ian D Walker. Overview of damped least-squares methods for inverse kinematics of robot manipulators. *Journal of Intelligent and Robotic Systems*, 14(1):43–68, 1995.

[133] J Lin, CC Lin, and H-S Lo. Pseudo-inverse jacobian control with grey relational analysis for robot manipulators mounted on oscillatory bases. *Journal of Sound and Vibration*, 326(3-5):421–437, 2009.

[134] Benjamin Parrell, Vikram Ramanarayanan, Srikantan S Nagarajan, and John F Houde. Facts: A hierarchical task-based control model of speech incorporating sensory feedback. In *Interspeech*, pages 1497–1501, 2018.

[135] Shinji Maeda. A digital simulation method of the vocal-tract system. *Speech Communication*, 1(3-4):199–229, 1982.

[136] Benjamin Parrell, Vikram Ramanarayanan, Srikantan Nagarajan, and John

Houde. The facts model of speech motor control: Fusing state estimation and task-based control. *PLoS computational biology*, 15(9):e1007321, 2019.

[137] Khalil Iskarous, Louis Goldstein, Douglas H Whalen, Mark Tiede, and Philip Rubin. Casy: The haskins configurable articulatory synthesizer. In *International Congress of Phonetic Sciences, Barcelona, Spain*, pages 185–188, 2003.

[138] Matías Zañartu, Gabriel E. Galindo, Byron D. Erath, Sean D. Peterson, George R. Wodicka, and Robert E. Hillman. Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction. *The Journal of the Acoustical Society of America*, 136(6): 3262–3271, 2014.

[139] Brad H. Story and Ingo R. Titze. Voice simulation with a body-cover model of the vocal folds. *The Journal of the Acoustical Society of America*, 97(2): 1249–1260, 1995.

[140] Jon Z Lin, Víctor M Espinoza, Katherine L Marks, Matías Zañartu, and Daryush D Mehta. Improved subglottal pressure estimation from neck-surface vibration in healthy speakers producing non-modal phonation. *IEEE Journal of Selected Topics in Signal Processing*, 14(2):449–460, 2019.

[141] Katherine L Marks, Jonathan Z Lin, James A Burns, Tiffiny A Hron,

Robert E Hillman, and Daryush D Mehta. Estimation of subglottal pressure from neck surface vibration in patients with voice disorders. *Journal of Speech, Language, and Hearing Research*, 63(7):2202–2218, 2020.

[142] Pedro Gómez-Vilda, Roberto Fernández-Baillo, Alberto Nieto, F Díaz, Francisco J Fernández-Camacho, Victoria Rodellar, Agustín Álvarez, and R Martínez. Evaluation of voice pathology based on the estimation of vocal fold biomechanical parameters. *Journal of Voice*, 21(4):450–476, 2007.

[143] Jonas Donhauser, Bogac Tur, and Michael Döllinger. Neural network-based estimation of biomechanical vocal fold parameters. *Frontiers in Physiology*, 15:1282574, 2024.

[144] Paavo Alku. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. *Speech communication*, 11(2-3):109–118, 1992.

[145] Manu Airaksinen, Tuomo Raitio, Brad Story, and Paavo Alku. Quasi closed phase glottal inverse filtering analysis with weighted linear prediction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22 (3):596–607, 2013.

[146] Rita R Patel, Shaheen N Awan, Julie Barkmeier-Kraemer, Mark Courey, Dimitar Deliyski, Tanya Eadie, Diane Paul, Jan G Švec, and Robert Hillman. Recommended protocols for instrumental assessment of voice: Ameri-

can speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function. *American journal of speech-language pathology*, 27(3):887–905, 2018.

[147] Ingo R Titze and Eric J Hunter. A two-dimensional biomechanical model of vocal fold posturing. *The Journal of the Acoustical Society of America*, 121(4):2254–2260, 2007.

[148] James Hillenbrand, Ronald A Cleveland, and Robert L Erickson. Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, 37(4):769–778, 1994.

[149] Rubén Fraile and Juan Ignacio Godino-Llorente. Cepstral peak prominence: A comprehensive analysis. *Biomedical Signal Processing and Control*, 14:42–54, 2014.

[150] Pascal Hecker, Nico Steckhan, Florian Eyben, Björn W. Schuller, and Bert Arnrich. Voice Analysis for Neurological Disorder Recognition–A Systematic Review and Perspective on Emerging Trends. *Frontiers in Digital Health*, 4:842301, 2022. doi: 10.3389/fdgth.2022.842301.

[151] Benjamin Schultz, Zaher Joukhadar, Maria Del Mar Quiroga, Usha Nattala, Gustavo Noffs, Sandra Rojas, Hannah Reece, Anneke Van Der Walt, and

Adam Vogel. The classification of neurodegenerative disease from acoustic speech data, 2021.

[152] Yu Chen Tai, Paucar G. Bryan, Francis Loayza, and Enrique Peláez. A voice analysis approach for recognizing Parkinson's disease patterns. *IFAC-PapersOnLine*, 54(15):382–387, 2021. doi: 10.1016/j.ifacol.2021.10.286.

[153] Ibrahim Karabayir, Samuel M. Goldman, Suguna Pappu, and Oguz Akbilgic. Gradient boosting for Parkinson's disease diagnosis from voice recordings. *BMC Medical Informatics and Decision Making*, 20(1):228, 2020. doi: 10.1186/s12911-020-01250-7.

[154] Ilias Tougui, Abdelilah Jilbab, and Jamal El Mhamdi. Machine Learning Smart System for Parkinson Disease Classification Using the Voice as a Biomarker. *Healthcare Informatics Research*, 28(3):210–221, 2022. doi: 10.4258/hir.2022.28.3.210.

[155] Anu Iyer, Aaron Kemp, Yasir Rahmatallah, Lakshmi Pillai, Aliyah Glover, Fred Prior, Linda Larson-Prior, and Tuhin Virmani. A machine learning method to process voice samples for identification of Parkinson's disease. *Scientific Reports*, 13(1):20615, 2023. doi: 10.1038/s41598-023-47568-w.

[156] Emiro J. Ibarra, Julián D. Arias-Londoño, Matías Zañartu, and Juan I. Godino-Llorente. Towards a Corpus (and Language)-Independent Screening

of Parkinson's Disease from Voice and Speech through Domain Adaptation. *Bioengineering*, 10(11):1316, 2023. doi: 10.3390/bioengineering10111316.

[157] Hadeel Ahmed Abd El Aal, Shereen A Taie, and Nashwa El-Bendary. An optimized rnn-lstm approach for parkinson's disease early detection using speech features. *Bulletin of Electrical Engineering and Informatics*, 10(5): 2503–2512, 2021.

[158] James R Williamson, Thomas F Quatieri, Brian S Helfer, Joseph Perricone, Satrajit S Ghosh, Gregory A Ciccarelli, and Daryush D Mehta. Segment-dependent dynamics in predicting parkinson's disease. In *Interspeech*, pages 518–522, 2015.

[159] Gregory Ciccarelli, Thomas F. Quatieri, and Satrajit S. Ghosh. Neurophysiological vocal source modeling for biomarkers of disease. In *Proceedings of the 17th Annual Conference of the International Speech Communication Association (INTERSPEECH 2016)*, pages 1541–1545, San Francisco, CA, USA, 2016.

[160] Thomas Francis Quatieri, Gregory Alan Ciccarelli, Satrajit S Ghosh, Christopher J Smalt, James R Williamson, Jeffrey Shane Palmer, et al. Assessing disorders through speech and a computational model, 2018. US Patent 10,127,929.

[161] Elaine Kearney, Alfonso Nieto-Castañón, Hasini R Weerathunge, Riccardo Falsini, Ayoub Daliri, Defne Abur, Kirrie J Ballard, Soo-Eun Chang, Sara-Ching Chao, Elizabeth S Heller Murray, et al. A simple 3-parameter model for examining adaptation in speech and voice production. *Frontiers in psychology*, 10:2995, 2020.

[162] N. D. Pah, M. A. Motin, and D. K. Kumar. Phonemes based detection of parkinson's disease for telehealth applications. *Scientific Reports*, 12(1): 9687, 2022.

[163] Paul Boersma and David Weenink. Praat: Doing phonetics by computer, 2015. URL `http://www.praat.org`.

[164] Paul Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam*, volume 17, pages 97–110, 1993.

[165] Jan G. Švec and Svante Granqvist. Tutorial and guidelines on measurement of sound pressure level in voice and speech. *Journal of Speech, Language, and Hearing Research*, 61(3):441–461, 2018. doi: 10.1044/2017_JSLHR-S-17-0095. URL `https://pubs.asha.org/doi/abs/10.1044/2017_JSLHR-S-17-0095`.

[166] Markus Iseli and Abeer Alwan. An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. In *2004 IEEE international conference on acoustics, speech, and signal processing*, volume 1, pages I–669. IEEE, 2004.

[167] Lloyd S Nelson. The anderson-darling test for normality. *Journal of Quality Technology*, 30(3):298–299, 1998.

[168] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Erlbaum, Hillsdale, NJ, 2nd ed. edition, 1988.

[169] Jeanine Romano, Jeffrey D Kromrey, Jesse Coraggio, and Jeff Skowronek. Appropriate statistics for ordinal level data: Should we really be using t-test and cohen'sd for evaluating group differences on the nsse and other surveys. In *annual meeting of the Florida Association of Institutional Research*, volume 177, 2006.

[170] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[171] Antonio Suppa, Giovanni Costantini, Francesco Asci, Pietro Di Leo, Mohammad Sami Al-Wardat, Giulia Di Lazzaro, Simona Scalise, Antonio Pisani, and Giovanni Saggio. Voice in Parkinson's Disease: A Machine

Learning Study. *Frontiers in Neurology*, 13:831428, 2022. doi: 10.3389/fneur.2022.831428.

[172] Igor Kononenko. Estimating attributes: Analysis and extensions of relief. In *European conference on machine learning*, pages 171–182. Springer, 1994.

[173] Marko Robnik-Šikonja and Igor Kononenko. Theoretical and empirical analysis of relieff and rrelieff. *Machine learning*, 53(1):23–69, 2003.

[174] Athanasios Tsanas, Max A Little, Patrick E McSharry, and Lorraine O Ramig. Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 59(5):1264–1271, 2012.

[175] C Okan Sakar, Gorkem Serbes, Ali Gunduz, Huseyin C Tunc, Hatice Nizam, Betul E Sakar, Melih Tutuncu, Tarkan Aydin, M Erdem Isenkul, and Hulya Apaydin. A comparative analysis of speech signal processing algorithms for parkinson's disease classification. *IEEE Access*, 7:22111–22118, 2019.

[176] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[177] Youri Maryn, Marc De Bodt, and Nelson Roy. The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders. *Journal of communication disorders*, 43(3):161–174, 2010.

[178] Olivia Murton, Robert Hillman, and Daryush Mehta. Cepstral peak prominence values for clinical voice evaluation. *American Journal of Speech-Language Pathology*, 29(3):1596–1607, 2020.

[179] Jules Fumel, Delphine Bahuaud, Ethan Weed, Riccardo Fusaroli, and Anahita Basirat. A systematic review and bayesian meta-analysis of acoustic measures of prosody in parkinson's disease. *Journal of Speech, Language, and Hearing Research*, 67(8):2548–2564, 2024.

[180] Shu Yang, Fengbo Wang, Liqiong Yang, Fan Xu, Man Luo, Xiaqing Chen, Xixi Feng, and Xianwei Zou. The physical significance of acoustic parameters and its clinical significance of dysarthria in parkinson's disease. *Scientific Reports*, 10(1):11776, 2020.

[181] Rita Chiaramonte and Marco Bonfiglio. Acoustic analysis of voice in parkinson's disease: a systematic review of voice disability and meta-analysis of studies. *Rev Neurol*, 70(11):393–405, 2020.

[182] Yasuhiro Tanaka, Masaki Nishio, and Seiji Niimi. Vocal acoustic characteristics of patients with parkinson's disease. *Folia Phoniatrica et logopaedica*, 63(5):223–230, 2011.

[183] Sabine Skodda, Wenke Visser, and Uwe Schlegel. Vowel articulation in parkinson's disease. *Journal of voice*, 25(4):467–472, 2011.

[184] Alexander M Goberman and Carl Coelho. Acoustic analysis of parkinsonian speech i: Speech characteristics and l-dopa therapy. *NeuroRehabilitation*, 17 (3):237–246, 2002.

[185] Jan Rusz, Roman Cmejla, Hana Ruzickova, and Evzen Ruzicka. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease. *The journal of the Acoustical Society of America*, 129(1):350–367, 2011.

[186] Máté Hireš, Matej Gazda, Peter Drotár, Nemuel Daniel Pah, Mohammod Abdul Motin, and Dinesh Kant Kumar. Convolutional neural network ensemble for Parkinson's disease detection from voice recordings. *Computers in Biology and Medicine*, 141:105021, 2022. doi: 10.1016/j.compbiomed.2021.105021.

[187] John Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 1988. doi: 10.1207/s15516709cog1202_4.

[188] Christopher G. Goetz and Glenn T. Stebbins. *Textbook of Clinical Neurology*. Saunders/Elsevier, Philadelphia, PA, third edition edition, 2007. ISBN 978-1416036180.

[189] J Ridley Stroop. Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6):643—662, 1935.

[190] Frank H Guenther. *Neural control of speech*. Mit Press, 2016.

[191] Susanne Fuchs and Amélie Rochet-Capellan. The respiratory foundations of spoken language. *Annual Review of Linguistics*, 7(7):13–30, 2021. ISSN 2333-9691. doi: 10.1146/annurev-linguistics-031720-103907. URL https://www.annualreviews.org/content/journals/10.1146/annurev-linguistics-031720-103907.

[192] Stanislas Boyer, Pierre-Vincent Paubel, Robert Ruiz, Radouane El Yagoubi, and Agnès Daurat. Human voice as a measure of mental load level. *Journal of Speech, Language, and Hearing Research*, 61(11):2722–2734, 2018. doi: 10.1044/2018_JSLHR-S-18-0066. URL https://pubs.asha.org/doi/abs/10.1044/2018_JSLHR-S-18-0066.

[193] Megan K. MacPherson. Cognitive load affects speech motor performance differently in older and younger adults. *Journal of Speech, Language, and Hearing Research*, 62(5):1258–1277, 2019. doi: 10.1044/2018_JSLHR-S-17-0222. URL https://pubs.asha.org/doi/abs/10.1044/2018_JSLHR-S-17-0222.

[194] Hans Rutger Bosker, Eva Reinisch, and Matthias J. Sjerps. Cognitive load makes speech sound fast, but does not modulate acoustic context effects. *Journal of Memory and Language*, 94:166–176, 2017. ISSN 0749-

596X. doi: https://doi.org/10.1016/j.jml.2016.12.002. URL https://www.sciencedirect.com/science/article/pii/S0749596X16302492.

[195] Selma Özaydın. Examination of Energy Based Voice Activity Detection Algorithms for Noisy Speech Signals. *Avrupa Bilim ve Teknoloji Dergisi*, pages 157–163, 2019. doi: 10.31590/ejosat.637741.

[196] Yan Zhang, Zhen-min Tang, Yan-ping Li, and Yang Luo. A Hierarchical Framework Approach for Voice Activity Detection and Speech Enhancement. *The Scientific World Journal*, 2014:723643, 2014. doi: 10.1155/2014/723643.

[197] Robert S. Witte and John S. Witte. *Statistics*. Wiley, 11th ed. edition, 2016.

[198] Dinesh K Chhetri, Juergen Neubauer, and Elazar Sofer. Influence of asymmetric recurrent laryngeal nerve stimulation on vibration, acoustics, and aerodynamics. *The Laryngoscope*, 124(11):2544–2550, 2014.

[199] Josué D. Martínez, Emiro J. Ibarra, Jesús A. Parra, Daryush D. Mehta, James T. Heaton, Robert E. Hillman, Michal J. Plocienniczak, Jameson C. Cooper, and Matías Zañartu. Toward acoustic-based normalization of laryngeal emg for improved interspeaker consistency in muscle-to-acoustic mapping. *Journal of Voice*, 2025. ISSN 0892-1997. doi: https://doi.

org/10.1016/j.jvoice.2025.08.014. URL https://www.sciencedirect.com/science/article/pii/S0892199725003339.

[200] Bernard Roubeau, Claude Chevrie-Muller, and Jean Lacau Saint Guily. Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta oto-laryngologica*, 117(3):459–464, 1997. doi: https://doi.org/10.3109/00016489709113421.

[201] Dinesh K Chhetri and Soo Jin Park. Interactions of subglottal pressure and neuromuscular activation on fundamental frequency and intensity. *The Laryngoscope*, 126(5):1123–1130, 2016.

[202] Staffan Björklund and Johan Sundberg. Relationship between subglottal pressure and sound pressure level in untrained voices. *Journal of Voice*, 30(1):15–20, 2016. ISSN 0892-1997. doi: https://doi.org/10.1016/j.jvoice.2015.03.006. URL https://www.sciencedirect.com/science/article/pii/S089219971500048X.

[203] Karen Ann Kochis-Jennings, Eileen M Finnegan, Henry T Hoffman, Sanyukta Jaiswal, and Darcey Hull. Cricothyroid muscle and thyroarytenoid muscle dominance in vocal register control: preliminary results. *Journal of Voice*, 28(5):652–e21, 2014. doi: https://doi.org/10.1016/j.jvoice.2014.01.017.

[204] Minoru Hirano, William Vennard, and John Ohala. Regulation of register, pitch and intensity of voice: An electromyographic investigation of intrinsic laryngeal muscles. *Folia phoniatrica et logopaedica*, 22(1):1–20, 1970. doi: https://doi.org/10.1159/000263363.

[205] Gabriel A Alzamendi, Rodrigo Manríquez, Paul J Hadwin, Jonathan J Deng, Sean D Peterson, Byron D Erath, Daryush D Mehta, Robert E Hillman, and Matías Zañartu. Bayesian estimation of vocal function measures using laryngeal high-speed videoendoscopy and glottal airflow estimates: An in vivo case study. *The Journal of the Acoustical Society of America*, 147 (5):EL434–EL439, 2020.

[206] Hye Rhyn Chung, Yoonjeong Lee, Neha K Reddy, Zhaoyan Zhang, and Dinesh K Chhetri. Effects of thyroarytenoid activation induced vibratory asymmetry on voice acoustics and perception. *The Laryngoscope*, 134(3): 1327–1332, 2024.

[207] Kenneth W Altman, Cory Atkinson, and Cathy Lazarus. Current and emerging concepts in muscle tension dysphonia: a 30-month review. *Journal of voice*, 19(2):261–267, 2005.

[208] Elvira Mendoza and Gloria Carballo. Acoustic analysis of induced vocal stressby means of cognitive workload tasks. *Journal of Voice*, 12(3):263–273, 1998.

[209] Daryush D Mehta, Jarrad H Van Stan, and Robert E Hillman. Relationships between vocal function measures derived from an acoustic microphone and a subglottal neck-surface accelerometer. *IEEE/ACM transactions on audio, speech, and language processing*, 24(4):659–668, 2016.

[210] Pascal van Lieshout, Boaz Ben-David, Melinda Lipski, and Aravind Namasivayam. The impact of threat and cognitive stress on speech motor control in people who stutter. *Journal of fluency disorders*, 40:93–109, 2014.

[211] Christopher J Poletto, Laura P Verdun, Robert Strominger, and Christy L Ludlow. Correspondence between laryngeal vocal fold movement and muscle activity during speech and nonspeech gestures. *Journal of Applied Physiology*, 97(3):858–866, 2004.

[212] Tobias Egner and Joy Hirsch. The neural correlates and functional integration of cognitive control in a stroop task. *Neuroimage*, 24(2):539–547, 2005.

[213] Colin M MacLeod. Half a century of research on the stroop effect: an integrative review. *Psychological bulletin*, 109(2):163, 1991.

[214] Defne Abur, Austeja Subaciute, Mara Kapsner-Smith, Roxanne K Segina, Lauren F Tracy, J Pieter Noordzij, and Cara E Stepp. Impaired auditory

discrimination and auditory-motor integration in hyperfunctional voice disorders. *Scientific Reports*, 11(1):13123, 2021.

[215] Hasini R Weerathunge, Jenny Vojtech, Courtney J Dunsmuir, Sarah J Cocroft, Manuel E Díaz-Cádiz, Victoria McKenna, and Cara E Stepp. Characterization of vocal motor control using laryngeal kinematics in individuals with hyperfunctional voice disorders. *Journal of Speech, Language, and Hearing Research*, 68(4):1743–1757, 2025.

[216] Diana Rose Becker, Sandeep Shelly, Dio Kavalieratos, Carissa Maira, and Amanda I Gillespie. Immediate effects of mindfulness meditation on the voice. *Journal of Voice*, 2022.

[217] Vanessa Veis Ribeiro, Amanda Gabriela de Oliveira, Jhonatan da Silva Vitor, Ana Carolina Ramos, Alcione Ghedini Brasolotto, and Kelly Cristina Alves Silverio. Effectiveness of voice therapy associated with electromyographic biofeedback in women with behavioral dysphonia: randomized placebo-controlled double-blind clinical trial. *Journal of Voice*, 33(3): 381–e11, 2019.

[218] Andrea Rohl, Stephanie Gutierrez, Karim Johari, Jeremy Greenlee, Kris Tjaden, and Angela Roberts. Speech dysfunction, cognition, and parkinson's disease. *Progress in brain research*, 269(1):153–173, 2022.

[219] Jesús A Parra, Carlos Calvache, Gabriel A Alzamendi, Emiro J Ibarra, Leonardo Soláque, Sean D Peterson, and Matías Zañartu. Asymmetric triangular body-cover model of the vocal folds with bilateral intrinsic muscle activation. *The Journal of the Acoustical Society of America*, 156(2):939–953, 2024.

[220] Ingo R Titze. Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, 85(4):1699–1707, 1989.