2020

# Acoustic and videoendoscopic techniques to improve voice assessment via relative fundamental frequency

BOSTON UNIVERSITY

COLLEGE OF ENGINEERING

Dissertation

# ACOUSTIC AND VIDEOENDOSCOPIC TECHNIQUES TO IMPROVE VOICE ASSESSMENT VIA RELATIVE FUNDAMENTAL FREQUENCY

by

**JENNIFER MICHELE VOJTECH**

B.S., University of Maryland, 2015
M.S., Boston University, 2019

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2020

Approved by

First Reader

Cara E. Stepp, Ph.D.
Associate Professor of Speech, Language & Hearing Sciences
Associate Professor of Biomedical Engineering
Associate Professor of Otolaryngology – Head and Neck Surgery

Second Reader

Kamal Sen, Ph.D.
Associate Professor of Biomedical Engineering

Third Reader

H. Steven Colburn, Ph.D.
Professor of Biomedical Engineering
Professor of Electrical and Computer Engineering

Fourth Reader

Melanie L. Matthies, Ph.D.
Associate Professor *Emerita* of Speech, Language & Hearing Sciences

Fifth Reader

Matías Zañartu, Ph.D.
Associate Professor of Electronic Engineering
Universidad Técnica Federico Santa María

## DEDICATION

*To my cats, Tyr and Lux*

*who were there for all of the memories, good and bad,*

*during which this dissertation was written*

*but could not have cared less about its completion*

## ACKNOWLEDGMENTS

First and foremost, I'd like to thank the person whose guidance and counsel was foundational to this work: Dr. Cara Stepp. It was your advice and support that first guided me into the world of speech science and, to be honest, I don't think I'll ever leave. As a research advisor, you challenged me to ask the difficult questions, think critically about the problem at hand, and remain steadfast in my pursuit of knowledge. As a friend, you always had an open ear, whether that be to listen to my latest idea or simply decompress over a drink (or several) at Cornwall's. You knew exactly when to push me and exactly when to let me grow. Looking back, I am so proud of what I accomplished in my research, but it would not have been possible without you.

I would also like to thank my dissertation committee: Dr. Matías Zañartu, Dr. Melanie Matthies, Dr. Steve Colburn, and Dr. Kamal Sen. Your expertise and guidance throughout these past years has been invaluable to say the least. Without your support, this dissertation work would not have been possible.

I want to express my gratitude toward Joshua Kline, Paola Contessa, Serge Roy, Gianluca De Luca, and Devi De Luca for making me feel like such a part of your community over the last year. I am also grateful for Bhawna Shiwani, Michael Chan, Michael Twardowski, Claire Mitchell, Tess Meier, John Chiodini, and John Letizi. I never could have expected such a level of support and encouragement, and I consider myself so lucky to call you all my colleagues. I owe a special shout-out to Bhawna Shiwani and Michael Chan for the endless debates about how Michael was wrong, heartfelt discussions about Kaldi, and—most of all—for being some of the best friends I

could ask for.

My graduate career at Boston University would not have been the same without the support and encouragement of my labmates, as well as Dr. Frank Guenther. Thank you, Frank, for the many Fridays spent at Cornwall's and accepting me for who I am even though I don't like birds. A huge thank you to (Drs.) Tory McKenna, Gabe Cler, and Liz Heller Murray for inspiring me to stay strong, be inquisitive, and own my confidence. Thank you, Tory, for always having a moment to chat about science and cats (mostly cats). You are one of the most humble, humorous, and sincere people I have ever met. Thanks to Gabe for always having a word of sage advice, regardless of how inconsequential the subject. I appreciate you for teaching me the ropes of EMG and for catching me when I made (so many) mistakes. Thanks as well to Liz—you are the most grounded person I know, and I have a tremendous amount of respect you for that. You were always there when I needed a shoulder to cry on, advice about something science-y, or simply just needed a break. Your confidence and strength are so inspiring. I also want to thank Daniel Buckley, Yeonggwang Park, Manuel Díaz-Cádiz, Matti Groll, and Hasini Weerathunge. There are no words to describe the love that each of you have shown me throughout the years, and it was with your unyielding support that I was able to make it to where I am now. Of course, this list would not be complete without Defne Abur. You are one of the most driven and dedicated people I have ever met, and I know you'll be successful in whatever you try to do. I'm so happy to call you my twin, my best friend, and the "difference" to my "just noticeable." You are one of my favorite people and I don't know what I would do without you, D.

Finally, I would like to acknowledge all of my family and friends. Thanks to my parents, brother, and fantastic extended family who have been at my side over these past years. I love you all more than words can express. I also want to thank my wonderful roommates who have had to deal with my antics for the past few years: Caity Sullivan, Colin Bianchi, Lauren Milling, and Konrad Ryba. There is no other group of humans I'd rather join (or start) a cult with. I also want to thank Cam Brody, Hali MacDonald, Liz Morlock, and Elise Jortberg—also known as my non-roommate roommates—who comprise some of the most supportive, generous people the world has to offer, and who I miss dearly whenever they are living at their own houses instead of at ours. And lastly, I want to thank Mitchell Bigelow, for without him, I would have never realized that we could jump higher, run faster, and dance more freely at night than under the sun. Thank you for providing me unconditional love and support all these years.

**ACOUSTIC AND VIDEOENDOSCOPIC TECHNIQUES TO IMPROVE VOICE ASSESSMENT VIA RELATIVE FUNDAMENTAL FREQUENCY**

**JENNIFER MICHELE VOJTECH**

Boston University College of Engineering, 2020

Major Professor: Cara E. Stepp, Ph.D., Assistant Professor of Speech, Language, and Hearing Sciences, Sargent College of Health and Rehabilitation Sciences; Assistant Professor of Biomedical Engineering, College of Engineering; Assistant Professor of Otolaryngology – Head and Neck Surgery, School of Medicine

## ABSTRACT

Quantitative measures of laryngeal muscle tension are needed to improve assessment and track clinical progress. Although relative fundamental frequency (RFF) shows promise as an acoustic estimate of laryngeal muscle tension, it is not yet transferable to the clinic. The purpose of this work was to refine algorithmic estimation of RFF, as well as to enhance the knowledge surrounding the physiological underpinnings of RFF. The first study used a large database of voice samples collected from 227 speakers with voice disorders and 256 typical speakers to evaluate the effects of fundamental frequency estimation techniques and voice sample characteristics on algorithmic RFF estimation. By refining fundamental frequency estimation using the Auditory Sawtooth Waveform Inspired Pitch Estimator—Prime (Auditory-SWIPE′) algorithm and accounting for sample characteristics via the acoustic measure, pitch strength, algorithmic errors related to the accuracy and precision of RFF were reduced by 88.4% and 17.3%, respectively. The second study sought to characterize the

physiological factors influencing acoustic outputs of RFF estimation. A group of 53 speakers with voice disorders and 69 typical speakers each produced the utterance, /ifi/, while simultaneous recordings were collected using a microphone and flexible nasendoscope. Acoustic features calculated via the microphone signal were examined in reference to the physiological initiation and termination of vocal fold vibration. The features that corresponded with these transitions were then implemented into the RFF algorithm, leading to significant improvements in the precision of the RFF algorithm to reflect the underlying physiological mechanisms for voicing offsets ($p < .001$, $V = .60$) and onsets ($p < .001$, $V = .54$) when compared to manual RFF estimation. The third study further elucidated the physiological underpinnings of RFF by examining the contribution of vocal fold abduction to RFF during intervocalic voicing offsets. Vocal fold abductory patterns were compared to RFF values in a subset of speakers from the second study, comprising young adults, older adults, and older adults with Parkinson's disease. Abductory patterns were not significantly different among the three groups; however, vocal fold abduction was observed to play a significant role in measures of RFF at voicing offset. By improving algorithmic estimation and elucidating aspects of the underlying physiology affecting RFF, this work adds to the utility of RFF for use in conjunction with current clinical techniques to assess laryngeal muscle tension.

## PREFACE

This dissertation is an exploration into refining algorithmic methods for calculating the acoustic measure, relative fundamental frequency (RFF). With advances in algorithmic development, the dissertation further investigates the physiological underpinnings of RFF. The dissertation is organized as three self-contained manuscripts (Chapters 2–4), preceded by a common foreword (Chapter 1). The authors and titles of these manuscripts can be found below. In this construction, there is some overlap between Chapters 2–4 and the Chapter 1. A final chapter (Chapter 5) provides a summary and synthesis of results for the three manuscripts (Chapters 2–4).

Chapter 2: Vojtech, J.M., Segina, R.K., Buckley, D.P., Kolin, K.R., Tardif, M.C., Noordzij, J.P., & Stepp, C.E. (2019). "Refining algorithmic estimation of relative fundamental frequency: Accounting for sample characteristics and fundamental frequency estimation method," *The Journal of the Acoustical Society of America*, 146(5), 3184-3202.

Chapter 3: Vojtech, J.M., Cilento, D., Luong, A., Noordzij, J.P., Jr., Park, Y., Diaz-Cadiz, M., Groll, M.D., Buckley, D.P., Noordzij, J.P., & Stepp, C. E. "The Relationship between Acoustic Features and Vocal Fold Vibratory Characteristics during Intervocalic Offsets and Onsets," *In Prep.*

Chapter 4: Vojtech, J. M., & Stepp, C. E. "The Relationship between Vocal Fold Abductory Kinematics and Relative Fundamental Frequency: An Analysis across Young Adults, Older Adults, and Adults with Parkinson's Disease," *In Prep.*

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

$\theta_{off}$ ..... Glottic Angle at Voicing Offset

Abbr ..... Abbreviation

ACO ..... Autocorrelation

A$f_o$ ..... Average Fundamental Frequency

ANCOVA ..... Analysis of Covariance

ANOVA ..... Analysis of Variance

A-P ..... Anterior-to-posterior

APS ..... Average Pitch Strength

aRFF ..... Semi-automated RFF Algorithm

aRFF-AP ..... Semi-automated RFF Algorithm with Auditory-SWIPE′ and Pitch Strength

aRFF-APH .. Semi-automated RFF Algorithm with Auditory-SWIPE′, Pitch Strength, and Physiologically Tuned Acoustic Features

Auditory-SWIPE' ..... Auditory Sawtooth Waveform Inspired Pitch Estimator—Prime

BPM ..... Beats Per Minute

CAPE-V ..... Consensus Auditory-Perceptual Evaluation of Voice

$cat_{off}$ ..... Pitch Strength Categories for Voicing Offset

$cat_{on}$ ..... Pitch Strength Categories for Voicing Onset

CI ..... Confidence Interval

CPP ..... Cepstral Peak Prominence

CSID ..... Cepstral Spectral Index of Dysphonia

CT ..... Cricothyroid

dB ..... Decibel

DBS ..... Deep brain Stimulation

EMG ..... Electromyography

$f_o$ ..... Fundamental Frequency

FPS ..... Frames Per Second

FVF ..... False Vocal Folds

GRBAS ..... Grade, Roughness, Breathiness, Asthenia, Strain

GUI ..... Graphical User Interface

HL ..... Hearing Level

HNR ..... Harmonics-to-noise Ratio

HSV ..... High-speed Videoendoscopy

IA ..... Interarytenoid

ICC ..... Intraclass Correlation Coefficients

LCA ..... Lateral Cricoarytenoid

LEMG ..... Laryngeal Electromyography

LHR ..... Low-to-high Ratio of Spectral Energy

LMT ..... Laryngeal Manual Therapy

LSVT LOUD ..... Lee Silverman Voice Treatment – LOUD

MBE ..... Mean Bias Error

MDS-UPDRS ..... Movement Disorder Society-sponsored Revision of the Unified Parkinson's Disease Rating Scale

MDS-UPDRS-III ..... Movement Disorder Society-sponsored Revision of the Unified

Parkinson's Disease Rating Scale: Part III, Motor Examination

M$f_o$ ..... Median Fundamental Frequency

MPS ..... Median Pitch Strength

N ..... Sample Size

NCCF ..... Normalized Cross-correlation Function

NXCO ..... Normalized Cross-correlation

NZC ..... Number of Zero Crossings

OAC ..... Older Adult Controls

OAwPD ..... Older Adults with Parkinson's Disease

OS ..... Overall Severity of Dysphonia

PAS ..... Phonatory Aerodynamic System

PCA ..... Posterior Cricoarytenoid

PD ..... Parkinson's Disease

PLR ..... Positive Likelihood Ratio

PRO ..... Patient-reported Outcome

PTP ..... Normalized Peak-to-peak Amplitude

QVP ..... Quick Vibratory Profile

RAPT ..... Robust Algorithm for Pitch Tracking

RFF ..... Relative Fundamental Frequency

RMS ..... Root-mean-square

RMSE ..... Root-mean-squared Error

ROC ..... Receiver Operating Characteristic

SD ..... Standard Deviation

SEMG ..... Surface Electromyography

SLE ..... Short-time Log Energy

SPL ..... Sound Pressure Level

ST ..... Semitone

STE ..... Short-time Energy

STM ..... Short-time Magnitude

TA ..... Thyroarytenoid

$T_{abd}$ ..... Duration of Vocal Fold Abduction

$t_{abd}$ ..... Start of Vocal Fold Abduction

$t_{add}$ ..... Termination of Vocal Fold Adduction

$t_{off}$ ..... Time of Voicing Offset

$t_{on}$ ..... Time of Voicing Onset

VALI ..... Voice-Vibratory Assessment with Laryngeal Imaging

VCV ..... Vowel–voiceless consonant–vowel

VH ..... Vocal Hyperfunction

WSS ..... Waveform Shape Similarity

$X^2$ ..... Chi-square

XCO ..... Cross-correlation

YAC ..... Young Adult Control

# CHAPTER 1. Introduction

During speech, listeners can attend to what a speaker is saying as well as *how* a speaker is saying it. This is because voice is a unique medium through which speakers are not only able to convey linguistic information, but also individualistic characteristics such as emotion, personality, and intent. An individual's voice is the product of intrinsic factors derived from their anatomy and physiology, as well as habitual factors adopted by the individual during their lifetime (e.g., sociolinguistic trends; Tiwari & Tiwari, 2012). These intrinsic and habitual factors converge in voice production via interactions between aerodynamic, structural, and acoustic mechanisms involving the lungs, larynx, and vocal tract (Zhang, 2016).

## *Voice Production*

The classic source-filter theory of voice production simplifies the complex interactions of aerodynamic, structural, and acoustic mechanisms necessary for speech using two components: a sound source and a filtering process (Fant, 1960; Stevens, 2005). Airflow passes from the lungs through narrow constrictions at or above the larynx to generate a sound source, which, in turn, is modified via articulatory mechanisms to produce speech. Located in the anterior neck, the larynx is a highly specialized structure that serves an integral role as a source of sound energy for human speech.

### *Anatomy and Physiology of the Larynx*

The larynx protects the airway, assists in swallowing, and enables voice production. It is composed of a framework of cartilages, bone, tissues, membranes, and muscles (Coleman, Zakowski, Gold, & Ramanathan, 2013; Young, Matsuzaki, & Sasaki,

2015). Complex interactions among these components modify the configuration of the larynx to regulate breathing, swallowing, and voicing.

## Vocal Folds as the Glottal Source

The vocal folds—located within the central region of the larynx—are essential structures for voice production. Each vocal fold is made up of multiple structural layers, with the deeper layers being less pliable than the more superficial layers (Ferrand, 2007; Hirano, 1974; Hixon, Weismer, & Hoit, 2018). The vocalis muscle is the deepest layer and is interdigitated with a fibrous ligament called the elastic conus; this interdigitation allows the two structures to act as a single unit during voicing. Covering the elastic conus is a mucous membrane that constitutes the lamina propria and squamous epithelium. The lamina propria consist of three morphologically different layers of connective tissue (Hirano, Kakita, Ohmaru, & Kurita, 1982), whereas the squamous epithelium is a single, thin epithelial layer. Distinct from these "true" vocal folds are the "false" vocal folds: the false vocal folds (ventricular folds) are thick folds of mucous membrane that lie superior to the true vocal folds (Agarwal, Scherer, & Hollien, 2003). These structures assist in lubricating the true vocal folds (Kutta, Steven, Kohla, Tillmann, & Paulsen, 2002), and—like the true vocal folds—have been shown to abduct (open) and adduct (close) during specific vocal gestures (e.g., throat singing; Lindestad, Sodersten, Merker, & Granqvist, 2001; Stager, Bielamowicz, Regnell, Gupta, & Barkmeier, 2000) as well as during pathological phonation (Arnold & Pinto, 1960; Lindestad, Blixt, Pahlberg-Olsson, & Hammarberg, 2004; Nasri et al., 1996; Von Doersten, Izdebski, Ross, & Cruz, 1992).

Each vocal fold comprises membranous and cartilaginous portions: the

membranous portion forms the anterior 55–65% of the length of the fold, whereas the cartilaginous portion corresponds to the posterior 35–45% of the length of the fold (Hirano, Kurita, Kiyokawa, & Sato, 1986). The membranous segment of the fold is bounded by the vocal ligament, a structure comprising the deeper of the lamina propria layers. On the other hand, two cartilaginous processes stemming from the inferior portion of the arytenoid cartilage ("vocal processes") form the cartilaginous segment of the fold (Ferrand, 2007; Hixon et al., 2018). The vocal folds are in a paramedial position during tidal breathing, leaving a slightly open space between the folds. This space is referred to as the glottis. The vocal folds may be forcefully abducted (opened) to a greater extent, thus expanding the glottis, when larger amounts of air are inhaled during deep breathing and sniffing gestures. During phonation, the vocal folds are adducted (closed) to the midline of the glottis.

The mechanical properties of the vocal folds during oscillation can be described by the cover-body theory of vocal fold vibration. In this model, the vocal folds act as a double-structured vibrator: the *body* corresponds to the vocalis muscle and elastic conus that operate as a single unit, whereas the *cover* refers to the mucous membrane (Hirano, 1974). This cover is made up of the epithelium and the intermediate and superficial layers of the lamina propria. The body has been shown to exhibit variable mechanical properties according to the degree of activation of the vocalis muscle, as well as mechanical changes from passive stretching via activations from other laryngeal muscles (e.g., cricothyroid; Vahabzadeh-Hagh, Zhang, & Chhetri, 2018). The mechanical properties of the mucous membrane differ from that of the body since it is only loosely connected to

the elastic conus, instead largely depending on the interactions of the vocalis muscle with other intrinsic laryngeal musculature (Hirano, 1974). The implications of this model are that different laryngeal adjustments are possible based on the unique relationships between the body and cover, and occur on a similar time scale as the articulators (e.g., tongue, lips, jaw, and velum). Yet these adjustments do not generate the vibratory motion of the vocal folds and, thus, do not lead to voice production; instead, vocal fold vibration depends on the aerodynamic conditions surrounding the vocal folds, as well as the configuration and mechanical properties of the folds (Story, 2015). The cover-body theory does not account for these conditions, which are necessary to initiate and sustain self-oscillation.

## Laryngeal Cartilages

There are nine cartilages that form the skeleton of the larynx. These include the unpaired cricoid, thyroid, and epiglottis as well as the paired arytenoids, corniculates, and cuneiforms (Coleman et al., 2013; Ferrand, 2007; Young et al., 2015). The laryngeal cartilages are attached to each other via membranes and ligaments that surround and protect the larynx.

### Unpaired Cartilages

The lower limit of the larynx is marked by the cricoid cartilage. As the most inferior of the laryngeal cartilages, the cricoid cartilage is a ring-shaped cartilage that connects to the tracheal rings of the airway. The cricoid cartilage also serves as an attachment point with the thyroid and arytenoid cartilages for ligaments and muscles to regulate respiration and voice production. The cricoid cartilage connects inferolaterally

and anteriorly with the thyroid cartilage as well as posterosuperiorly with the arytenoid cartilages via these membranes and ligaments.

Just superior to the cricoid cartilage lies the thyroid cartilage. The thyroid cartilage is a shield-like structure with two laminae that fuse together at the laryngeal prominence (commonly known as the Adam's apple). The true vocal folds attach to the inner surface of the thyroid just below the superior surface of the laryngeal prominence. This intersection point of the true vocal folds on this surface is referred to as the *anterior commissure*. In addition to its connection with the cricoid cartilage, the thyroid cartilage attaches to the hyoid bone by the thyrohyoid membrane to anchor the larynx during respiration and phonation (Coleman et al., 2013).

Marking the entryway to the laryngeal vestibule, the epiglottis is the most superior of the laryngeal cartilages. The epiglottis is a leaf-like cartilaginous structure attached to the thyroid cartilage via the thyroepiglottic ligament and is laterally bordered by triangular folds of mucous membrane extending into the arytenoid cartilages (aryepiglottic folds). The epiglottis and aryepiglottic folds both cover the upper airway to prevent foreign bodies from entering during swallowing (Young et al., 2015). Specifically, the epiglottis sits in an otherwise upright position to allow for breathing, but retroflexes over the entrance of the larynx during swallowing to both protect the airway and guide prandial material toward the esophagus (Ferrand, 2007). Similarly, the aryepiglottic folds adduct to prevent unwanted aspiration while swallowing (Standring, Borley, & Gray, 2008).

**Paired Cartilages**

The arytenoids are three-sided pyramidal cartilages that function in airway protection and voice production. The arytenoid cartilages are attached to the cricoid cartilage at a ball-and-socket joint referred to as the cricoarytenoid joint; as a result, the arytenoids articulate with the cricoid ring. The base of each arytenoid cartilage is characterized by a vocal process, which attaches to the vocal ligament, and a muscle process, which attaches to intrinsic laryngeal musculature. Because of its attachment to the vocal ligament, the arytenoids can articulate to promote vocal fold tensing, relaxation, or approximation to alter voice production (Coleman et al., 2013; Ferrand, 2007; Hixon et al., 2018).

In addition to the arytenoids, there are two other pairs of laryngeal cartilages: the corniculate cartilages and the cuneiform cartilages. The corniculate cartilages are small, conical structures that articulate with the apex of the arytenoids to assist in vocal fold approximation (Jacob, 2007, p. 205). The cuneiform cartilages are thin, elongated structures that are also small relative to the other laryngeal cartilages. The cuneiforms lie within either side of the posterior portion of the aryepiglottic folds, providing rigidity to the folds (Coleman et al., 2013).

## Laryngeal Musculature

Throughout this complex network of membranes and ligaments are muscles that enable motion within and around the larynx. These muscles are often referred to as *intrinsic* and *extrinsic* muscle groups. Muscles categorized as “intrinsic” originate within the larynx and have insertions that fall between the laryngeal cartilages, whereas muscles

categorized as “extrinsic” have one insert in the laryngeal cartilages and the other in adjacent structures. The intrinsic laryngeal muscles are directly responsible for inducing changes in the length and tension of the vocal folds as well as modifying glottal positioning. The extrinsic laryngeal muscles work to raise, lower, or stabilize the larynx during speech and swallowing movements.

**Intrinsic Laryngeal Muscles**

The intrinsic laryngeal muscles comprise five major muscle pairs that differentially affect the biomechanical state of the vocal folds. These muscles include the thyroarytenoid (TA), lateral cricoarytenoid (LCA), interarytenoid (IA), posterior cricoarytenoid (PCA), and cricothyroid (CT). The LCA, TA, and IA act as adductors to close the vocal folds, whereas the PCA acts as an abductor to open the vocal folds (Ferrand, 2007). The CT functions to lengthen and stretch the vocal folds which, in turn, alters the vibratory rate of the vocal folds (Chhetri, Neubauer, Sofer, & Berry, 2014). The distinct roles of each intrinsic muscle in voice production are described in detail below.

The three adductors (LCA, TA, and IA) differentially act to approximate the vocal folds. The LCA is a paired muscle that acts as the principal adductor by pulling the vocal processes inward and downward to medially compress the vocal folds. The IA is an unpaired muscle that assists in vocal fold adduction by pulling the arytenoid cartilages together to bring the posterior portion of the vocal folds together (Andaloro & La Mantia, 2019; Choi, Ye, & Berke, 1995). Also considered an adductor muscle, the TA is a bilaterally paired muscle that comprises the main mass of each vocal fold. The medial region of the TA is often referred to as the vocalis muscle, described above as the body of

the vocal fold. The TA is anteriorly attached to the internal surface of the thyroid cartilage and posteriorly attached to the vocal and muscular processes of the arytenoids. As a result, other intrinsic and extrinsic laryngeal muscles that alter the tension of the vocal folds thus influence the status of the TA (Hixon et al., 2018). Contraction of the TA contributes to the approximation of the vocal folds (particularly the membranous portion of the folds), as well as shortens the vocalis to increase vocal fold body stiffness (Choi, Berke, Ye, & Kreiman, 1993b; Hixon et al., 2018; Sataloff, Heman-Ackah, & Hawkshaw, 2007) to, in turn, increase the rate of vibration of the vocal folds and stabilize the onset of phonation (Chhetri & Neubauer, 2015; Choi et al., 1993b).

As the sole abductor of the group, the PCA induces an outward rotation of the arytenoids on the cricoid cartilage to open the vocal folds. The primary function of PCA is to abduct the vocal folds respiratory tasks such as inspiration (Hast, 1967b). However, the PCA also promotes devoicing by supporting the larynx against pulls from the CT and adductor muscles during phonation (Choi, Berke, Ye, & Kreiman, 1993a; Faaborg-Andersen, 1957; Fujita, Ludlow, Woodson, & Naunton, 1989; Hirano, 1988).

The main function of the CT is to lengthen and stretch the vocal folds, which, in turn, increases the stiffness of the body and cover as well as the vibratory rate of the vocal folds (Chhetri et al., 2014). The CT muscle has two components: the pars recta, with vertically oriented muscle fibers, and the pars oblique, with obliquely upward oriented muscle fibers. The pars recta and pars oblique simultaneously act to displace the joint that connects the cricoid and thyroid cartilages (cricothyroid joint) vertically (pars recta) and horizontally (pars oblique; Hong, Kim, & Kim, 2001; Hong et al., 1998).

Displacement of this joint increases the distance between the arytenoid cartilages and the thyroid cartilage; since the vocal folds are attached anteriorly to the thyroid and posteriorly to the vocal processes of the arytenoids, an increase in distance between these cartilages, in turn, passively lengthens the TA (Hixon et al., 2018; Hong et al., 2001). Thus, the pars recta and pars oblique act to lengthen and tense the vocal folds during phonation and are the primary means of pitch variation during voicing.

**Extrinsic Laryngeal Muscles**

The role of the extrinsic laryngeal muscles in voice production is to modulate laryngeal height and tilt. The extrinsic laryngeal muscles are categorized based on their attachment points relative to the hyoid bone: *suprahyoid* muscles attach above the hyoid bone whereas *infrahyoid* muscles attach below the hyoid bone. The suprahyoid muscles include the anterior and posterior digastrics, stylohyoid, mylohyoid, geniohyoid, and hyoglossus muscles. These muscles contribute to elevating the larynx by anteriorly (anterior digastric, mylohyoid, hyoglossus, and geniohyoid) or posteriorly (posterior digastric and stylohyoid) pulling the hyoid bone (Broniatowski et al., 1999; Sataloff et al., 2007; Suárez-Quintanilla, Fernández Cabrera, & Sharma, 2019). The infrahyoid muscles include the thyrohyoid, omohyoid, sternothyroid, and sternohyoid. Contraction of the thyrohyoid draws the thyroid and hyoid bone closer together, thereby elevating the larynx; conversely, contraction of the omohyoid, sternothyroid, and sternohyoid each contribute to lowering the larynx (Ferrand, 2007).

Changes in laryngeal elevation via the extrinsic laryngeal muscles have been shown to indirectly affect the vibratory rate of the vocal folds (Ueda, Oyama, Harvey, &

Ogura, 1972). Specifically, vertical laryngeal movements affect vocal fold length and tension through rotation of the cricoid cartilage because of cervical lordosis (i.e., excessive inward curvature of the spine; Honda, Hirai, Masaki, & Shimada, 1999). Altering the length and tension of the vocal folds, in turn, modifies vocal fold vibratory rate.

### *Glottal Dynamics during Voice Production*

Stevens (2005) describes two types of sound sources that are produced within the larynx. These sources include the glottal phonation source and aspiration noise. These sound sources are not mutually exclusive, yet are generated through different processes. Whereas the glottal phonation source arises from quasiperiodic pulses of air generated from airflow traveling through the vibrating vocal folds, aspiration noise arises from turbulent airflow through slightly abducted vocal folds (Mehta, 2006; Stevens, 2005). This turbulent airflow acts as a stochastic excitation source to introduce noise into the voice production system (Kent & Read, 2002).

Voice production via the glottal phonation source occurs because of aerodynamic forces and vocal fold mechanical properties. Theoretical models of intraglottal aerodynamics suggest that the forces responsible for separating the adducted vocal folds occur in the form of a centerline glottal jet that travels from the lungs and makes contact with the inferior-medial surface of the folds (Khosla, Muruguppan, Gutmark, & Scherer, 2007); the vocal folds diverge as the glottal jet passes through the glottis. Recent modeling in hemilarynges has shown that as the folds are pushed apart, the glottal jet undergoes flow separation (Khosla et al., 2007; Oren, Khosla, & Gutmark, 2014). These

models have also shown that the velocity of the jet relates to the motion of the laryngeal walls and the magnitude of subglottal pressure, such that flow separation vortices have been demonstrated at high subglottal pressures (Oren et al., 2014).

As the column of air pressure travels vertically through the glottis, the intraglottal geometry of the cover can undergo wave propagation as the result of phase differences in tissue compression and rarefaction (Berke & Gerratt, 1993). Mucosal wave propagation is the result of subglottal air pressure against the vocal folds, wherein the medial-lateral and superior-inferior surfaces oscillate along with the main body of the folds as they separate (Berry, Montequin, & Tayama, 2001; Doellinger, Berry, & Berke, 2005; Krausert et al., 2011). Oscillations occur within the cover of the vocal fold due to the mucous membrane being a mechanically pliable structure as compared to the body of the vocal fold (Ferrand, 2007).

When the vocal folds meet at the midline during oscillatory motion, the passage of air through the glottis is temporarily halted until air pressure from the lungs once again pushes the vocal folds apart. This vibratory cycle repeats while airflow from the lungs passes through the glottis at a high enough pressure (subglottal pressure) to drive the vocal folds apart and induce vocal fold oscillations (Finck & Lejeune, 2010; Stevens, 2005). The intermittent closure of the vocal folds transforms the airflow into a series of (glottal) pulses that constitute the glottal sound source for phonation. The lowest frequency of the glottal source spectrum is its fundamental frequency ($f_o$). The $f_o$ of the glottal source is related to the rate of vocal fold vibration, and is perceptually correlated with vocal pitch (Hixon et al., 2018). Higher frequencies are also contained within the

spectrum and comprise integer multiples of the $f_o$ (harmonics). These harmonics are produced by the collisions of the vocal folds as well as the diversion of acoustic energy toward the vocal folds to distort the glottal airflow.

### *Voiced and Unvoiced Speech Sounds*

Human speech comprises voiced, unvoiced, and mixed speech sounds, referred to as phonemes. The source for voiced phonemes (e.g., /ɔɪ/ as in "voice") corresponds to the vibrating vocal folds, whereas the source for unvoiced phonemes (e.g., /s/ as in "voice") originates from airflow passing through a constriction in the vocal tract across the pharynx, oral, and/or nasal cavities. Mixed speech sounds necessitate a combination of voiced and unvoiced sound sources (e.g., /v/ as in "voice"). In each of these scenarios, the vocal tract acts as a resonator to filter specific frequency bands while attenuating others (Fant, 1960; Ferrand, 2007). Due to the different sizes and shapes of each of the supraglottal regions (e.g., pharynx, oral cavity, nasal cavity), the vocal tract acts as a broadly tuned resonator with multiple, different resonant frequencies (Ferrand, 2007; Hixon et al., 2018). Based on the configuration of the vocal tract, specific frequencies are amplified as others are dampened to alter the quality of the sound as it emerges at the lips (Ferrand, 2007). As a result, a wide range of speech sounds may be produced based on the source of the sound and the configuration of the vocal tract (Fant, 1960).

### *Voice Disorders*

Voice disorders arise when an individual feels that their voice does not meet their daily physical, social, emotional, and/or professional needs (Verdolini & Ramig, 2001). These disorders are prevalent throughout the world, with approximately one third of

adults reporting problems using their voice at some point during their lifetime in the United States alone (Bainbridge, Roy, Losonczy, Hoffman, & Cohen, 2017; Bhattacharyya, 2014; Roy, Merrill, Gray, & Smith, 2005). Nearly a quarter of these individuals further report recurrent issues (Roy et al., 2005). Considering that around 28 million individuals in the United States rely on their voice in order to successfully carry out their job (Verdolini et al., 2001), having a voice disorder can have a substantial impact on one's life.

Voice disorders have been historically characterized as voice production that deviates from a speaker's expectations, whether it be voice quality, pitch, and/or loudness differing from those of a similar background (e.g., age, gender, culture, geographical location; Ramig & Verdolini, 1998), or due to functional and/or structural changes to the laryngeal mechanism that prevent the speaker from meeting daily voice needs (Stemple, Roy, & Klaben, 2018). These scenarios are not mutually exclusive: individuals who describe functional or structural issues (e.g., pain in the larynx or neck, globus sensation, dryness, and/or a need to cough during voice use) may also exhibit signs of inappropriate voice quality, pitch, or loudness. It is estimated that approximately 10–50% of cases referred to multidisciplinary voice clinics include some sort of tension component (Dworkin-Valenti, Stachler, Stern, & Amjad, 2018; Roy, 2003). Because the proper regulation of laryngeal muscle tension is vital for voice production, it is important to be able to comprehensively assess tension using clinical outcome measures.

### *Laryngeal Muscle Tension*

The regulation of laryngeal muscle tension is necessary to produce voice.

Extrinsic laryngeal muscle tension is directly responsible for raising and lowering the larynx, whereas intrinsic laryngeal muscle tension is responsible for tensing, abducting, and adducting the vocal folds (Boone, McFarlane, Von Berg, & Zraick, 2014, p. 42). Increased vocal exertion (effort) is thought to be a byproduct of higher cervical muscle tension (Hunter et al., 2020) and has been associated with a strained or "strangled" voice quality in some speakers (Baldner, Doll, & van Mersbergen, 2015; Borg, 1982; Brandt, Ruder, & Shipp, 1969; Lagier et al., 2010; McKenna, Diaz-Cadiz, Shembel, Enos, & Stepp, 2018a; Mooshammer, 2010).

The etiologies associated with excessive laryngeal muscle forces are diverse, ranging from overuse and/or misuse of the laryngeal mechanism in the absence of organic pathology, pathological changes in the vocal fold tissue (e.g., nodules, polyps, contact ulcers), and neurological disorders affecting the laryngeal mechanism (Boone et al., 2014, p. 120; Ghassemi et al., 2014; Hillman, Holmberg, Perkell, Walsh, & Vaughan, 1989). It is thought that each of these etiologies necessitate functional changes in laryngeal muscle tension to compensate for the additional effort required to maintain adequate phonation.

### *Excessive Laryngeal Tension in Clinical Populations*

Hypertonicity of the laryngeal mechanism is thought to be a prevalent characteristic in a range of voice disorders and is a critical target of many therapeutic interventions (e.g., circumlaryngeal massage). Although the specific pathophysiology of excessive laryngeal muscle tension may vary across speakers, it is a prevalent characteristic of many functional, structural, and/or neurological disorders, including

adductor focal laryngeal dystonia (Aronson & Bless, 2009; Nash & Ludlow, 1996), unilateral vocal fold paralysis (Neel et al., 1994; Pinho, Pontes, Gadelha, & Biasi, 1999), hyperfunctional voice disorders (Hillman et al., 1989), and Parkinson's disease (Gallena, Smith, Zeffiro, & Ludlow, 2001). As vocal hyperfunction and Parkinson's disease will be the focus of this dissertation, a discussion of laryngeal muscle tension in relation to vocal hyperfunction and Parkinson's disease is included below.

## Vocal Hyperfunction

Vocal hyperfunction (VH) is a common feature exhibited in individuals with voice disorders. VH is described as excessive or imbalanced tension in the laryngeal musculature and is associated with daily vocal overuse and/or misuse (e.g., yelling; Hillman et al., 1989). Hyperfunctional vocal behaviors may occur in the presence or absence of organic pathology. These behaviors are present as either the primary cause of a voice disorder or as a compensatory adaptation to glottal insufficiency, and as such, are often sorted as "non-phonotraumatic VH" and "phonotraumatic VH," respectively.

Non-phonotraumatic VH (also referred to as non-adducted VH or primary VH) is characterized as persistent dysphonia and excessive tension of the laryngeal and circumlaryngeal areas in the absence of vocal fold tissue trauma (Bhattacharyya, 2014; Boone et al., 2014, p. 66; Hillman et al., 1989; Mehta et al., 2015). These behaviors are also associated with stiff and tensed vocal folds without complete vocal fold adduction (Hillman et al., 1989). Additional behaviors may include elevated hyolaryngeal position, decreased space between the hyoid bone and laryngeal cartilage, increased extrinsic laryngeal muscle activation, and excessive supraglottal constriction (Lowell, Kelley,

Awan, Colton, & Chan, 2012a). Speakers that present with non-phonotraumatic VH are often diagnosed with "muscle tension dysphonia (MTD)," or "primary MTD;" these diagnoses describe excessive tension of the (para)laryngeal musculature with various contributing etiologies (Altman, Atkinson, & Lazarus, 2005; Van Houtte, Claeys, D'haeseleer, Wuyts, & Van Lierde, 2013).

Phonotraumatic VH (also referred to as adducted VH or secondary VH) is associated with hyperfunctional behaviors in the presence of vocal fold lesions (e.g., nodules, polyps; Mehta et al., 2015; Titze, Svec, & Popolo, 2003). Vocal fold tissue trauma is thought to lead to incomplete glottic closure, which, in turn, elicits increases in laryngeal muscle tension and subglottal pressure to assist in vocal fold closure (Hillman et al., 1989). As a result of these increases, phonotraumatic VH is associated with stiff and tightly approximated vocal folds that contribute to increased vocal fold collision forces (Espinoza, Zañartu, Van Stan, Mehta, & Hillman, 2017), and ultimately, more trauma to the vocal fold tissues. In many cases, however, it is unclear if the overuse and misuse of the vocal mechanism elicited structural changes to the vocal folds, and/or if the structural changes elicited an aberrant compensatory response (i.e., excessive laryngeal muscle tension) to maintain phonation (Ghassemi et al., 2014).

**Laryngeal Muscle Tension in Vocal Hyperfunction**

Hyperfunctional vocal behaviors encompass a broad range of symptoms, including (but not limited to) tension in the paralaryngeal musculature (Dworkin, Meleca, & Abkarian, 2000a), elevated laryngeal positioning (Morrison, 1997; Roy, Ford, & Bless, 1996), glottal insufficiency (Hillman et al., 1989), hyperadduction of the true and/or false

vocal folds (Higgins, Chait, & Schulte, 1999), and supraglottic compression (Hočevar-Boltežar, Janko, & Zargi, 1998; Stager et al., 2000; Stepp, Heaton, Jette, Burns, & Hillman, 2010a). These symptoms may lead to a voice that sounds rough, breathy, and/or strained, as well as periods of vocal fry, low vocal pitch, and low dynamic range (Dworkin et al., 2000a; Morrison, 1997; Morrison, Nichol, & Rammage, 1986; Morrison, Rammage, Belisle, Pullan, & Nichol, 1983).

Although it is presumed that VH affects both intrinsic and extrinsic laryngeal musculature, there is a lack of objective evidence supporting elevated tension in these muscle groups. This discordance in the literature may be, in part, because there is no gold-standard measure for assessing the presence or severity of laryngeal muscle tension. Instead, much of the early literature describing the role of the intrinsic laryngeal muscles in hyperfunctional vocal behaviors was based on suppositions about the observed laryngoscopic images. For instance, Morrison et al. (1983) attributed the presence of posterior glottal gap to increased muscle tension in the posterior cricoarytenoid during phonation. Furthermore, there is also disagreement in the hypothesized role of the intrinsic laryngeal muscle tension in VH: prior work has suggested that tension manifests as specific glottic and supraglottic contraction patterns that can be observed through laryngoscopic imaging (Koufman & Blalock, 1991; Morrison et al., 1986; Morrison & Rammage, 1993; Morrison et al., 1983), whereas other sources indicate that there is not a specific glottic configuration that is uniquely associated with increased tension (Aronson, 1990). Even if there were an agreed upon, unique glottic configuration observed in VH, there has been no objective, quantitative means of confirmation that the degree of

intrinsic laryngeal muscle tension is associated with the observed supraglottic activity. This is likely because perturbing the intrinsic laryngeal muscles and structures to estimate tension (via intramuscular electromyography; see *Laryngeal EMG* for details) may alter typical muscle function. As such, there has been no direct comparisons of intrinsic laryngeal muscle tension between vocally healthy speakers and speakers with VH.

Studies examining extrinsic laryngeal muscle tension in VH are more prevalent in the literature, likely a result of the relative ease of using surface electromyography (sEMG; see *Surface EMG* for details about this technique) to non-invasively assess the perioral, suprahyoid, and paralaryngeal muscles. There is conflicting evidence regarding whether speakers with VH exhibit increased extrinsic laryngeal muscle activity compared to vocally healthy speakers. Redenbaugh and Reich (1989) demonstrated greater mean (normalized) sEMG activity in speakers with VH when recording from a single electrode placed on the anterior neck. Yet the sample size examined in the study was small (seven speakers with VH and seven vocally healthy speakers) and varied in age, sex, and clinical presentation. In a more recent study, Hočevar-Boltežar et al. (1998) recorded sEMG activity from 18 pairs of differential electrodes placed on the face and neck musculature of 11 speakers with VH (nodules, MTD) and five vocally healthy speakers. Significant differences were demonstrated in mean sEMG activity between the two groups; however, the sEMG signals examined in the study were not normalized. Without normalizing the sEMG signal, it is difficult to interpret and generalize variations in sEMG activity that may be otherwise confounded by sEMG electrode configuration and contact, as well as participant neck mass. Due to the limitations in experimental methodology of these two

works, it is difficult to generalize study findings to our theoretical understanding of extrinsic laryngeal muscle activity in VH.

More recently, Van Houtte et al. (2013) examined differences in extrinsic laryngeal muscle activity between vocally healthy speakers and speakers with MTD. The sample sizes included in this study were larger and more controlled than those of the aforementioned works (18 speakers with MTD, 44 vocally healthy speakers), and the resulting sEMG signals were normalized to a reference contraction prior to comparison. The authors found no significant differences in sEMG activity between groups. In a similar study, Stepp et al. (2011b) compared sEMG activity between 10 vocally healthy individuals, 10 professionally trained singers with nodules, and eight non-singers with nodules. Although sternocleidomastoid activity during the initiation of the vowel, /a/, was statistically significantly greater in non-singers with nodules, the authors ultimately found that sEMG activity could not effectively discriminate nodule and control groups. Stepp et al. (2010a) examined sEMG activity of speakers with VH before and after injection laryngoplasty, a surgical procedure for correcting glottal insufficiency via injecting material into the vocal fold. As glottal insufficiency has been observed in speakers with VH, it has been postulated that increased activity of the laryngeal muscles could be used to achieve glottal closure. Yet the authors saw no significant changes in sEMG activity pre- to post-procedure, despite observing reductions in supraglottic compression. Taken together, the findings of these studies suggest that sEMG may not be a suitable technique for assessing differences in extrinsic laryngeal muscle tension between speakers with and without VH. More importantly, there is a lack of evidence to suggest that the extrinsic

laryngeal muscle tension exhibited in these individuals differs from that of typical speakers.

Despite the substantial theoretical framework describing elevated laryngeal muscle tension in VH, the current literature base does not support differences in this tension between speakers with and without VH. In addition to EMG, other techniques have been implemented to assess laryngeal muscle tension; one such example is manual palpation, in which tension of the extrinsic laryngeal and other superficial neck musculature are examined through visual and tactile inputs. However, these methods suffer from poor validity and reliability and, moreover, do not assess tension of the intrinsic laryngeal muscles (see *Manual Palpation* for more details). Further investigations are thus needed to objectively quantify the presence and degree of both intrinsic and extrinsic laryngeal muscle tension in these groups.

**Assessing and Treating Vocal Hyperfunction**

Treatment for VH typically aims to identify and reduce maladaptive phonatory behaviors that may be related to laryngeal muscle tension (Holmberg, Hillman, Hammarberg, Sodersten, & Doyle, 2001). Methods of treatment may include indirect approaches to modify cognitive, behavioral, psychological, and physical environments, as well as direct approaches to manipulate vocal behavior through motor execution, somatosensory feedback, and auditory feedback (Roy et al., 2001; Thomas & Stemple, 2007).

Indirect treatments include two components: patient education and counseling. Patient education opens a line of discussion between the clinician and patient to

characterize normative voice production and how a voice disorder may impact it. In this intervention, the patient is tasked with learning strategies to modify their vocal health (e.g., diet; Roy et al., 2001; Van Stan, Roy, Awan, Stemple, & Hillman, 2015b). On the other hand, counseling employs coping strategies, stress management techniques, and therapeutic interventions to identify and modify negative psychosocial influences that may impact vocal health (Van Stan et al., 2015b).

Whereas indirect treatments aim to inform the patient about vocal hygiene, direct treatments focus on modifying some combination of auditory, vocal functional, somatosensory, musculoskeletal, and respiratory behaviors (Van Stan et al., 2015b). Targeting musculoskeletal behaviors are a large focus of therapies for individuals with non-phonotraumatic VH, as increased or imbalanced laryngeal muscle tension is a primary etiological concern. As such, Manual Circumlaryngeal Therapy (Dromey, Nissen, Roy, & Merrill, 2008; Khoddami, Ansari, & Jalaie, 2015; Roy, Bless, Heisey, & Ford, 1997; Roy & Leeper, 1993; Roy, Nissen, Dromey, & Sapir, 2009) and Laryngeal Manual Therapy (Mathieson, 2011; Mathieson et al., 2009) are two established programs that employ palpatory techniques to reduce this tension (see *Manual Palpation* for an overview of manual palpation). Other sources of treatment in VH have also shown success, including vocal function exercises (Nguyen & Kenny, 2009; Pedrosa, Pontes, Pontes, Behlau, & Peccin, 2016), resonant voice therapy (Roy et al., 2003; Watts, Hamilton, Toles, Childs, & Mau, 2019) and semi-occluded vocal tract exercises (Guzman et al., 2015; Guzman et al., 2016; Titze, 2006), and/or a combination of respiratory and phonoarticulatory methods (e.g., Accent Method; Fex, Fex, Shiromoto, & Hirano, 1994;

Kotby, Shiromoto, & Hirano, 1993; Liang et al., 2014). Besides behavioral therapy, direct treatments may include surgery to remove benign fibrovascular lesions or pharmacological interventions (e.g., topical lidocaine; Dworkin, Meleca, Simpson, & Garfield, 2000b). These treatments are often implemented on a patient-specific basis, and may comprise multiple aspects of indirect and/or direct therapies.

In general, voice therapy that includes some form of direct intervention has demonstrated better outcomes compared to a vocal hygiene program or no intervention (Carding, Horsley, & Docherty, 1999; Desjardins, Halstead, Cooke, & Bonilha, 2017; Ogawa & Inohara, 2018). Unfortunately, the heterogeneity in outcome measures employed in individual studies makes it challenging to compare specific interventions across studies. In their review, Desjardins et al. (2017) report that pre- to post-treatment outcomes in the literature have been assessed via a myriad of patient-reported outcomes, auditory-perceptual judgments of voice quality, laryngeal imaging, and acoustic analyses. This lack of standardization of outcomes measures is also paralleled by the failure of many studies to control for patients' motivational and behavioral characteristics, such as adherence to therapy and voice use. While it may be difficult to specifically control for these characteristics, there is evidence to suggest that therapy outcomes are a result of patient adherence (van Leer & Connor, 2015; Ziegler, Verdolini Abbott, Johns, Klein, & Hapner, 2014). Assessing treatment efficacy in individuals with phonotraumatic VH is even more challenging since the source of improvement is not always clear (Ogawa et al., 2018). For instance, an individual with a vocal fold polyp may exhibit improved voice quality and shrunken size of the polyp, yet it is unknown whether these benefits are due

to vocal training, vocal hygiene education, or another source. As such, the clinical meaningfulness of voice therapy in VH has not been fully characterized.

## Parkinson's Disease

Parkinson's disease (PD) is a progressive neurodegenerative disorder that involves numerous neurotransmitter pathways across the central and peripheral nervous systems (Braak et al., 2003; Schapira, Chaudhuri, & Jenner, 2017). PD is primarily known for its cardinal motor symptoms of muscle rigidity (tension), bradykinesia, tremor, and postural instability (Shahed & Jankovic, 2007), but also manifests through non-motor symptoms, including mood disorders (e.g., apathy, anxiety), pain, sleep disturbances, urinary/bowel symptoms, hallucinations, and dementia (Gallagher & Schrag, 2012; Schapira et al., 2017). It is estimated that—in addition to these symptoms—up to 90% of individuals with PD further develop a motor speech disorder called hypokinetic dysarthria (Darley, Aronson, & Brown, 1969; Robbins, Logemann, & Kirshner, 1986).

### Speech Symptoms in Parkinson's Disease

Hypokinetic dysarthria predominantly manifests as reduced loudness (Canter, 1965; Goberman, Coelho, & Robb, 2002; Logemann, Fisher, Boshes, & Blonsky, 1978; Metter & Hanson, 1986; Zwirner, Murry, & Woodson, 1991) and pitch variability (Bowen, Hands, Pradhan, & Stepp, 2013; Zwirner et al., 1991), as well as several other symptoms resulting from detrimental changes to the respiratory, laryngeal, articulatory, and resonatory subsystems. Respiratory symptoms include reduced vital capacity (De Letter et al., 2007), as well as impaired speech breathing (i.e., fewer words, less time producing speech per breath, faster interpause speech rate; Solomon & Hixon, 1993) and

tidal breathing (i.e., faster breathing rate, decreased minute ventilation rate; Solomon et al., 1993; Vercueil, Linard, Wuyam, Pollak, & Benchetrit, 1999). Laryngeal symptoms not only manifest as impaired prosody (e.g., monoloudness and monopitch; Holmes, Oates, Phyland, & Hughes, 2000), but also as abnormal voice quality (Zwirner & Barnes Gary, 1992) and intrinsic laryngeal muscle rigidity (Gallena et al., 2001; Zarzur, Duprat, Cataldo, Ciampi, & Fonoff, 2013; Zarzur, Duprat, Shinzato, & Eckley, 2007). Individuals with PD may also exhibit vocal fold bowing (Blumin, Pcolinsky, & Atkins, 2004), glottal insufficiency (Stelzig, Hochhaus, Gall, & Henneberg, 1999; Yuceturk, Yilmaz, Egrilmez, & Karaca, 2002), reduced vocal fold abductory and adductory movements (Perju-Dumbrava et al., 2017; Stelzig et al., 1999), atypical vocal fold vibratory (e.g., phase amplitude and symmetry; Perez, Ramig, Smith, & Dromey, 1996; Yuceturk et al., 2002), and abnormal mucosal wave characteristics (Stelzig et al., 1999; Yuceturk et al., 2002). Reports of vocal tremor have also been attributed to PD (Holmes et al., 2000; Logemann et al., 1978; Perez et al., 1996; Stelzig et al., 1999); however, the evidence regarding the source of the tremor is unclear. Articulatory and resonatory symptoms comprise imprecise articulation (Skodda, Grönheit, & Schlegel, 2012), reduced vowel space area (Skodda, Grönheit, Mancinelli, & Schlegel, 2013; Whitfield & Goberman, 2014), and velopharyngeal incompetence (Hoodin & Gilbert, 1989; Robbins et al., 1986).

**Neurophysiological Mechanisms of Speech Symptoms in PD**

To date, the specific neurophysiological mechanisms contributing to speech symptoms in PD are unknown. Prior work suggests that speech symptoms observed in the respiratory, laryngeal, articulatory, and resonatory domains arise from deteriorations in

motor control (Henderson, Trojanowski, & Lee, 2019; Jankovic, 2008; Kwan & Whitehill, 2011). Motor deficits in these subsystems has been primarily attributed to progressive dopaminergic depletion in the substantia nigra (Chu & Kordower, 2007; Dauer & Przedborski, 2003), which limits the ability of the basal ganglia to coordinate neural motor signals. Yet there is also evidence to suggest that deficits in sensorimotor integration (i.e., how sensory information is transformed into motor actions) lead to the speech symptoms observed in PD.

PD is neuropathologically characterized by the presence of Lewy bodies, or abnormal proteinaceous aggregates that develop inside nerve cells. These aggregates are primarily composed of the synaptic protein, α-Synuclein, and have been identified in PD in the sensory neurons innervating the mouth, pharynx, and larynx (Mu et al., 2015). Since the ability to reach a desired speech target is partially dependent on somatosensory feedback (Houde & Nagarajan, 2011; Lametti, Nasir, & Ostry, 2012; Larson, Altman, Liu, & Hain, 2008; Tourville & Guenther, 2011), Lewy-type synucleinopathy in these sensory nerves suggests that impairments in sensorimotor integration may be a contributing factor to speech symptoms in PD. Indeed, prior work has demonstrated reduced somatosensation in PD in response to air bursts in the laryngeal mechanism (Hammer & Barlow, 2010).

Another mechanism in support of abnormal sensorimotor integration is impaired auditory feedback. Abnormal responses to perturbations in auditory feedback have been demonstrated in the laryngeal (Abur et al., 2018; Liu, Wang, Metman, & Larson, 2012; Mollaei, Shiller, Baum, & Gracco, 2016) and articulatory (Mollaei et al., 2016; Mollaei,

Shiller, & Gracco, 2013) subsystems in PD. Because the magnitude of responses to brief, unanticipated perturbations in auditory feedback (thought to engage feedback mechanisms in speech motor control; Houde et al., 2011; Lametti et al., 2012; Larson et al., 2008; Tourville et al., 2011) are different between laryngeal (Liu et al., 2012; Mollaei et al., 2016) and articulatory (Mollaei et al., 2016) subsystems, it is likely that multiple neural regions contribute to impaired speech motor control in PD. These findings are in support of impaired sensorimotor integration during speech production as a possible etiology for the speech symptoms exhibited in PD.

It is important to note that much of the research examining speech symptoms in PD has focused on the physical limitations that arise from impaired motor control (i.e., due to hallmark motor symptoms of tremor, rigidity, bradykinesia, and postural instability). Yet these limitations do not appear to purely be the result of an inability to achieve motor targets; for instance, acoustic and auditory-perceptual measures of vocal loudness have been shown to improve in PD when an individual is externally cued for loudness (Ramig et al., 2001). Instead, the evidence described here suggests that these limitations are co-occurring with a neutrally mediated change—that is, impaired sensorimotor integration—wherein there is an impairment in the way that the brain processes sensory information for the desired motor output. Subsequently, there are therapeutic techniques that rely on volitional control to overcome some of these deficits.

**Laryngeal Muscle Tension in Parkinson's Disease**

Muscle tension is one of the hallmark motor symptoms of PD (Berardelli et al., 2018; Shahed et al., 2007; Sprenger & Poewe, 2013). Tension in PD is also referred to as

*muscle rigidity*, and is characterized by increased muscle stiffness during mobilization. Although the specific mechanisms contributing to muscle tension in PD have not been fully characterized, it is suspected to be the product of physical modifications of the muscles (leading to muscle fiber hypertrophy or atrophy; Dietz, Quintern, & Berger, 1981; Edstrom, 1968; Mu et al., 2012; Rossi et al., 1996; Watts, Wiegner, & Young, 1986) and neural processes. The presence of muscle tension in PD has been documented throughout the body, including in the upper limbs (Cantello et al., 1991; Cantello, Gianelli, Civardi, & Mutani, 1995; Edstrom, 1970; Meara & Cody, 1993; Prochazka et al., 1997; Robichaud et al., 2009; Watts et al., 1986), lower limbs (Berardelli, Sabra, & Hallett, 1983; Rossi et al., 1996), and axial muscles (e.g., neck, trunk, hips; Anastasopoulos, Maurer, Nasios, & Mergner, 2009; Gurfinkel et al., 2006; Kroonenberg et al., 2006; Mak, Wong, & Hui-Chan, 2007; Nagumo & Hirayama, 1993, 1996). Increased baseline muscle activity has also been identified in the oropharyngeal muscles: a reduction in orofacial muscle activity was observed following levodopa administration (Leanderson, Meyerson, & Persson, 1971; Nakano, Zubick, & Tyler, 1973), has been shown to coincide with improvements to speech articulation (Wolfe, Garvin, Bacon, & Waldrop, 1975). More recently, investigations into muscle tension in PD in the laryngeal muscles have been underway.

Intrinsic laryngeal muscle tension has been reported in speakers with PD. Zarzur et al. (2007) reported that 19 of 26 participants with PD exhibited hypercontractility of the TA and CT at baseline, described as spontaneous activity during voice rest. The authors compared these results to those of 26 age-matched controls, showing that mean

activity at baseline was significantly higher in those with PD ($p$ = .004). To further characterize hypercontractility in PD, Zarzur et al. (2013) assessed the TA and CT muscles in a larger sample size. The authors split 94 participants into disease severity groups based on Hoehn-Yahr stage (Hoehn & Yahr, 1967): 57 participants were considered "mild" (stage I and II), 21 as "moderate" (stage III), and 16 as "severe" (stage IV and V). Hypercontractility at baseline was identified in 86 of the 94 participants (50 of 57 mild cases, 20 of 21 moderate cases, 16 of 16 severe cases). No significant effect of disease severity or age was reported. Taken together, these studies indicate that the TA and CT exhibit active contractile patterns at rest in PD.

Laryngeal muscle tension may be a contributing factor to the speech symptoms observed in PD. In particular, a study by Gallena et al. (2001) examined the relationship between intrinsic laryngeal muscle activity and perceptual measures of speech impairment during the initiation and termination of voicing. Although no significant group differences were found in the muscle activity of individuals with and without PD, there was a significant relationship between increased TA and CT muscle activity and degree of speech impairment. The authors also noted a relationship between individuals who exhibited increased TA and CT activity and vocal fold bowing, a suspected byproduct of hypercontracted TA and PCA muscles and a loss of CT lengthening (Hanson, Gerratt, & Ward, 1984). After administering levodopa medication, a marked reduction in TA activity was met with improvements to overall speech and voice proficiency. These findings indicate a relationship between laryngeal muscle tension and observed speech impairments in PD; however, the specific mechanisms associated with

this relationship are, to date, unknown.

**Assessing and Treating Speech Symptoms in Parkinson's Disease**

Speech is primarily evaluated in individuals with PD via subjective assessment by a trained technician conducting the motor examination section of the Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS; Stebbins & Goetz, 1998; Goetz et al., 2008). The MDS-UPDRS comprises four parts—non-motor experiences of daily living, motor experiences of daily living, motor examination, and motor complications—that incorporate responses from the patient and/or their caregiver, as well as a clinical investigator. Different aspects of the severity and progress of PD are quantified on a 5-point Likert scale. The scale was validated for reliability, demonstrating high internal consistency (Cronbach's alpha = .79–.93; Goetz et al., 2008). Interestingly, the MDS-UPDRS incorporate another scale, called the Hoehn-Yahr scale, which can be performed as part of the motor examination or independently. The Hoehn-Yahr scale was developed to describe the general motor progression of PD (Hoehn et al., 1967); a clinical investigator ascribes a score (I-V) to the overall severity of motor symptoms based on the observed level of clinical disability.

Dopaminergic therapy is considered the gold standard for alleviating motor symptoms in PD (Ferreira et al., 2012). Levodopa medication is the typical method of administration of dopaminergic therapy, as it is thought that levodopa could slow the degeneration of residual dopamine neurons in PD (Fahn, 1996; Olanow et al., 2004). However, there is conflicting evidence in the literature regarding improved speech symptoms following levodopa administration. Respiratory symptoms such as reduced

vital capacity and impaired speech and tidal breathing have been shown to normalize in PD with medication in some studies (De Letter et al., 2007; Vercueil et al., 1999), but not change in others (Solomon et al., 1993). Although laryngeal rigidity has been shown to decrease following dopaminergic treatment (Gallena et al., 2001; Jiang, Lin, Wang, & Hanson, 1999a), the impact on acoustic metrics of prosody (e.g., standard deviation of $f_o$ to assess vocal pitch variability) is equivocal. For instance, Skodda, Visser, and Schlegel (2010) did not see significant changes in intonation ($f_o$ variability) and phonation (mean $f_o$) before and after medication. Azevedo, Cardoso, and Reis (2003) and Bowen et al. (2013), on the other hand, saw improvements in $f_o$ variability in PD with medication. Articulatory symptoms have been shown to improve through increased vowel articulation (Skodda et al., 2010), decreased orofacial rigidity (Cahill et al., 1998; Leanderson et al., 1971; Svensson, Henningson, & Karlsson, 1993), and reduced dysfluencies (Tykalová et al., 2015); yet, other studies report no changes in PD with medication (Goberman & Blomgren, 2003; Lowit, Dobinson, Timmins, Howell, & Kröger, 2010). Resonatory speech symptoms have not been characterized before and after the administration of levodopa medication; as such, it is unclear how dopaminergic medicine offsets these symptoms.

Besides medication, treatment for speech symptoms in PD include surgical and behavioral therapy. Surgical therapy includes the use of deep brain stimulation (DBS) to alleviate motor symptoms (Deuschl et al., 2006; Fasano, Daniele, & Albanese, 2012). In this treatment, electrodes are implanted in the basal ganglia to stimulate certain nuclei. Behavioral therapy, on the other hand, focuses on using external cues to improve acoustic

and auditory-perceptual measures of speech. Although there has been some reported success in alleviating global motor symptoms in PD via DBS (Deuschl et al., 2006; Fasano et al., 2012; Limousin et al., 1998), the effects of DBS on speech symptoms is controversial, with some studies reporting detriments to speech after DBS. This includes reduced MDS-UPDRS speech scores (Gervais-Bernard et al., 2009; Kleiner-Fisman et al., 2003), vowel space area (Sidtis, Alken, Tagliati, Alterman, & Van Lancker Sidtis, 2016) and intelligibility (Tripoliti et al., 2011; Yorkston, Beukelman, & Traynor, 1984). Two methods of behavioral therapy include the Lee Silverman Voice Therapy (LSVT LOUD), in which patients are instructed to focus on producing a loud, clear voice (Cannito et al., 2012; Ramig, Fox, & Sapir, 2004, 2008; Saffarian, Amiri Shavaki, Shahidi, Hadavi, & Jafari, 2019; Sapir, Ramig, & Fox, 2011; Spielman, Ramig, Mahler, Halpern, & Gavin, 2007) and SPEAK OUT!, which instructs the patient to speak with intent (Boutsen, Park, Dvorak, & Cid, 2018). The therapeutic effects from LSVT LOUD have been shown to last for up to two years (Wight & Miller, 2015), but long-term data for SPEAK OUT! is not yet available. Overall, these intensive treatment programs have demonstrated immediate improvements in sound pressure level; however, long-term effectiveness may be affected by the progressive nature of PD and limited understanding of the neurophysiological mechanisms contributing to speech symptoms in PD (Broadfoot, Abur, Hoffmeister, Stepp, & Ciucci, 2019).

### *Current State of Clinical Assessments of Laryngeal Tension*

Clinical assessments of laryngeal muscle tension can be classified into two groups: non-instrumental methods that do not require equipment and instrumental

methods that use tools for assessment. Non-instrumental methods include case history, patient-reported outcomes, auditory-perceptual judgments of the voice, and manual palpations of the extrinsic laryngeal and other neck musculature. Importantly, these methods are not sensitive to tension of the intrinsic laryngeal muscles. Instrumental approaches encompass laryngeal visualizations, as well as aerodynamic, electroglottographic, electromyographic, accelerometric, and acoustic signal analysis techniques that may capture aspects of intrinsic and/or extrinsic laryngeal muscle activity. Despite having many techniques available to assess muscle tension, many of these methods fall short in terms of validity, reliability, and/or specificity. A discussion of the advantages and disadvantages of these methods in relation to laryngeal muscle tension is included below.

## Non-Instrumental Assessments

### Case History

Patient case history is a time-honored technique used to gather information about the presenting complaint. A case history typically includes information about how the patient describes the voice problem, including the onset and variability of symptoms. It may also include patient medical status and history, including daily habits, past surgeries, medications and treatments (e.g., voice therapy), and how stress and other psychological factors may be influencing the voice. Many individuals with excessive or imbalanced laryngeal muscle tension report a history of smoking and/or organic triggers (e.g., reflux), and report concerns in using their voice (Kridgen, 2019; Morrison et al., 1986). Throat pain, neck/shoulder tightness, vocal effort or fatigue, and intensified symptoms following

extended voice use are each associated with excessive and/or imbalanced laryngeal muscle tension (Morrison et al., 1986; Roy et al., 1996). Although case history can provide insight into the etiology and pathology of the voice complaint, the subjective nature of this method means that it does not provide direct information about laryngeal muscle tension. Therefore, case history is largely limited to monitoring and evaluating voice change over time from the patient's point of view.

**Patient-reported Outcomes**

Patient-reported outcome (PRO) measures systematically capture the experiences of the patient without interpretation from other individuals (e.g., clinicians). Although subjective in nature, these approaches are appealing since many voice problems are clinically complex and manifest differently across individuals. Popular examples of PRO instruments include the Vocal Fatigue Index (Nanjundeswaran, Jacobson, Gartner-Schmidt, & Verdolini Abbott, 2015), the Voice Handicap Index (Jacobson et al., 1997), and the Voice-Related Quality of Life questionnaire (Hogikyan & Sethuraman, 1999; Karnell et al., 2007).

There are two pitfalls when applying PRO instruments to assess laryngeal muscle tension. First, PRO measures do not collect direct information about muscle tension. Although a patient may report a sense of discomfort or describe a problem that could be related to excessive tension, PRO instruments are restricted to the psychosocial consequences of voice complaints, as well as how different individuals are affected by the same voice problem. For instance, a patient could report "tightness" of the throat when completing the Vocal Tract Discomfort scale (Mathieson et al., 2009); however, it

is unclear whether this tightness is a result of excessive tension or inflammation. Second, a review of 32 voice-related PRO instruments revealed that only 20 PRO measures (62.5%) showed adequate reliability (e.g., test-retest, internal consistency) and only 3 PRO measures (9.4%) showed sufficient longitudinal validity (Francis et al., 2017). Longitudinal validity was qualified as a demonstrated responsiveness to change as well as adequate test-retest reliability, and included the Voice Outcome Survey (test-retest reliability: $r = .87$, $p < .001$; Richard, Robert, & William, 1999), Voice-related Quality of Life (test-retest reliability: $r = .93$, $p < .001$; Hogikyan et al., 1999), and Linear Analogue Self-Assessments of Voice Quality (test-retest reliability: $ICC > .54$ for all 16 scale items; Llewellyn-Thomas et al., 1984). Interestingly, none of the 32 PRO measures offered a statistical justification for interpreting severity scores. It is clear that—while PRO approaches are appealing for providing a patient's unique perspective on their voice problem—caution must be used when selecting, collecting, and interpreting PRO measures. As with case history, PRO measures may be used to provide useful insights into the possible etiology and pathology of a tension-based voice complaint; however, these measures should be combined with additional forms of assessment for a more comprehensive analysis.

**Auditory-perceptual Assessments**

Auditory-perceptual assessments of voice quality are performed by clinicians to quantify the severity of auditory-perceptual attributes of voice problems. During these evaluations, a clinician listens to and critically judges a person's vocal output to determine whether vocal symptoms are consistent with a referral diagnosis, and if these

symptoms are exhibited consistently or intermittently. In these assessments, the presence of laryngeal muscle tension is often regarded as *vocal strain* (Askenfelt & Hammarberg, 1986; Dejonckere et al., 1996; Jafari et al., 2017; Lowell et al., 2012a).

Vocal strain is defined as the perception of excessive vocal effort during phonation (Hirano, 1981; Kempster, Gerratt, Verdolini Abbott, Barkmeier-Kraemer, & Hillman, 2009). Vocal effort, in turn, is the perceived exertion of a vocalist to a perceived communication scenario (i.e., vocal demand; Baldner et al., 2015; Borg, 1982; Hunter et al., 2020) and has been linked to dry throat, odynophonia (i.e., pain in using voice), and vocal fatigue (McCabe & Titze, 2002). Strain is specifically thought to be related to the degree of compression of the vocal folds and hypertonicity in or around the larynx (Askenfelt et al., 1986; Lowell et al., 2012a) and can occur either as a primary feature of the voice disorder or as a result of the individual attempting to compensate for unrelated pathologies (e.g., vocal fold paralysis; Lowell et al., 2012).

Two well-established rating scales that can capture vocal strain are the GRBAS (Hirano, 1981) and the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster et al., 2009). The GRBAS scale contains subscales to assess five core attributes of voice: grade (G), roughness (R), breathiness (B), asthenia (A), and strain (S). These attributes are scored on a four-point ordinal scale as either 0 (normal or none of the quality), 1 (mild), 2 (moderate), or 3 (severe). The CAPE-V scale also encompasses these attributes, except for asthenia (weakness). However, the CAPE-V also assesses pitch and loudness, and further, includes optional space to analyze two non-predetermined parameters (e.g., diplophonia, aphonia, tremor). CAPE-V attributes are evaluated by

considering the degree of perceived deviance from "normal" on a continuous visual analog scale. GRBAS and CAPE-V scales both allow experimenters to judge whether the deviance of each attribute was consistent or intermittent throughout the voice evaluation.

Although the GRBAS and CAPE-V scales are similar in that both instruments allow experimenters to assess vocal strain, the method of analyzing these attributes are distinct. The GRBAS scale uses an ordinal scale that does not allow for parametric statistical analysis, whereas the CAPE-V operates on a visual analog scale. Because of this, the GRBAS is considered less sensitive in evaluating subtle differences in voice quality (Nemr et al., 2012; Wuyts, De Bodt, & Van de Heyning, 1999). The GRBAS scale also fails to supply instructions regarding the vocal tasks that should be used. These concerns and others led to the development of the CAPE-V as a more standardized practice for auditory-perceptual evaluations. The CAPE-V uses a continuous visual analog scale to allow for parametric analysis, and moreover, the CAPE-V specifies vocal tasks (sustained vowels, scripted sentences, and spontaneous speech) to minimize variability in elicitation and analysis.

Of the auditory-perceptual features assessed in clinical voice quality assessments, vocal strain is considered one of the least reliable features. Dejonckere et al. (1996) examined the reliability of GRBAS parameters collected from two experienced clinicians in a large subset of voice samples across different institutes. The authors combined asthenia and strain into a single parameter, termed "tonus." Interrater reliability was calculated across 943 voice samples via Spearman's rank-correlation coefficients, resulting in a reliability of $r_s$ = .87 for grade, $r_s$ = .70 for roughness, $r_s$ = .69 for

breathiness, and $r_s$ = .65 for tonus. Similarly, intrarater reliability was calculated across 80 repeated voice samples, producing a reliability of $r_s$ = .89 for grade, $r_s$ = .74 for roughness, $r_s$ = .78 for breathiness, and $r_s$ = .68 for tonus. Despite showing promising reliability, the authors noted that behavioral aspects of tonus were observably difficult to tease apart, hence combining asthenia and strain. The authors further noted that variability in interrater reliability were somewhat reduced by rater training and experience.

In a similar study, Zraick et al. (2011) examined the intrarater reliability of a larger group of raters when carrying out GRBAS and CAPE-V evaluations. The rater group included 21 speech-language pathologists with more than five years of clinical voice experience. For the 21 raters, intrarater reliability (assessed via Spearman's rank-correlation coefficients) when using the GRBAS scale produced reliability scores of $r_s$ = .65 for grade, $r_s$ = .67 for roughness, $r_s$ = .67 for breathiness, $r_s$ = .69 for asthenia, and $r_s$ = .53 for strain. In a separate analysis, the authors considered the number of raters with intrarater reliability scores above a cut-off of .70, resulting in 4 raters for grade, 9 raters for roughness, 11 raters for breathiness, 8 raters for asthenia, and 3 raters for strain. When using the CAPE-V scale, intrarater reliability was found to be $r_s$ =.57 for overall severity (with only 2 raters with $r_s$ > .70), $r_s$ = .77 for roughness (14 raters with $r_s$ > .70), $r_s$ = .82 for breathiness (17 raters with $r_s$ > .70), $r_s$ = .35 for strain (0 raters with $r_s$ > .70), $r_s$ = .78 for loudness (7 raters with $r_s$ > .70), and $r_s$ = .64 for pitch (15 raters with $r_s$ > .70). Except for the percept of strain, all CAPE-V intrarater reliability scores were greater than those of the GRBAS. Most importantly, strain resulted in the lowest intrarater reliability

for both scales, as well as the least number of raters with reliability above .70.

In the same study, Zraick et al. (2011) compared interrater reliability between the GRBAS and CAPE-V. For the 21 raters, interrater reliability was low-to-moderate for the GRBAS (.66 for grade, .56 for roughness, .59 for breathiness, .58 for asthenia, .48 for strain) and negligible-to-strong for the CAPE-V (.76 for overall severity, .62 for roughness, .60 for breathiness, .56 for strain, .54 for loudness, .28 for pitch). In general, reliability was greater for CAPE-V ratings than for corresponding GRBAS ratings. Both scales showed good reliability for overall severity of voice/grade; however, raters found strain to be difficult to assess using either scale.

Although laryngeal muscle tension is most often associated with the auditory-perceptual quality of strain, perceptually assessing vocal strain is difficult. It is possible that excessive or imbalanced laryngeal muscle forces manifest through multiple voice percepts, as it is rare that deviant voice production varies along a single dimension of quality (Aronson et al., 2009; Lowell et al., 2012a). For instance, roughness and breathiness are common percepts of voice quality included in both the GRBAS and CAPE-V, as described above. Roughness refers to perceived irregularity in the voicing source (Hirano, 1981; Kempster et al., 2009) stemming from fluctuations in amplitude and/or frequency of vocal fold vibration (Dejonckere, Obbens, de Moor, & Wieneke, 1993; Hirano, 1981); these irregularities lead to the production of a crackly or creaky voice (Bassich & Ludlow, 1986; Borrie & Delfino, 2017). Breathiness, on the other hand, is thought to be related to turbulent noise in the signal from insufficient glottal closure (Ferrer, Haderlein, Maryn, de Bodt, & Nöth, 2018), and is associated with the perception

of audible air escape during voicing (Askenfelt et al., 1986; Hirano, 1981). In principle, these percepts are independent regarding their pathophysiology and acoustical manifestation. Yet roughness and breathiness have been shown to be highly correlated with one another (Kreiman & Gerratt, 2000; Kreiman, Gerratt, & Berke, 1994), and acoustically, the harmonics-to-noise ratio is related to both roughness (de Krom, 1995; Eskenazi, Childers, & Hicks, 1990; Ferrand, 2007; Martin, Fitch, & Wolfe, 1995) and breathiness (Castillo-Guerra & Ruiz, 2009; de Krom, 1995; Heman-Ackah, Michael, & Goding, 2002; Martin et al., 1995; Samlan, Story, & Bunton, 2013; Shrivastav & Sapienza, 2003). It is likely that the underlying etiologies leading to roughness or breathiness of the voice may be similar (Ferrer et al., 2018), and moreover, that both breathiness and roughness can co-occur in the same individual (Kreiman et al., 1994; Lowell et al., 2012a). Thus, it is possible that abnormal laryngeal muscle tension manifests through multiple voice quality percepts for some individuals.

The reliability and validity of auditory-perceptual assessments to assess laryngeal muscle tension remain questionable. Although auditory-perceptual evaluations fail to quantify the degree of laryngeal muscle tension present in the system, these examinations may still provide some insight into tension within the laryngeal mechanism. These assessments may be useful to monitor and track changes in voice quality in an individual over time, but have been recommended for use in conjunction with additional methods of clinical voice evaluation (Oates, 2009).

**Manual Palpation**

Manual laryngeal palpation techniques necessitate the use of visual and tactile inputs to assess the laryngeal spaces and the extrinsic laryngeal and other superficial neck musculature (Altman et al., 2005; Hirano, 1981). Laryngeal palpation is useful for evaluating the tension of the extrinsic laryngeal and other superficial neck musculature; it is a safe technique that requires no equipment and has no reported side effects (Khoddami, Ansari, Izadi, & Talebian Moghadam, 2013). Many laryngeal palpation methods are purely qualitative and have no reported validity or reliability regarding the criteria for judgement. These methods mainly assess laryngeal elevation—one of the key features of excessive laryngeal tension (Lowell, Kelley, Colton, Smith, & Portnoy, 2012b)—as well as pain in response to pressure, resistance in response to movement, lateral mobility, tenderness, and hypertonicity (Altman et al., 2005; Khoddami et al., 2013; Morrison, 1997; Morrison et al., 1986; Roy, 2008; Roy et al., 1996; Roy et al., 1993; Rubin, Blake, & Mathieson, 2007; Rubin, Lieberman, & Harris, 2000; Van Lierde, De Bodt, Dhaeseleer, Wuyts, & Claeys, 2010).

Few manual palpation methods include a grading system to quantify criteria for judgement. Two popular scales that are specific to extrinsic laryngeal muscle tension include the 4-point scale from Angsuwarangsee and Morrison (2002) to assess tension and the 5-point scale from Mathieson and colleagues (2009) to evaluate muscle resistance and position. The former of these scales was modified from prior work by Lieberman (1998) to include a grading system, and was evaluated in a selection of 465 dysphonic patients. The authors found good interrater reliability in the assessment of

suprahyoid, thyrohyoid, and cricothyroid muscles, as well as a correlation between thyrohyoid tension and muscle misuse dysphonia (assessed via subjective interpretations of case history and visualization of the laryngeal mechanism; Morrison et al., 1993). The scale from Mathieson and colleagues, called the Laryngeal Manual Therapy (LMT) scale, was validated in 10 individuals with MTD; because only one investigator (a speech-language pathologist) performed the evaluation, interrater reliability was not obtained. The authors did not report any validity metrics.

Although these manual palpation schemes successfully quantify clinical findings of excessive muscle tension, there are conflicting views as to the reliability and validity of these scales. For instance, the Polish version of the LMT scale was administered to a group of 51 female speakers with disordered voices (16 with nodules, 35 with MTD) and a group 50 female control speakers (Woźnicka, Niebudek-Bogusz, Morawska, Wiktorowicz, & Śliwińska-Kowalska, 2017). Significant relationships were identified between the LMT scale and the Voice Handicap Index (Jacobson et al., 1997), GRBAS auditory-perceptual scale (Hirano, 1981), and acoustically derived maximum phonation time; however, the authors did not assess the validity of the scale using objective tools that—unlike the Voice Handicap Index and GRBAS scale—do not assume that raters are correct in their evaluation (Jafari et al., 2020). Upon comparing LMT scores to objective findings, the reliability and validity of these scales seems questionable. Lowell et al. (2012b) compared LMT scores to radiographic findings of hyoid position, laryngeal position, and hyolaryngeal space during phonation in 20 individuals with and without MTD. The authors found low-to-moderate significant correlations between total LMT

score and radiographic hyoid and laryngeal positions, but no correlation between LMT score of laryngeal position and radiographic laryngeal position. Similarly, a study by Stepp et al. (2011a) compared the LMT scale and the scale by Angsuwarangsee et al. (2002) to objective findings of (para)laryngeal muscle activity (see 0). The authors found low interrater reliability as well as low correlations between manual palpation grades and muscle activity (Stepp et al., 2011a). Taken together, the results of these studies indicate conflicting evidence regarding the generalizability of manual palpation techniques.

In more recent years, there has been a push to design manual palpation scales that use descriptive and instructive text to inform practitioners how to validly, reliably, and quickly assess tensioned structures (Khoddami et al., 2015). Jafari et al. (2020) sought to develop such a scale by drawing information from other palpatory scales and from the opinions of a panel of speech-language pathologists, otolaryngologists, and physical therapists with more than four years of experience with clinical voice disorders. In doing so, the authors developed a 45-item scale called the "laryngeal palpatory scale," or "LPS," to evaluate pain, posture, muscle tenderness and tightness, laryngeal and hyoid position, range of movement, and thyrohyoid spacing. The scale includes clinician-based ratings (on a four-point Likert scale), as well as inputs from the patient regarding their own assessment of pain in the anterior/posterior neck and tenderness of the muscles. When assessing scale reliability in a group of 55 patients with MTD, the authors saw that the weighted kappa for the 45 items ranged from .41 (moderate inter-rater agreement) to 1.0 (nearly perfect inter-rater agreement) across two experienced and blinded speech-language pathologists. This scale shows promise for quantitatively assessing the head,

neck, and shoulders, which—despite a known clinical relation between cervical problems and laryngeal muscle tension (Hülse, 1991; Kooijman et al., 2005)—have not all been included in a palpatory scale to date. However, there are three main limitations with this scale: 1) it is unclear how level of training and experience affect rater outcomes, 2) the accuracy of the scale was not validated against objective findings (e.g., extrinsic laryngeal muscle activity), and 3) despite arguments of the need for a quick and easy-to-administer test, the time required to carry out the 45-item scale is estimated to be 15 minutes.

Manual palpation is a valuable tool to provide insight into extrinsic laryngeal muscle tension. Yet the current state of laryngeal palpation techniques suffers from a lack of standardization. Despite being an easy, direct, non-instrumental assessment method, it is well-known that laryngeal manual palpation schemes depend on the skill and experience of the practitioner (Khoddami et al., 2015). Research is ongoing to develop a comprehensive, clinically useful tool for assessing tensioned structures; however, more work must be done to fully characterize the reliability, validity, and efficiency of such a tool.

## Instrumental Assessments

### Laryngeal Imaging

During laryngoscopic imaging, a device is inserted via the oral or nasal passages to visualize the vocal folds. Visualizing the laryngeal mechanism is an important step to identify laryngoscopic features that may indicate some presence of excessive tension. There are two main types of laryngoscopy: indirect and direct. Indirect laryngoscopy

necessitates the insertion of a laryngeal mirror into the oropharynx to reflect the image of the vocal folds, whereas direct laryngoscopy requires the insertion of an endoscope either through the nose or mouth transmit light to the vocal folds and receive the image back (Colton, Casper, & Leonard, 2011, pp. 223-24). Direct laryngoscopic techniques can be divided into *rigid* laryngoscopy, in which the endoscope is inserted into the oropharynx, and *flexible* laryngoscopy, in which the endoscope is passed transnasally and into the hypopharynx. A rigid laryngoscope placed in the oral cavity requires the technician to physically hold the tongue in a protruded position, such that the patient is unable to articulate any speech sounds other than vowels. Those with sensitive gag reflexes or limited jaw/neck mobility may not tolerate rigid laryngoscopic examinations. On the other hand, a flexible laryngoscope is passed through the nasal cavity. Subsequently, the patient may speak freely since the tongue is not restrained.

*Indirect vs. Direct Laryngoscopy*

Indirect and direct (rigid, flexible) laryngoscopic techniques each have advantages and disadvantages. For instance, indirect and rigid laryngoscopy restrict the speech that can be evaluated since the tongue must be restrained to visualize the vocal folds (Hartnick & Zeitels, 2005). On the other hand, a flexible laryngoscope may be inserted through the nasal cavity to allow the patient to speak without their tongue being restrained. These examinations are generally performed using distal chip or fiberoptic technology. By not restraining the tongue, flexible laryngoscopy allows clinicians to visualize the functionality of the laryngeal mechanism across various speech tasks and phonemes. This includes visualizing the phonatory and non-phonatory supraglottal

behaviors of the larynx (Colton et al., 2011, p. 226). Although direct laryngoscopy is more invasive, it is advantageous over rigid and flexible endoscopes since a full image of the vocal folds can be recorded; this contrasts with the use of laryngeal mirrors that often do not enable the visualization of the anterior commissure.

Direct laryngoscopies can be performed under continuous light, stroboscopy, or high-speed video imaging. Using continuous light allows for the evaluation of structure and gross function; however, these images are traditionally captured at 30 frames-per-second (fps), whereas the vocal folds typically vibrate around 80–1000 cycles per second (Hz) during speech according to age, sex, psychological state, loudness, speaking task, and environment (Aronson et al., 2009, p. 143; Baken & Orlikoff, 2000; Titze, 1994; Woo, 2009, pp. 11-17). As 30 fps is too slow to capture the individual vocal fold oscillations during typical vocalizations, stroboscopy is often employed in conjunction with videoendoscopy (Deliyski et al., 2008) to examine vocal fold vibratory function.

*Stroboscopy*

Stroboscopy emits a pulse of light at a rate that can be controlled either by the clinician or by the $f_o$ of the vocalization (from a laryngeal microphone placed on the surface of the anterior neck, approximating the thyroid cartilage). The light pulses are emitted at a rate slightly greater or less than the vocal vibrational frequency in order to sample different points in the vibratory cycle (Colton et al., 2011, p. 228). This makes the vocal folds appear to be vibrating in slow motion, as different phases of the vibratory cycle are captured across multiple periods and concatenated into a single video stream (Deliyski et al., 2008). Videostroboscopy is useful for detecting vocal fold vibratory

asymmetry (phase or amplitude), abnormal glottal closure, presence and regularity of the mucosal wave, supraglottic compression, and organic pathologies such as lesions (Hsiao, Liu, Hsu, Lee, & Lin, 2001; Morrison, Rammage, & Emami, 1999; Morrison et al., 1986).

One drawback of stroboscopy, however, is that aperiodic vibrations may cause the strobe light to become asynchronized with vocal fold movements. Since even healthy voices are considered quasiperiodic at best (Rabiner, 1977), many individuals are unable to reap the benefits of stroboscopy for laryngeal imaging. Reports state that 17–63% of recordings are considered invalid due to an inability of the strobe light to synchronize with the $f_o$ of the vocalization (Patel, Dailey, & Bless, 2008; Woo, Colton, Casper, & Brewer, 1991).

*High-speed Videoendoscopy*

Laryngeal high-speed videoendoscopy (HSV) is an alternative technique for assessing vocal fold vibratory function. This method uses high-speed (≥1000 fps) endoscopic imaging techniques to capture an accurate representation of the true vibratory motion of the vocal folds (Powell et al., 2016). Laryngeal HSV can be performed using a rigid or flexible endoscope, both of which enable laryngeal visualization. Rigid endoscopes are a popular choice for providing a minimally distorted, brightly illuminated view of the laryngeal and pharyngeal anatomy. Since the scope is inserted through the mouth, only sustained vowels can be captured using this method. Flexible endoscopes may be inserted transnasally such that the vocal folds can be examined during connected speech. However, flexible scopes typically provide less light than their rigid counterparts

and are susceptible to field distortion (Eller et al., 2008; Popolo, 2017). Because flexible endoscopes require a small barrel diameter to successfully pass through the nasal passages and pharynx, the camera lens at the end of the scope must be a wide-angle lens. As a result of this "barreling effect," objects may appear overly rounded or bent compared to those observed using a rigid endoscope (Eller et al., 2008). The use of flexible or rigid endoscopes may be preferred depending on the patient population, desired stimuli (e.g., sustained vowels versus connected speech), and intended recording parameters (e.g., frame rate, color).

Laryngeal HSV has been applied to investigate vocal fold vibratory characteristics across individuals with and without voice disorders (Döllinger et al., 2012; Patel et al., 2008; Samlan, Kunduk, Ikuma, Black, & Lane, 2018; Tsuji et al., 2014), as well as before and after surgical intervention (Mehta et al., 2012b; Powell et al., 2019). Because laryngeal HSV does not depend on the synchronization of light with the estimated $f_o$ of the vocalization (Patel et al., 2008), even voices characterized by aperiodicity (i.e., with an undistinguishable or inconsistent $f_o$) can be captured and analyzed. The temporal resolution of laryngeal HSV is advantageous over that of videostroboscopy for identifying nonstationary laryngeal dynamics such as phonatory offset or onset (Deliyski et al., 2008). Despite the promise of laryngeal HSV, however, its adoption rate in voice clinics remains low. This is likely due to high cost, low spatial resolution, and short recording duration compared to videostroboscopy, as well as a lack of a commercially available system to purchase.

*Features of Excessive Laryngeal Muscle Tension*

Under continuous light, observable features that may relate to excessive tension include posterior glottic opening (gap), organic pathologies (vocal fold nodules, polyps), diffuse and/or localized laryngeal edema or swelling, false vocal fold adduction, irregular degree and/or symmetry of arytenoid excursion during vocal fold adduction and abduction, and supraglottic hyperfunction in the anteroposterior and mediolateral planes (Aronson et al., 2009; Morrison et al., 1986, p. 154). Videostroboscopy and HSV enable the visualization of vocal fold vibratory patterns; vibratory cycles that are predominately closed phase, irregular in terms of vocal fold phase symmetry and/or periodicity, and exhibit an interruption of the mucosal wave are among features often associated with excessive laryngeal tension (Aronson et al., 2009, p. 155). Perhaps most importantly, videostroboscopy and HSV are useful for assessing supraglottic compression as a feature of excessive supralaryngeal muscle activation.

Supraglottic compression (also referred to as *supraglottic constriction*) refers to the degree of tightening of the supraglottic structures (Patel et al., 2018; Poburka, Patel, & Bless, 2017), and is typically associated with temporary obstruction of the view of the true vocal folds. Anterior-to-posterior supraglottic compression (A-P compression) occurs when the arytenoid cartilages are drawn toward the petiole of the epiglottis, whereas medial compression occurs as the adduction of the false vocal folds (FVF compression; Stager et al., 2000). Supraglottic compression is a clinical feature that may be observed in those with voice disorders characterized by excessive laryngeal muscle tension, such as MTD (Behrman, Dahl, Abramson, & Schutte, 2003; Garaycochea,

Navarrete, del Río, & Fernández, 2019; Morrison et al., 1993; Ogawa et al., 2005; Stager et al., 2000) and vocal fold nodules (Behrman et al., 2003). Because supraglottic compression can vary across speech sounds (Dabirmoghaddam et al., 2019), assessing supraglottic compression as "present" or "absent" may not be a useful metric. Instead, the *degree* of supraglottic compression present in the laryngeal mechanism may be more useful.

During videostroboscopic or HSV examinations, supraglottic compression can be assessed using a standardized scale. Indeed, ratings of supraglottic compression have been shown to not statistically significantly differ between stroboscopic imaging at 30 fps and HSV at 4000 fps (Zacharias, Deliyski, & Gerlach, 2018). However, the Voice-Vibratory Assessment with Laryngeal Imaging (VALI; Poburka et al., 2017) scale trains clinicians to make reliable visual-perceptual judgments of supraglottic compression, among other features of vibratory motion (e.g., amplitude, mucosal wave, phase closure) based on imaging modality. Inter- and intrarater reliability were greater for HSV analysis of A-P and FVF compression, ranging from $ICC$ = .85–.89 for interrater and $r_s$ = .28–.84 for intrarater reliability; this is compared to $ICC$ = .75–.93 for interrater and $r_s$ = .19–.39 for intrarater reliability when using stroboscopy.

More quantitative estimates of supraglottic compression have also been attempted. Behrman et al. (2003) quantitatively assessed the normalized pixel area of the glottis to assess A-P and FVF compression, ultimately observing statistically significant differences in A-P compression between individuals with and without MTD. However, Stepp et al. (2010a) later carried out this methodology to compare the pixel-based

estimates of supraglottic activity to visual-perceptual judgments obtained using a 5-point Likert scale. Results of this study showed a lack of significant correlations between visual-perceptual and quantitative measures of A-P and FVF compression. The authors also identified crucial errors in the theoretical bases from Behrman et al. (2003) that were used to obtain the pixel-based estimates of supraglottic activity. Thus, although assessing supraglottic compression can provide insight into the degree of laryngeal muscle tension present in the system, further research is needed to develop valid, quantitative methods to comprehensively estimate supraglottic compression.

**Aerodynamics**

The assessment of aerodynamics includes measurement of air volume, flow and pressure during phonation. Subglottal air pressure, glottal flow rate, phonation threshold pressure, and maximum phonation time are commonly extracted to monitor change of voice, identify laryngeal abnormalities, and describe laryngeal function (Mehta & Hillman, 2008; Scherer, 1991). Invasive aerodynamic techniques (tracheal puncture, transnasal pressure transducer) allow for direct collection of aerodynamic measures. Yet these methods are not ideal since they can be painful, time-consuming, and/or uncomfortable for the patient. Non-invasive techniques, on the other hand, indirectly assess glottal parameters by capturing intraoral air pressure (e.g., via subject- or mechanical-controlled labial interruptions to create a continuous, enclosed system) or circumferential changes in the rib cage and abdomen (i.e., inductance plethysmography).

Investigations into the validity and reliability of indirect aerodynamics measures to clinically assess laryngeal muscle tension is ongoing. Many studies highlight indirect

measures of *subglottal pressure*, a driving pressure that builds up below the adducted vocal folds until it exceeds the resistance of the folds and sets them into oscillatory motion. Subglottal pressure is thought to be increased in those with a voice disorder characterized by excessive laryngeal tension, increased vocal fold stiffness, and abnormal vocal fold adduction (Hillman et al., 1989; Hillman, Montgomery, & Zeitels, 1997; Netsell, Lotz, & Shaughnessy, 1984). For instance, subglottal pressure estimates of typical speakers have been shown to differ individuals with VH (Dastolfo, Gartner-Schmidt, Yu, Carnes, & Gillespie, 2016; Espinoza et al., 2017; Hillman et al., 1989; Holmberg, Doyle, Perkell, Hammarberg, & Hillman, 2003; Kuo, Holmberg, & Hillman, 1999; Zheng et al., 2012) and Parkinson's disease (Jiang et al., 1999b; Murdoch, Manning, Theodoros, & Thompson, 1997). Studies have also reported increases in subglottal pressure during intentional modulations in vocal effort of typical speakers (Lien, Michener, Eadie, & Stepp, 2015b; McKenna, Llico, Mehta, Perkell, & Stepp, 2017; Rosenthal, Lowell, & Colton, 2014).

Previous work indicates a strong relationship between indirect and direct measures of subglottal pressure (Bard, Slavit, McCaffrey, & Lipton, 1992; Hertegard, Gauffin, & Lindestad, 1995; Löfqvist, Carlborg, & Kitzing, 1982), which would suggest that indirect, non-invasive measures could be used in clinical assessments of subglottal pressure. Subglottal pressure is often measured indirectly as the intraoral air pressure of a vowel when produced subsequent to a bilabial stop consonant (e.g., /pi/; Löfqvist et al., 1982; McKenna et al., 2017). This configuration requires the velopharyngeal port and lips to be closed, but the vocal folds remain abducted such that pulmonary pressure may

equalize above and below the glottis at the time of lip opening (i.e., to initiate the vowel). Moreover, work from Plant and Hillel (1998) argues that assessing intraoral pressure during the production of a bilabial stop consonant may not accurately reflect changes in subglottal pressure of the corresponding vowel. Thus, mechanical interruptions (e.g., balloon valve) during the production of a sustained vowel may be used as an alternative to the labial method. Lamb, Schultz, Scholp, Wendel, and Jiang (2020) recently showed that the mechanical method led to a significantly greater test-retest reliability of subglottal pressure estimates than the labial method. This is likely because subglottal pressure estimates obtained via the labial method are subject to variability from human inconsistencies, as the participant controls when their lips open and close. Despite work showing that the mechanical method leads to more reliable estimates of subglottal pressure, the labial method remains the most commonly used technique. As a result, further research must be carried out to standardize a non-invasive method that both validly and reliably assesses subglottal pressure.

**Electroglottography**

Electroglottography captures the electrical conductance across two electrodes placed on either side of the thyroid cartilage. Electrical conductance is greater when the vocal folds are contacting than when the glottis is open; as such, changes in vocal fold contact area during phonation alter the captured conductance to provide insight into the glottal source (Askenfelt et al., 1986; Childers, Hicks, Moore, & Alsaka, 1986; Herbst, 2019).

Electroglottography has been explored as a non-invasive measure of intravocal

fold impact stress (Peterson, Verdolini-Marston, Barkmeier, & Hoffman, 1994; Verdolini, Chan, Titze, Hess, & Bierhals, 1998). This method assesses the proportion of time that the vocal folds are closing or opening with respect to the pitch period of the vocalization. The proportion of time that the vocal folds are *closing* in a single pitch period ("closed quotient") has been shown to positively correlate with vocal fold impact stress. This closed quotient has also been shown to distinguish individuals with MTD from those with typical voices (Hosokawa et al., 2012; Ogawa et al., 2014) and from those with vocal fold lesions (e.g., nodules; Hosokawa et al., 2012). The proportion of time that the vocal folds are *open* in a single pitch period ("quasi-open quotient") has also been investigated as a possible discriminatory metric, but with little success (Szkiełkowska, Krasnodębska, Miaśkiewicz, & Skarżyński, 2018).

Overall, electroglottography shows some promise for evaluating vocal fold impact stress. Yet further investigations must be undertaken to comprehensively assess the utility of electroglottography for assessing laryngeal muscle tension, as research in this area is sparse. At present, the limitations of using electroglottography may outweigh its benefits for assessing tension. For instance, this method suffers from speech-induced artifacts caused by vertical movements of the larynx—wherein the vocal folds move in and out of the field captured by the neck electrodes—as well as contractions from the neck muscles (Colton & Conture, 1990). These artifacts may cause considerable variability in electroglottographic waveform shape. Yet waveform shape may also be affected by errors in detecting vocal fold tissue contact, irregular vocal fold vibratory motion, or mucous on/around the vocal folds (Childers, Hicks, Moore, Eskenazi, & Lalwani, 1990). Since

there are no standardized methods of using electroglottography to assess laryngeal muscle tension, care must be exercised to avoid results that are confounded by these instrumental-, participant-, and speech-related artifacts.

**Electromyography**

Electromyography (EMG) is a technique that captures the electrical activity of muscles. In this method, sensors placed intramuscularly (laryngeal EMG) or at the surface of the skin (surface EMG) capture depolarized zones of muscle fibers during muscle contractions.

*Laryngeal EMG*

Laryngeal EMG (LEMG) has been used to examine how combinations of the five major intrinsic laryngeal muscles are involved in voice control (Gay, Hirose, Strome, & Sawashima, 1972; Hirano & Ohala, 1969a; Poletto, Verdun, Strominger, & Ludlow, 2004), and to gain understanding of laryngeal biomechanics (Hast, 1966, 1967a; Hirano et al., 1969a; Hirano, Ohala, & Vennard, 1969b). For instance, a large focus of LEMG-based research has been to elucidate the activity of the CT and TA muscles as they pertain to vocal pitch and intensity; work in this area has shown that the contraction force of the CT and TA muscles jointly increase with increases in vocal pitch and intensity (Lindestad, Fritzell, & Persson, 1991). Further research has implicated these muscles in reflexive control of voice $f_o$ (Liu, Behroozmand, Bove, & Larson, 2011). Unfortunately, there has been no objective, quantitative evidence of increased intrinsic laryngeal muscle in those diagnosed a voice disorder characterized (via other assessment methods) by excessive laryngeal muscle tension compared to typical speakers. This is likely, in part,

because perturbing the intrinsic laryngeal muscles and structures using LEMG could alter typical muscle function.

*Surface EMG*

Surface EMG (sEMG) is a non-invasive alternative that captures myoelectric activity via electrode(s) placed on the surface of the skin. Although sEMG is not able to detect sufficient activity from the intrinsic laryngeal muscles due to their distance from the surface of the skin, sEMG is able to detect activity from the extrinsic laryngeal musculature. One drawback of using sEMG to examine anterior neck musculature is that the small, overlapping, and interdigitated nature of these muscles makes it difficult to isolate the electrical activity of a single targeted muscle. Surface EMG does exhibit some key advantages, however, as this technique is non-invasive, easy to implement, and provides an objective view of myoelectric activity.

There has been some evidence to suggest that sEMG is useful for assessing extrinsic laryngeal muscle tension. For instance, Redenbaugh et al. (1989) reported increased myoelectric activity of the anterior neck muscles (likely comprising the sternohyoid and omohyoid; Loucks, Poletto, Saxon, & Ludlow, 2005) in speakers with MTD when directly compared to vocally healthy controls. However, the method of data collection in this study was relatively primitive, with sEMG activity processed in real-time and on-screen values recorded by hand. In a similar study, Hočevar-Boltežar et al. (1998) reported increased myoelectric activity of the perioral and anterior neck muscles in speakers with MTD both before and during phonation. Yet it must be noted that sEMG activity was not normalized before comparing activity levels across conditions and

participants in either study. Both studies used root-mean-squared amplitude to compare sEMG activity across electrode sites.

Yet there have also been reports that sEMG activity is not distinct between speakers with hyperfunctional voices and vocally healthy controls. Work from Stepp et al. (2011b) compared myoelectric activity of the thyrohyoid, omohyoid, sternohyoid, and CT[1] muscles between normal speakers and both singers and non-singers with vocal nodules. The authors indicated that muscle activity was not significantly different between groups, but suggested a potential use of sEMG for identifying inappropriate phonatory behaviors in individuals with vocal nodules (e.g., increased activity of the extrinsic laryngeal musculature prior to phonation). In a separate study, Stepp et al. (2011a) assessed the utility of sEMG for detecting changes in myoelectric activity across a session of voice therapy in individuals with VH, ultimately reporting a lack of reliable changes over the singular session of therapy. Contrary to findings from Redenbaugh et al. (1989), a more recent study from Van Houtte et al. (2013) that sEMG activity was not discriminable between individuals with MTD and vocally healthy controls. The authors pointed out that the type of electrodes, pathophysiology and etiology of the disorder, and the emotion state of the participant were all confounding factors.

As an alternative to comparing differences in sEMG amplitude, some studies have explored the utility of beta coherence as a metric for evaluating extrinsic laryngeal muscle activity. Beta coherence is thought to represent the transmissions from the

[1] The authors acknowledged that it was unlikely for the cricothyroid to have contributed to the resulting myoelectric signal since it is a deep, intrinsic laryngeal muscle.

primary motor cortex to spinal motor neurons (Salenius, Portin, Kajola, Salmelin, & Hari, 1997). Neck intermuscular beta coherence (NIBcoh) could therefore elucidate patterns of EMG activity between electrode positions as a means of assessing potential imbalances between laryngeal muscle forces. Stepp, Hillman, and Heaton (2010c) investigated NIBcoh by targeting the activity of the 1) thyrohyoid, omohyoid, and sternohyoid, and 2) sternohyoid (contralateral) and cricothyroid. The authors found that mean NIBcoh was significantly reduced in speakers with vocal nodules when directly compared to vocally healthy controls. In a similar study comprising vocally healthy speakers, Stepp, Hillman, and Heaton (2011c) saw a significant reduction in NIBcoh when speakers were instructed to mimic a strained, hyperfunctional voice. Although NIBcoh shows promise for distinguishing between typical and hyperfunctional voices, more work must be done to determine the sensitivity and specificity of this metric across different *degrees* of laryngeal muscle tension.

Surface EMG is a non-invasive method that can easily be incorporated into clinical assessments of laryngeal muscle tension. Yet it must be noted that many of these studies regard reduced muscle activity as a reduction in muscle tension even though sEMG activity does not provide a direct measure of muscle tension (Roberts & Gabaldón, 2008). This is because muscle tension is a function of muscle length and velocity, whereas the sEMG signal only comprises myoelectric activity near the electrode site. NIBcoh, on the other hand, provides insight into muscle imbalance rather than directly measuring muscle tension. Additionally, variations in study design and sensor configuration (e.g., type and placement) make it difficult to directly compare the findings

of studies utilizing sEMG to assess extrinsic laryngeal muscle tension. For these reasons, the use of sEMG must be critically evaluated in its use for assessing extrinsic laryngeal muscle tension.

**Accelerometry**

Accelerometry for voice assessment includes placing a vibratory transducer at the base of the neck, just superior to the sternal notch. Neck-placed accelerometers act as vibratory transducers that sense phonation-related neck-skin vibration, and as a result, are not coupled to the air like microphones (Hillman, Heaton, Masaki, Zeitels, & Cheyne, 2006; Zanartu et al., 2009). These devices are advantageous in monitoring phonation as accelerometers are relatively impervious to environmental noise and are less affected by supraglottal vocal tract resonances and aspiration noise (Cheyne, Hanson, Genereux, Stevens, & Hillman, 2003). Specifically, neck-surface accelerometers are thus capable of passively rejecting airborne sound that may be undesirable in acoustic analysis (e.g., ambient room noise; Coleman, 1988). Still, the signals collected via accelerometers and microphones both capture information of the glottal source: accelerometer-derived measures of voice $f_o$, jitter (cycle-to-cycle perturbations in cycle period), and cepstral peak prominence strongly correlate to those measures when collected via a microphone signal (Mehta, Van Stan, & Hillman, 2016; Szabo, Hammarberg, Hakansson, & Sodersten, 2001).

Prior work has shown that neck-surface accelerometry can be used to differentiate between typical and hyperfunctional voices, as well as to track changes in vocal effort over time. When used in ambulatory voice monitoring, Ghassemi et al. (2014) observed

that measures of sound pressure level and voice $f_o$ from the neck-surface accelerometer signal distinguished 22 of 24 speakers (12 typical speakers and 12 speakers with phonotraumatic VH). Cortés et al. (2018) used an accelerometer signal to derive parameters relating to glottal airflow. The authors tested this approach in 48 speakers with and 48 speakers without phonotraumatic VH, demonstrating that seven parameters relating to peak-to-peak glottal airflow (i.e., mean, minimum, median, logarithm versions of mean, minimum, median, and kurtosis) and the difference between the magnitudes of the first two harmonics could be used to classify whether or not a speaker was diagnosed with phonotraumatic VH. A recent study by Van Stan et al. (2020) used a larger group of speakers than these previous studies to further identify measures for discriminating daily vocal behaviors between typical speakers and those with phonotraumatic VH. The authors compared week-long patterns of voice use between 90 typical speakers and 90 speakers with phonotraumatic VH and demonstrated that phonotraumatic VH was related to higher sound pressure levels and more abrupt glottal closure. A negatively skewed sound pressure level supported clinical impressions that speakers with phonotraumatic VH talk louder more often (rather than on average) than typical speakers. The authors also hypothesized that more abrupt glottal closure may be the result of hyperadduction to compensate for glottal insufficiency.

Neck-surface accelerometry has also been related to aerodynamic measures of subglottal pressure. Specifically, McKenna et al. (2017) demonstrated that the magnitude of the neck-surface accelerometer signal was related to subglottal pressure (non-invasively derived via intraoral estimates), as well as speaking intensity during variations

in vocal effort. In a separate study, McKenna et al. (2018a) demonstrated that a neck-surface accelerometer-derived measure of subglottal pressure was significantly predictive of self-ratings of vocal effort, wherein a greater rating of effort corresponded to higher subglottal pressure.

Together, these findings suggest that neck-surface accelerometry may be useful in the assessment of laryngeal muscle tension. However, further investigations must be conducted to identify a single estimator of laryngeal muscle tension. Unfortunately, signal quality is highly affected by accelerometer sensor placement, thickness of neck adipose tissue, and movement distortions (Behrman, 2005; Mehta, Zanartu, Feng, Cheyne, & Hillman, 2012a; Popolo, Svec, & Titze, 2005).

**Acoustics**

Acoustic analysis is a common method employed in voice assessments since data can be non-invasively collected via a microphone. Speech signals captured using a microphone can be analyzed to provide insight into basic metrics of the glottal source (timing, frequency, and amplitude of vocal fold vibration) and phonation (e.g., $f_o$, vocal intensity, and phonation duration). There have been numerous attempts to identify acoustic metrics that relate to muscle tension in the laryngeal mechanism. To date, however, a single acoustic indicator specific to laryngeal muscle tension does not exist.

*Time- and Amplitude-based Measures*

Clinical assessments of laryngeal muscle tension typically comprise measures of sound pressure level and $f_o$, as well as acoustic perturbation measures (e.g., jitter, shimmer, harmonics-to-noise ratio; Aronson et al., 2009, pp. 236-43; Colton et al., 2011,

p. 239; Mehta et al., 2016). These current clinical assessment methods objectively quantify vocal function as it relates to vocal loudness, pitch, and quality.

Vocal Sound Pressure Level

Vocal sound pressure level (SPL) is a measure of radiated power from the mouth ("vocal power") and reflects properties of the voice source and vocal tract (Gramming, 1991; Švec & Granqvist, 2018). SPL is quantified in decibels (dB SPL) relative to a known reference for sound pressure in a given medium (e.g., air). Prior work indicates a positive relationship between vocal SPL and both perceived loudness and effort of the voice (Brandt et al., 1969; Rosenthal et al., 2014). Moreover, mean SPL is a significant factor in self- and listener-perceptual ratings of vocal effort (McKenna & Stepp, 2018b). Yet there is evidence to suggest that the relationship between mean SPL and vocal effort is different between individuals with and without a voice disorder characterized by excessive laryngeal tension. For instance, there is a strong relationship between subglottal pressure and vocal SPL in typical speakers (Baker, Ramig, Sapir, Luschei, & Smith, 2001; Bjorklund & Sundberg, 2016; Fryd, Van Stan, Hillman, & Mehta, 2016; Marks, Lin, Fox, Toles, & Mehta, 2019); however, those with voice disorders characterized by excessive laryngeal muscle tension have been shown to exhibit increases in subglottal pressure without concurrent increases in vocal SPL (Hillman et al., 1989). Research suggests that individuals with stiff or heavy vocal folds (as in phonotraumatic VH) may leverage subglottal pressure to improve vocal fold vibratory amplitude (Zanartu et al., 2014), leading to increased subglottal pressure and perceived vocal effort *without* increasing mean SPL out of the normative range (Espinoza et al., 2017; Friedman,

Hillman, Landau-Zemer, Burns, & Zeitels, 2013; Hillman et al., 1989; Zeitels, Burns, Lopez-Guerra, Anderson, & Hillman, 2008).

The specific relationship between vocal SPL and laryngeal muscle tension is less clear. One study found that individuals with MTD spoke at a significantly lower vocal SPL than typical speakers (66.95 dB SPL in MTD versus 68.37 dB SPL in typical voices; Belsky, Rothenberger, Gillespie, & Gartner-Schmidt, 2020). On the other hand, Van Stan et al. (2015a) determined that vocal SPL is not significantly different between speakers with and without phonotraumatic lesions (84.4 dB SPL in phonotraumatic VH versus 83.6 dB SPL in typical voices). The differences observed between phonotraumatic and non-phonotraumatic VH may reflect differences in phonatory mechanisms associated with the disorder. Although both phonotraumatic and non-phonotraumatic VH have been linked to lower vocal efficiency compared to typical speakers, phonotraumatic VH is also associated an increased likelihood of trauma to the vocal fold tissues (Espinoza et al., 2017). Yet the discrepancy in average SPL values between these studies also indicates that methodology is a crucial factor to consider when interpreting results: Van Stan and colleagues (2015) obtained data from ambulatory monitoring to enhance ecological validity, whereas Belsky et al. (2020) collected data while participants wore a facemask to simultaneously collect acoustic and aerodynamic data via the KayPENTAX PAS Model 6600. The notion of differences in methodology is also supported in the fact that the reported SPL means of typical speakers from Belsky et al. (2020) are lower compared than normative adult values when using the KayPENTAX PAS Model 6600 (Zraick, Smith-Olinde, & Shotts, 2012). Vocal SPL is a useful measure to consider when

evaluating voice disorders; however, the relationships between vocal SPL and both vocal effort and laryngeal muscle tension must be interpreted with caution.

Fundamental Frequency

Voice $f_o$ is based on the length, mass, and tension of the vocal folds (Van Den Berg, 1958). Increases in mean $f_o$ have been associated with increased activity of the TA (Titze, Luschei, & Hirano, 1989) and CT (Löfqvist, Baer, McGarr, & Story, 1989) muscles, which are known to alter the configuration of the vocal folds (see *Intrinsic Laryngeal Muscles*). This relationship, however, is not straightforward. When stimulated in isolation, the TA and CT muscles have each been associated with increases in mean $f_o$, whereas concurrent TA and CT stimulation results in antagonistic activity that may result in either no change or decreases to mean $f_o$ (Chhetri, Neubauer, & Berry, 2012; Lowell & Story, 2006; Titze et al., 1989; Yin & Zhang, 2013). The extrinsic laryngeal muscles (primarily the sternothyroid and thyrohyoid) also influence voice $f_o$ by altering the position of the hyoid and thyroid cartilages (see *Extrinsic Laryngeal Muscles*). Previous work has shown that speakers with VH exhibit increased values of mean $f_o$ over the course of the day (Ghassemi et al., 2014), but reduced mean $f_o$ following a successful course of vocal therapy (Kennard, Lieberman, Saaid, & Rolfe, 2015; Roy et al., 1997). Despite clear connections between laryngeal muscle tension and voice $f_o$, a wide range of $f_o$ values have been reported across the spectrum of vocal function (Holmberg et al., 2003; Mehta et al., 2015; Redenbaugh et al., 1989; Van Stan et al., 2015a); as such, voice $f_o$ is a non-specific indicator of laryngeal muscle tension.

Perturbation Measures

Jitter, shimmer, and harmonics-to-noise ratio (HNR) are common perturbation measures included in acoustic voice assessments. Jitter refers to cycle-to-cycle perturbations in $f_o$, and is a measure of frequency instability. Shimmer, on the other hand, refers to cycle-to-cycle perturbations in amplitude, and is thus a measure of amplitude instability. HNR is an indication of the ratio of harmonic energy to noise in the speech signal. These measures have the potential to provide an estimate of quality of sustained vowel productions. However, prior work indicates poor test-retest reliability of these measures in dysphonic voices, with *ICC* = .46 for jitter, .40 for shimmer, and .33 for HNR (Carding et al., 2004). Test-retest reliability was substantially greater for non-dysphonic voices, with moderate reliabilities of *ICC* = .73 for jitter, .55 for shimmer, and .68 for HNR. Yet these values may be influenced by speaking voice pitch (measured as $f_o$) and loudness (measured as vocal SPL), as increased $f_o$ and SPL have each been shown to result in reduced jitter and shimmer (Brockmann-Bauser, Bohlender, & Mehta, 2018; Brockmann-Bauser & Drinnan, 2011; Brockmann, Drinnan, Storck, & Carding, 2011; Gelfer, 1995) and increased HNR (Brockmann-Bauser et al., 2018). As such, these measures may give clinically useful measurements of subtle changes in non-dysphonic or mildly dysphonic patients; however, the reliability of these measures grows more questionable in more dysphonic voices.

Although jitter, shimmer, and HNR may provide some clinical insights into phonatory quality, the clinical utility of these measures across a broad range of vocal function is low. Titze (1995) describes three types of voice signals encountered in

acoustic voice analysis: Type I, which are nearly periodic signals; Type II, or signals with strong subharmonics and/or modulating frequencies that mask the presence of a single, obvious $f_o$; and Type III, which are signals with no apparent periodic structure. Because jitter, shimmer, and HNR each rely on the periodic structure of the acoustic signal, these measures are generally restricted for use with only Type I and some Type II signals. Yet the level of aperiodicity in a voice signal typically increases when voice problems are present (Eadie & Doyle, 2005; Titze, 1995), such that these perturbation measures cannot be reliably used to assess more dysphonic signals. As a result, these measures are often used in conjunction with other methods of voice assessment (e.g., laryngoscopy) to provide a more comprehensive evaluation of the voice.

*Spectral- and Cepstral-based Measures*

Laryngeal muscle tension has been related to a collection of spectral- and cepstral-based measures, including the low-to-high ratio of spectral energy (L/H ratio), standard deviation of the L/H ratio (L/H ratio SD), cepstral peak prominence (CPP), the standard deviation of CPP (CPP SD), and the cepstral spectral index of dysphonia (CSID). Spectral and cepstral measures are both derived from the spectral distribution of acoustic energy. Spectral measures (L/H ratio, L/H ratio SD) reflect the distribution of spectral energy within the acoustic waveform, whereas cepstral measures (CPP, CPP SD, CSID) reflect the distribution of energy at harmonically related frequencies and the regularity of harmonic peaks (Hillenbrand & Houde, 1996).

The spectrum of the acoustic waveform is first obtained via the fast Fourier transform of the time-based signal; from here, the distribution of spectral energy across

frequency can be analyzed. The L/H ratio is a measure of spectral tilt, and is calculated by comparing the spectral energy below and above an established cut-off frequency. A popular cut-off frequency is 4 kHz, as it is thought that the energy above 2–3 kHz is susceptible to the effects of high-frequency aspiration noise stemming from posterior glottal gap size (Klatt & Klatt, 1990). Larger posterior glottal gap sizes are associated with an increased escape of turbulent air, resulting in increased energy above 4 kHz and a voice that may be perceived as excessively breathy (Hillenbrand, Cleveland, & Erickson, 1994; Hillenbrand et al., 1996). The L/H ratio is a promising metric since hyperfunctional voices (Holmberg et al., 2001; Patel, Liu, Galatsanos, & Bless, 2011) have been considered as being perceptually breathy (although breathiness is more often associated with hypofunctional voice disorders; Watts & Awan, 2011). Indeed, the L/H ratio at a cut-off frequency of 4 kHz has been shown to successfully distinguish perceptually dysphonic speakers from typical speakers (Awan, Roy, Jette, Meltzner, & Hillman, 2010; Lowell, Colton, Kelley, & Mizia, 2013).

Employing the L/H ratio shows promise as an indicator of laryngeal muscle tension, the specific relationship between tension and the L/H ratio is unclear. Recent work suggests that the L/H ratio and the L/H ratio SD cannot be used to distinguish typical speakers from those diagnosed with MTD (Belsky et al., 2020). Additional work comparing L/H ratio SD values between normative and dysphonic speakers found minimal differences (Lowell et al., 2013). It is possible that these findings are due to the non-specificity of the L/H ratio, as this metric may also be affected by vocal fold vibratory characteristics that—in addition to increased aspiration noise—could reduce the

ratio. For instance, a lower L/H ratio could result from increased energy in the upper harmonics *or* from increased energy from aspiration noise.

Issues employing the L/H ratio to assess voice depends on the cut-off frequency used. Prior work examining the L/H ratio before and after behavioral therapy in PD found that the L/H ratio at a cut-off frequency of 4 kHz was not sufficient to differentiate the effects of behavioral therapy in PD (Alharbi, Cannito, Buder, & Awan, 2019). The authors adjusted the cut-off frequency to 2 kHz for males and 2.5 kHz for females to focus on the first two formant frequencies in the spectrum; by using an adjusted L/H ratio, a significant difference was found before and after behavioral therapy (LSVT) in PD. It is unclear whether this change was a result of actually honing in on the first two formant frequencies, or if other characteristics of vocal fold vibration and aspiration noise were differentiated because of the change. Specifically, the third formant hovers around 1.7–3 kHz for adult men and 1.9–3.4 kHz for adult women, because it varies according to the target phoneme (Hillenbrand, Getty, Clark, & Wheeler, 1995). This range can impact the distribution of energy and inaccurately inflate high-frequency energy in some cases. Additionally, the L/H ratio is a non-specific measure of dysphonia; for instance, a decrease in the L/H ratio could be the result of an increase in breathiness (to increase high-frequency spectral energy) or in pressed voice (due to higher harmonics). Thus, although L/H ratio and L/H ratio SD may provide some insight into the spectral composition of voices characterized by excessive laryngeal muscle tension, these metrics cannot distinguish the contributions from specific physiological mechanisms.

The cepstrum is usually obtained by computing the fast Fourier transform of the

logarithm of the power spectrum. In other words, the cepstrum is the spectral representation of the spectrum. The dominant peak in the cepstrum corresponds to the fundamental period of the spectrum, just as the dominant frequency of the spectrum corresponds to the $f_o$ of the voice signal. A highly periodic voice will have a strong peak at the $f_o$ of the original spectrum (and integer multiples of this $f_o$) and a strong peak at the fundamental period of the cepstrum (measured here in terms of "quefrency" instead of frequency). Cepstral analyses do not rely on specific time-based information from the acoustic signal (e.g., time between glottal pulses). Consequently, cepstral analyses are particularly useful when it is too difficult to extract time-based measures (e.g., voice $f_o$), such as in a severely dysphonic speaker.

CPP was first introduced as an acoustic correlate of breathiness by Hillenbrand et al. (1994), and was later modified by Hillenbrand et al. (1996) via adding smoothing operations to the temporal and cepstral domains ("smoothed" CPP, or CPPs). CPP is a measure of cepstral peak amplitude when normalized to overall amplitude; it is computed by (1) constructing a linear regression line through the cepstrum to represent average sound energy, (2) locating the peak in the cepstrum that exhibits the largest amplitude, and (3) calculating CPP as the distance between the magnitude of this peak and the regression line (Heman-Ackah et al., 2003; Hillenbrand et al., 1994). CPP is associated with the $f_o$ of the original spectrum and is affected by the periodicity of the original signal. A periodic voice signal will display a well-defined harmonic structure, and as a result, exhibit a more prominent cepstral peak than an aperiodic voice signal (Hillenbrand et al., 1994). For this reason, CPP demonstrates promise in the analysis of disordered

voices.

CPP has demonstrated use as an acoustic marker of dysphonia. Prior work indicates that CPP is a strong correlate to overall severity of dysphonia (Awan & Roy, 2005, 2009; Awan, Roy, & Dromey, 2009; Eadie & Baylor, 2006) and can successfully differentiate between typical speakers and those with disordered voices (Awan et al., 2005; Eadie et al., 2005; Heman-Ackah et al., 2014; Lowell et al., 2012a; Shim, Jung, Koul, & Ko, 2016). It has been reported that dysphonic voices may be characterized by lower CPP values (Awan et al., 2005; Awan et al., 2010; Shim et al., 2016) and more specifically, a strained voice quality (Lowell et al., 2012a) than typical speakers. However, the responsiveness of CPP to changes in laryngeal muscle tension remains equivocal. Work by Rosenthal et al. (2014) suggests that CPP *is* responsive to tension. When they instructed 18 typical speakers to modulate their level of vocal effort (comfortable, minimum, maximal), the authors noted that increased vocal effort led to an increase in CPP and decrease in CPP SD, and further, that decreased vocal effort resulted in a decrease in CPP and decrease in CPP SD. However, work from McKenna et al. (2018b) found that CPP was not salient to changes in vocal effort for speakers *or* listeners.

The conflicting results of these studies may be due to vocal SPL acting as a confounding factor: increases in vocal SPL have been shown to increase CPP values in male speakers (Awan, Giovinco, & Owens, 2012). Indeed, Rosenthal et al. (2014) male speakers were reported to carry out each speaking task at an average intensity 3–7 dB higher than female speakers in the current study. However, it is also unclear how task

instructions and cueing affected participant strategies for modulating their voice in these studies.

CPP and CPP SD have shown promise for assessing PD. A recent study by Alharbi et al. (2019) evaluated the ability of these metrics for differentiating between speaker voices pre-/post-LSVT. Both CPP ($p$ = .006) and CPP SD ($p$ = .007) significantly increased from pre- to post-treatment, indicating an effect of increased vocal intensity and/or improved periodicity of the voice in post-treatment recordings. Although the primary objective of LSVT is for individuals to focus on speaking with a clear, loud voice, only a weak correlation between vocal intensity and CPP was found. This suggests that vocal intensity may have contributed to increases in CPP and CPP SD, but that the significant increase in CPP and CPP SD were likely not a result of increases in vocal intensity alone. Overall, these results are in support of utilizing CPP and CPP SD to acoustically characterize voice changes in PD.

In addition to the L/H ratio and CPP, the Cepstral/Spectral Index of Dysphonia (CSID) has been used as an acoustic marker of dysphonia. The CSID is a composite index that combines CPP, CPP SD, L/H ratio, and L/H ratio SD into a weighted formula that takes speech stimuli (sustained vowel or continuous speech) and speaker sex into account (Awan & Roy, 2006; Awan et al., 2010; Watts & Awan, 2015). The CSID can differentiate between normative and dysphonic, strained voice samples (Lowell et al., 2012a; Shim et al., 2016; Watts et al., 2015), as well as between voice samples collected in speakers with PD pre- to post-behavior therapy (Alharbi et al., 2019). Moreover, CSID-estimated severity has been demonstrated to be a strong correlate of listener

perceptual ratings of severity in voice samples collected from dysphonic speakers before and after voice treatment (behavior, medical, and/or surgical; Peterson et al., 2013). As such, the CSID offers an objective, non-invasive means of quantifying dysphonia severity.

*Limitations with Current Acoustic Methods*

Given the non-invasive nature of acquiring acoustic signals, it may be a promising modality for developing an objective estimator of laryngeal muscle tension. However, current clinical acoustic measures are limited in their ability to specifically assess laryngeal muscle tension. This may be because the steps for developing and validating acoustic measures are often difficult, as acoustic measures require robust testing (i.e., sufficient methodology, user interfaces, and sample sizes) prior to their implementation for routine clinical use (Roy et al., 2013). In recent years, relative fundamental frequency has been suggested as an indicator of laryngeal muscle tension. This non-invasive, objective, acoustic measure is in the process of undergoing robust testing to evaluate and refine its ability to track changes in laryngeal muscle tension within an individual over time, as well as differentiate levels of laryngeal muscle tension across individuals. For these reasons, relative fundamental frequency may be a promising clinical outcome measure for tracking baseline laryngeal muscle tension.

## *Relative Fundamental Frequency (RFF)*

Relative fundamental frequency (RFF) has been proposed as an acoustic estimate of laryngeal muscle tension. RFF examines instantaneous $f_o$ of vocalic devoicing and voicing gestures preceding and following a voiceless consonant. Changes in $f_o$ during

these gestures are dependent on the vibration of the vocal folds; this vibratory rate, in turn, is a function of vocal fold length, mass, and tension (Van Den Berg, 1958). Changes in RFF during vocalic devoicing and voicing gestures have been shown to form a characteristic pattern: Voicing offset in typical speakers is characterized by a relatively stable, if not slightly decreasing, RFF trend (Goberman & Blomgren, 2008; Robb & Smith, 2002), whereas voicing onset is marked by a decreasing RFF trend (Robb et al., 2002; Watson, 1998). The formation of this pattern has been attributed to interactions of laryngeal muscle tension, vocal fold kinematics, and changes in airflow (Löfqvist et al., 1989; Stepp, Merchant, Heaton, & Hillman, 2011d; Stevens, 1977; Van Den Berg, 1958; Watson, 1998). There is evidence to support that laryngeal muscle tension is transiently elevated before, during, and after the production of the voiceless consonant to inhibit vocal fold vibration (Löfqvist et al., 1989; Stevens, 1977). As glottal abduction commonly occurs for voiceless sounds, it is postulated that vocal fold abductory kinematics act in concert with elevated muscle tension to achieve devoicing during the transition into the voiceless consonant (Watson, 1998); however, the specific contributions of abduction to RFF are unclear. The transition out of the voiceless consonant is then thought to occur as a result of the interplay between increases in laryngeal muscle tension and airflow from the preceding vowel (Watson, 1998), in addition to vocal fold adductory kinematics necessary to bring the vocal folds together and reinitiate vibration.

### *RFF in Clinical Populations*

Recent work has shown that RFF correlates with severity of vocal symptoms in speakers with dysphonia (Eadie & Stepp, 2013) and can distinguish between typical voices and those characterized by excessive laryngeal muscle tension, including PD (Goberman et al., 2008; Stepp, 2013), adductor laryngeal dystonia (Eadie et al., 2013), and VH (Heller Murray et al., 2017; Stepp, Hillman, & Heaton, 2010b; Stepp et al., 2011d). Specifically, individuals with voice disorders characterized by excessive laryngeal muscle tension have lower average RFF values than age-matched typical speakers, perhaps due to increased baseline laryngeal muscle tension that impedes their ability to use tension as a strategy for devoicing (voicing offset) and reinitiating voicing (voicing onset). RFF also normalizes in individuals with VH following voice therapy (i.e., functional intervention; Stepp et al., 2011c), but not in individuals with vocal nodules or polyps following therapy (i.e., structural intervention; Stepp et al., 2010), suggesting that RFF is reflects functional changes to the voice rather than structural changes. When employing RFF to investigate functional differences between phonotraumatic and non-phonotraumatic VH, RFF values were found to significantly differ between the two subtypes, indicating that differences between phonotraumatic and non-phonotraumatic VH may be functional as well as structural (Heller Murray et al., 2017). Overall, these results support the use of RFF for differentiating the degree of laryngeal muscle tension between individuals, as well as tracking changes in laryngeal muscle tension within an individual over time.

RFF may also demonstrate utility in quantifying the degree of laryngeal muscle

tension present. Vocalic cycles closest to the voiceless consonant are the most sensitive to changes in laryngeal muscle tension (Stepp et al., 2010b; Stepp et al., 2011d) and are also predictors of laryngeal stiffness (McKenna, Heller Murray, Lien, & Stepp, 2016), a proposed biomechanical correlate of intrinsic laryngeal muscle tension (Shiller, Laboissiere, & Ostry, 2002). RFF also significantly correlates with listener perceptual judgements of dysphonia severity (Roy, Fetrow, Merrill, & Dromey, 2016; Stepp, Sawin, & Eadie, 2012), and can be manipulated by typical speakers to achieve values similar to those observed in individuals with voice disorders characterized by excessive laryngeal muscle tension (Lien et al., 2015b; McKenna et al., 2018b). With refinements to make the measure less time-consuming to calculate, RFF may be a useful clinical outcome measure in the assessment of laryngeal muscle tension.

### *Methods for Estimating RFF*

RFF can be calculated in two ways: manually or semi-automatically. Currently, the gold-standard method of RFF estimation is manual analysis. This requires a trained technician to use an acoustic software, such as Praat (Boersma, 2001), to visualize the acoustic signal in both time and frequency domains. The technician uses this information to locate the ten vocal cycles prior to and following the production of the voiceless consonant within a vowel–voiceless consonant–vowel (VCV) utterance. The specific steps that RFF technicians are trained to perform are as follows: (1) visual examination of the acoustic waveform during the transition from the vowel to the voiceless consonant (i.e., voicing offset) or from the voiceless consonant to the vowel (i.e., voicing onset), (2) identification of the vocal cycle that marks the boundary between voiced and voiceless

speech segments, (3) extraction of glottal pulses corresponding to the ten vocal cycles closest to the voiceless consonant, and (4) calculation of vocal cycle periods from the extracted glottal pulses, (5) calculation of the instantaneous $f_o$ of each vocal cycle via taking the inverse of each cycle period, and (6) calculating RFF by normalizing the instantaneous $f_o$ values to an approximate steady-state $f_o$ ($f_o^{ss}$), as in **Eq. 1.1**):

The $f_o^{ss}$ corresponds to the vocal cycle closest to the midpoint of each vowel; this corresponds to the $f_o$ of the first cycle for voicing offset RFF values and the tenth cycle for voicing onset RFF values. Resulting RFF values are in the normalized unit, semitones (ST), which allows for comparison within and across speakers.

$$\text{RFF (ST)} = 12 \times \log_2 \left( \frac{f_o}{f_o^{ss}} \right) \quad \textbf{[1.1]}$$

The second step of this process is considered the most tedious, as the technician must use trial-and-error techniques to identify the vocal cycle marking the termination or initiation of voicing; however, this step is arguably the most important since the voiced/unvoiced boundary is used to further identify eleven glottal pulse timings that correspond to the edges of the ten vocal cycles identified in the third step.

Unfortunately, clinical implementation of RFF via manual estimation is unfeasible. At least six of these RFF speech sequences are needed for a single reliable RFF estimate, averaging around 20–40 minutes of manual analysis per reliable RFF extraction (Eadie et al., 2013). To minimize the need for inefficient, manual intervention by trained technicians, a semi-automated RFF algorithm was developed using rule-based signal processing techniques (Lien, 2015; Lien et al., 2017).

The current semi-automated algorithm estimates RFF in six steps: (1) identification of the fricatives and vowels in each production according to the acoustic signal via high-to-low energy ratios, (2) estimation of average $f_o$ via autocorrelation of the vowel, (3) identification of peaks and troughs of potential vocal cycles pertaining to the vowel, (4) identification of boundaries between each vowel and the voiceless consonant via the normalized peak-to-peak amplitude, number of zero crossings, and waveform shape similarity, (5) rejection of instances that do not meet specified criteria (e.g., less than 10 onset or offset cycles, glottalization, misarticulation, voicing during the voiceless consonant), and (6) RFF calculation. Within this algorithm, all steps are fully automated except for step (1), wherein the location of the voiceless consonant may require manual intervention. To identify potential vocal cycles in the vowel of a VCV production, the utterance is first band-pass filtered according to the average $f_o$ of the speaker. A sliding window that is constructed using the inverse of the average $f_o$ of the speaker then shifts from the identified midpoint of the voiceless consonant in the first step and moves toward the vowel of interest. Potential vocal cycles are identified by leveraging three acoustic features obtained from each sliding window: normalized peak-to-peak amplitude, number of zero crossings, and waveform shape similarity. Of the six steps in this algorithm design, step (4) is especially prone to error due to the complexity of $f_o$ estimation at voicing offsets and onsets (Quatieri, 2008); as a result, there remains a need to further improve the algorithm for clinical and investigational use.

### *Purpose of the Current Work*

Quantitative measures of laryngeal muscle tension are needed to improve voice assessment and track clinical progress. RFF has shown promise as an acoustic estimator of laryngeal muscle tension, yet there remains a need to refine the semi-automated RFF algorithm for clinical and investigational use. The purpose of this work was to improve the accuracy and precision of the RFF algorithms for the objective quantification of laryngeal muscle tension, and to use the refined algorithm to determine the role of vocal fold abductory kinematics in estimates of RFF. This dissertation comprises three studies to achieve these goals, which (1) examine the impacts of sample characteristics and $f_o$ estimation method on the correspondence between semi-automated and manual RFF estimates, (2) elucidate the relationship of acoustic features and vocal fold vibratory characteristics during intervocalic offsets and onsets, and (3) determine the association between vocal fold abductory kinematics and RFF across a range of voice types. This work aims to improve the clinical applicability of RFF related to estimating laryngeal muscle tension for use in conjunction with current clinical voice assessment techniques.

# CHAPTER 2. Refining Algorithmic Estimation of Relative Fundamental Frequency by Accounting for Sample Characteristics and Fundamental Frequency Estimation Method


## *Abstract*

**Purpose:** The purpose of this study was to evaluate the impact of voice sample characteristics and fundamental frequency ($f_o$) estimation techniques on the correspondence between automated and gold-standard manual relative fundamental frequency (RFF) estimates.

**Methods:** Acoustic recordings were collected from individuals with ($N$ = 227) and without ($N$ = 256) voice disorders. Four common $f_o$ estimation methods (Auditory-SWIPE′, Halcyon, RAPT, YIN) were evaluated against the autocorrelation method currently implemented in the RFF algorithm. Using a training set (1158 samples), sample categories were constructed using pitch strength. RFF algorithm parameters were then tuned to each pitch strength category. From here, RFF values were recalculated in a test set (291 samples) using these category-specific thresholds. Algorithmically extracted RFF values were evaluated against manually extracted RFF values using mean bias error (MBE) and root-mean-square error (RMSE).

**Results:** The RFF algorithms with Auditory-SWIPE′ for $f_o$ estimation led to the greatest correspondence with manual RFF and was implemented in concert with category-specific thresholds. Refining $f_o$ estimation and accounting for sample characteristics led to increased correspondence with manual RFF ($MBE$ = 0.01 ST, $RMSE$ = 0.28 ST) compared to the unmodified algorithm ($MBE$ = 0.90 ST, $RMSE$ = 0.34 ST),

reducing the MBE and RMSE of semi-automated RFF estimates by 88.4% and 17.3%, respectively.

**Conclusions:** Refining the method of $f_o$ estimation using Auditory-SWIPE′ and accounting for sample characteristics via pitch strength categories led to improvements in the precision and accuracy of semi-automated RFF estimates. Future work should investigate additional metrics to use in order to objectively quantify the variations in voice sample characteristics that may be encountered in clinical populations.

## *Background*

Relative fundamental frequency (RFF) is an acoustic measure that has demonstrated feasibility in assessing and tracking laryngeal muscle tension. RFF is measured by examining instantaneous changes in fundamental frequency ($f_o$) during voicing transitions. In a vowel–voiceless consonant–vowel (VCV) production, this corresponds to the ten vocal cycles that mark the transition into and out of the voiceless consonant (see **Fig. 2.1**). The instantaneous $f_o$ of these ten cycles before and after the voiceless consonant are normalized to the steady-state $f_o$ of the nearest vowel. Changes in

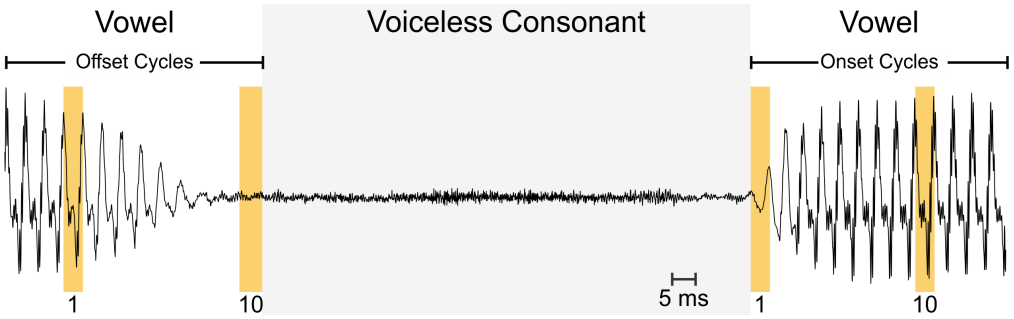


**Figure 2.1. Acoustic waveform of a vowel–voiceless consonant–vowel production, with voicing offset and voicing onset cycles identified. The first and tenth cycles of each voiced sonorant are highlighted. Voicing offset cycles are normalized to offset cycle 1, whereas voicing onset cycles are normalized to onset cycle 10.**

RFF during these transitions into and out of voicing form a characteristic pattern in typical speakers that has been attributed to interactions of laryngeal muscle tension, vocal fold kinematics, and changes in airflow (Löfqvist et al., 1989; Stepp et al., 2011d; Stevens, 1977; Van Den Berg, 1958; Watson, 1998).

Since RFF is thought to relate to laryngeal muscle tension, research efforts have explored the possibility of using RFF as an acoustic indicator of tension. Recent work has shown that RFF is capable of differentiating between healthy and disordered voices characterized by excessive laryngeal tension, including vocal hyperfunction (Stepp et al., 2010b; Stepp et al., 2011d), adductor laryngeal dystonia (Eadie et al., 2013), and Parkinson's disease (Goberman et al., 2008; Stepp, 2013). RFF has also been shown to correlate with auditory-perceptual judgements of dysphonia severity (Roy et al., 2016; Stepp et al., 2012), which encompasses multiple dimensions of voice quality—including breathiness, roughness, and strain—and to quantify the degree of laryngeal tension. Despite the promise of RFF for assessing laryngeal muscle tension, however, clinical implementation remains laborious.

### *Issues with Manual RFF Estimation*

The gold-standard method for estimating RFF requires a technician to subjectively evaluate the acoustic speech waveform using the acoustic analysis software, Praat (Boersma, 2001). Currently, the manual calculation of RFF cannot be implemented into clinical practice because of the extensive time needed for estimation. A single reliable RFF estimate requires approximately 20–40 minutes of analysis time from a trained technician. The majority of this time is spent exercising trial and error to identify

the vocal cycle marking the termination or initiation of voicing. Selecting this "boundary cycle" and then ensuring that adjacent glottal pulse timings are appropriately estimated can be challenging since $f_o$ estimation is particularly difficult near voicing offsets (voiced-to-unvoiced transitions) and onsets (unvoiced-to-voiced transitions; Quatieri, 2008). This is because vocal cycles nearest to the voiceless consonant may be masked by environmental noise or by concurrent aspiration and frication from coarticulation. Boundary cycle selection grows more difficult since the $f_o$ estimation method used in Praat implies stationarity, meaning that it does not change with time. This is problematic since instantaneous $f_o$ is expected to change when transitioning between voiced and unvoiced speech. As a result, locating the boundary cycle and extracting instantaneous $f_o$ is often a time-consuming process during manual RFF analysis.

Considering that at least six RFF speech sequences are needed for a reliable RFF estimate (Eadie et al., 2013; Lien et al., 2017), trained technicians must perform this tedious boundary selection step a minimum of 12 times (6 voicing offset, 6 voicing onset) to achieve a reliable RFF estimate for one speaker. As such, semi-automated RFF estimation was developed to mitigate the time-consuming nature of manual RFF estimation (Lien, 2015; Lien et al., 2017).

### *Issues with Semi-automated RFF Estimation*

The current semi-automated algorithm (called "aRFF") uses signal processing techniques to estimate RFF. The aRFF algorithm estimates $f_o$ via autocorrelation to identify potential vocal cycles in the vowel of a vowel–voiceless consonant–vowel (VCV) production. Instead of relying on manual intervention to identify the boundary

cycle, the aRFF algorithm employs rule-based acoustic analyses. This algorithm has known limitations, however. Although the aRFF algorithm minimizes the need for manual intervention—thus, expediting the RFF computation process—the accuracy of semi-automated RFF values was found to vary across a wide range of voice signals (Lien et al., 2017). It is possible that some of the observed variations in RFF accuracy are the result of unrefined signal processing techniques employed within the aRFF algorithm, including inaccurate $f_o$ estimation and a failure to account for voice sample characteristics.

## Effects of $f_o$ Estimation Method

Accurate estimates of $f_o$ are necessary for two steps of the aRFF algorithm: to estimate the average $f_o$ of the speaker, and to calculate the period between potential vocal cycles. Typical $f_o$ estimation techniques operate under the assumption that (i) vocal fold vibration generally varies a small percentage from one period to the next, and (ii) the configuration of the vocal tract varies at a much slower rate than that of the vocal folds (Talkin, 1995). The former assumption is important for identifying physiologically possible $f_o$ estimates, and the latter is necessary to further assume that the speech sound being produced will not change from one cycle to another. These methods often operate by comparing a segment of a voice signal with another segment that has been shifted in time, or by examining the frequency content of the signal.

The aRFF algorithm relies on autocorrelation-based $f_o$ estimation to calculate the mean $f_o$ of each vowel in a VCV production. This method compares a segment of signal to a delayed copy of itself as a function of the delay. Autocorrelation was originally

selected for the RFF algorithm because Praat also uses autocorrelation for $f_o$ estimation. This method is favorable because it provides high resolution in the time domain and is of low computational complexity.

Autocorrelation may not be the best $f_o$ estimator for computing RFF because it assumes signal periodicity and requires a significant timeframe to examine physiological $f_o$ ranges encountered in the human voice. Performing simple autocorrelation analyses on dysphonic voices may lead to increases in measurement error since the periodicity of the human voice is considered quasiperiodic at best (Rabiner, 1977). This is problematic since the level of aperiodicity in a voice signal often increases when voice problems are present (Eadie & Doyle, 2005; Titze, 1995). Autocorrelation also requires 2–3 complete pitch periods to develop an accurate estimate of $f_o$ (Rabiner, 1977); however, RFF specifically captures rapid changes in $f_o$ while transitioning into and out of voiced speech, which may further lead to estimation inaccuracies and poor cycle-to-cycle resolution.

Alternative methods of estimating $f_o$ have not previously been explored for semi-automated RFF calculations. Although manual RFF estimation faces the same difficulties in using autocorrelation as an $f_o$ estimation method in Praat, manual estimation allows a trained technician to subjectively make decisions about the boundary and adjacent cycles to (time-intensively) bypass these issues. Thus, an investigation examining the effects of different $f_o$ estimation methods is warranted to determine whether the shortcomings of using autocorrelation for semi-automated $f_o$ estimation can be overcome by a different technique.

## Effects of Voice Sample Characteristics

During the development of the aRFF algorithm, semi-automated RFF estimates were compared against manual estimates to gauge algorithmic accuracy (Lien et al., 2017). The algorithm was trained ($N = 126$) and tested ($N = 64$) on voice signals that varied in recording location and equipment, speaker diagnosis, and speaker dysphonia severity. This group included 36 typical speakers and 154 speakers with disordered voices.

When testing the aRFF algorithm, Lien et al. (2017) found that the relationship between RFF estimation methods was dependent on voice sample characteristics, noting dysphonia severity and signal quality as potential factors influencing this relationship. The authors noted that speech signals recorded from a waiting area or quiet room of a voice clinic resulted in a poorer correlation (.82 versus .91) and greater root-mean-squared error (0.37 ST versus 0.28 ST) between semi-automated and manual RFF estimates than those recorded in a sound-treated room. Yet the samples recorded in a voice clinic were also, on average, more dysphonic. The trends observed in signal quality and dysphonia severity may have been a result of participant recruitment logistics (e.g., more dysphonic participants were recorded in a voice clinic rather than in a research laboratory), but may have also been a result of interactions among participant, room, and task. For instance, it is possible that room acoustics affected speaker levels of vocal effort, comfort, control, and clarity (Bottalico, Graetzer, & Hunter, 2016).

Although the authors did not evaluate the effects of dysphonia severity and/or signal quality on RFF, it is possible that these sample characteristics led to the observed

variations in algorithm performance. In order to identify the vocal cycles nearest to the voiceless consonant, the aRFF algorithm first leverages a set of acoustic features to identify the boundary between voiced and voiceless segments. A sliding window based on the speaker's average $f_0$ navigates from the midpoint of the voiceless consonant toward the vowel. Within each window of time, three acoustic features are calculated: normalized peak-to-peak amplitude (PTP), number of zero crossings (NZC), and waveform shape similarity (WSS). PTP is computed as the range of the amplitude of the windowed speech signal, NZC is the number of sign changes of the windowed speech signal, and WSS is the normalized sum of square error between the current window of time and the previous window of time. If a positive or negative peak is identified in the region of the voiceless consonant, PTP is expected to be low and both NZC and WSS are expected to be high (Lien, 2015).

After the window navigates through the vowel, the aRFF algorithm stops collecting acoustic features and instead applies rule-based signal processing techniques on the three acoustic feature vectors to locate the boundary between the vowel and voiceless consonant. The algorithm assumes that the largest change in the three acoustic feature vectors will occur at this boundary. To locate this change, the algorithm identifies the feature value that maximizes the effect size between the left and right components of each acoustic feature vector. To help in identifying this boundary, a single set of constant threshold multipliers are applied to each acoustic feature vector. These constant threshold multipliers were identified by choosing values that minimized the overall difference between manual and semi-automated RFF estimates across their training set of

heterogeneous voice samples (Lien et al., 2017). The cycle index corresponding to the identified feature value is selected as a boundary cycle candidate, and the median of these candidates is chosen as the boundary cycle.

The problem in using a threshold set that has been optimized across heterogeneous voice samples is that Lien and colleagues (2017) found that boundary cycle identification varies according to the voice samples used to train and test the algorithm. This is likely because any singular threshold set may not be the best thresholds for an individual's voice samples. For instance, a typical voice sample may be more periodic than many of the samples used to train the semi-automated algorithm to calculate RFF. In this scenario, the thresholds that identify the boundary cycle are tuned to more aperiodic signals, but the criteria for choosing the boundary cycle may differ for more periodic signals.

It is possible that the thresholds required to minimize the difference between manual and semi-automated RFF estimates vary across a wide range of speech signals. This may be, in part, due to vocal cycle masking and a lack of $f_0$ stationarity at voicing offsets and onsets. If so, it is likely that manual RFF estimation is less impacted by these complications since trained technicians visualize the acoustic waveform and subjectively choose the boundary cycle, employing trial-and-error techniques when necessary. Since voice sample characteristics have been shown to affect the performance of the aRFF algorithm, it is necessary to take these differences into consideration prior to RFF computation.

### *Purpose of the Current Study*

The specific purpose of the study was to investigate the effects of $f_o$ estimation method and voice sample characteristics (dysphonia severity, signal quality) on the correspondence between manual and semi-automated RFF values. To assess the effects of $f_o$ estimation method on this correspondence, each of five popular $f_o$ estimation methods were used for semi-automated RFF estimation and compared against manual RFF values. To examine the effects of voice sample characteristics, a training set of voice samples was used to tune algorithmic parameters according to voice sample attributes. The correspondence between manual and semi-automated RFF values was then evaluated in an independent test set of voice samples. The results of this study aimed to improve the clinical applicability of using RFF as an estimator of laryngeal muscle tension.

## ***Methods***

### *Participants*

A total of 483 participants were recruited for the current study. All participants provided informed, written consent with the Boston University or University of Washington Institutional Review Board.

#### Typical Speakers

A group of 256 individuals without voice disorders (152 female, 104 male) aged 18–100 years of age ($M$ = 37.6 years, $SD$ = 22.3 years) were recruited to participate in the study. All participants without voice disorders were healthy adult speakers of English, and had no history of speech, language, hearing, neurological, pulmonary, or voice disorders.

Speakers with Voice Disorders

A group of 227 individuals with disordered voices (148 female, 79 male) aged 18–84 years of age ($M$ = 52.9 years, $SD$ = 17.7 years) were also recruited to participate in the study. Participants within this group were either diagnosed with idiopathic Parkinson's disease by a neurologist, or were diagnosed with a voice disorder by a board-certified laryngologist. All individuals with Parkinson's disease were recorded while on their typical carbidopa/levodopa medication schedule. Individuals who used deep brain stimulation devices were requested to turn their device off for the duration of the study.

Twenty primary voice-related problems were described by the 227 individuals with disordered voices. These problems ranged from muscle tension dysphonia ($N$ = 83) to upper respiratory infection ($N$ = 1). A detailed list of these voice-related problems and the frequency to which they were reported are included in **Table 2.1**. A broad range of vocal function was included in the current group to represent a sample of the populations that may be seen in clinical practice.

**Table 2.1. Frequency of primary voice-related problems for speakers with disordered voices.**

| Primary Voice-related Problem | Frequency of Problem |
|---|---|
| Cyst(s) | 3 |
| Dysphagia | 3 |
| Ear, Nose, and/or Throat Infection | 3 |
| Edema | 4 |
| Gastroesophageal Reflux | 3 |
| Globus Sensation | 1 |
| Granuloma | 4 |
| Laryngeal Trauma | 3 |
| Muscle Tension Dysphonia | 83 |
| Nodules | 20 |
| Papilloma | 2 |
| Paradoxical Vocal Fold Motion | 1 |
| Parkinson's Disease | 74 |
| Polyp | 6 |
| Presbylarynges | 1 |
| Spasmodic Dysphonia | 6 |
| Upper Respiratory Infection | 1 |
| Vocal Fold Atrophy | 3 |
| Vocal Fold Paralysis | 2 |
| Vocal Fold Scarring | 4 |

### Dysphonia Severity

A speech-language pathologist specializing in voice disorders assessed the overall severity of dysphonia (OS; 0–100) of each participant using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster et al., 2009). Sentences for analysis included (i) "Only we feel you do fail in new fallen dew," and (ii) either "We all found a wee fly on my food on Monday" ($N = 443$) or "To rock out, Molly shows zoo cats as they take all food" ($N = 40$) based on the availability of pre-recorded stimuli. Sentences with VCV words loaded with /f/ (e.g., "do fail," "all found," "all food") to resemble the RFF stimuli produced by participants in the study. Both sentences were blindly evaluated for OS by the speech-language pathologist, yielding two OS scores. A final OS score for each participant was obtained by averaging the scores from each sentence. The average OS score for speakers without voice disorders was 11.5 ($SD = 8.1$, *range* = 0–44.6), and that of speakers with disordered voices was 22.1 ($SD = 20.0$, *range* = 0–100).

The speech-language pathologist reanalyzed 15% of speakers in a separate sitting to ensure adequate intrarater reliability. Average OS ratings were collected for the randomly selected speakers, and were compared to previously made ratings. The Pearson's product-moment correlation coefficient was calculated on the ratings using the statistical package R (Version 3.2.4), yielding an intrarater reliability of $r = .96$.

## *Recording Procedures*

### Recording Environment and Equipment

Participants were recruited for the study over the course of seven years from 2011–2018. Each participant was recorded in one of three locations: Boston University

(226 female, 134 male), Boston Medical Center (17 female, 7 male), or University of Washington (1 female, 22 male). Participants run at Boston University were recorded in a quiet room or sound-treated room using a condenser headset microphone (model SM35XLR; Shure, Niles, IL), whereas those run at Boston Medical Center were recorded in a waiting area or quiet room of a voice clinic using a dynamic headset microphone (model WH20XLR; Shure, Niles, IL), and those run at University of Washington were recorded in a quiet room using a dynamic headset microphone (model WH20XLR; Shure, Niles, IL). All microphone signals were sampled at 44.1 kHz with 16-bit resolution, and were placed 45° from the midline of the vermilion and 7–10 cm from the corner of the lips.

Of the 483 participants, 335 (207 speakers without voice disorders, 128 speakers with voice disorders) were recorded in a sound-treated room, and 148 (49 speakers without voice disorders, 99 speakers with voice disorders) in a quiet room or waiting area. Headset microphones and room characteristics (e.g., reflection, noise, and reverberation) were not standardized across recording locations, and may also account for some variability across recording equipment and settings available to research laboratories and in clinical practice.

## Speaker Training

Participants were trained to produce three sets of nonsense words that each comprised three repetitions of a VCV production loaded with the voiceless consonant, /f/. The selected utterances were /ɑfɑ/, /ifi/, and /ufu/. In between each set, participants were instructed to take a breath, resulting in nine VCV productions per string: /ɑfɑ ɑfɑ ɑfɑ/,

breath, /ifi ifi ifi/, breath, /ufu ufu ufu/.

For this task, VCV productions were selected as stimuli rather than running speech (e.g., “Only we feel you do fail in new fallen dew” as in the OS ratings) to minimize intraspeaker variability (Lien et al., 2014). Similarly, uniform utterances (i.e., the same vowel surrounding the voiceless consonant) loaded with /f/ were selected as stimuli to minimize individual variations within speaker.

Although sound pressure level was not standardized across participant, each participant was instructed to speak using their typical pitch and loudness and to refrain from chanting or singing the production; if a VCV production was sung, the participant was instructed to repeat the set. Further, if any of the VCV sets were misarticulated or glottalized, participants were instructed to repeat the set.

### *Data Analysis*

#### Overview

The database examined in this study included 4347 VCV productions (1449 voice samples) collected from 483 independent speakers. **Fig. 2.2** depicts the composition of this database. Manual RFF estimation—described in the next section—was conducted on all 4347 VCV productions. The aRFF algorithm was then used to perform semi-automated RFF estimation on the entire database.

Simple random sampling was implemented to divide the database into training (80%) and test (20%), ensuring low bias in model performance (Kuhn & Johnson, 2013; Reitermanova, 2010). The training set comprised 3474 VCV productions from 386 independent speakers, whereas the test set consisted of 873 VCV productions produced

by 96 speakers (see **Fig. 2.2**). Samples from the training set were used to investigate the effects of different $f_o$ estimation techniques and voice sample characteristics on the semi-automated RFF algorithm. Because simple random sampling can lead to high variance in model performance, *k*-fold cross validation was performed on the training set to quantify this variation. Resulting parameters were used to tune the RFF algorithm, which was then evaluated in the independent test set.

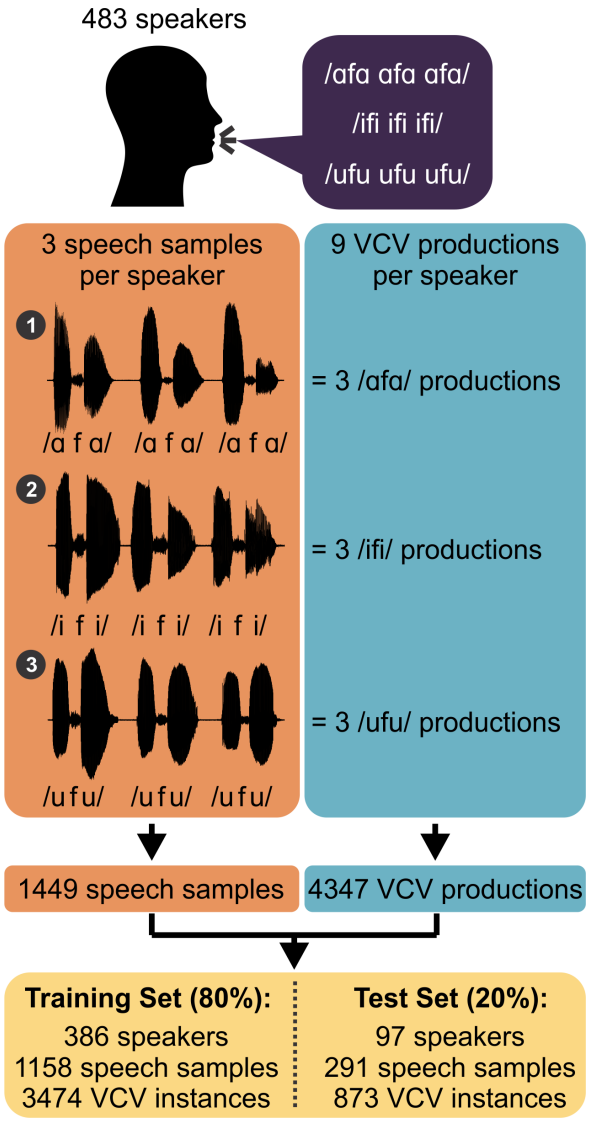


**Figure 2.2. Voice sample collection flowchart. Speakers produced three repetitions each of vowel–voiceless consonant–vowel (VCV) utterances /ɑfɑ/, /ifi/, and /ufu/.**

## Manual RFF Estimation

### Effects of the Number of Trained Technicians on Manual RFF Estimates

Although a single technician has been shown to have high internal reliability when manually extracting RFF values (Lien et al., 2015a; Lien et al., 2015b), the ability

to distinguish the vocal cycle closest to the voiceless consonant (i.e., offset cycle 10 and onset cycle 1) can vary across technicians and ultimately, influence RFF cycle values. Thus, an analysis into the effects of the number of RFF technicians on resulting offset cycle 10 and onset cycle 1 values was conducted on a subset of 88 speakers from Lien et al. (2017). The subset of 88 speakers was selected because each speaker's nine VCV productions had been rated by three trained technicians, thereby enabling a comparison between one, two, and three technicians on resulting RFF values. The average deviation of mean RFF values from three trained technicians—considered in this analysis as the gold standard—was assessed as a function of the number of technicians in the subset.

**Fig. 2.3** shows that the deviation of RFF values of each speaker between 1 and 3 technicians differ by an average of 0.30 ST (95% CI = 0.26–0.36 ST) for voicing offset and 0.38 ST (95% CI = 0.31–0.44 ST) for voicing onset. The deviation of RFF values between 2 and 3 technicians differ by an average of 0.15 ST (95% CI = 0.13–0.18 ST) for voicing offset and 0.19 ST (95% CI = 0.16–0.22 ST) for voicing onset. These values are less than those reported as clinically meaningful changes after a course of voice therapy; specifically, Stepp et al. (2011d) reported that average RFF values

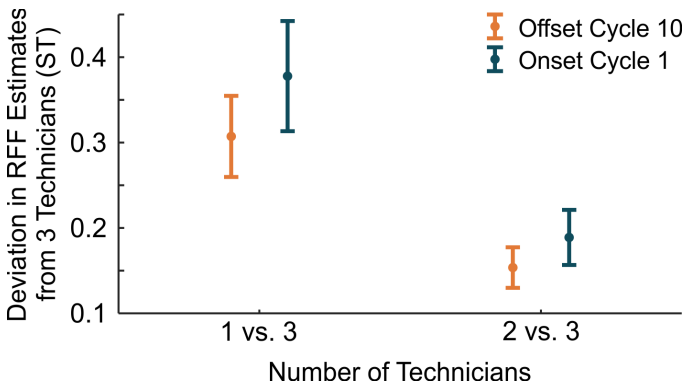


**Figure 2.3. Average deviation of mean relative fundamental frequency (RFF) values from the gold-standard of three trained technicians, as a function of the number of technicians in the speaker subset. Error bars indicate 95% confidence intervals.**

increased after voice therapy by 0.5 ST for voicing offset cycle 10 and 0.81 ST for voicing onset cycle 1. The results of this analysis confirm that one technician is sufficient to reliably calculate RFF. However, these findings also suggest that using two technicians to estimate RFF will lead to a smaller error than when only using one technician. As such, two technicians were chosen as sufficient for carrying out manual RFF estimation on each participant of the current dataset.

**Technician Training Paradigm**

In the current study, technicians were trained to perform manual RFF estimation using a dataset and training protocol described in Vojtech and Heller Murray (2019). This training regimen was developed to guide technicians through manual RFF estimation using Praat and Microsoft Excel (Microsoft, Redmond, WA). In brief, technicians were trained to: (1) determine an appropriate pitch range for the speaker, and alter the range from default (male: 60–300 Hz, female: 90–500 Hz) when necessary, (2) locate voicing offset or onset of the acoustic signal, (3) identify the boundary cycle distinguishing the voiced consonant from vowel, (4) isolate the nine vocal cycles adjacent to the boundary cycle, and (5) extract and export the glottal pulse timings corresponding to these ten vocal cycles into Excel.

Technicians were provided an Excel template that automatically calculated the period of each vocal cycle as the difference between adjacent pulse timings. The instantaneous $f_0$ of each vocal cycle was then automatically computed as the inverse of the cycle period, and RFF was automatically calculated as this instantaneous $f_0$ when normalized to the reference $f_0$ ($f_o^{ref}$; offset cycle 10 for voicing offset and onset cycle 1 for

voicing onset) using **Eq. 2.1**:

$$\text{RFF (ST)} = 12 \times \log_2 \left( \frac{f_o}{f_o^{ref}} \right) \quad \textbf{[2.1]}$$

Technicians were further trained to examine the acoustic waveform in the time and frequency domains, as well as the resulting RFF values (if applicable), to determine whether an offset or onset instance should be rejected. Examples of rejection criteria include glottalization, voicing of the voiceless consonant, or failure to reach steady state. Using this training paradigm, technicians were empirically considered to be reliable if they obtained ≥ .93 interrater reliability with the first author of the training regimen.

**Gold-standard RFF Computation**

Two trained technicians were assigned to manually calculate RFF on each of the nine VCV productions per speaker. Average RFF values were computed across both technicians to serve as the estimated RFF values for each speaker. Due to the availability of technicians to perform manual RFF, a total of eight trained technicians (referred to A–H) completed manual RFF estimates throughout the course of data collection. Three trained technicians (F–H) completed a version of the training regimen prior to 2015, as described in Lien (2015); the remaining five technicians (A–E) completed training after 2015. **Table 2.2** shows the number of speakers that each of the eight trained technicians rated and the number of speakers that overlapped with other raters. Of the 483 speakers, 437 were rated by two trained technicians (technicians A–H), and 46 were rated by three trained technicians (only technicians F–H).

**Table 2.2. Number of speakers for which eight trained technicians manually computed relative fundamental frequency. The matrix shows common speakers analyzed between technicians, whereas the diagonal (bolded) describes the number of speakers a single technician rated in total.**

| Technician | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| A | **103** | | | | | | | |
| B | 93 | **278** | | | | | | |
| C | 1 | 92 | **188** | | | | | |
| D | 0 | 79 | 9 | **91** | | | | |
| E | 9 | 0 | 86 | 3 | **99** | | | |
| F | 0 | 0 | 0 | 0 | 0 | **96** | | |
| G | 0 | 1 | 0 | 0 | 1 | 95 | **97** | |
| H | 0 | 13 | 0 | 0 | 0 | 47 | 46 | **60** |

Interrater reliability was conducted on the RFF estimates using two-way random intraclass correlation coefficients for consistency of agreement (ICC). Intrarater reliability was evaluated by computing the Pearson correlation coefficients within each technician when asked to re-estimate 15% of their samples in a different sitting (Lien et al., 2017; Lien et al., 2015b). The average interrater reliability was computed as *ICC* = .92 (*SD* = .05, *range* = .82–.99), and the average intrarater reliability was calculated as *r* = .92 (*SD* = .04, *range* = .87–.99).

## Semi-automated RFF Estimation

Semi-automated RFF estimation was first performed on all 4347 VCV productions using the aRFF algorithm in MATLAB (version 9.3; The MathWorks, Natick, MA). Further analyses were performed in MATLAB by adding functionality to the aRFF algorithm that evaluates the effects of $f_o$ estimation method and sample characteristics on resulting RFF values.

## Method of $f_o$ Estimation

### *Choice of $f_o$ Estimation Techniques*

In addition to the autocorrelation method implemented in the aRFF algorithm (see section Effects of *fo* Estimation Method for more details about the use of the autocorrelation function for $f_o$ estimation), four additional $f_o$ estimators were identified to evaluate the accuracy and precision of semi-automated RFF values: Auditory-SWIPE′, YIN, Halcyon, and RAPT.

#### Auditory-SWIPE′

Auditory-SWIPE′ operates in the frequency domain to estimate $f_o$. It is based on the SWIPE (Sawtooth-Waveform Inspired Pitch Estimator; Camacho, 2007) algorithm, which measures the similarity between the square-root of the spectrum of a speech signal and that of a sawtooth waveform. A sawtooth waveform is constructed across the desired range of $f_o$ values to examine, and the correlation between the speech signal and sawtooth waveform is measured. The $f_o$ of the sawtooth waveform that results in the highest correlation is considered the $f_o$ of the examined signal. The degree of this correlation (0–1) is the "pitch strength," described further in *Dysphonia Severity*.

The Auditory-SWIPE′ algorithm[2] builds upon the original SWIPE algorithm by introducing negative weights between harmonics, as well as a kernel that only considers the first and prime harmonics of the signal (SWIPE′; Camacho, 2007). These modifications were implemented to minimize subharmonic errors. Additionally, the

[2] The Auditory-SWIPE′ algorithm may be downloaded from: octolinker-demo.now.sh/saul-calderonramirez/Aud-SWIPE-P/tree/master/Aud-SWIPE_MATLAB

Auditory-SWIPE′ uses an auditory processing front-end (Auditory-SWIPE; Camacho, 2012), including outer and middle ear transfer functions and a cochlear filter bank, to recover potentially missing harmonics. Auditory-SWIPE′ is favored over the basic autocorrelation function because it adds additional steps to account for missing harmonics, inharmonic signals, and selecting subharmonics of the true $f_o$. Therefore, Auditory-SWIPE′ is more computationally complex than the autocorrelation function.

YIN

The YIN algorithm (de Cheveigne & Kawahara, 2002) is based on the autocorrelation function for estimating $f_o$. YIN is more advantageous, however, in that it makes use of additional steps to reduce the errors seen when solely using the autocorrelation function. These additional steps include a difference function, cumulative mean normalized difference function, absolute threshold, parabolic interpolation, and best local estimate. A difference function is used to reduce harmonic and subharmonic errors, whereas the cumulative mean normalized difference function is used to reduce the sensitivity of the autocorrelation function to amplitude changes, avoid an upper limit of the $f_o$ search range, and normalize the function as a pre-processing step for the thresholding step. An absolute threshold is then implemented as a means of avoiding subharmonics being chosen as the estimated $f_o$. These three steps are effective when the $f_o$ candidate is an integer multiple of the sampling frequency (e.g., an $f_o$ candidate of 150 Hz detected from a voice signal sampled at 44.1 kHz). Parabolic interpolation is effective if this is not that case, instead relying on the spectral properties of the autocorrelation function to fine-tune period candidates. Finally, a best local estimate method is

implemented to identify a more precise estimate of $f_o$ from the analysis points identified in previous steps.

As with Auditory-SWIPE′, YIN is advantageous over the simple autocorrelation function because it includes steps for reducing subharmonic errors, as well as a method for tuning the choice of $f_o$. Despite these steps, the accuracy of YIN is still limited by a high dependence on sampling frequency and generates a greater computational cost than the simple autocorrelation function (Sukhostat & Imamverdiyev, 2015).

Halcyon

The Halcyon algorithm is an $f_o$ estimation technique developed by Azarov, Vashkevich, and Petrovsky (2016). Halcyon generates $f_o$ candidates based on a normalized cross-correlation function (NCCF) that has been altered to improve time resolution. The Halcyon algorithm operates by cycling through the $f_o$ range of interest. At each $f_o$ candidate, the speech signal undergoes the following process: (1) the signal is resampled to a specified multiple of the $f_o$ candidate, (2) the resampled signal is energy normalized, (3) a set of instantaneous parameters are derived for each $f_o$ candidate for use as inputs to a NCCF, (4) the instantaneous NCCF is evaluated, (5) resulting $f_o$ candidate values are weighted to penalize low-frequency candidates, (6) a rough $f_o$ estimate is extracted via dynamic programming, and (7) a fine $f_o$ estimate is computed using the instantaneous parameters of the rough $f_o$ estimate via a weighted sum.

Halcyon may be favored over simple autocorrelation since local maxima derived from the NCCF tend to be more prominent and less affected by rapid variations in signal amplitude. Moreover, Halcyon uses an altered version of the NCCF as well as a

weighting function to reduce the impact of harmonic mixing and to penalize low-frequency candidates. Due to these advantages, Halcyon remains more computationally taxing than autocorrelation.

RAPT

RAPT (Robust Algorithm for Pitch Tracking; Talkin, 1995) operates in the time domain and, like the Halcyon algorithm, employs the NCCF to estimate $f_o$. RAPT operates on two versions of the speech signal: the first version of the signal is unaltered, whereas the second version is at a significantly reduced sampling rate. The NCCF is first calculated for all time delays (lags) within the $f_o$ range of interest using the low-rate signal. Indices corresponding to local maxima are then used as input lags for the NCCF of the regular-rate signal. Local maxima from the NCCF of the regular-rate signal are selected as $f_o$ candidates. These candidates are then compared to each other using rule-based signal processing techniques (e.g., voicing tends to change states at low $f_o$ values, amplitude tends to increase at the onset of voicing and decrease at the offset of voicing) to produce an estimate of $f_o$ to characterize the signal.

Similar to the Halcyon algorithm, RAPT improves basic autocorrelation by making use of the NCCF to generate $f_o$ candidates. Although RAPT is less computationally complex than the Halcyon algorithm, it is still more computationally intensive than the autocorrelation implemented in the aRFF algorithm. Furthermore, the RAPT algorithm suffers when local maxima occur at double or half the "true" lag value.

Performance of Selected $f_o$ Estimation Methods in the Literature

The methods described above were selected from many $f_o$ detection algorithms due to their superior performance and evaluation in the literature. A brief discussion of comparing the selected $f_o$ detectors is included here:

Despite its implementation in speech research, comparisons of the Auditory-SWIPE′ version of SWIPE against other $f_o$ estimation methods have not been widely conducted. Comparisons of SWIPE and SWIPE′, on the other hand, are prevalent in the literature. For instance, Jouvet and Laprie (2017) compared 15 $f_o$ estimators on simulated and real noisy speech data. Among the assessed algorithms were autocorrelation, RAPT, SWIPE, and YIN. Overall, the authors found that the magnitude of errors produced by RAPT were lower than those produced by YIN and autocorrelation function on both real and simulated noisy data. SWIPE did not perform well on simulated speech data, and as such, was not considered for examination on real speech data. On the other hand, Camacho and Harris (2008) found SWIPE′ to outperform 12 other $f_o$ detectors (including autocorrelation, RAPT, and YIN) on each of three databases, including the Disordered Voice Database (Model 4337; KayPENTAX, Lincoln Park, NJ), Keele pitch database (Plante, Meyer, & Ainsworth, 1995), and musical instruments samples (Fritts, 1994). In a smaller examination of $f_o$ estimators, Sukhostat et al. (2015) found that YIN resulted in the most superior performance against two other time-domain methods (autocorrelation, average magnitude difference function) in three different noise types (babble, car, and white) and separately at different signal-to-noise levels (clean, -5, 0, 15, and 20 dB).

Although not widely implemented, Halcyon has been compared to other popular

$f_o$ detectors on real and simulated speech and shows great promise. Azarov et al. (2016) found that the Halcyon algorithm resulted in the least overall error in natural speech when compared to RAPT, YIN, and SWIPE′; however, SWIPE′ produced the most accurate results regarding mean fine pitch error (percentage of voiced frames with estimated $f_o$ within ±20% of true $f_o$) in natural female speech. When comparing these four algorithms on simulated data, the Halcyon algorithm also led to the least overall error in $f_o$, with YIN leading to the worst performance. SWIPE′ and RAPT performed similarly in terms of gross $f_o$ error (percentage of voiced frames with estimated $f_o$ greater than ±20% of true $f_o$), but RAPT produced less error compared to SWIPE′ when considering mean fine pitch error.

It is clear that the selected $f_o$ estimators perform differently on different databases (e.g., natural versus simulated speech). The findings support the notion that $f_o$ estimation method should be investigated to determine if autocorrelation is sufficient, or whether there is a more favorable method to maximize correspondence between manual and semi-automated RFF estimates.

*Assessment of $f_o$ Estimation Accuracy*

A method of $f_o$ calculation is necessary to estimate the average $f_o$ of the speaker and periods of potential vocal cycles. The average $f_o$ of the speaker is used to generate a window that slides along the acoustic waveform; within each window of time, the algorithm estimates PTP, NZC, and WSS and locates peaks and troughs in amplitude. The period between the collected peaks and troughs are calculated and recorded as a vector of vocal cycle candidates. From here, the effect size of each acoustic feature vector

is maximized with respect to time, and the corresponding vector index is considered the boundary cycle. The vocal cycle candidate located at the boundary cycle index, along with the adjacent nine vocal cycle candidates, are then used to calculate RFF.

Two simulations were created to mimic average $f_o$ and vocal cycle period estimation steps. In the first simulation, the midpoint of the voiceless consonant in a VCV production was provided as an input to the aRFF algorithm. Using the entire VCV production, the algorithm estimated the average $f_o$ of the speaker to create a sliding window for estimating acoustic features and capturing peaks and troughs in amplitude. The algorithm then calculated the pitch period of vocal cycle candidates from the collected peaks. This simulation evaluated both $f_o$-dependent steps of the aRFF algorithm.

In the second simulation, the manually defined indices of the boundary cycle were provided to solely evaluate the ability of the $f_o$ estimation method to calculate the pitch period of vocal cycle candidates. Doing so removes the step of measuring acoustic features and collecting peaks and troughs that do not necessarily pertain to the ten cycles immediately adjacent to the voiceless consonant. In this way, errors from incorrect boundary cycle identification are ignored.

RFF was calculated on a subset of 180 speech samples (9 speech samples from 20 participants) in each of the aforementioned simulations: (1) when the approximate midpoint of the voiceless consonant was provided, and (2) when the manually defined indices of the boundary cycle were provided. A small subset of speech samples was chosen for this analysis to include a range of pitch strength values. Of the subset of 20 participants, 15 speakers were recorded in a sound-attenuated room, whereas the

remaining five were recorded in a quiet room or waiting area of a voice clinic. Moreover, 11 of the 20 individuals were diagnosed with a voice disorder. The samples provided pitch strength values ranging from 0.04 to 0.51 ($M$ = 0.33, $SD$ = 0.12). To compute RFF, each $f_o$ estimation was implemented with specific input parameters:

- Autocorrelation: Minimum $f_o$ (50 Hz) and maximum $f_o$ (400 Hz)
- Auditory-SWIPE′: Minimum $f_o$ (50 Hz), maximum $f_o$ (400 Hz), time interval (0.001 s), Hann window overlap proportion (0.6), pitch strength threshold (0), spectrum step size (1/32), and resolution (1/32 steps per octave)
- YIN: Minimum $f_o$ (50 Hz) and maximum $f_o$ (400 Hz)
- Halcyon: Minimum $f_o$ (50 Hz), maximum $f_o$ (400 Hz), number of $f_o$ candidates (100), number of cycles within time window (4), number of harmonics (8), number of shifts for instant phase estimates (1), resolution (8 frames per step), dynamic step progression (21 samples)
- RAPT: Minimum $f_o$ (50 Hz) and maximum $f_o$ (400 Hz)

**Accounting for Voice Sample Characteristics**

As noted previously, the performance of the aRFF algorithm was observed to vary across a broad range of voice signals, with its authors specifically citing dysphonia severity and signal quality as observed sources of error. Since the thresholds used to locate the boundary cycle were determined from a heterogeneous group of voice samples, it is possible that the thresholds contributed to the variability in errors observed between manual and semi-automated RFF estimates.

One method of reducing this variability may be to develop categories based on

signal quality and dysphonia severity. In this way, RFF estimation would take place via category-specific thresholds instead of using a single set of thresholds for all voice samples. To develop category-specific thresholds, signal quality and dysphonia severity were first quantified. An automated rejection criterion was then developed to bypass samples with poor signal characteristics (e.g., high dysphonia severity or bad signal quality). The boundary between voiced and voiceless speech was then examined across the spectrum of signal quality and dysphonia severity to create voice sample categories and tune threshold parameters. These steps are described in detail below.

*Quantification of Dysphonia Severity and Signal Quality*

Signal Quality

Signal quality is determined by features of signal acquisition and the room conditions in which a voice sample is recorded. Numerous factors may affect the quality of a voice signal, including speaker characteristics (e.g., distance from the microphone, loudness) and recording environment. As such, the degree of room reverberation, degree of room noise, and proximity of the acquisition to reflecting surfaces must be taken into consideration when capturing acoustic signals (Titze, 1995).

The signal-to-noise ratio (SNR) has been widely used to capture global signal quality. SNR compares the power of a target stimulus to the power of background noise, and is defined as the ratio of signal intensity (computed as root-mean-square; RMS) to noise intensity, as shown in **Eq. 2.2**:

$$\text{Signal-to-noise Ratio (dB)} = 20 \cdot \log_{10}\left(\frac{\text{RMS}_{\text{signal}}}{\text{RMS}_{\text{noise}}}\right) \qquad \textbf{[2.2]}$$

Dysphonia Severity

Dysphonia severity is a term that describes a perceptual judgment relating to the degree of perceived vocal dysfunction. It has been referred to as "overall voice quality," "grade," and "overall severity (OS)," among other names. Although there are multiple auditory-perceptual evaluation methods for characterizing dysphonia severity, the CAPE-V's OS attribute was selected for the current study to remain consistent with the methods employed in the aRFF algorithm.

In the CAPE-V, OS is considered an auditory-perceptual attribute describing the global, integrated impression of vocal deviance (Kempster et al., 2009). Many acoustic parameters are sensitive to OS; however, the validity and clinical utility of these measures have been widely disputed. For instance, perturbation measures of jitter and/or shimmer have been frequently employed to characterize OS (Núñez-Batalla, Díaz-Fresno, Álvarez-Fernández, Muñoz Cordero, & Llorente Pendás, 2017); however, these measures demonstrate poor reliability (Carding et al., 2004; Deliyski, Shaw, Evans, & Vesselinov, 2006) and are generally restricted for use with only type 1 (nearly periodic) and some type 2 (strong modulations/subharmonics that approach the $f_o$ in energy) voices (Titze, 1995).

More recently, researchers have turned toward spectral analysis of the voice signal to characterize OS. The advantage of using spectral measures for voice analysis is that estimates of aperiodicity can be obtained in the absence of detecting specific vocal cycle boundaries. One measure, called "pitch strength," has been shown to be sensitive to OS (Eddins, Anand, Camacho, & Shrivastav, 2016; Kopf et al., 2017; Shrivastav, Eddins,

& Anand, 2012), and as such, may be used to objectively quantify sample characteristics. Pitch strength describes the saliency of pitch sensation and can be calculated using the Auditory-SWIPE′ model as the spectral similarity between a voice signal and a sawtooth waveform with missing non-prime harmonics at the same estimated $f_o$ as the voice signal (Camacho, 2012; Camacho et al., 2008). Calculated as the degree of correlation between the spectrums (from 0 to 1), sounds with higher pitch sensations result in higher pitch strengths, whereas sounds with lower pitch sensations result in lower pitch strengths.

Pitch strength has been implemented in the objective assessment of voice quality due to its versatility across voice signal types (Kopf et al., 2017). For instance, a perceptually breathy speech signal may be classified as containing some level of stochastic noise due to the turbulence surrounding the airflow jet when the voice is produced. Despite lacking an obvious $f_o$, the signal may still elicit a pitch sensation, and therefore, a non-zero pitch strength. Indeed, pitch strength has been shown to be correlated with perceptual judgments of voice quality (Eddins et al., 2016; Shrivastav et al., 2012) and has recently been proposed as a treatment outcome for dysphonia (Kopf et al., 2017). Because of this, pitch strength may be a viable, objective measure that can assess overall severity of dysphonia.

Relationship between Pitch Strength and Signal-to-Noise Ratio

Although pitch strength and SNR have been employed to characterize voice and signal quality, respectively, it is possible that these measures exhibit some degree of collinearity. If so, it would be redundant to use two separate measures to characterize dysphonia severity and signal quality. As an example, introducing noise into a speech

signal—such as environmental noise or the noisy by-product of a turbulent airstream generated at the glottis—will reduce the SNR and may, in turn, offset the pitch strength of the signal.

To examine the relationship between pitch strength and SNR, noise was added to voice samples from the same subgroup of 20 speakers examined in the $f_o$ estimate analysis. Two types of noise were selected: multi-speaker babble and ambient room noise. Multi-speaker babble was chosen to emulate noise that may be heard in a clinical environment, such as a waiting area or examination room. It consisted of four healthy male speakers and four healthy female speakers who were not included in the speaker dataset. Ambient room noise, on the other hand, was selected to simulate the magnetic noise sources that may exist in laboratory and clinical environments (e.g., fluorescent lights). This noise source was constructed as the sum of sine waves at integer multiples of 60 Hz (i.e., the $f_o$ associated with the mains' hum in the United States; Cowan, 1993, p. 155). Noise was added to the 60 voice samples from 20 speakers at SNRs of -5 dB to +50 dB using each noise source type. Resulting SNR and pitch strength values were extracted to assess the relationship between SNR and pitch strength.

For this analysis, SNR was reassessed as the root-mean-square of the vowels (signal) compared to that of the first and last 50 ms of the voice sample (noise), and pitch strength was calculated from the vowels (signal). Methodology from Lien (2015) was implemented to isolate the vowels (/a/, /i/, or /u/) from voiceless consonants and periods of silence: (1) the waveform was filtered using a low-pass $5^{th}$ order Butterworth filter with a corner frequency of 3.4 kHz, (2) the filtered waveform was smoothed using a 50-

ms moving average filter, and (3) the smoothed waveform was normalized via subtracting the mean and dividing by the standard deviation. The root-mean-square was then calculated from the vowel(s) and compared to the 100-ms interval of noise. Pitch strength, on the other hand, was calculated across the entire waveform using Auditory-SWIPE′. The default output of the algorithm was a pitch strength contour, which was then averaged across the vowel(s) to produce a single pitch strength estimate.

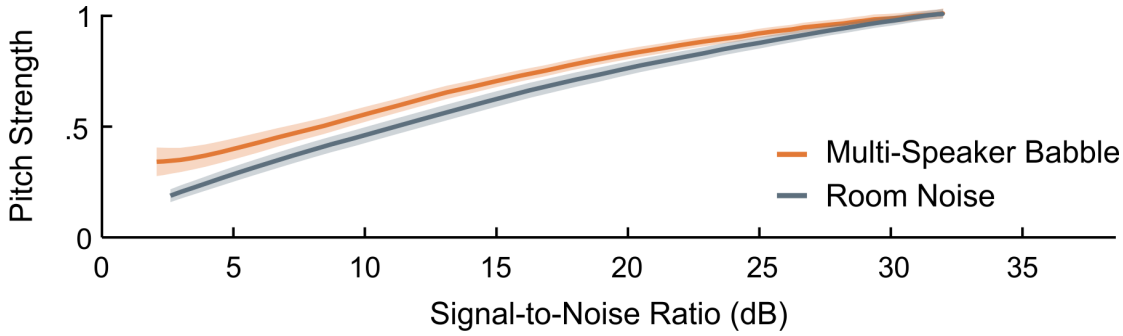


**Figure 2.4. Relationship between pitch strength and signal-to-noise ratio when multi-speaker babble (orange) and room noise (gray) were differentially added to voice samples.**

The correlation between SNR and pitch strength was $r = .990$ when examining multi-speaker babble and $r = .996$ when examining ambient room noise (see **Fig. 2.4**). These results indicate a strong relationship between pitch strength and SNR, wherein pitch strength increases with increasing SNR. This suggests that pitch strength not only provides information about dysphonia severity, but signal quality as well (i.e., pitch strength increases to some degree as signal quality increases). Because of this relationship, pitch strength was chosen as a singular parameter to describe dysphonia severity and signal quality.

*Development of Category-specific Thresholds*

Pitch strength was calculated for the 1157 VCV productions of the training set using the Auditory-SWIPE′ algorithm. The default output of the algorithm was a pitch strength contour, which was then averaged across the vowel(s) to produce a single pitch strength estimate. Pitch strength estimates were used to develop category-specific thresholds via the following steps:

Automated Sample Rejection

A rejection criterion was created to eliminate samples with pitch strength values considered too low (i.e., little-to-no pitch sensation) to accurately analyze. This criterion was developed by constructing a receiver operating characteristic (ROC) curve to determine the discriminatory ability of pitch strength to distinguish between VCV productions that were rejected versus retained during manual RFF analysis. A pitch strength criterion was then selected by maximizing the probability of rejecting a sample that would also be rejected through manual analysis, but minimizing the probability of rejecting a sample that would be retained through manual analysis (i.e., maximum positive likelihood ratio, or PLR). Thus, any VCV productions (including both offset and onset instances) with a pitch strength value below this criterion would be excluded from further RFF analyses.

Boundary Cycle Shifts

The goal of sample category creation is to tune the parameters necessary for identifying the boundary between voiced and voiceless speech segments (i.e., voicing offset cycle 10, voicing onset cycle 1). To tune the parameters for semi-automated RFF

estimation, the average discrepancy in boundary cycle identification between the aRFF algorithm and manual estimates must be quantified. While manual RFF estimation techniques allow the technician to subjectively evaluate where this boundary occurs, the aRFF algorithm leverages a set of acoustic features (PTP, NZC, WSS) to identify the boundary.

In the aRFF algorithm, PTP, NZC, and WSS are calculated as a sliding window travels from the midpoint of the voiceless consonant into the vowel. After various criteria are met to confirm that the sliding window has successfully reached the vowel, the three acoustic feature vectors are examined using rule-based signal processing. The aRFF algorithm assumes that a state transition will occur in feature values upon reaching the vowel, and the location where this inflection in feature values occurs is considered the boundary cycle. **Fig. 2.5** shows a schematic of the ideal state transition of an acoustic feature vector; in this schematic, the boundary cycle marks a change in feature magnitude that occurs when transitioning from the voiceless consonant to the vowel.

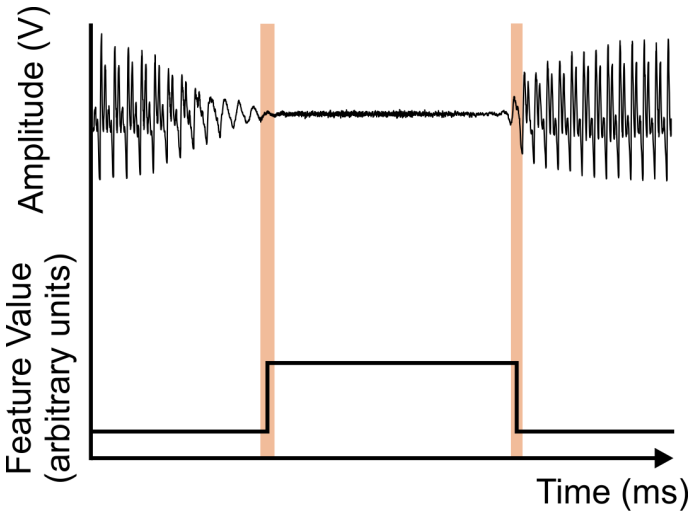


**Figure 2.5. Schematic of ideal feature plots for voicing offset and voice onset. The upper panel shows an acoustic waveform, and the lower panel shows an ideal feature vector calculated from the acoustic waveform. Highlighted segments mark the offset (left) and onset (right) boundary cycles, described as a marked transition in acoustic feature values between voiced and voiceless components.**

This logic for identifying the boundary cycle is enacted in the aRFF algorithm by maximizing the effect size of each acoustic feature vector. It is thus assumed that either side of the boundary cycle contains stable feature values corresponding to the vowel and voiceless consonant. The vector index that maximizes the effect size is then chosen as the boundary cycle candidate for that feature, and the median of these candidate indices is selected as the ultimate boundary cycle index.

Prior to evaluating the discrepancy between manual and semi-automated boundary cycle selections, all VCV productions with pitch strength values (averaged between offset and onset instances of a VCV production) below the automated rejection cut-off were removed from further analysis. Average PTP, NZC, and WSS values were calculated from remaining offset and onset instances as a function of the average number of pitch periods away from the manual, or "true," boundary cycle. For each acoustic feature, methodology from the aRFF algorithm was implemented to maximize the effect size of these feature vectors within ±2 pitch periods from the true boundary cycle. The pitch period cycle that elicited the greatest effect size was considered the automated, or "predicted," boundary cycle candidate for that feature.

Filtered versions of PTP, NZC, and WSS were also examined to assess whether filtering the acoustic signal enhanced the correspondence between true and predicted boundary cycle indices. Specifically, the aRFF algorithm uses a version of the microphone signal that is band-pass filtered 3 ST above and below the average $f_o$ of the speaker to identify peaks and troughs in signal amplitude. This was done to attenuate frequencies that were not directly associated with the $f_o$ of the speaker. Thus, predicted

boundary cycle candidates, slope directions, and corresponding effect sizes were extracted from the filtered versions of PTP, NZC, and WSS. Raw and filtered feature counterparts were compared within each pair; the version that led to the least discrepancy between true and predicted boundary cycle indices and/or greatest effect size was selected to represent the acoustic feature.

For the selected feature versions, the predicted boundary cycle candidate represents the average error (in number of vocal cycles) between manual and semi-automated RFF estimation. The acoustic feature slope across the predicted boundary cycle corresponds to the direction of the error (i.e., toward or away from the voiceless consonant).

Category Creation

Pitch strength estimates of samples that exceeded the automated rejection criterion were used to construct voice sample categories. For these samples, acoustic feature values computed at the predicted boundary cycle were evaluated with respect to pitch strength. Trends were visually inspected to identify local extrema, inflection points, and standard deviations. Pitch strength categories were manually identified by choosing pitch strength levels that represented consistent increases, decreases, or stable feature values. The discriminatory ability of pitch strength to distinguish between features at the true versus predicted boundary was assessed for each chosen category. An optimal feature threshold was selected by maximizing specificity and sensitivity for each offset and onset acoustic feature using the Youden index (Youden, 1950).

Concatenating Category Components

Thus far in the development of category-specific thresholds, four pieces of information were collected from the 1158 voice samples (3474 VCV productions) of the training set: (a) automated rejection criterion, (b) pitch strength category cut-offs, (c) feature-specific offset and onset category thresholds, (d) feature-specific magnitude and direction of error from the true boundary cycle. This information was implemented into the aRFF algorithm as follows:

1. If the average pitch strength of the VCV production fell below (a), the production was rejected. Otherwise, the offset and onset instances of the VCV production were classified using (b) and proceed through further analyses.
2. For both offset and onset instances, a boundary cycle candidate was estimated for each offset or onset acoustic feature vector by maximizing the effect size of the vector (unchanged from the aRFF algorithm).
3. The acoustic feature value at the candidate index was compared to (c); if the value did not exceed corresponding threshold, then the instance as not considered as marking the correct boundary cycle and needed to be shifted.
4. If the boundary cycle candidate needed to be shifted, the decision was adjusted using (d).
5. The median of the three boundary cycle candidates was extracted as the ultimate boundary cycle for the offset or onset instance (unchanged from the aRFF algorithm).

**Metrics of Algorithmic Performance**

To examine algorithm performance, two measures of error were selected to evaluate the correspondence between manual and semi-automated estimates. These measures were mean bias error (MBE; **Eq. 2.3**) and root-mean-square error (RMSE; **Eq. 2.4**):

$$\text{Mean bias error (MBE; ST)} = \frac{1}{n}\sum_{i=1}^{n}\left(RFF_{alg} - RFF_{man}\right) \qquad \textbf{[2.3]}$$

$$\text{Root-mean-square error (RMSE; ST)} = \sqrt{\overline{\left(RFF_{alg} - RFF_{man}\right)^2}} \qquad \textbf{[2.4]}$$

MBE was chosen to measure the average bias of semi-automated RFF estimates when compared to manual counterparts; positive MBE values suggest that RFF is being overestimated by the algorithm. On the other hand, RMSE was selected to examine the spread of errors between manual and semi-automated RFF values; larger RMSE values suggest a greater discrepancy between manual and semi-automated estimates.

To assess the accuracy of $f_o$ estimation, the number of erroneous rejections were used in conjunction with MBE and RMSE; these rejections were tabulated as type I or type II errors. Type I errors constituted offset or onset instances that were rejected by the algorithm, but not during manual analysis. Type II errors comprised offset or onset instances that were not rejected by the algorithm, but were rejected during manual analysis (e.g., due to glottalization). Resulting MBE, RMSE, and Type I/II errors were examined for each $f_o$ estimation method. The goal was to determine which method contributed the least amount of error in reference to manual RFF and use it as the primary $f_o$ estimation method in a new version of the aRFF algorithm.

**Validation and Performance**

RFF was first recalculated in the training set using the semi-automated algorithm with updated $f_o$ estimation method and category-specific thresholds. To assess the variation in model performance based on the samples used to train the algorithm, a *k*-fold cross-validation was performed. In this analysis, $k = 10$ was chosen to provide an appropriate estimate of model performance and ensure that the model was not over-fitted (Kuhn et al., 2013). The training dataset was therefore split into *k*-training (347 speakers/1041 voice samples/3123 VCV productions) and *k*-validation (39 speakers/117 speech samples/351 VCV productions) datasets. Algorithm performance was quantified as the average MBE and RMSE error across the 10 folds.

RFF was then calculated in the test set using the aRFF algorithm with refined $f_o$ estimation (pending that autocorrelation was not selected as the optimal $f_o$ estimation method), with category-specific thresholds, and with both refined $f_o$ estimation and category-specific thresholds. RFF values from each algorithm were compared against manual RFF counterparts using MBE and RMSE.

To determine the effectiveness of using pitch strength to account for signal quality and overall severity of dysphonia, resulting MBE and RMSE values for the algorithm with refined $f_o$ estimation and category-specific thresholds were compared against these subjective characteristics. Welch's tests were performed to compare the variances of MBE and RMSE values across signal quality (i.e., recorded in a sound-attenuated room or in a quiet room/waiting area of a voice clinic). Pearson's product-moment correlation coefficients were examined to compare MBE and RMSE values against overall severity

of dysphonia.

## *Results*

### *Evaluation of $f_o$ Estimation Accuracy*

**Table 2.3** shows the errors (RMSE, MBE, Type I/II) for 20 speakers (60 VCV productions) when using the aRFF algorithm in conjunction with each $f_o$ estimation technique. The range of errors produced by the five $f_o$ estimation methods was greater when using the true boundary cycle as a reference (*RMSE*: 1.45 ST, *MBE*: 1.63 ST) than when using the voiceless consonant as a reference (*RMSE*: 0.11 ST, *MBE*: 0.07 ST). The aRFF algorithm using Auditory-SWIPE′ (*RMSE* = 0.52 ST; *MBE* = -0.20 ST) and Halcyon (*RMSE* = 0.81 ST; *MBE* = 0.03 ST) for $f_o$ estimation resulted in the greatest correspondence to manual RFF estimates when the true boundary cycle was provided as a reference. In this scenario, none of the algorithms erroneously rejected VCV productions. The aRFF algorithm using YIN (*RMSE* = 0.39 ST) and RAPT (*MBE* = 0.02 ST) for $f_o$ estimation resulted in the least error when the midpoint of the voiceless consonant was

**Table 2.3. Comparison of fundamental frequency ($f_o$) estimation methods when provided with the manually determined time point corresponding to the vocal cycle closest to the voiceless consonant, and when provided only with the midpoint of the voiceless consonant.**

| Method of $f_o$ Estimation | True Boundary Cycle as Reference | | | | Voiceless Consonant as Reference | | | |
|---|---|---|---|---|---|---|---|---|
| | Error (ST) | | Type I/II Errors | | Error (ST) | | Type I/II Errors | |
| | RMSE | MBE | Offset | Onset | RMSE | MBE | Offset | Onset |
| Autocorrelation | 1.06 | 0.50 | 0 | 0 | 0.43 | 0.09 | 54 | 90 |
| Halcyon | 0.81 | 0.03 | 0 | 0 | 0.41 | 0.06 | 60 | 92 |
| Auditory-SWIPE′ | 0.52 | -0.20 | 0 | 0 | 0.43 | 0.04 | 52 | 98 |
| RAPT | 1.97 | -1.13 | 0 | 0 | 0.50 | 0.02 | 72 | 94 |
| YIN | 0.80 | -0.19 | 0 | 0 | 0.39 | 0.05 | 59 | 100 |

*Note. ST = semitone, RMSE = root-mean-square error, MBE = mean bias error.*

provided as a reference. However, implementing RAPT and YIN each resulted in the largest number of Type I and Type II errors (72 offset and 94 onset for RAPT; 59 offset and 100 onset for YIN). When considering RMSE, MBE, and erroneous rejections together when provided the midpoint of the voiceless consonant as a reference, Auditory-SWIPE′ resulted in the best performance (*RMSE* = 0.43 ST; *MBE* = 0.04 ST; 52 erroneous offset rejections, 98 erroneous onset rejections).

In both scenarios, Auditory-SWIPE′ and Halcyon demonstrate comparable accuracy. Analyzing the performance of each $f_o$ estimation method when provided with indices for the true boundary cycle was conducted to simulate the downstream effects of the acoustic features accurately identifying the boundary between the voiced and voiceless segments. As such, this scenario was weighed more heavily than when the midpoint of the voiceless consonant was provided. With this in mind, Auditory-SWIPE′ exhibited superior performance, as well as respectable performance when only provided with the midpoint of the voiceless consonant. Considering not only these results, but also its functionality for computing the pitch strength contour necessary to account for sample characteristics, Auditory-SWIPE′ was selected for $f_o$ estimation in the refined version of the algorithm, called "aRFF-A" (**aRFF** with **A**uditory-SWIPE′).

*Evaluation of Category-specific Thresholds*

Automated Sample Rejection

**Fig. 2.6** shows the distribution of pitch strength values from the 3474 VCV productions of the training set. The average pitch strength of the training set was .33 (*SD* = .08, *range* = .01–.54). Of the 3474 VCV productions, 3271 offset and 2854 onset instances were considered valid during manual analysis, whereas 203 offset and 620 onset instances were rejected. An ROC curve was constructed to examine the discriminatory ability of pitch strength to distinguish valid and rejected instances (see **Fig. 2.7**). The resulting area under the ROC curve was .73 (*95% CI* = .71–.75). Using the maximum PLR (100% specificity, 4% sensitivity), a pitch strength threshold of .05 was selected as rejection criterion. Thus, speech samples with a pitch strength of .05 or lower would be rejected prior to

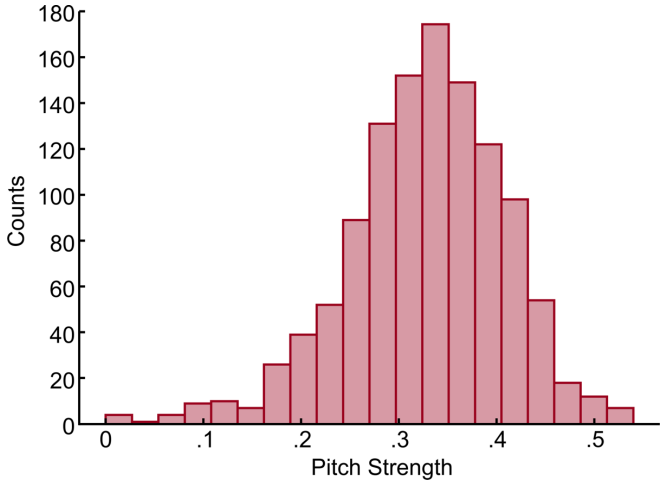


**Figure 2.6. Histogram of pitch strength values for the 3474 vowel–voiceless consonant–vowel productions of the training set.**

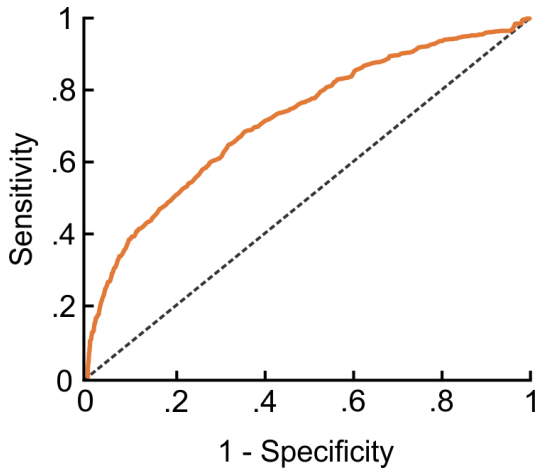


**Figure 2.7. Receiver operating characteristic curve of pitch strength values for relative fundamental frequency instances rejected during manual analysis. The dashed line is indicative of no discrimination.**

RFF cycle analysis. Further analysis includes 3270 offset instances and 2853 onset instances that were not excluded due to manual rejection or low (<.05) pitch strength values.

**Boundary Cycle Shifts**

**Fig. 2.8** shows the relationship between acoustic features and the true boundary cycle for the training dataset (3270 offset instances, 2853 onset instances). Acoustic features were calculated using raw and band-pass filtered versions of the microphone signal. For both versions of the signal, normalized peak-to-peak amplitude increased toward the vowel for voicing offset (negative pitch period distance) and onset (positive

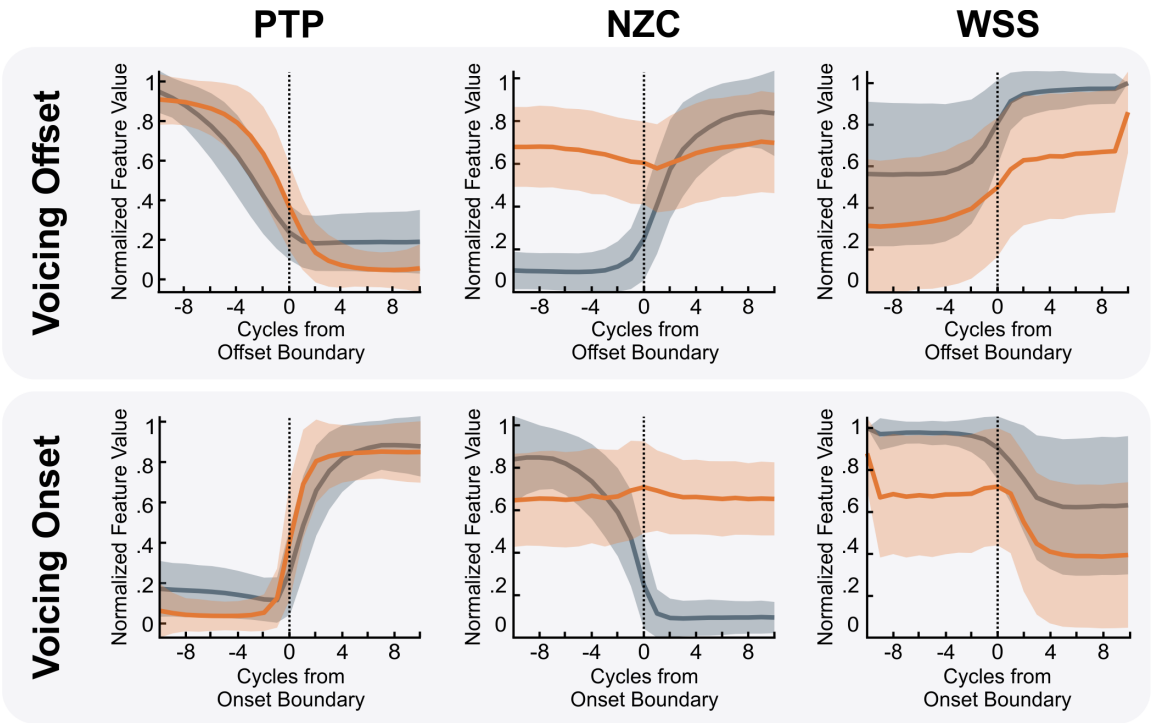


**Figure 2.8. Peak-to-peak amplitude (PTP), number of zero crossings (NZC), and waveform shape similarity (WSS) as a function of the number of pitch periods from the true boundary cycle (dashed vertical line). Offset cycle 10 for voicing offset is shown in the upper panels, and onset cycle 1 for voicing onset is shown in the lower panels. Features are calculated using raw (gray) and band-pass filtered (orange) versions of the microphone signal. Solid lines indicate mean values and shaded regions indicate standard deviation.**

pitch period distance), yet was relatively stable during the voiceless consonant. Zero crossing trends were not well-defined when using the filtered version of the microphone signal. Using the raw microphone signal, the number of zero crossings increased towards the voiceless consonant for voicing offset and onset, but were stable during the vowel. When calculated using either version of the microphone signal, waveform shape similarity—calculated in reference to the voiceless consonant—matched the trends observed in number of zero crossings for voicing offset and onset. However, these trends were observably less pronounced when using the filtered microphone signal.

**Fig. 2.9** shows the results of the Cohen's *d* analysis to identify the average discrepancy between boundary cycle candidates ("predicted" boundary cycle) and the

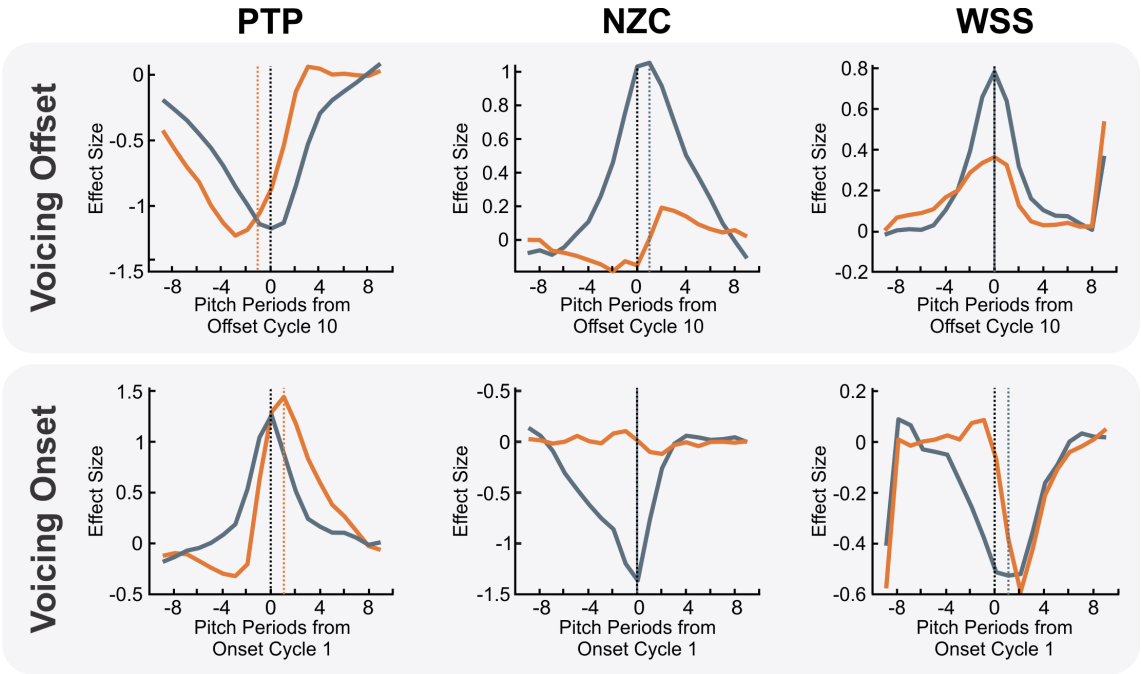


**Figure 2.9. Cohen's *d* effect sizes computed across cycles for normalized peak-to-peak amplitude (PTP), number of zero crossings (NZC), and waveform shape similarity (WSS). Trends are shown as a function of the number of pitch periods from the true boundary cycle (dashed vertical line). Offset cycle 10 for voicing offset is shown in the upper panels, and onset cycle 1 for voicing onset is shown in the lower panels. Features are calculated using raw (gray) and band-pass filtered (orange) versions of the microphone signal. Vocal cycles that elicit the maximum effect size are denoted by a gray or orange dashed line.**

true boundary cycle. Effect sizes were calculated by analyzing the mean feature values across cycles; for instance, a Cohen's *d* value at the true boundary (denoted by the black dashed lines in **Fig. 2.9**) would be computed by calculating the effect size between -1 and +1 pitch periods from the true boundary cycle. The filtered microphone signal was selected for calculating PTP since larger effect sizes were elicited when comparing feature values across cycles. The raw microphone signal was selected for calculating NZC and WSS since filtering the microphone signal reduced the effect size of feature values across cycles.

On average, the boundary cycle candidates predicted using filtered PTP and raw WSS were located one cycle closer to the vowel than the true boundary cycle for voicing onset. This would suggest that the candidate should be shifted backward in time by one vocal cycle. The boundary cycle candidate predicted using filtered PTP for voicing offset was also one vocal cycle closer to the vowel, implicating a shift forward in time by one

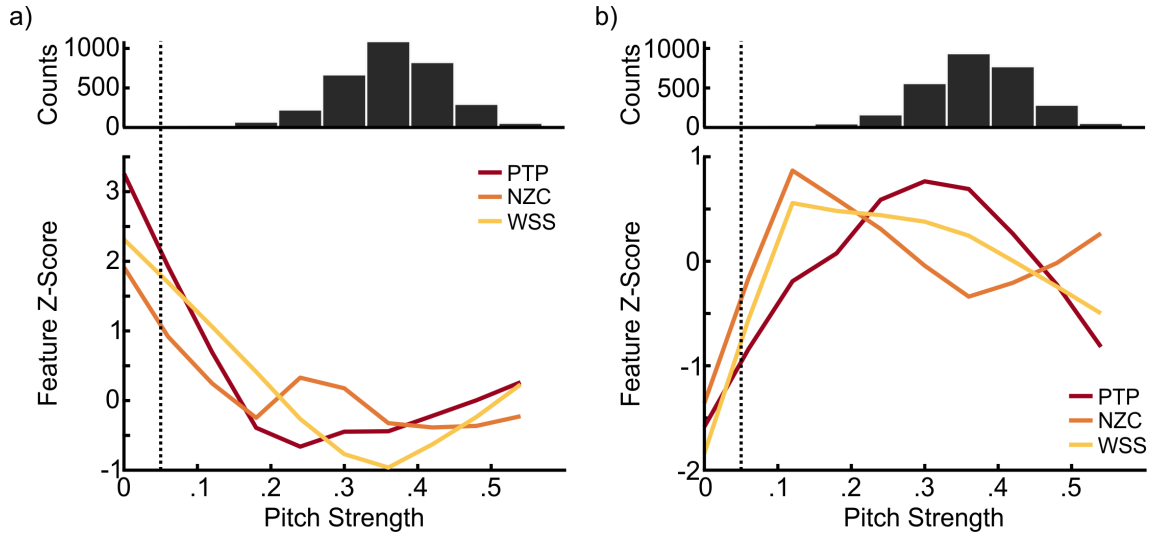


**Figure 2.10. Distribution of normalized feature values (via z-scores) across pitch strength for (a) voicing offset and (b) voicing onset. Upper panels show sample counts per pitch strength bin.**

vocal cycle. Boundary cycle candidates identified using the raw NZC were typically one vocal cycle closer to the voiceless consonant for voicing offset than the true boundary cycle, corresponding to a shift backward in time by one cycle for voicing offset.

**Category Creation**

Voice sample categories were created by manually visualizing trends in acoustic feature values at the true boundary cycle across pitch strength for 3270 offset instances and 2853 onset instances (see **Fig. 2.10**). Four pitch strength cut-offs were empirically selected to describe the trends in acoustic feature values for voicing offset: .15, .25, .35, and .45. Thus, in addition to the rejection criteria of .05, five pitch strength categories resulted for voicing offset, as follows:

$$cat_{off}(S)=\begin{cases} 1, & .05 < PS \leq .15 \\ 2, & .15 < PS \leq .25 \\ 3, & .25 < PS \leq .35 \\ 4, & .35 < PS \leq .45 \\ 5, & PS > .45 \end{cases} \quad \textbf{[2.5]}$$

In **Eq. 2.5**, pitch strength is denoted by the variable S, and the speech sample category is described by $cat_{off}$. Similar to voicing offset, manual examination of the three acoustic features resulted in four pitch strength cut-offs for voicing onset: .15, .25, .35, and .55. Five categories resulted for voicing onset ($cat_{on}$) as a function of pitch strength (PS):

$$cat_{on}(S)=\begin{cases} 1, & .05 < PS \leq .15 \\ 2, & .15 < PS \leq .25 \\ 3, & .25 < PS \leq .35 \\ 4, & .35 < PS \leq .55 \\ 5, & PS > .55 \end{cases} \quad \textbf{[2.6]}$$

Within each category, optimal acoustic feature thresholds were identified by

determining the discriminatory ability of pitch strength to distinguish PTP, NZC, and WSS feature values between the true and predicted boundary cycles (see **Fig. 2.11**). For each offset and onset feature category, the Youden index was identified as the threshold to determine whether a boundary cycle candidate should be shifted. Resulting feature thresholds are shown in **Table 2.4**.

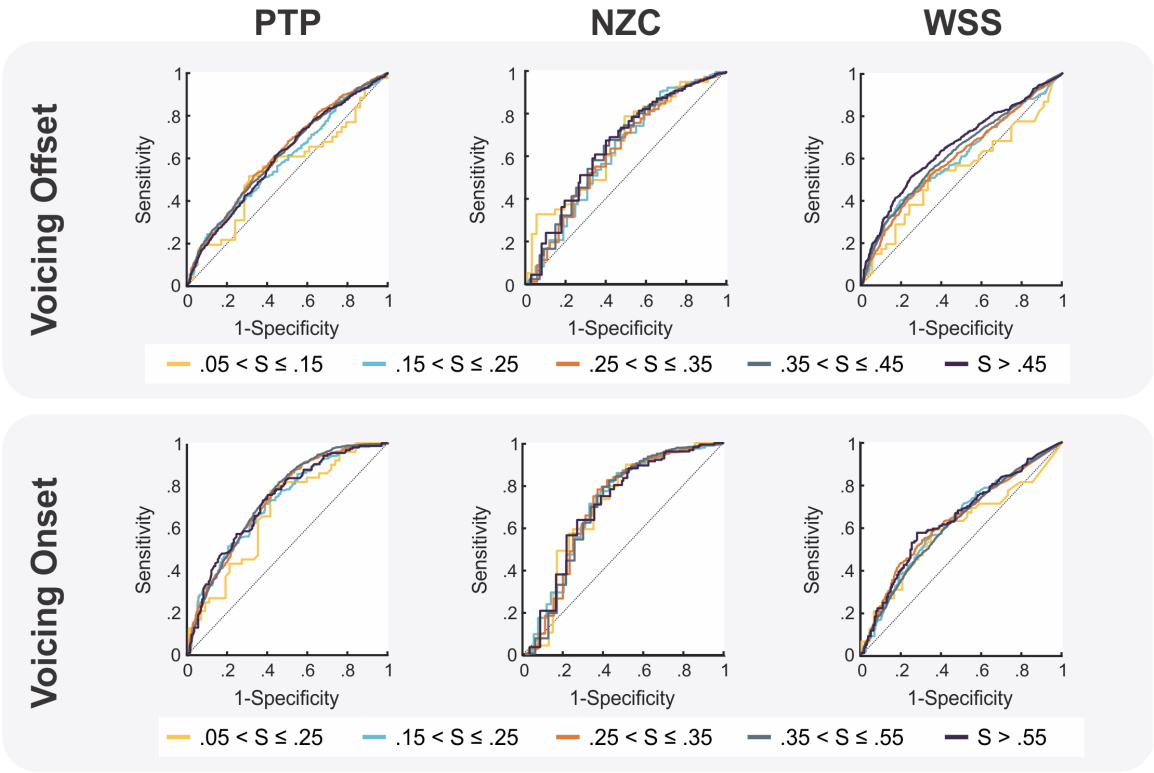


**Figure 2.11. Discriminatory ability of pitch strength (S) categories to distinguish acoustic features at the true versus predicted boundary cycle. Normalized peak-to-peak amplitude (PTP; left panels), number of zero crossings (NZC; middle panels), and waveform shape similarity (WSS; right panels) are shown for voicing offset (top panels) and voicing onset (bottom panels) for the pitch strength categories.**

## *Algorithm Performance*

### Training Set Performance

The effects of implementing the pitch strength-tuned categories (i.e., categorization of samples, acoustic feature thresholds, boundary cycle shifts) were

**Table 2.4. Optimal thresholds obtained at the Youden index from the receiver operating characteristic curves for normalized peak-to-peak amplitude (PTP), number of zero crossings (NZC), and waveform shape similarity (WSS).**

| Acoustic Feature | Optimal Thresholds by Pitch Strength (PS) Category | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Voicing Offset | | | | | Voicing Onset | | | | |
| | .05 < PS ≤ .15 | .15 < PS ≤ .25 | .25 < PS ≤ .35 | .35 < PS ≤ .45 | PS > .45 | .05 < PS ≤ .15 | .15 < PS ≤ .25 | .25 < PS ≤ .35 | .35 < PS ≤ .55 | PS > .55 |
| PTP | 0.140 | 0.100 | 0.192 | 0.177 | 0.182 | 0.224 | 0.192 | 0.231 | 0.180 | 0.128 |
| NZC | 18 | 24 | 17 | 11 | 10 | 20 | 13 | 13 | 11 | 5 |
| WSS | 0.861 | 0.784 | 0.822 | 0.765 | 0.668 | 0.964 | 0.942 | 0.884 | 0.954 | 0.836 |

evaluated on the training set. Algorithm performance in selecting the true boundary cycle was compared between the aRFF algorithm (autocorrelation for $f_o$ estimation, does not account for voice sample characteristics), the algorithm with Auditory-SWIPE′ for $f_o$ estimation ("aRFF-A"), and the algorithm with Auditory-SWIPE′ for $f_o$ estimation and pitch strength-tuned sample categories ("aRFF-**AP**" for **A**uditory-SWIPE′ and **p**itch strength).

Out of 3270 instances to classify for voicing offset (**Fig. 2.12, top row**), the aRFF-AP algorithm resulted in the largest number of correctly identified boundary cycles ($N$ = 1503), followed by aRFF-A ($N$ = 1399) then aRFF ($N$ = 1349). When considering the instances for which the predicted boundary cycle did not match the true boundary cycle, the majority of misclassifications occurred closer to the vowel for aRFF ($N$ = 1692), aRFF-A ($N$ = 1636), and aRFF-AP ($N$ = 1584).

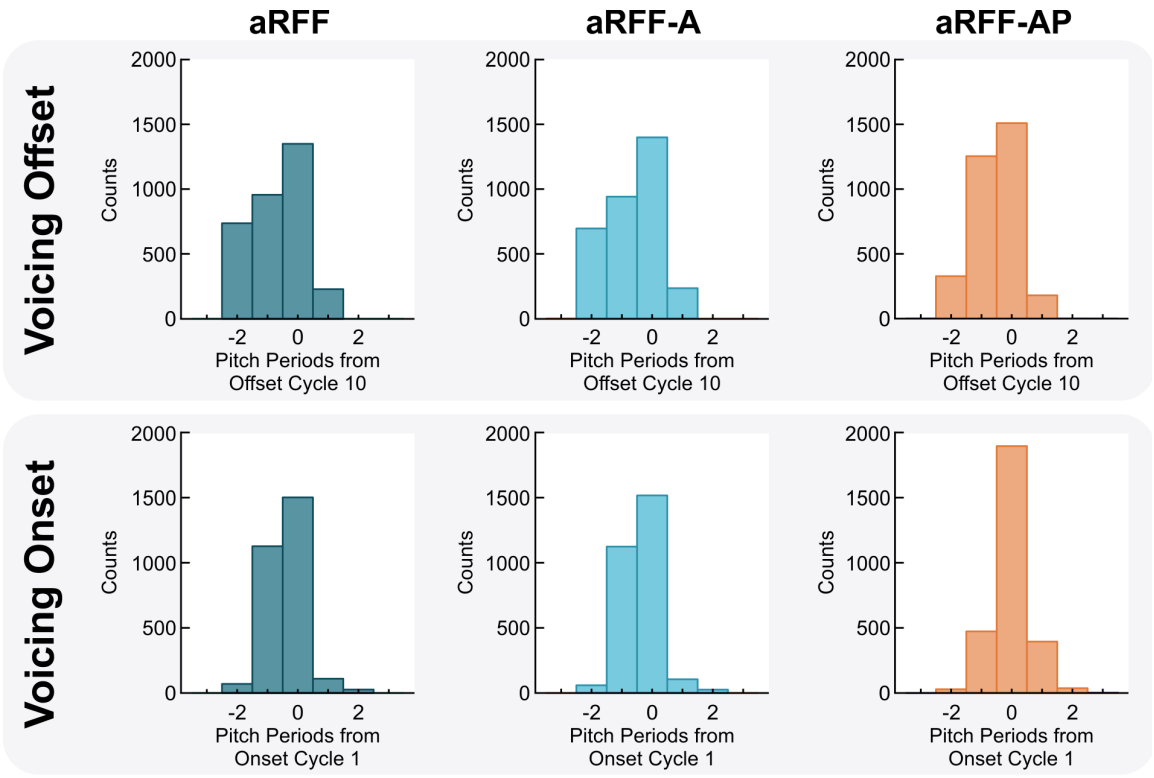


**Figure 2.12. Boundary cycle identification by each of the semi-automated RFF algorithms. Cycle classification is measured as a function of average pitch periods from the true boundary cycle (offset cycle 10 for voicing offset and onset cycle 1 for voicing onset). Results for voicing offset are shown in the upper panels and for voicing onset in the lower panels.**

Out of 2853 instances to classify for voicing onset (**Fig. 2.12, bottom row**), aRFF-AP resulted in the greatest number of correctly identified cycles ($N = 1896$). The aRFF and aRFF-A algorithms produced similar results, with 1502 correctly identified cycles for aRFF and 1517 for aRFF-A. Dissimilar from voicing offset, a great majority of misclassified boundary cycles were identified as occurring closer to the voiceless consonant for aRFF ($N = 1197$) and aRFF-A ($N = 1184$). However, results for aRFF-AP showed a more even split for misclassified cycles: of the 937 misidentified cycles, the boundary cycle was identified as occurring closer to the vowel in 504 instances, whereas it was identified as being closer to the voiceless consonant in 434 instances.

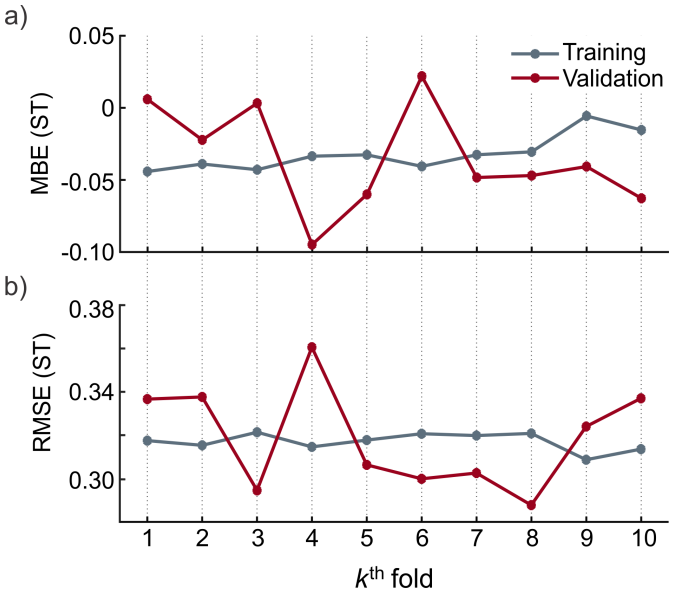


**Figure 2.13. Results of the 10-fold cross-validation examining (a) mean bias error (MBE) and (b) root-mean-squared error (RMSE) for *k*-training (gray) and *k*-validation (red) sets.**

As a next step, *k*-fold cross-validation was performed on all 3474 VCV productions to assess whether category and threshold parameters were overfit to the data (see **Fig. 2.13**). The cross-validation estimate of prediction error was averaged across $k = 10$ folds, resulting in an MBE of -0.03 ST ($SD = 0.01$ ST) and RMSE of 0.31 ($SD = 0.004$ ST) of the *k*-training set ($N = 1042$), and an MBE of -0.03 ST ($SD = 0.04$ ST) and RMSE of 0.32 ST ($SD = 0.02$ ST) in the *k*-validation set ($N = 116$). Given the small discrepancy between error estimates of the *k*-training and *k*-validation sets, it was determined that the constructed model was not overfit to the training data, and parameters were retained to finalize the aRFF-AP algorithm.

## Test Set Performance

### Distribution of Pitch Strength Categories

**Table 2.5** shows the distribution of voice samples in the test set (873 VCV productions) by the pitch strength categories described in **Eq. 2.5** and **Eq. 2.6**. In general, more onset instances were rejected than offset instances for each factor. Results are further discussed by speaker factor.

**Table 2.5. Distribution of pitch strength categories for voicing offset and onset instances in the test set (873 vowel–voiceless consonant–vowel productions from 291 speech samples). Values are shown as a percentage (%) of the total number of productions (N) and do not reflect speech samples that were rejected during pre-processing.**

| Speaker Factor | N | Voicing Offset (N = 286) | | | | | | Voicing Onset (N = 290) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | % of Samples/Pitch Strength Category | | | | | | % of Samples/Pitch Strength Category | | | | | |
| | | *Rej* | *1* | *2* | *3* | *4* | *5* | *Rej* | *1* | *2* | *3* | *4* | *5* |
| **Sex** | | | | | | | | | | | | | |
| Male | 279 | 0 | 2.9 | 12.5 | 40.5 | 35.1 | 9.0 | 0.7 | 9.3 | 11.8 | 26.2 | 49.8 | 2.2 |
| Female | 594 | 1.9 | 1.7 | 7.1 | 25.6 | 43.4 | 20.4 | 2.5 | 4.9 | 8.8 | 19.0 | 61.3 | 3.5 |
| **Group** | | | | | | | | | | | | | |
| Voice Disorder | 423 | 2.6 | 3.3 | 12.1 | 33.3 | 35.5 | 13.2 | 3.1 | 7.3 | 11.1 | 23.4 | 54.1 | 0.9 |
| No Voice Disorder | 450 | 0 | 0.9 | 5.8 | 27.6 | 45.8 | 20.0 | 0.9 | 5.3 | 8.4 | 19.3 | 60.9 | 5.1 |
| **Location** | | | | | | | | | | | | | |
| Quiet Room | 306 | 2.9 | 3.6 | 13.4 | 41.5 | 30.4 | 8.2 | 3.3 | 10.5 | 13.7 | 25.5 | 45.4 | 1.6 |
| Sound Booth | 567 | 0.4 | 1.2 | 6.3 | 24.4 | 46.4 | 21.3 | 1.2 | 4.1 | 7.6 | 19.0 | 64.2 | 3.9 |

*Note. Rej = Rejected due to pitch strength values <.05.*

A larger proportion of female voices were rejected (1.9% for voicing offset, 2.5% for voicing onset) due to low pitch strength values (<.05) than male voices (0% for voicing offset, 0.7% for voicing onset). To this end, a greater proportion of VCV productions from female speakers fell within the higher pitch strength categories (i.e., categories 4 and 5) than male speakers for both voicing offset and onset (offset: female = 63.8%, male = 44.1%; onset: female = 64.8%, male = 52.0%).

By speaker group, a lower percentage of voice samples were rejected from typical speakers (0% for voicing offset, 0.9% for voicing onset) than from speakers with disordered voices (2.6% for voicing offset, 3.1% for voicing onset). A larger percentage of typical speakers ("no voice disorder") also exhibited pitch strength values above .35 (i.e., categories 4 or 5) for offset and onset VCV productions (offset: 65.8%, onset: 66.0%) than speakers with disordered voices ("voice disorder;" offset: 48.7%, onset:

55.1%).

When taking recording location into account, a greater proportion of samples were rejected when recorded in a quiet room or waiting area (2.9% for voicing offset, 3.3% for voicing onset) compared to in a sound-attenuated room (0.4% for voicing offset, 1.2% for voicing onset). Of 306 VCV productions recorded in a quiet room or waiting area of a voice clinic, the majority of these productions (179 offset instances, 152 onset instances) exhibited a pitch strength value below .35 (i.e., categories 1–3). Of the 567 VCV productions recorded in a sound-attenuated room, however, more than 50% of these productions (384 offset instances, 386 onset instances) were classified as having a pitch strength above .35 (i.e., categories 4 or 5). Finally, a greater proportion of speakers recorded in a sound-attenuated room resulted in higher pitch strength categories (offset: 67.7%, onset: 68.1%) than those recorded in a quiet room or waiting area (offset: 38.6%, onset: 47.1%).

**Comparison to Manual RFF Estimates**

RFF was computed in the independent test set (873 VCV productions) described above in **Table 2.5** when using each of the semi-automated algorithms (aRFF, aRFF-A, aRFF-AP). The aRFF-AP algorithm resulted in the least error when compared to manual RFF estimates (*MBE* = 0.01 ST, *RMSE* = 0.28 ST; see **Table 2.6**), followed by aRFF-A (*MBE* = 0.08 ST, *RMSE* = 0.32 ST) then aRFF (*MBE* = 0.09 ST, *RMSE* = 0.34 ST).

When examining these errors across RFF cycle (see **Fig. 2.14**), the MBE of offset cycles 2–10 substantially decrease when using aRFF-AP rather than aRFF or aRFF-A. The MBE of onset cycle 2 improves when using aRFF-AP, but approach similar values

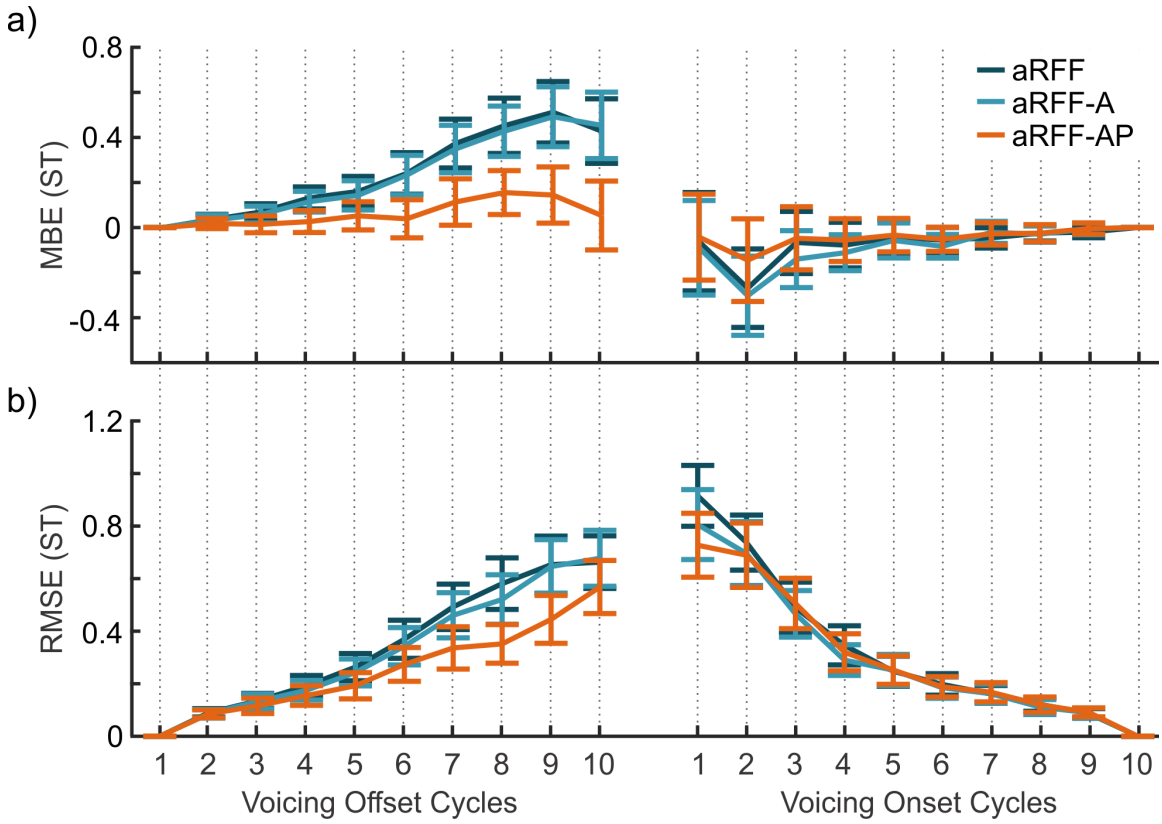


**Figure 2.14. Resulting (a) mean bias error (MBE) and (b) root-mean-squared error (RMSE) of for aRFF (dark blue), aRFF-A (light blue), and aRFF-AP (orange) algorithms across vocal cycles.**

**Table 2.6. Comparison of manual and automated relative fundamental frequency estimates by algorithm version, computed using a test set of 291 speech samples. Error values are shown as mean (95% confidence interval).**

| Algorithm Version | Mean Bias Error (ST) | Root-mean-square Error (ST) |
|---|---|---|
| aRFF | 0.09 (0.07–0.11) | 0.34 (0.32–0.36) |
| aRFF-A | 0.08 (0.05–0.10) | 0.32 (0.29–0.34) |
| aRFF-AP | 0.01 (-0.01–0.03) | 0.28 (0.26–0.30) |

for onset cycles 1 and 3–9. RMSE values also decrease for offset cycles 2–10 and onset cycle 1 when using aRFF-AP compared to aRFF and aRFF-A. Taking these findings into account, aRFF-AP leads to the greatest correspondence to manual RFF estimates.

**Comparison to Voice Sample Characteristics**

The Welch's test examining MBE values against signal quality (i.e., recorded in a quiet room or waiting area versus sound-attenuated room) revealed that recording location produced a medium significant effect ($p$ = .04, $d$ = 0.47) on RFF values produced from the aRFF-AP algorithm (Witte et al., 2010, p. 383). The average MBE was larger for sound samples recorded in a quiet room or waiting area ($M$ = 0.08 ST, $SD$ = 0.23 ST) compared to those recorded in a sound-attenuated room ($M$ = -0.02 ST, $SD$ = 0.21 ST). However, the Welch's test examining RMSE values across recording locations showed that recording location was not a significant factor ($p$ = .25), with the average RMSE for sound samples recorded in a quiet room or waiting area ($M$ = 0.31 ST, $SD$ = 0.18 ST) similar to that of sound samples recorded in a sound-attenuated room ($M$ = 0.27 ST, $SD$ = 0.16 ST). Pearson product-moment correlation coefficients conducted for MBE and RMSE against overall severity of dysphonia elicited $r$ = -.08 ($p$ = .44) and $r$ = .44 ($p$ < .001), respectively.

**Algorithmic Run Time**

Because the semi-automated RFF algorithm was designed, in part, to mitigate the time-intensive nature of manual RFF estimation, $f_o$ processing time was also evaluated. This *post hoc* analysis was performed because autocorrelation was removed from the aRFF algorithm in favor of Auditory-SWIPE′. The runtimes necessary to compute the $f_o$ contour of test set samples were compared between autocorrelation (used in the aRFF algorithm) and Auditory-SWIPE′ (used in the aRFF-A and aRFF-AP algorithms). On average, autocorrelation required 0.28 seconds ($SD$ = 0.11 seconds) to process each voice

sample containing three VCV productions, whereas Auditory-SWIPE′ required 3.59 seconds (*SD* = 1.34 seconds).

### *Discussion*

This study sought to examine the impacts of $f_o$ estimation method and voice sample characteristics on the semi-automated RFF algorithm. To carry out this work, a large database of RFF stimuli collected across a wide range of voice signals were analyzed. The samples exhibited a large degree of vocal function and were recorded in a variety of locations, including the waiting areas of a voice clinic, in quiet rooms, and in sound-attenuated rooms. Five $f_o$ estimation techniques were compared within the RFF algorithm to determine which method yielded the greatest correspondence with gold-standard, manual RFF estimates. From this analysis, Auditory-SWIPE′ was implemented in the RFF algorithm to replace the previous $f_o$ estimation method, autocorrelation. The effects of voice sample characteristics were then quantified using the acoustic measure, pitch strength. Categories based on pitch strength values were developed, and RFF algorithm parameters were tuned to each category. Semi-automated RFF estimates were then calculated using the category-specific thresholds and compared against manual values in an independent test set of voice samples.

The results of the current study show that refining the method of $f_o$ estimation and accounting for the variation in voice sample characteristics increases the correspondence between manual and semi-automated RFF estimates. MBE and RMSE were employed to provide insight into the accuracy and precision of semi-automated RFF values, respectively. Using these metrics, it was determined that the refined RFF algorithm will,

on average, generate a positively biased systematic error of 0.01 ST, with a spread of error values that approach 0.28 ST.

The errors seen after refining the RFF algorithm are smaller than the meaningful changes in RFF discussed in the literature. After undergoing voice therapy, individuals with hyperfunctional voices were found to produce RFF values comparable to those obtained from typical speakers (Stepp et al., 2011d). The largest observed changes in RFF values were in the boundary cycles: on average, voicing offset cycle 10 increased by +0.5 ST, and voicing onset cycle 1 increased by +0.81 ST. The mean accuracy of RFF values when using the refined algorithm were +0.05 ST for voicing offset cycle 10 and -0.04 ST for voicing onset cycle 1. These results suggest that the MBE associated with the refined RFF algorithm is on the order of one magnitude smaller than the increases in RFF observed by Stepp et al. (2011d). Users can therefore expect that clinically meaningful changes in RFF will not be masked by errors associated with using the semi-automated algorithm to compute RFF.

Autocorrelation was replaced as the $f_o$ estimator in the aRFF algorithm in favor of Auditory-SWIPE′. To obtain lower MBE and RMSE errors, Auditory-SWIPE′ is more computationally complex than autocorrelation; this algorithm switch led to a nearly 13-fold increase in runtime for calculating the $f_o$ contour of three VCV productions. Although more processing time is required to compute $f_o$, the trade-off for more accurate $f_o$ estimation is justified to improve the clinical viability of the aRFF algorithm. It is also worth considering that this large increase in runtime may be, in part, because Auditory-SWIPE′ also calculates the pitch strength contour of the signal in addition to the $f_o$

contour. The necessity of using pitch strength to categorize voice samples in the aRFF-AP algorithm further substantiates the switch from simple autocorrelation to the more computationally intensive Auditory-SWIPE′ algorithm.

Pitch strength was used to quantitatively account for variations in signal quality and overall severity of dysphonia. With sample pitch strength estimates, RFF is calculated using rule-based processing rather than through subjective, specific sample characteristics such as clinical diagnosis or recording location. When examining how resulting errors compare to these subjective sample characteristics, it was found that the aRFF-AP algorithm was differentially affected by sample characteristics. For instance, the spread of RFF estimates was relatively similar across recording environments, suggesting that the precision of RFF estimates was not affected by signal quality. On the other hand, the accuracy of RFF estimates was lower for samples recorded in a quiet room; these findings indicate that the RFF values were affected by signal quality despite using pitch strength to account for sample characteristics. Users of the aRFF-AP algorithm can expect systematic errors to occur, on average, on the order of 0.08 ST for samples recorded in a quiet room and of -0.02 ST for samples recorded in a sound-attenuated room.

Errors resulting from the aRFF-AP algorithm were found to be affected by variations in overall severity of dysphonia. A very weak, negative relationship was found between overall severity of dysphonia and MBE. This indicates that the accuracy of an RFF estimate from the aRFF-AP algorithm would not be substantially different from manual estimates as a function of overall severity of dysphonia. On the other hand, a

moderate, positive relationship was found between overall severity of dysphonia and RMSE. These findings suggest that the precision of RFF values may be positively related to the overall severity of dysphonia of the speaker.

### *Limitations and Future Directions*

In the current study, the aRFF algorithm was refined to increase correspondence with manual RFF estimates; however, neither of the error metrics computed between manual and semi-automated RFF estimates reached zero. This may be because pitch strength failed to comprehensively account for variations in overall severity of dysphonia and signal acquisition quality, as demonstrated by the weak relationship between overall severity of dysphonia and MBE. Future investigations should examine additional or alternative acoustic metrics to account for the diversity in clinical sample characteristics. Examples of such metrics may include cepstral peak prominence to assess speaker-related sample characteristics (Anand, Kopf, Shrivastav, & Eddins, 2019a).

The source of these non-zero errors may also be attributed to difficulties in $f_o$ estimation. Of the $f_o$ detection methods examined in the current study, Auditory-SWIPE′ was shown to be the best choice for $f_o$ estimation. Yet the approach used to select the best $f_o$ estimation method (i.e., augmenting the aRFF algorithm with each $f_o$ estimation method and comparing the resulting RFF values to manual RFF values) is limited. As only five relatively well-established $f_o$ estimation methods were compared, it is possible that other $f_o$ estimation algorithms not examined here (e.g., nearly defect-free algorithm; Kawahara, de Cheveigné, Banno, Takahashi, & Irino, 2005) may result in greater algorithmic RFF accuracy.

In addition to comparing a larger set of $f_o$ estimation methods, it may be interesting to explore the utility of fusing $f_o$ detection methods or using adaptive techniques to estimate $f_o$. For instance, Tsanas et al. (2014) showed that using an adaptive Kalman filter framework led to improvements in $f_o$ estimation accuracy from a sustained /a/ over nine previously established methods (including Auditory-SWIPE′). Importantly, the authors assessed $f_o$ estimation accuracy relative to ground-truth values that were either calculated from glottal cycles detected via electroglottography or from synthetic signals with pre-determined $f_o$ values. As both manual and semi-automated RFF estimation rely on the acoustic signal to compute $f_o$, it is thus also important to consider that manual RFF estimation may not be a true reflection of $f_o$. Future work should therefore use electroglottography and/or numerical modeling to compare and validate $f_o$ values from both algorithmic and manual RFF methods.

Within this vein, it is also possible that non-zero error metrics were obtained because manual estimation is not a true gold standard for RFF values. If so, it may not be necessary to remove errors between manual and semi-automated RFF estimates. Manual RFF is derived using microphone (Eadie et al., 2013; Goberman et al., 2008; Robb et al., 2002; Stepp, 2013; Stepp et al., 2010b; Stepp et al., 2011d; Stepp et al., 2012; Watson, 1998; Watson & Schlauch, 2008) or accelerometer signals (Lien et al., 2015a). However, there may be a discrepancy between these signals and the physiological initiation or termination of voicing at the vocal fold level. Trained technicians implement trial-and-error to identify this physiological boundary via manual RFF estimation. Due to the subjective nature of their process, the selected boundary may not be the true initiation or

termination of voicing. The semi-automated RFF algorithm makes use of three acoustic features to identify this transition point. Yet it is unclear as to how these features relate to the physiological vibrations of the vocal folds during the transition into and out of voicing. As a result, investigation into the physiological relevance of RFF via manual and semi-automated techniques is warranted.

Although the current study details preliminary steps taken to refine the semi-automated algorithm for RFF estimation, further investigation is warranted to continue to enhance accuracy and versatility across a broad range of vocal function. Specifically, the sample distribution analyzed in this study may not be fully representative of clinical practice. For instance, Martins et al. (2016) reports a substantial prevalence of vocal polyps in adults with voice disorders (12% of 2019 adults analyzed); however, only 3% of the population examined in the current study was diagnosed with vocal polyps (see **Table 2.1**). Furthermore, nearly 37% of the speakers with voice disorders analyzed in the current study were diagnosed with Parkinson's disease and approximately 33% were diagnosed with muscle tension dysphonia. Because a substantial portion of our sample group consisted of these individuals, it is possible that our results are biased towards speakers with Parkinson's disease and speakers with muscle tension dysphonia. As such, future studies consider the prevalence of voice disorders in the examined population and make these representative of those seen in clinical practice. Doing so will enhance the clinical relevance of using RFF to acoustically examine vocal function.

It is also unclear whether the heterogeneity of the equipment used to capture speech acoustics played a role in the differences seen in RFF accuracy in terms of signal

acquisition quality. In particular, it was hypothesized that signal acquisition quality was a feature of acoustic speech samples that affected the accuracy of RFF estimates; however, signal acquisition quality was examined solely in terms of whether the speech sample was recorded in a sound-attenuated room versus a quiet room or waiting area. As such, future work should also take into account the equipment used to record speech (e.g., microphone) and the characteristics of the recording environment (e.g., background noise levels, reverberation) when examining signal acquisition quality.

### *Conclusions*

RFF has shown promise as an acoustic measure for assessing and tracking laryngeal muscle tension; however, semi-automated RFF is not yet transferable to the clinic due to instability across a range of vocal signals that would be typically encountered. Thus, the impacts of $f_0$ estimation method and sample characteristics on the correspondence between automated and gold standard manual RFF estimates was evaluated. Upon refining the $f_0$ estimation method using the Auditory-SWIPE′ algorithm, in conjunction with accounting for sample characteristics via pitch strength categories, the accuracy and precision of semi-automated RFF estimates increased by 88.4% and 17.3%, respectively. These findings highlight the importance of considering the broad range of vocal function that may be encountered in clinical populations.

# CHAPTER 3. The Relationship between Acoustic Features and Vocal Fold Vibratory Characteristics during Intervocalic Offsets and Onsets


## *Abstract*

**Purpose:** The aim of this study was to elucidate the physiological factors influencing acoustic outputs of RFF estimation.

**Methods:** Sixty-nine vocally healthy adults (33 female, 36 male; *M* = 43.2 years, *SD* = 23.1 years) and fifty-three adults with disordered voices (29 female, 24 male; *M* = 49.5 years, *SD* = 18.4 years) produced strings of the utterance, /ifi/, while altering their vocal rate and vocal effort. Simultaneous recordings were made using a microphone and flexible nasendoscope. The initiation (voicing onset) and termination (voicing offset) of vocal fold vibration were identified through laryngoscopic images. A series of acoustic features were examined in reference to these time points, and the acoustic features that best coincided with voicing offset and onset were then implemented within the semi-automated RFF algorithm ("aRFF-APH"). The accuracy of the aRFF-APH algorithm in identifying these physiological transitions in voicing was then assessed against (1) the current version of the semi-automated RFF algorithm ("aRFF-AP"), and (2) manual RFF estimation, the current gold-standard technique for calculating RFF. Algorithmic accuracy was measured as the discrepancy between the physiological transition (boundary) and acoustically determined boundary. Chi-square tests of independence were performed to investigate the association between the three RFF estimation methods and accuracy in identifying the physiological boundary cycle.

**Results:** The association between RFF estimation methods and the accuracy of identifying the physiological boundary cycle was significant for both voicing offsets ($p <$ .001, $V = .53$) and voicing onsets ($p < .001$, $V = .51$). The aRFF-APH algorithm led to the greatest overall correspondence between acoustically and physiologically identified boundary cycles. Of 7709 /ifi/ productions, 76.9% of boundary cycles were accurately identified when using the aRFF-APH algorithm (5567 offset, 6290 onset), compared to only 71.4% when using the aRFF-AP algorithm (5016 offset, 5984 onset) and 20.2% when using manual estimation (984 offset, 2137 onset).

**Conclusions:** Incorporating acoustic features that corresponded to the physiological termination and initiation of vocal fold vibration led to improvements in algorithmic accuracy. By reducing the discrepancy between acoustically and physiologically determined voicing boundaries, improvements in the precision of using RFF were shown to reflect the underlying physiological mechanisms for voicing offsets and onsets. Future work should validate the aRFF-APH algorithm in a larger speaker dataset that encompasses a broader range of vocal function.

### *Background*

Relative fundamental frequency (RFF) has been proposed as an acoustic metric that estimate the degree of laryngeal muscle tension. RFF is calculated from short-term changes in instantaneous fundamental frequency ($f_o$) as a speaker devoices (i.e., voiced-to-unvoiced transition, or "voicing offset") and reinitiates voicing (i.e., unvoiced-to-voiced transition, or "voicing onset"). These transitions may be captured in a vowel–voiceless consonant–vowel (VCV) production by estimating the instantaneous $f_o$ of the ten voiced cycles preceding and following the voiceless consonant, respectively (see **Fig. 3.1**). These $f_o$ values are then normalized to a steady-state $f_o$ of the nearest vowel ($f_o^{ref}$) to produce an RFF estimate in semitones (ST), as shown in **Eq. 3.1**:

$$\text{RFF (ST)} = 12 \times \log_2\left(\frac{f_o}{f_o^{ref}}\right) \qquad \textbf{[3.1]}$$

Currently, the gold-standard method of computing RFF is through manual estimation techniques using Praat software (Boersma, 2001), which employ simple autocorrelation to calculate $f_o$. Autocorrelation operates by comparing a segment of the

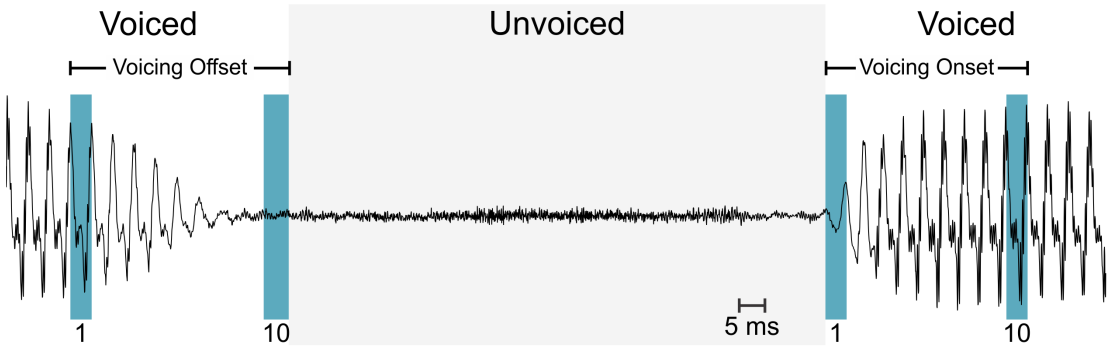


**Figure 3.1. Acoustic waveform of the nonsense word /ifi/, with /i/ segments marked as "voiced" and the /f/ segment marked as "unvoiced" (shaded gray). Intervocalic transitions labeled as voicing offset (/i/ to /f/) and voicing onset (/f/ to /i/). The first and tenth vocal cycles are highlighted for each transition.**

voice signal with an offset by a certain period to provide insight into potential $f_o$ values. Although this method is fast and provides high temporal resolution, simple autocorrelation suffers from assumptions of signal periodicity. Moreover, $f_o$ estimation via autocorrelation requires 2–3 complete pitch periods to examine the physiological $f_o$ ranges encountered in speech. These characteristics are not ideal for estimating $f_o$ during the voicing offset and onset transitions examined in RFF, which specifically capture rapid changes in $f_o$ (characterized by a lack of $f_o$ stationarity; Quatieri, 2008). Using autocorrelation for manual RFF estimation may therefore lead to $f_o$ estimation inaccuracies and poor cycle-to-cycle resolution.

More recent investigations into RFF have resulted in a semi-automated RFF (aRFF) algorithm (Lien, 2015; Lien et al., 2017). Similar to manual RFF estimation in Praat, however, the aRFF algorithm uses autocorrelation to track $f_o$. Specifically, the aRFF algorithm leverages average $f_o$ estimates to create a sliding window that navigates across the acoustic signal in time, collecting potential vocal cycles. Due to the aforementioned shortcomings in using autocorrelation for $f_o$ estimation, Vojtech et al. (2019b; see also *Chapter 2*) compared the effects of different $f_o$ estimation techniques on resulting RFF estimates. The authors determined that $f_o$ estimation via the Auditory-SWIPE′ (Camacho, 2012) method increased the correspondence between semi-automated and manual RFF estimates compared to simple autocorrelation. The results of this work led to a refined version of the aRFF algorithm that not only employs Auditory-SWIPE′ for $f_o$ estimation, but also accounts for differences in voice sample characteristics using the acoustic metric, pitch strength (Camacho et al., 2008; Kopf et al., 2017). As such, the

improved aRFF algorithm is called the “aRFF-AP” algorithm.

In both manual and semi-automated RFF estimation methods, the most tedious step of the RFF computational process is identifying the boundary between voiced and unvoiced speech. As RFF depends on the termination and initiation of voicing within a VCV production, these points in time must be identified from the acoustic signal prior to collecting vocal cycles for estimating RFF. Manual RFF estimation relies on trial-and-error techniques of trained technicians to locate this boundary (requiring 20–40 minutes of analysis time per RFF estimate), whereas the aRFF and aRFF-AP algorithms take advantage of a faster, more objective approach. Specifically, Lien (2015) proposed acoustic feature extraction to identify desired vocal cycles from the voiced-to-unvoiced transitions (and vice versa), which was implemented in aRFF and aRFF-AP. During the sliding window process, acoustic features are calculated per window of time. After collecting a sufficient amount of feature values, the algorithms examine each feature to determine where a state transition occurs. In other words, the aRFF and aRFF-AP algorithms assume that each acoustic feature will exhibit a substantial change in feature values over time and that this change will occur at the boundary between voiced and unvoiced segments. This logic was implemented in both algorithms by maximizing the effect size of each acoustic vector, with both sides of the transition point containing stable values pertaining either to the vowel or voiceless consonant. The index that maximized the effect size of the vector was considered the “boundary cycle” for that acoustic feature. The algorithm then took the median index of the three boundary cycle candidates as the ultimate boundary cycle.

Although semi-automated methods are more objective than manual techniques in identifying the voiced/unvoiced boundary, current methods for locating the boundary remain questionable. In particular, three acoustic features are employed in the aRFF and aRFF-AP algorithms: normalized peak-to-peak amplitude, number of zero crossings, and waveform shape similarity. However, it is unclear whether these are the best choice of acoustic features to mark the initiation and termination of vibration. Whether the boundary cycle identified using these acoustic features actually corresponds to the *physiological* beginning or end of vocal fold vibration requires further inquiry and validation, as both manual and semi-automated RFF methods rely on the acoustic signal as recorded using a microphone. While microphones are able to provide indirect information about the vibration of the vocal folds, these vibrations may be masked to some extent by supraglottic resonances, coarticulation, and radiation. For instance, the vocal cycles closest to the voiceless consonant may be masked by the burst of high frequency energy in frication or aspiration as a result of coarticulation. Thus, in addition to a lack of $f_o$ stationarity during vocal fold offset and onset transitions, signal masking may introduce difficulties in identifying the initiation or termination of vocal fold vibration. Therefore, the uncertainty in boundary cycle identification using the microphone signal warrants further investigation to inform the implementation of acoustic features used in the semi-automated RFF algorithm for more accurate representations of voicing offsets and onsets.

Laryngeal imaging is one technique for visualizing the vibrations of the vocal folds. During laryngoscopic imaging, a device is inserted via the oral or nasal passages to

visualize vocal fold anatomy and physiology. To incorporate vibrational information from laryngeal imaging to RFF estimations, flexible laryngoscopy may be used to visualize the vibrating vocal folds during VCV transitions; using this modality, a device is passed transnasally rather than orally, allowing participants to speak and articulate freely while images of the glottal source are captured. In doing so, the physiological initiation and termination of vocal fold vibration during VCV productions can be recorded for analysis against acoustically derived time points.

The issue with using conventional laryngeal imaging to examine instantaneous $f_o$ is that these laryngoscopic systems record at a frame rate of 30 frames per second (fps). This rate is much too low to observe basic vocal fold vibratory motion, as the average rate of vocal fold vibration in adults is 85–255 cycles per second (Hz) during modal phonation (Baken et al., 2000, p. 156). Performing conventional videoendoscopy will thus not provide sufficient information related to vocal fold vibratory behaviors. One solution to this frame rate issue is to use high-speed videoendoscopy (HSV).

HSV operates using significantly higher frame rates compared to conventional videoendoscopy and videostroboscopy, making it suitable for assessing instantaneous changes in $f_o$. By sampling at frame rates higher than the typical modal speaking rate, HSV is able to capture the true vibratory behavior of the vocal folds, including aperiodic vibration (Deliyski, 2010; Döllinger et al., 2012). As such, HSV can be used to record the fast cycle-to-cycle changes in the vocal fold vibratory behavior during voicing offset and onset (Braunschweig, Flaschka, Schelhorn-Neise, & Döllinger, 2008; Ikuma, Kunduk, & McWhorter, 2013; Kunduk, Yan, McWhorter, & Bless, 2006).

Prior work has used HSV to investigate voicing offsets and onsets relative to the acoustic signal. Specifically, Patel, Forrest, and Hedges (2017) simultaneously captured the acoustic signal and laryngoscopic images in vocally healthy adults. Using the laryngoscopic images, the authors computed the glottic angle waveform to represent the oscillatory onset and offset behavior of the vocal folds. Patel et al. (2017) observed that the onset of the acoustic signal was significantly related to the vocal fold oscillatory onset measures, whereas the offset of the acoustic signal was related to oscillatory offset, as well as the first instance of incomplete glottal closure and complete cessation of vocal fold vibration. The results of this work indicated a tight coupling between the acoustic signal and the physiological vibrations of the vocal folds; however, this relationship may not be generalizable to the acoustic outputs typically examined with RFF. The authors recorded laryngeal images using a rigid laryngoscope as vocally healthy speakers repeated /hi hi hi/ at their typical pitch and loudness. Distinct from other fricatives, /h/ is called a *voiceless glottal fricative*, as it is produced at the level of the glottis. Transitioning from a vowel to /h/ (and vice versa) may require different mechanisms than when transitioning from a vowel to a voiceless obstruent produced via oral constrictions (e.g., /f/, /s/, /ʃ/, /p/, /t/, /k/). This difference could affect the relationship between oscillatory events obtained from the laryngoscopic images and from the acoustic signal. Additionally, the participants in their study were limited to vocally healthy adults, whereas the target population of RFF includes speakers with voice disorders characterized by excessive laryngeal muscle tension. As such, additional investigations must be carried out to examine voicing offsets and onsets in the context of vocally

healthy and disordered speakers.

There are potential advantages for implementing HSV to investigate the relationship between acoustic features and vocal fold vibratory characteristics during voicing offset and onsets. Since the semi-automated RFF algorithms assume that a state transition will occur in acoustic feature values at the boundary between voiced and unvoiced segments, identifying acoustic features that exhibit this trend at the physiological initiation and termination of vocal fold vibration may be useful to enhance the clinical relevance of RFF. Doing so will provide more comprehensive insights into the use of RFF as an objective, acoustic indicator of laryngeal muscle tension.

### *Purpose of the Current Study*

At present, the semi-automated RFF algorithm requires further development prior to widespread implementation in clinical practice. The overarching purpose of the current study was to characterize the physiological factors influencing acoustic outputs within semi-automated RFF algorithms in order to improve the relevance and applicability of RFF in clinical settings. Therefore, individuals with and without voice disorders were enrolled across a wide age range to investigate the relationship between acoustic features and vocal fold vibratory characteristics during intervocalic voicing offset and onsets. To carry out this aim, acoustic features were first identified that corresponded with the physiological initiation and/or termination of vocal fold vibration. The aRFF-AP algorithm (see *Chapter 2*) was further refined by modifying algorithmic parameters corresponding to the HSV-tuned acoustic feature set ("aRFF-APH"). The accuracy of manual and semi-automated (aRFF-AP, aRFF-APH) estimation methods in identifying

the physiological transition in voicing was then compared to assess the physiological relevance of RFF.

## *Methods*

### *Participants*

A total of 122 participants were enrolled for the current study. All participants provided informed, written consent with the Boston University Institutional Review Board. Participants over the age of 50 were administered the Montreal Cognitive Assessment (MoCA) to determine cognitive status. An *a priori* cut-off of ≥21 was set to ensure all included participants had the capacity to consent to the study tasks (Dalrymple-Alford et al., 2010). Participants were designated as vocally healthy, or "typical," speakers or speakers with disordered voices; this latter group comprised adults diagnosed with a voice disorder as well as adults with Parkinson's disease. These two groups are described in detail below.

#### Typical Speakers

Sixty-nine vocally healthy individuals (33 female, 36 male) aged 18–91 years of age ($M$ = 43.2 years, $SD$ = 23.1 years) were recruited to participate in the study. All typical speakers were fluent in English, and had no history of speech, language, hearing, neurological, or voice problems. Participants had no trained singing experience beyond grade school in order to minimize variability in phonatory behaviors that may occur when differentiating between singers and non-singers (Stepp et al., 2011b). All were non-smokers, and were screened by a certified voice-specializing speech-language pathologist for healthy vocal function via auditory-perceptual assessment and flexible nasendoscopic

laryngeal imaging.

Speakers with Disordered Voices

Fifty-three individuals with disordered voices (29 female, 24 male) aged 19–75 years of age (*M* = 49.5 years, *SD* = 18.4 years) were recruited to participate in the study. All speakers were fluent in English and reported no history of hearing problems. Participants within this group were either diagnosed with idiopathic Parkinson's disease (PD) by a neurologist, or were diagnosed with a voice disorder by a board-certified laryngologist. All individuals with Parkinson's disease were recorded while on their typical carbidopa/levodopa medication schedule. Individuals who used deep brain stimulation devices were requested to turn their device off for the duration of the data collection.

**Table 3.1** shows the demographic information for participants with disordered voices. Of the 53 participants, 25 individuals (6 female, 19 male) were diagnosed with PD. The average time since diagnosis was 7 years (SD = 5.8, range = 0–24). The Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) was administered to each participant with PD to determine the extent of both motor and non-motor complications; each examination was administered and scored per protocol by a certified MDS-UPDRS administrator. The severity of motor complications as assessed via the MDS-UPDRS were, on average, moderate (*M* = 48.8, *SD* = 20.5), and ranged from mild to severe (*range* = 13–91; Martínez-Martín et al., 2015). The mean Hoehn-Yahr score was 2.1 (*SD* = 1.1) and ranged from 0 (no disability) to 4 (severe disability; Goetz et al., 2004; Hoehn et al.,

1967). The remaining 28 individuals (23 female, 5 male) were diagnosed with a voice disorder, including muscle tension dysphonia (20/28), nodules (4/28), polyp (2/28), scarring (1/28), or a lesion of unknown type on the vocal folds (1/28).

**Table 3.1. Demographic information of participants with disordered voices.**

| Participant | Sex | Age | Dx | CAPE-V OS | Years Post-Dx | MDS-UPDRS-III | Hoehn-Yahr Scale |
|---|---|---|---|---|---|---|---|
| VD1 | F | 23 | MTD | 0.9 | | | |
| VD2 | F | 26 | MTD | 1.3 | | | |
| VD3 | F | 19 | Nodules | 1.6 | | | |
| VD4 | M | 30 | MTD | 3.3 | | | |
| VD5 | F | 25 | MTD | 4.0 | | | |
| VD6 | M | 59 | PD | 4.0 | 2 | 23 | 2 |
| VD7 | F | 40 | MTD | 4.6 | | | |
| VD8 | F | 54 | MTD | 4.8 | | | |
| VD9 | M | 68 | PD | 5.0 | 6 | 38 | 2 |
| VD10 | F | 21 | MTD | 5.0 | | | |
| VD11 | F | 35 | MTD | 5.1 | | | |
| VD12 | F | 62 | PD | 5.6 | 9 | 49 | 3 |
| VD13 | M | 49 | PD | 5.8 | 7 | 47 | 1 |
| VD14 | M | 67 | PD | 6.4 | 4 | 63 | 3 |
| VD15 | M | 73 | PD | 6.8 | 3 | 23 | 1 |
| VD16 | M | 50 | PD | 7.1 | 0 | 17 | 0 |
| VD17 | F | 24 | MTD | 7.4 | | | |
| VD18 | F | 60 | MTD | 7.4 | | | |
| VD19 | M | 32 | MTD | 8.0 | | | |
| VD20 | F | 29 | MTD | 8.1 | | | |
| VD21 | M | 68 | PD | 8.5 | 1 | 52 | 3 |
| VD22 | F | 27 | MTD | 8.7 | | | |
| VD23 | F | 23 | Lesion | 9.0 | | | |
| VD24 | F | 68 | MTD | 9.3 | | | |
| VD25 | F | 51 | PD | 9.7 | 5 | 13 | 0 |
| VD26 | M | 62 | PD | 10.0 | 3 | 50 | 2 |
| VD27 | M | 45 | PD | 10.4 | 10 | 51 | 2 |
| VD28 | F | 57 | MTD | 10.7 | | | |
| VD29 | F | 57 | MTD | 10.7 | | | |
| VD30 | F | 22 | Nodules | 10.8 | | | |
| VD31 | F | 39 | Scarring | 12.0 | | | |
| VD32 | F | 26 | Nodules | 14.0 | | | |
| VD33 | F | 70 | PD | 14.7 | 6 | 77 | 4 |
| VD34 | F | 65 | PD | 15.4 | 10 | 48 | 3 |
| VD35 | F | 32 | MTD | 15.5 | | | |
| VD36 | M | 55 | PD | 18.4 | 21 | 49 | 3 |
| VD37 | M | 40 | MTD | 19.2 | | | |
| VD38 | M | 67 | PD | 19.4 | 2 | 38 | 2 |
| VD39 | M | 75 | PD | 22.1 | 1.5 | 68 | 2 |
| VD40 | F | 35 | MTD | 26.8 | | | |
| VD41 | F | 21 | Nodules | 26.9 | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| VD42 | M | 62 | PD | 27.9 | 13 | 47 | 2 |
| VD43 | M | 65 | PD | 28.3 | 1 | 35 | 0 |
| VD44 | M | 60 | PD | 30.1 | 7 | 54 | 2 |
| VD45 | F | 74 | PD | 30.6 | 24 | 59 | 2 |
| VD46 | F | 67 | MTD | 32.5 | | | |
| VD47 | F | 73 | PD | 33.3 | 8 | 52 | 2 |
| VD48 | M | 73 | PD | 33.6 | 9 | 19 | 1 |
| VD49 | M | 43 | PD | 35.8 | 5 | 91 | 3 |
| VD50 | M | 70 | Polyp | 38.3 | | | |
| VD51 | M | 48 | Polyp | 38.5 | | | |
| VD52 | M | 72 | PD | 40.9 | 7 | 81 | 4 |
| VD53 | M | 67 | PD | 51.3 | 11 | 77 | 3 |

*Note. Dx = Diagnosis, CAPE-V OS = Consensus of Auditory-Perceptual Evaluation of Voice, Overall Severity of Dysphonia, PD = Parkinson's disease, MDS-UPDRS-III = Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale: Part III, Motor Examination, MTD = Muscle tension dysphonia.*

## Hearing Status

Hearing screening data were collected for 114 of 122 participants. The eight participants for which this data was not collected were vocally healthy young adults who reported no history of hearing disorders. Of the 114 participants, 27 were vocally healthy young adults, 34 were vocally healthy older adults, 28 were adults diagnosed with a voice disorder, and 25 were adults with PD. All vocally healthy young adults passed a hearing screening of pulsed pure tones (Burk & Wiley, 2004) at frequencies of 125, 250, 500, 1000, 2000, and 4000 Hz under 25 dB HL in both ears (American Speech-Language-Hearing Association, 2005).

Of the remaining 87 participants (vocally healthy older adults, adults diagnosed with a voice disorder, and adults with PD), 77 passed the hearing screening of pulsed pure tones at frequencies of 125, 250, 500 and 1000 under 25 dB HL and 2000 and 4000 Hz under 40 dB HL in at least one ear (Schow, 1991). One vocally healthy older adult demonstrated a threshold of 55 dB HL at 4000 Hz in both ears, and one vocally healthy older adult (who was 91 years of age) could not hear frequencies beyond 2000 Hz at any

hearing level. One participant with PD (VD9 in **Table 3.1**) passed at all frequencies in at least one ear except for 2000 Hz, instead showing a threshold of 45 dB HL. Four participants with PD passed at all frequencies below 4000Hz, but two participants (VD39, VD43) had a threshold of 45 dB HL and two participants (VD42, VD47) had a threshold of 50 dB HL at 4000 Hz. One participant with PD (VD52) and one participant with MTD (VD24) passed at frequencies below 2000 Hz, but demonstrated thresholds of 50 dB HL for 2000 Hz and 75 dB HL for 4000 Hz. Finally, one participant with PD (VD38) wore hearing aids during the course of the study, and demonstrated thresholds of 45 dB HL at 125 Hz and 30 dB HL at 1000 Hz.

Dysphonia Severity

A speech-language pathologist specializing in voice disorders assessed the overall severity of dysphonia (OS; 0–100) of each participant using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster et al., 2009). As described in detail in *Chapter 2*, sentences for analysis included "Only we feel you do fail in new fallen dew," and "We all found a wee fly on my food on Monday." Both sentences were blindly evaluated for OS by the speech-language pathologist, yielding two OS scores. The average OS score was computed for each speaker. The speech-language pathologist reanalyzed 15% of speakers in a separate sitting to ensure adequate intrarater reliability. The Pearson's product-moment correlation coefficient was calculated on the ratings using the statistical package R (Version 3.2.4), yielding an intrarater reliability of $r = .96$. From this analysis, the average OS for typical speakers was 8.3 ($SD = 6.7$, *range* = 0.6–34.2), and that of speakers with disordered voices was 15.6 ($SD = 12.4$, *range* = 0.9–51.3). The

overall demographic information, including OS, for vocally healthy speakers (young adults, older adults) and speakers with disordered voices (adults with a voice disorder, adults with Parkinson's disease) are included in **Table 3.2.**

**Table 3.2. Overall demographic information for the 122 speakers.**

| Cohort | Sex | | Age | | | Overall Severity of Dysphonia | | |
|---|---|---|---|---|---|---|---|---|
| | M | F | Mean | SD | Range | Mean | SD | Range |
| Young Adults | 18 | 17 | 22.8 | 5.5 | 18–31 | 5.4 | 3.8 | 0.6–23.5 |
| Older Adults | 18 | 16 | 65.6 | 10.8 | 41–91 | 11.4 | 7.7 | 1.7–34.2 |
| Adults with Voice Disorder | 5 | 23 | 37.5 | 16.1 | 19–70 | 12.3 | 10.7 | 0.9–38.5 |
| Adults with Parkinson's Disease | 19 | 6 | 63.0 | 9.4 | 43–75 | 19.2 | 13.3 | 4.0–51.3 |

*Recording Procedures*

Participants received training to produce iterations of the utterance /ifi/, which comprised four /ifi/ productions, a pause, followed by four /ifi/productions (for a total of eight /ifi/s). The phonemes /i/ and /f/ were chosen for the VCV production since the token /i/ provides an open pharynx for better laryngeal view under endoscopy (McKenna et al., 2016), and the token /f/ minimizes within-speaker variation in the acoustic signal (Lien, Gattuccio, & Stepp, 2014). Subsequently, individuals were trained to produce /ifi/ strings at varying speeds (in beats-per-minute; BPM) and levels of effort to alter the stiffness of their laryngeal musculature (Stepp, Hillman, & Heaton, 2010d). Stiffness was modulated via speed and effort in order to generate voice with varying degrees of tension (McKenna et al., 2016). In doing so, the relationship between acoustically derived signal features and RFF could be investigated across a wide range tension. A metronome was used to train these vocal speeds: slow rate at 50 BPM, regular rate at 65 BPM, and fast rate at 80 BPM. Participants were then trained to produce /ifi/ strings at varying levels of effort. In

order to elicit different levels of vocal effort, participants were cued using methodology described by McKenna et al. (2018b), which instructed participants to "increase your effort during your speech as if you are trying to push your air out," while maintaining comfortable speaking rate and volume. Mild effort was described as "mildly more effort than your regular speaking voice," moderate effort as "more effort than mild," and maximum effort as "as much effort as you can while still having a voice."

Following training, participants were seated in a sound-attenuated booth and instrumented with recording equipment, including: a microphone, neck-surface accelerometer, and flexible endoscope. A directional headset microphone (Shure SM35 XLR) was placed 45° from the midline and 7 cm from the lips. A neck-surface accelerometer (BU series 21771 from Knowles Electronic, Itasca, IL) was placed on the anterior neck, superior to the thyroid notch and inferior to the cricoid cartilage using double-sided adhesive. For this study, a directional microphone was selected to reduce the impacts of noise emitted by the light source from the flexible endoscopic equipment. Microphone and accelerometer signals were pre-amplified (Xenyx Behringer 802 Preamplifier) and digitized at 30 kHz (National Instruments 6312 USB).

A flexible routine endoscope (Pentax, Model FNL-10RP3, 3.5-mm) was then passed transnasally over the soft palate and into the hypopharynx for laryngeal visualization. In cases in which participant anatomy or comfort interfered with image acquisition using the routine endoscope, a flexible slim endoscope (Pentax, Model FNL-7RP3, 2.4-mm) was used. A numbing agent was not administered so as to not affect laryngeal function (Dworkin et al., 2000b), but a nasal decongestant was offered to

minimize participant discomfort as the endoscope was passed through the nasal cavity. To record images of the larynx, the endoscope was attached to a camera (FASTCAM Mini AX100l; Model 540K-C-16GB; 256 × 256 pixels) with a 40-mm optical lens adapter. A steady xenon light was used for imaging (300 W KayPENTAX Model 7162B). Video images were acquired at a frame rate of 1 kHz using Photron Fastcam Viewer software (v.3.6.6) in order to track the fundamental frequency of vibration of the vocal folds, which is estimated to be 85–255 Hz during modal phonation in adults (Baken et al., 2000, p. 156), as well as the gross abductory and adductory gestures, which occur within 104–227 ms (Dailey et al., 2005). Recording was trigged using a custom MATLAB (version 9.3; The MathWorks, Natick, MA) algorithm that automatically time-aligned the video images with the microphone and accelerometer signals.

During the endoscopy procedure, participants were instructed to produce the eight ifi/ repetitions for each recording. Conditions were cued in the following order: slow rate, regular rate, fast rate, mild effort, moderate effort, maximum effort. Participants completed a minimum of two recordings per condition; however, recordings were repeated in the event that the vocal folds were not adequately captured (e.g., obstruction by the epiglottis). To further minimize participant discomfort during the procedure, the length of the endoscopic examination was approximately 5–10 minutes. The total experimental time (including consent, training, setup, and recording) required approximately 1–2 hours.

## *Data Analysis*

### High-speed Video Processing

**Reliability Training**

Individual VCV production usability and HSV data processing was performed by nine trained technicians. Prior to processing experimental data, the technicians underwent a training scheme described by McKenna et al. (2018a). This first included glottic angle identification training on flexible laryngoscopic images at a conventional frame rate of 30 fps, recorded using a halogen light source. The identified glottic angles (extending from the anterior commissure along the medial vocal fold edge to the vocal process) were compared to angle markings made previously by a gold-standard technician, and were required to meet two-way mixed-effects intraclass correlation coefficients (ICC) for consistency of agreement ≥ .80 (Diaz-Cadiz, McKenna, Vojtech, & Stepp, 2019). The resulting average reliability for the nine technicians was *ICC(3,1)* = .89 (*SD* = .01, *range* = .88–.91).

Technicians then completed training to use a semi-automated glottic angle tracking algorithm. This algorithm was developed in MATLAB to track the glottic angle over time within VCV productions and is described in detail in Diaz-Cadiz et al. (2019). In brief, the algorithm first takes microphone and accelerometer signals as inputs to an event detector that identifies VCV productions from the recordings. The technician is then prompted to choose a VCV production for examination; then, the algorithm carries out an automated glottic angle extraction process to identify the glottis, segment vocal fold edges, and estimate the glottic angle over time. The result of this three-step process

is a glottic angle waveform for the VCV production, which is shown in a graphical user interface (GUI) alongside time-aligned high-speed video frames, microphone and accelerometer signals, and glottal angular velocity traces. If the technician does not agree with the results of the automated algorithm, they may manually mark glottic angles for the VCV production at a downsampled rate of 50 Hz. The automated glottic angle extraction procedure then runs again, this time using the manual glottic angle data as a reference. Within the glottic angle tracking training, technicians were required to meet reliability standards of $ICC(3,1) \geq .80$ compared to a gold-standard technician, described in Diaz-Cadiz et al. (2019). The resulting average reliability of the nine technicians was $ICC(3,1) = .85$ ($SD = .04$, *range* = .80–.91). Following the training, the technicians analyzed VCV productions of the experimental data.

**VCV Usability**

The first step of experimental data processing required technicians to determine whether each /ifi/ production was "usable" based on manual inspection of the laryngoscopic recordings. For instance, if the glottis was obstructed (e.g., by the epiglottis) or if video quality was too poor to resolve the glottis, then the VCV production was considered unusable since the glottic angle could not be estimated. Such productions were rejected from further analysis. Due to the recording limitations of the high-speed imaging, the synchronized microphone, accelerometer, and HSV recordings were restricted in duration to 7.940 seconds when the 3.5-mm endoscope was used and 8.734 seconds when the 2.4-mm endoscope was used. Because of these pre-defined parameters in the current study, it was possible that /ifi/ productions at the end of the recording were

incompletely captured; these VCV productions were also considered unusable and excluded from further analysis. Finally, manual intervention was implemented if algorithmic estimates of the glottic angle waveform was deemed inappropriate by the technician; if errors still persisted following manual intervention, the technicians were instructed to mark the instance as unusable.

**Experimental Data Processing**

Nine technicians used the semi-automated algorithm to calculate the glottic angle waveform for each /ifi/ production ($N = 10776$). Within this analysis, a single technician determined whether the /ifi/ production was usable and, if so, obtained a quantitative estimate of the glottic angle for the production. The technicians accepted the automated results in 55.7% of cases (6005 of 10776), whereas the technicians accepted the automated results only after performing manual glottic angle estimation in 15.9% of cases (1717 of 10776). The technicians discarded the remaining 29.3% of productions that were unusable (10.9% of cases, 1178 of 10776) or could not be determined by the algorithm either before or after manual-assisted glottic angle estimation (17.4% of cases, 1876 of 10776). This analysis resulted in 7709 usable VCV productions for further processing. Algorithmic reliability was not assessed in the current study since prior work indicates that the algorithm yields good reliability ($ICC \geq .8$) compared to manual glottic angle estimates (Diaz-Cadiz et al., 2019); however, the initial data processing was then rechecked by a second technician.

Since the goal of the current analysis was to examine RFF in reference to the physiological termination or initiation of vocal fold vibration (rather than estimated

transition points using the acoustic signal), a series of kinematic time points were then extracted from each /ifi/ production to mark these transitions. Technicians were presented with a MATLAB GUI (see **Fig. 3.2**) showing time-aligned high-speed video frames, the microphone signal, the previously extracted glottic angle waveform, and a quick vibratory profile (QVP). The QVP is a one-dimensional waveform that captures the vibration of the vocal folds, in addition to non-glottal activities such as camera or epiglottic motion (Ikuma et al., 2013). The QVP was included in this analysis as an alternative to the glottal area waveform due to its sensitivity to HSV imagery and superior ability to track the harmonic motion of the vibrating vocal folds. Specifically, the QVP is sensitive to changes in light intensity of the image—as opposed to only being sensitive to vocal fold vibrations like the glottic angle waveform—such that the QVP waveform is ideal for identifying (i) the vibrating glottis in images of poor resolution, and (ii) the time window containing the transition between voiced and unvoiced segments (Ikuma et al., 2013).

In the current analysis, the QVP was calculated with methodology and suggested parameters as described in Ikuma et al. (2013). In brief, the HSV frame was first centered over the glottis using methodology from the semi-automated glottic angle extraction algorithm (Diaz-Cadiz et al., 2019). Vertical and horizontal profiles of the HSV frames were then calculated using an observation duration of 0.02 seconds in order to sufficiently capture a minimum $f_o$ of 50 Hz. From here, the two HSV frame profiles were summated to produce the QVP. The resulting QVP profile was then high-pass filtered using a $7^{th}$ order Butterworth to attenuate low frequency noise below a cut-off frequency

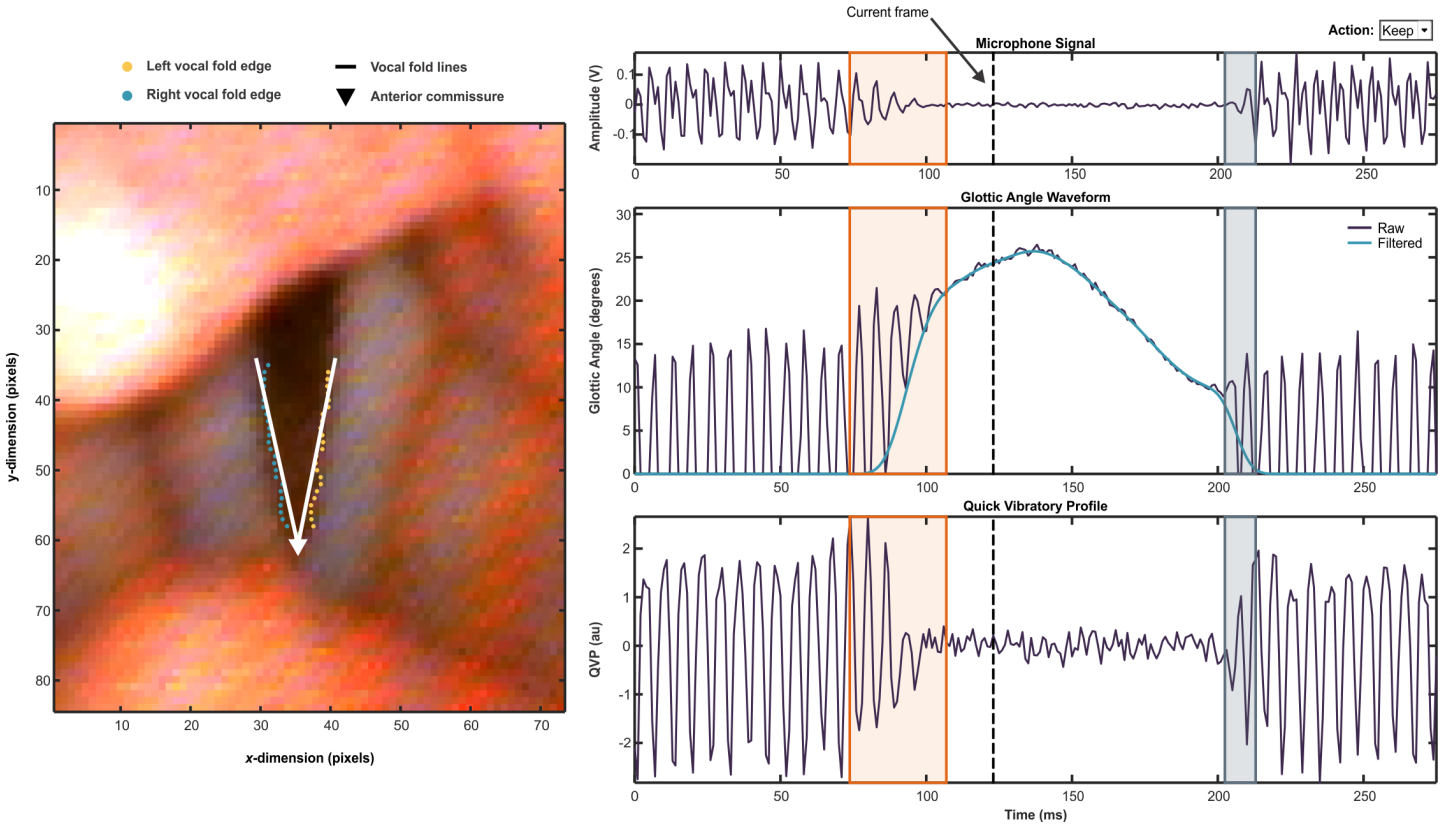


**Figure 3.2. Graphical user interface shown to technicians in order to extract kinematic time points.**
The left panel shows the current frame of the high-speed video, with left (yellow) and right (green) vocal folds as well as fitted vocal fold angle trajectory (white). The right panels show the raw microphone signal (top), raw and filtered glottic angle waveform (middle; see Diaz-Cadiz et al., 2019 for more details), and the quick vibratory profile (bottom). The location of the current frame is highlighted using a black dashed line across the right-hand plots. Technicians were instructed to move and resize the abductory (orange) and adductory (gray) boxes to align with the pertinent kinematic time points. In this example, the left border of the abductory box corresponds to the start of abduction and the right border of the abductory box corresponds to the time of voicing offset. Similarly, the left border of the adductory box represents the time of voicing onset and the right border of the adductory box corresponds to the termination of adduction. If the technician was not satisfied with the glottic angle waveform and/or quick vibratory profile, the technician could set the production to be re-processed by selecting "Redo" in the "Action" dropdown bar (upper right); otherwise, the production was set to "Keep" (as shown).

With the MATLAB GUI described above, a total of three technicians used the time-aligned microphone signal, glottic angle waveform, and QVP to identify four kinematic timing metrics using methodology described in Park et al. (Under Review):

- Start of abduction ($t_{abd}$): Last full or maximum contact of the vocal folds during voicing offset
- Time of voicing offset ($t_{off}$): Termination of the last vibratory cycle before the voiceless consonant
- Time of voicing onset ($t_{on}$): Initiation of the first vibratory cycle after the voiceless consonant
- Termination of adduction ($t_{add}$): First full or maximum contact of the vocal folds during voicing onset

In the event that the arytenoid cartilages obstructed the view of the vocal folds during voicing offset (e.g., due to supraglottic constriction), $t_{abd}$ was considered as the time in which the arytenoid cartilages began to move away from one another; similarly, if the arytenoid cartilages blocked the vocal folds during voicing onset, $t_{add}$ was considered as the time at which the arytenoid cartilages stopped moving toward one another. In the event that the vocal folds exhibited an abrupt closure at the start of voicing onset (i.e., prior to vocal fold vibration), $t_{on}$ was extracted as the time point immediately before the point of abrupt vocal fold closure.

Technicians were instructed to use the glottic angle waveform and QVP to identify these four time points, then corroborate the selected indices via manual visualization of the raw HSV images. This process was carried out to minimize errors

that may occur in the event that the glottic angle waveform failed to capture small glottal gaps during vibratory cycle phases or if the QVP was confounded by lighting artifacts (e.g., intensity saturation due to the epiglottis coming into view). Of note, the microphone signal was included within the GUI in the event that the glottic angle waveform and QVP both failed to properly track the vibrations of the vocal folds. In such instances, the technicians were able to select "Redo" from a drop-down menu (labeled "Action" in the right-hand corner of **Fig. 3.2**) to indicate that the production needed to be rejected or reprocessed using methodology from *VCV Usability*. Productions from which the technician successfully identified the four time metrics were marked as "Keep."

The technicians each reanalyzed 10% of participants in a separate sitting to ensure adequate intrarater reliability. The three technicians also analyzed the HSV images of the same participant to assess interrater reliability. Intrarater reliability was assessed via two-way mixed-effects ICCs for absolute agreement, whereas interrater reliability was computed using two-way mixed-effects ICCs for consistency of agreement (single measures). The reliability of each technician in extracting the four kinematic time points is shown in **Table 3.3**. Intrarater reliability ranged from moderate to excellent (.70–.99), with an overall mean reliability of .98 (95% CI = .97–1.0). Average interrater reliability

**Table 3.3. Reliability of kinematic time point extraction for three trained technicians.**

| Measure | Intrarater Reliability | | | Interrater Reliability |
|---|---|---|---|---|
| | Technician 1 | Technician 2 | Technician 3 | Mean (95% CI) |
| $t_{abd}$ | .70 | .98 | .99 | .75 (.67–.82) |
| $t_{off}$ | .86 | .99 | .99 | .91 (.86–.96) |
| $t_{on}$ | .95 | .99 | .99 | .97 (.96–.99) |
| $t_{add}$ | .95 | .99 | .99 | .97 (.96–.99) |

*Note.* $t_{abd}$ *= start of abduction,* $t_{off}$ *= time of voicing offset,* $t_{on}$ *= time of voicing onset,* $t_{add}$ *= termination of adduction.*

for the four kinematic time measure ranged from good to excellent (.75–.97), producing an overall reliability of .90 (95% CI = .86 –.94).

An error analysis was then performed on the 7709 VCV productions to determine the resolution error in capturing cycle-to-cycle changes in $f_o$ at a sampling rate of 1 kHz. To do so, $t_{on}$ and $t_{off}$ were used to localize the initiation and termination of vocal fold vibration for voicing onset and offset, respectively. The 10 glottal pulses adjacent to $t_{on}$ and $t_{off}$ were identified from the QVP using a custom peak detector in MATLAB (version 9.3). Vocal cycle durations were then calculated for steady-state vocal cycles (offset cycle 1, onset cycle 10) and boundary cycles (offset cycle 10, onset cycle 1). Cycle periods were then compared to the sampling period (0.001 s) to quantify the proportion of the sampling period that is accounted for in a single vocal cycle. On average, the sampling period accounted for 17.0% (*SD* = 6.4%, *range* = 5.1–40.0%) of a single vocal cycle. For voicing offset, the sampling period constituted 17.4% (*SD* = 5.8%) of offset cycle 1 and 15.9% (*SD* = 6.7%) of offset cycle 10. For voicing onset, the sampling period made up 17.3% (*SD* = 5.1%) of onset cycle 1 and 17.2% (*SD* = 5.3%) of onset cycle 10. These results indicate that the sampling period used here was sufficient to identify cycle-to-cycle changes in $f_o$ without introducing aliasing. As such, the resolution of the QVP was deemed appropriate for identifying the approximate initiation and termination of vocal fold vibration for voicing onset and offset.

## Manual RFF Estimation

Using methodology described in detail in *Chapter 2*, two trained technicians carried out manual RFF estimation on each participant (7709 total VCV productions).

Due to the availability of technicians to perform manual RFF techniques, five trained technicians, who met interrater reliability criterion ≥ .93[3], were assigned to manually estimate RFF throughout the course of data collection. **Table 3.4** shows the number of speakers that each of the five technicians rated. Mean RFF values were computed across technicians to use as the gold-standard for RFF estimates. Technicians used the manual RFF rejection criteria detailed in Vojtech and Heller Murray (2019a) to determine whether an offset and/or onset instance should be rejected. Examples of such criteria include glottalization, misarticulation, or voicing of the /f/.

**Table 3.4. Number of speakers for which each of five trained technicians manually computed relative fundamental frequency.**

| Technician | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | **38** | | | | |
| 2 | 5 | **82** | | | |
| 3 | 19 | 53 | **80** | | |
| 4 | 14 | 13 | 2 | **29** | |
| 5 | 0 | 11 | 6 | 0 | **17** |

*Note. The matrix shows common speakers analyzed between technicians, whereas the diagonal (bolded) describes the number of speakers a single technician rated in total.*

Intrarater reliability was assessed via Pearson correlation coefficients within each technician when instructed to reanalyze 20% of participants in a separate sitting, whereas interrater reliability was computed via two-way mixed-effects ICCs for consistency of agreement. The average intrarater reliability was calculated as $r$ = .90 (*SD* = .05, *range* = .84–.97), and the average interrater reliability was computed as *ICC(3,1)* = .93 (*SD* = .04, *range* = .87–.98). Rater reliability was also examined by assessing the difference between selected boundary cycles (i.e., voicing offset cycle 10, voicing onset cycle 1) of original

[3] The dataset used to train individuals in manual relative fundamental frequency estimation is a separate dataset from that described here and may be downloaded from https://sites.bu.edu/stepplab/research/rff/ (Last viewed May 30, 2019).

and reanalyzed samples. Errors in boundary cycle selection were quantified as the magnitude of the average number of vocal cycles between original and reanalyzed samples. The mean intrarater error was 0.64 vocal cycles (*SD* = 0.44 cycles), with errors in boundary cycle selection ranging from 0 to 5 vocal cycles. The mean interrater error was 0.71 vocal cycles (*SD* = 0.41 cycles), with errors in boundary cycle selection across technicians ranging from 0 to 6 vocal cycles.

## Semi-automated RFF Estimation

Semi-automated RFF estimation was first performed on all 7709 VCV productions using the aRFF-AP algorithm in MATLAB (version 9.3). The relationship between acoustic features and the physiological vibrations of the vocal folds was then examined. First, a literature review was conducted to select a set of acoustic features that showed promise in distinguishing voiced segments from voiceless segments, as is the goal of the acoustic features in the semi-automated RFF algorithm. From here, the acoustic feature set was reduced to reflect the features that best corresponded with the termination ($t_{off}$) or initiation ($t_{on}$) of voicing. The features were then implemented in the aRFF-AP algorithm (now "aRFF-APH") to enhance the physiological relevance of RFF.

### Acoustic Feature Selection

In the aRFF and aRFF-AP algorithms, acoustic feature trends are examined to identify a state transition in feature values that marks the boundary cycle—that is, the vocal cycle that marks the transition between voiced and voiceless speech segments (also called voiced/unvoiced detection). The boundary cycle is offset cycle 10 for voicing offset and onset cycle 1 for voicing onset (see **Fig. 3.1**). In both algorithm versions,

normalized peak-to-peak amplitude, number of zero crossings, and waveform shape similarity are used in this process; however, it is not clear whether these acoustic features are the best choice for performing voiced/unvoiced detection since they were selected to increase correspondence with manual RFF. As both manual and semi-automated RFF are computed using the acoustic signal, there may be other features that better correspond to the true vibrations of the vocal folds (which are obtained in the current study using HSV). As such, additional features were identified that could be used in voiced/unvoiced detection. This resulted in the inclusion of 15 additional acoustic features to investigate in regard to classifying voiced and unvoiced speech segments: (1) autocorrelation; (2,3) simple and normalized cross-correlation; (4,5) average and median pitch strength; (6,7,8) average, median, and standard deviation of voice $f_o$; (9,10) mean and standard deviation of cepstral peak prominence; (11) high-to-low ratio of spectral energy; (12) short-time energy; (13) short-time log energy; (14) short-time magnitude; and (15) signal-to-noise ratio. **Table 3.5** provides a description of each of acoustic features, along with the proposed hypotheses in feature values when used for voiced/unvoiced detection.

**Table 3.5. Acoustic measures for classifying voiced and unvoiced speech segments, with abbreviations (Abbr). Rows that are shaded yellow indicate that the acoustic feature was included in the aRFF and aRFF-AP algorithms.**

| Feature Name | Abbr. | Definition and Rationale |
|---|---|---|
| Autocorrelation | ACO | ACO is a comparison of a segment of a voice signal to a delayed copy of itself as a function of the delay, and is often used in $f_o$ estimation and voiced/unvoiced classification (Camacho, 2007; Jalil, Butt, & Malik, 2013; Nandhini & Shenbagavalli, 2014). As more periodic signals elicit higher ACO values, it was expected that ACO values of the vocal cycles during the vowel /i/ would be greater than those calculated from voiceless consonant /f/. |

| | | |
|---|---|---|
| Mean Cepstral Peak Prominence | CPP | CCP is a correlate to overall severity of dysphonia (Heman-Ackah et al., 2014) and reflects the distribution of energy at harmonically related frequencies (Hillenbrand et al., 1996). It is calculated as the magnitude of the peak with the highest amplitude in the cepstrum (i.e., the Fourier transform of the power spectrum, representing the spectral representation of the spectrum). Using CPP, quasiperiodic vocal cycles during the vowel /i/ may be differentiated from aspiration noise of the /f/. As higher CPP values are associated with more periodic signals, CPP was hypothesized to be greater in /i/ than /f/. |
| Average Pitch Strength | APS | APS is a correlate to overall severity of dysphonia (Kopf et al., 2017; Shrivastav et al., 2012) and has been implemented to discriminate voice signal types (Anand, Kopf, Shrivastav, & Eddins, 2019b). Using Auditory-SWIPE′, pitch strength is calculated by correlating a voice signal with a sawtooth waveform constructed across a range of possible $f_o$ values; the $f_o$ value that elicits the greatest correlation is considered the $f_o$ of the signal, and the degree of this correlation is the pitch strength. APS is then calculated as the average pitch strength of the window. Because vocal cycles are characterized using $f_o$, it is expected that the correlation between the sawtooth waveform and voice signal would be greater in voiced segments than unvoiced segments. As such, APS was hypothesized to be greater in /i/ than /f/. |
| Average Voice $f_o$ | $Af_o$ | A$f_o$ is the acoustic correlate of vocal pitch, and is calculated in the current study using the Auditory-SWIPE′ algorithm (described above in APS). It was expected that A$f_o$ would exhibit similar trends to resulting RFF measures, wherein A$f_o$ would remain stable or decrease during voicing offset, then increase during voicing onset. |
| Cross-correlation | XCO | XCO is a comparison of a segment of a voice signal with a different segment of the signal. As with the ACO, XCO is often implemented in $f_o$ estimation and voiced/unvoiced classification (Camacho, 2007; Ghaemmaghami, Baker, Vogt, & Sridharan, 2010; Samad, Hussain, & Fah, 2000). It was expected that quasiperiodic voiced cycles of the vowel /i/ would elicit higher XCO values compared to the aspiration and frication noise of the consonant /f/. |
| Low-to-high ratio of spectral energy | LHR | LHR is a measure of spectral tilt, and is calculated by comparing spectral energy above and below a specified frequency. Using a cut-off frequency of 4 kHz (Hillenbrand et al., 1994; Hillenbrand et al., 1996), the LHR may be able to distinguish harmonic energy due to the vowel, /i/, from high-frequency aspiration and frication noise (above 2–3 kHz) that may occur when producing the voiceless consonant, /f/. As such, it was hypothesized that larger LHR values will occur in the /i/ than the /f/ of /ifi/ productions. |
| Median Pitch Strength | MPS | MPS was included in the current study as an alternative to APS. Similarly, it was expected that MPS would be greater in voiced |

| | | |
|---|---|---|
| | | segments corresponding to the vowel /i/ than unvoiced segments of the consonant /f/. |
| Median Voice $f_o$ | $Mf_o$ | M$f_o$ was incorporated as an alternative to A$f_o$. M$f_o$ was hypothesized to be different between /i/ and /f/ segments, with values corresponding to /f/ segments exhibiting lower, more variable values due to errors in $f_o$ estimation within unvoiced segments. |
| Normalized Cross-correlation | NXCO | A variant of the simple XCO, NXCO also compares a segment of a voice signal with a different segment of the signal. NXCO is often considered more robust than XCO in voiced/unvoiced classification since the amplitude of the compared windows are normalized, thereby removing differences in signal amplitude as a factor. It was expected in the current study that voiced cycles of the vowel /i/ would elicit higher NXCO values compared to windows of the voiceless consonant /f/ due to increased periodicity. |
| Normalized Peak-to-peak Amplitude | PTP | PTP is computed as the range of the amplitude of a windowed voice signal. Because vowels generally exhibit higher amplitudes than consonants, it was postulated that the amplitude of the vowel (/i/ in the current study) would be greater than that of /f/, leading to higher PTP values. |
| Number of Zero Crossings | NZC | NZC refers to the number of sign changes of the windowed voice signal. It was expected that NZC would be greater in the voiceless consonant compared to the vowel due to stochastic aspiration and frication noise in producing /f/. |
| Short-time Energy | STE | STE is the energy of a short voice segment, wherein high energy would result from a voiced signal segment and lower energy would correspond to a unvoiced signal segment (Dong, Liu, Zhou, & Cai, 2002; Jalil et al., 2013; Swee, Salleh, & Jamaludin, 2010). In the context of the current study, it was expected that STE values of the vowel /i/ would be substantially greater than those of the voiceless consonant, /f/. |
| Short-time Log Energy | SLE | SLE is a common parameter in automated speech recognition systems and is calculated as the logarithm of the energy of a short voice segment. It is used in the current study as an alternative to STE. As with STE, it was expected that /i/ segments would elicit higher SLE values than /f/ segments due to greater signal energy occurring within the vowel than voiceless consonant. |
| Short-time Magnitude | STM | STM is the magnitude of a short voice segment, wherein high magnitudes would refer to a voiced signal segment and lower magnitudes would refer to an unvoiced signal segment (Dong et al., 2002; Jalil et al., 2013; Swee et al., 2010). STM was expected to be greater in windows of time pertaining to the vowel /i/ than of the voiceless consonant, /f/. |

| | | |
|---|---|---|
| Signal-to-noise Ratio | SNR | SNR is an estimate of the power of a signal compared to that of a segment of noise. In the current study, SNR was postulated to be greater in windows containing vocal cycles of the vowel, /i/, compared to windows containing aspiration and frication noise of the voiceless consonant, /f/. |
| Standard Deviation of Cepstral Peak Prominence | SD CPP | Calculated as the standard deviation of CPP values within a window, it was expected that SD CPP would be greater in /f/ segments than /i/ segments due to variations in signal periodicity as a result of aspiration and frication noise. |
| Standard Deviation of Voice $f_o$ | SD $f_o$ | Calculated as the standard deviation of $f_o$ values within a window, it was expected that SD $f_o$ would be greater in /f/ segments than /i/ segments due to errors in $f_o$ estimation (as the unvoiced segments would not have a valid $f_o$ value). |
| Waveform Shape Similarity | WSS | WSS is computed as the normalized sum of square error between the current window of time and the previous window of time. It is calculated in reference to a window of time in the voiceless consonant, such that another window in the voiceless consonant would elicit higher WSS values than would a window in the vowel. |

Thirteen of the 18 features were calculated directly from the microphone signal: ACO, CPP, XCO, LHR, NXCO, NZC, PTP, STE, SLE, STM, SNR, SD CPP, and WSS. The remaining five features were calculated using a processed version of the microphone signal. Specifically, Auditory-SWIPE′ (Camacho, 2012; Camacho et al., 2008)—the $f_o$ estimation method used in the aRFF-AP algorithm—was used to extract the $f_o$ contour and pitch strength contour from the microphone signal of each /ifi/ production. Three features were calculated from the $f_o$ contour (A$f_o$, M$f_o$, SD $f_o$), and two features were computed using the pitch strength contour (APS, MPS).

In addition to examining the 13 acoustic features extracted from the raw microphone signal, filtered versions of these features were also considered. The aRFF and aRFF-AP algorithms employ a version of the microphone signal when band-pass filtered ±3 ST around the average $f_o$ of the speaker to identify peaks and troughs in signal

amplitude. The aRFF-AP algorithm also used this filtered version of the signal to compute PTP (whereas NZC and WSS were calculated using the raw microphone signal). With filtering, a total of 31 acoustic features were considered for further analysis.

**Feature Set Reduction**

The acoustic feature set was examined to (i) remove features that did not appropriately capture the transition between voiced and unvoiced segments, and (ii) reduce multicollinearity amongst the selected features. To do so, the discrepancy in boundary cycle identification (i.e., voicing offset cycle 10, voicing onset cycle 1) was first quantified between HSV-derived voicing transitions and acoustic features operating on the microphone signal.

Acoustic features were assessed by simulating a sliding window process to calculate features across time, ranging from the midpoint of the voiceless consonant and into the vowel. The sliding window was positioned using methodology described in Vojtech et al. (2019; see *Chapter 2*) such that features were computed as a function of the number of pitch periods[4] away from the "true" boundary cycle. In this previous work, the true boundary cycle was in reference to vocal cycles that were identified through manual RFF estimation. Here, however, the true boundary cycle was set to reference the time of voicing offset (i.e., $t_{off}$) and the time of voicing onset (i.e., $t_{on}$) to investigate the relationship between these acoustic features and the physiologically derived termination and initiation of vocal fold vibration, respectively. Acoustic features were analyzed as a

[4] "Pitch period" refers to the duration of one glottal cycle, and was computed per /ifi/ production using the average $f_o$ determined using Auditory-SWIPE′.

function of ±10 pitch periods from the true boundary cycle to comprehensively examine trends in feature values. In this way, each /ifi/ production resulted in 21 feature values (i.e., one feature value for each pitch period) for each of the 31 acoustic features. The feature values were then visually inspected to determine which acoustic features failed to exhibit a substantial change in feature magnitude during the transition between the voiceless consonant and vowel; such features were removed from subsequent analysis.

Useful acoustic features were then input into a stepwise binary logistic regression to determine the probability of feature values corresponding to a voiced or unvoiced segment. In this model, acoustic feature values were input as continuous predictors when calculated -10 to +10 average pitch periods away from the true boundary cycle. This resulted in 21 feature values for each /ifi/ production for each acoustic feature. The response variable corresponded to whether the segment analyzed was voiced (1) or unvoiced (0). For voicing offset, -10 to 0 pitch periods from the true boundary cycle were considered voiced, whereas 1 to 10 pitch periods from the true boundary cycle were considered unvoiced. For voicing onset, -10 to -1 were considered unvoiced and 0 to 10 were considered voiced. Importantly, the data values for each feature were assumed independent in the regression model to identify which features were significantly related to voicing status rather than to create a regression equation for predicting voicing status. Variable significance was set to $p < .05$. Highly correlated features (variable inflation factor > 10) were removed from the model to reduce multicollinearity. Acoustic features that exhibited significant predictive effects and were sufficiently independent were retained for further algorithmic refinement.

**Algorithmic Modifications**

After identifying an acoustic feature set, the features were implemented into the semi-automated RFF algorithms. In order to do this, the methodology for refining these algorithms from *Chapter 2* was adapted to the current study. First, the pitch strength rejection criterion set for the aRFF-AP algorithms was carried over to the current study, such that VCV productions with an average pitch strength < .05 were rejected from further analysis. The remaining productions were then examined to identify potential vocal cycles. As described in detail in *Chapter 2*, a sliding window based on the speaker's average $f_o$ advanced from the voiceless consonant and into the vowel of interest (either to assess voicing offset or voicing onset); within each window of time, the set of acoustic features were calculated. In the current study, the selected acoustic features were computed rather than the acoustic feature set used in the aRFF and aRFF-AP algorithms (i.e., PTP, NZC, WSS). Rule-based signal processing techniques were then implemented to identify the boundary cycle that separates the vowel from the voiceless consonant. To locate this cycle, the algorithm identified a feature value that maximized the effect size between left and right components of each acoustic feature vector (i.e., such that the vector could be split into a "voiced" segment and "voiceless" segment). The cycle index that corresponded to this identified feature value was selected as the boundary cycle candidate for that feature. The median of these candidates was then calculated as the boundary cycle.

Distinct from the rule-based signal processing techniques of the aRFF-AP algorithms, pitch strength categories were not implemented in the present study. As such,

the boundary cycle identification methods used in aRFF-APH algorithms more closely mirrored that of the original aRFF algorithms (see Lien, 2015 for more details). To summarize, the aRFF-APH algorithms comprised the (a) physiologically relevant acoustic features and (b) pitch strength rejection criterion, such that:

1. If the average pitch strength of the VCV production did not meet (b), the production was rejected.

2. For both voicing offset and onset, boundary cycle candidates were identified for (a) using effect size methodology implemented in the aRFF and aRFF-AP algorithms.

3. The median of the boundary cycle candidates was calculated as the predicted boundary cycle for the offset or onset instance.

*Algorithmic Performance*

To assess the effectiveness of introducing physiologically relevant acoustic features into the semi-automated RFF algorithms, the capacity of the aRFF-APH algorithms to locate the true boundary cycle (referenced to $t_{off}$ for voicing offset and $t_{on}$ for voicing onset) was compared against that of manual estimation and the aRFF-AP algorithms. The 7709 VCV productions from the 122-speaker dataset were first examined using each RFF estimation method. The accuracy of each method in identifying the true boundary cycle was quantified as the distance between the true boundary cycle (relative to $t_{off}$ for voicing offset and $t_{on}$ for voicing onset) and the estimated boundary cycle for each voicing offset and onset instance when using each RFF estimation method. This distance was measured in average pitch periods from the true boundary cycle, as described in *Acoustic Feature Selection*. The distance between true and estimated

boundary cycles was then compared across RFF estimation methods to determine which method corresponded most closely to the vibratory characteristics of the vocal folds.

### *Statistical Analysis*

To determine whether there was a relationship between RFF estimation method (manual, aRFF-AP, aRFF-APH) and resulting boundary cycle classification accuracy, two chi-square tests of independence were performed (one for voicing offset and onset for voicing onset). In each analysis, a contingency table was developed to describe the frequency of correctly classified boundary cycles—wherein the distance between true and estimated boundary cycles was zero—versus misclassified boundary cycles (i.e., non-zero distance between true and estimated boundary cycles). Significance was set to *a priori* to $p < .05$. Cramer's V was used to assess effect sizes of significant associations. *Post hoc* chi-square tests of independence were then performed for pairwise comparisons of the three RFF estimation methods using a Bonferroni-adjusted $p$ value of (.05/3 =) .017.

## ***Results***

### *Acoustic Feature Selection*

**Fig. 3.3** shows the relationship between acoustic features and the true boundary cycle (relative to $t_{off}$) for 7709 voicing offset instances. Similarly, **Fig. 3.4** shows this relationship (relative to $t_{on}$) for 7709 voicing onset instances. Acoustic features calculated directly from the microphone signal are shown when calculated from the raw signal as well as from the band-pass filtered signal. It may be observed that acoustic features that depend on signal energy (PTP, LHR, STE, STM, SLE, SNR) are greatest when calculated

within a window of time pertaining to a vowel rather than the voiceless consonant. Similar trends are exhibited by the majority of features that depend on signal periodicity (CPP, XCO, NXCO, ACO) and $f_o$ (APS, MPS, A$f_o$, M$f_o$). The opposite is true for NZC

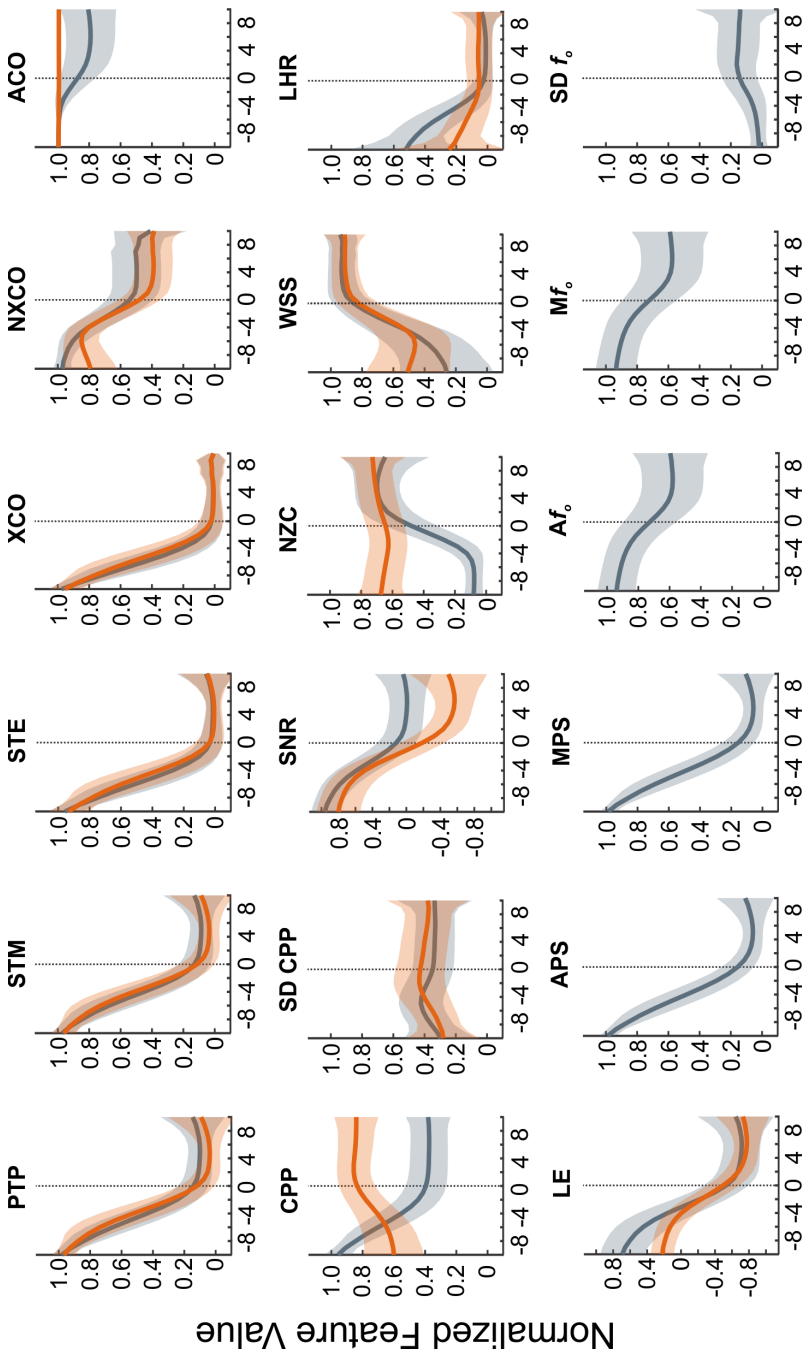


**Figure 3.3. Normalized feature values (blue) with respect to distance (pitch periods) from the true boundary cycle for voicing offset.**
Features calculated from band-pass filtered microphone signal are overlaid in green (when applicable). Top row: normalized peak-to-peak amplitude (PTP), short-time magnitude (STM), short-time energy (STE), cross-correlation (XCO), normalized cross-correlation (NXCO), autocorrelation (ACO). Middle row: mean and standard deviation of cepstral peak prominence (CPP, SD CPP), signal-to-noise ratio (SNR), number of zero crossings (NZC), waveform shape similarity (WSS), low-to-high ratio of spectral energy (LHR). Bottom row: log energy (LE), average and median pitch strength (APS, MPS), average, median, and standard deviation of $f_o$ (A$f_o$, M$f_o$, SD $f_o$). Solid lines indicate mean values and shaded regions indicate standard deviation.

(when calculated from the raw microphone signal) and WSS, which are both greatest when calculated within a window of time pertaining to the voiceless consonant. The

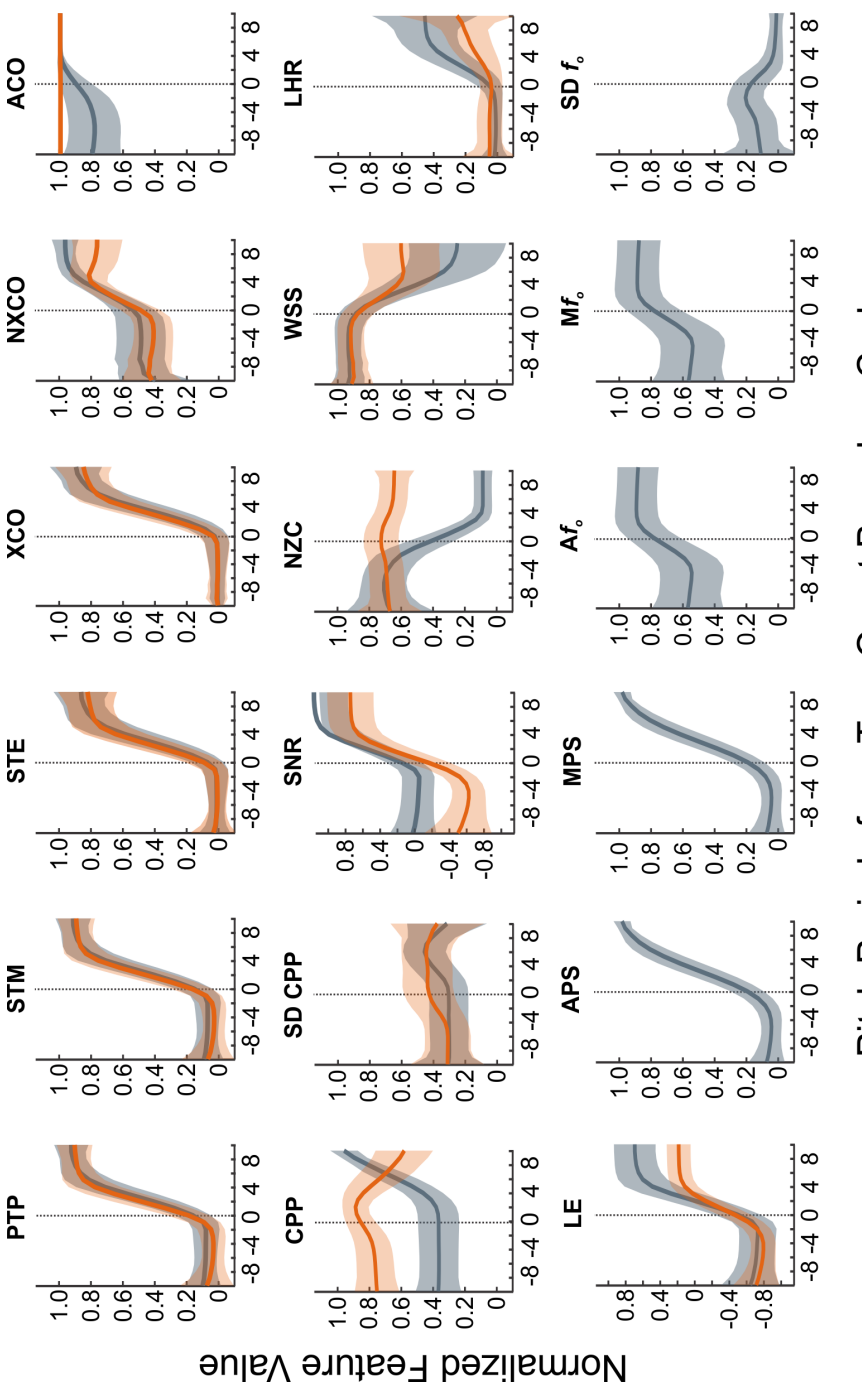


**Figure 3.4. Normalized feature values (blue) with respect to distance (pitch periods) from the true boundary cycle for voicing onset.**
Features calculated from band-pass filtered microphone signal are overlaid in green (when applicable). Top row: normalized peak-to-peak amplitude (PTP), short-time magnitude (STM), short-time energy (STE), cross-correlation (XCO), normalized cross-correlation (NXCO), autocorrelation (ACO). Middle row: mean and standard deviation of cepstral peak prominence (CPP, SD CPP), signal-to-noise ratio (SNR), number of zero crossings (NZC), waveform shape similarity (WSS), low-to-high ratio of spectral energy (LHR). Bottom row: log energy (LE), average and median pitch strength (APS, MPS), average, median, and standard deviation of $f_o$ (A$f_o$, M$f_o$, SD $f_o$). Solid lines indicate mean values and shaded regions indicate standard deviation.

features that examine the standard deviation of values (SD $f_o$ and SD CPP) exhibit a similar yet less pronounced trend to that of NZC and WSS.

Interestingly, the majority of raw and filtered counterparts showed similar trends for both voicing offset and onset. This included PTP, STM, STE, XCO, NXCO, WSS, LHR, SD CPP, SNR, and LE. For NZC, CPP, and ACO, however, raw and filtered signal counterparts exhibited discrepant trends. In particular, NZC was low in the vowel and increased toward the voiceless consonant (positive pitch period distance in **Fig. 3.3** and negative pitch period distance in **Fig. 3.4**) when calculated using the raw microphone signal, but was relatively constant when estimated using the filtered signal. ACO exhibited a similar trend in stationarity when calculated from the band-pass filtered signal. CPP, on the other hand, demonstrated higher values nearest the vowel when calculated using the raw microphone signal, but the opposite was true when using the band-pass filtered signal.

Manual inspection of these 31 features resulted in the removal of the filtered NZC, raw and filtered ACO, filtered CPP, filtered LHR, raw and filtered SD CPP, and SD $f_o$ due to a lack of discrimination between voiced and unvoiced segments. All further analyses were completed using the remaining 23 features.

*Stepwise Binary Logistic Regression*

**Table 3.6** shows that filtered WSS, M$f_o$, CPP, NZC, STE, APS, NXCO, and XCO were all significant predictors of voicing status for voicing offset ($p < .05$). When using these eight features, the model for voicing offset accounted for 61.8% of the variance in voicing status (adjusted $R^2 = 61.8\%$), with an area under the receiver operating

**Table 3.6. Summary of significant variables in the stepwise binary logistic regression statistical model.**

| Model | Acoustic Feature | Coef | SE Coef | *z* | *p* | Odds Ratio |
|---|---|---|---|---|---|---|
| Offset | Constant | 0.12 | 0.07 | 1.68 | .09 | — |
| | Filtered WSS | -1.51 | 0.05 | -30.04 | <.001 | 0.22 |
| | $Mf_o$ | 1.43 | 0.04 | 34.26 | <.001 | 4.19 |
| | CPP | 1.20 | 0.06 | 19.53 | <.001 | 3.34 |
| | NZC | -3.30 | 0.04 | -78.53 | <.001 | 0.04 |
| | STE | -5.69 | 0.15 | -38.04 | <.001 | 0.01 |
| | APS | 9.26 | 0.12 | 78.81 | <.001 | 10535.03 |
| | NXCO | -0.84 | 0.05 | -16.92 | <.001 | 0.43 |
| | XCO | 1.01 | 0.16 | 6.35 | <.001 | 2.74 |
| Onset | Constant | -2.18 | 0.10 | -22.46 | <.001 | — |
| | Filtered WSS | 1.43 | 0.08 | 18.66 | <.001 | 4.20 |
| | $Mf_o$ | 2.16 | 0.06 | 39.29 | <.001 | 8.65 |
| | CPP | 1.09 | 0.08 | 12.95 | <.001 | 2.98 |
| | NZC | -2.56 | 0.06 | -41.07 | <.001 | 0.08 |
| | APS | 8.93 | 0.15 | 59.13 | <.001 | 7527.75 |
| | SNR | 0.51 | 0.06 | 8.92 | <.001 | 1.67 |
| | Filtered STE | -3.67 | 0.10 | -36.61 | <.001 | 0.026 |
| | Filtered LE | 3.23 | 0.07 | 46.45 | <.001 | 25.33 |

*Note. "Filtered" refers to using the band-pass filtered version of the microphone signal to calculate the corresponding acoustic feature.*

characteristic (ROC) curve of .96. Inspection of the coefficients and corresponding odds ratios indicated that the log odds of voicing decreased per one unit increase in STE, NXCO, NZC, or filtered WSS (i.e., negative coefficient). On the other hand, the log odds of voicing increased per one unit increase in M$f_o$, CPP, APS, or XCO (i.e., positive coefficient). For voicing onset, the stepwise binary logistic regression revealed that filtered WSS, M$f_o$, CPP, NZC, APS, SNR, filtered STE, and filtered LE were all significant predictors of voicing status ($p < .05$; see **Table 3.6**). The model for voicing onset accounted for 76.0% of the variance in voicing status (adjusted $R^2 = 76.0\%$), with an area under the ROC curve of .98. The model for voicing onset indicated that the log odds of voicing decreased per one unit increase in NZC or filtered STE. The log odds of

voicing increased per unit increase in filtered WSS, M$f_o$, CPP, APS, SNR, or filtered LE. The resulting acoustic features were then incorporated into the aRFF-APH algorithms to identify the boundary cycle of voicing.

*Algorithmic Performance*

The comparison of aRFF-APH, aRFF-AP, and manual RFF estimation techniques in identifying the true boundary cycle is shown in **Fig. 3.5**. Out of 7709 offset instances (see **Fig. 3.5a**), the aRFF-APH algorithms resulted in the greatest number of correctly identified boundary cycles ($N$ = 5567, 72.2% of instances), followed by aRFF-AP ($N$ = 5016, 65.1% of instances) then manual ($N$ = 984, 12.8% of instances). For each RFF estimation method, the majority of offset

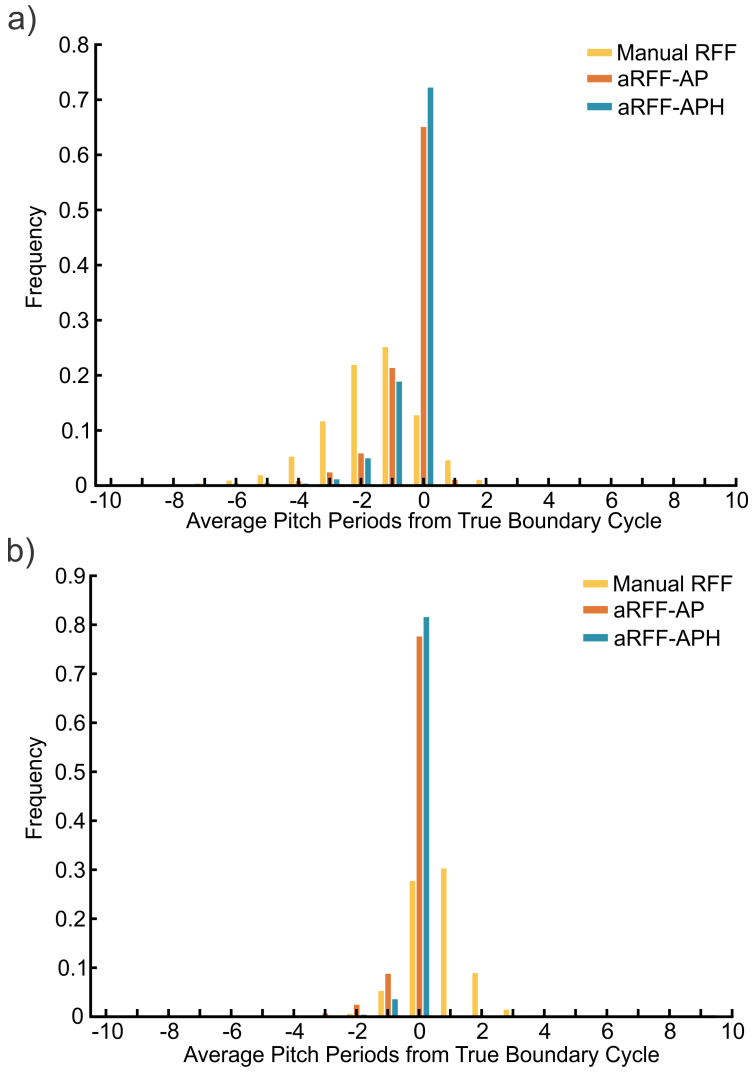


**Figure 3.5. Boundary cycle identification of each relative fundamental frequency estimation method (manual, aRFF-AP, aRFF-APH). For (a) voicing offset and (b) voicing onset. Results for manual RFF estimation are shown in yellow, for aRFF-AP are shown in orange, and for aRFF-APH are shown in blue.**

misclassifications occurred closer to the vowel, including 67.4% of total instances when using manual RFF, 25.7% when using aRFF-AP, and 31.6% when using aRFF-APH. The aRFF-APH algorithm rejected the least number of offset instances ($N$ = 154; 2%), followed by the aRFF-AP algorithm ($N$ = 160; 2.1%), then manual estimation ($N$ = 838; 10.9%). A total of eight offset instances were automatically rejected by the aRFF-AP and aRFF-APH algorithms due to pitch strength values < .05. The remainder of these rejections were due to errors in identifying voiced cycles ($N$ = 150 for aRFF-AP, $N$ = 146 for aRFF-APH), or post-processing of resulting RFF values (e.g., glottalization; $N$ = 2 for aRFF-AP, $N$ = 0 for aRFF-APH).

Out of 7709 onset instances (see **Fig. 3.5b**), the aRFF-APH algorithms also resulted in the largest amount of correctly identified boundary cycles ($N$ = 6290, 81.6% of instances). The aRFF-AP algorithm produced the second greatest number of correctly identified cycles ($N$ = 5984, 77.8% of instances), followed by manual RFF ($N$ = 2137, 27.7% of instances). For voicing onset, the majority of misclassifications for the semi-automated RFF algorithms (aRFF-AP, aRFF-APH) occurred within the voiceless consonant (12.2% for aRFF-AP and 4.0% for aRFF-APH). Misclassifications using manual RFF estimation were more concentrated within the vowel (41.2% of onset instances). The aRFF-AP algorithm rejected the least number of onset instances ($N$ = 782; 10.1%), followed by the aRFF-APH algorithm (N = 1107; 14.4%) and manual RFF ($N$ = 1913; 24.8%). A total of 236 onset instances were automatically rejected by the aRFF-AP and aRFF-APH algorithms due to a pitch strength < .05; the remainder of these rejections were due to errors in identifying voiced cycles ($N$ = 531 for aRFF-AP, $N$ = 851 for aRFF-

APH), or post-processing of resulting RFF values (e.g., glottalization; $N = 15$ for aRFF-AP, $N = 20$ for aRFF-APH).

The results of the chi-square tests of independence are shown in **Table 3.7**. This analysis showed that the relation of RFF estimation method and ability to identify the true boundary cycle was significant, resulting in a large effect sizes for both voicing offset ($p < .001$, $V = .53$) and voicing onset ($p < .001$, $V = .51$). The relation of manual and aRFF-AP methods with boundary cycle accuracy had a significant, large effect for voicing offset ($p < .001$, $V = .54$) and voicing onset ($p < .001$, $V = .50$), wherein aRFF-AP was more likely to correctly identify the true boundary cycle than manual estimation. Similarly, the relation of manual and aRFF-APH method with boundary cycle accuracy had a significant, large effect for both offset ($p < .001$, $V = .60$) and onset ($p < .001$, $V = .54$), such that aRFF-AP was more likely to correctly identify the true boundary cycle. Finally, the relation of semi-automated RFF algorithms (aRFF-AP, aRFF-APH) was

**Table 3.7. Chi-square ($X^2$) tests of independence to examine the association between RFF estimation method and accuracy of boundary cycle identification for voicing offset (top model) and onset (bottom model).**

| Model | RFF Estimation Methods | *df* | *N* | $X^2$ | *p* | *V* | Effect Size Interpretation |
|---|---|---|---|---|---|---|---|
| Offset | Manual vs. aRFF-AP vs. aRFF-APH | 2 | 23127 | 6497.0 | <.001 | .53 | Large |
| | Manual vs. aRFF-AP | 1 | 15418 | 4435.7 | <.001 | .54 | Large |
| | Manual vs. aRFF-APH | 1 | 15418 | 5575.0 | <.001 | .60 | Large |
| | aRFF-AP vs. aRFF-APH | 1 | 15418 | 91.5 | <.001 | .08 | Negligible |
| Onset | Manual vs. aRFF-AP vs. aRFF-APH | 2 | 23127 | 5917.8 | <.001 | .51 | Large |
| | Manual vs. aRFF-AP | 1 | 15418 | 3850.5 | <.001 | .50 | Large |
| | Manual vs. aRFF-APH | 1 | 15418 | 4513.8 | <.001 | .54 | Large |
| | aRFF-AP vs. aRFF-APH | 1 | 15418 | 37.4 | <.001 | .05 | Negligible |

*Note. Effect size interpretations of Cramer's V are based on criteria from Cohen (1988).*

significant for both voicing offset and onset ($p < .001$); however, the size of this effect was negligible ($V = .08$ for offset and $V = .05$ onset).

### *Discussion*

The goal of this study was to conduct an exploratory analysis toward understanding the physiological factors that influence acoustic outputs within the RFF algorithm. To do so, a large set of speakers produced the utterance, /ifi/, while altering vocal rate and vocal effort. Acoustic signals were collected via a microphone in conjunction with laryngeal images via a flexible nasendoscope. The resulting database of voiced–unvoiced–voiced productions were used to determine the relationships between a range of acoustic features and the termination (voicing offset) and initiation (voicing onset) of vocal fold vibration. A stepwise binary logistic regression was conducted to identify the acoustic features that best coincided with the time of voicing offset and/or onset. After implementing these features into the semi-automated RFF algorithm ("aRFF-APH"), algorithmic performance was assessed by quantifying the distance between algorithmically and physiologically identified boundary cycles (i.e., the vocal cycles immediately adjacent to the intervocalic fricative, or voicing offset cycle 10 and voicing onset cycle 1). This accuracy was compared against that of two other methods of calculating RFF: manual RFF estimation and semi-automated RFF estimation in the absence of physiologically determined acoustic features ("aRFF-AP").

The results of this work indicate that incorporating acoustic features that coincide well with voicing transitions leads to increased correspondence between the algorithmic and physiologic boundary cycles. By examining the discrepancy in boundary selection

between RFF estimation methods, the aRFF-APH algorithm led to the greatest overall percentage of correctly identified boundary cycles (71.8%) compared to that of the aRFF-AP algorithm (66.5%) and manual estimation (21.4%) methods. Indeed, the aRFF-APH algorithm was significantly more likely to identify the physiological boundary cycle than the aRFF-AP algorithm or manual estimation.

Despite the promising results obtained for the aRFF-APH algorithm, however, the aRFF-AP algorithm remains the gold-standard method for semi-automatically estimating RFF. This is because the aRFF-AP algorithm was developed to increase the *clinical applicability* of RFF, whereas the aRFF-APH algorithm was developed to improve the *physiological relevance* of RFF. Due to the inherent differences in the aims of these works, pitch strength-tuned algorithm parameters were not developed in the present study. Whereas the dataset used in *Chapter 2* to develop the aRFF-AP algorithms comprised a broad range of vocal function (across the spectrum of dysphonia severity and recording conditions), the sample used in the current study was more limited. The participants in the current study exhibited a with a narrower range of diagnoses (57% typical, 16% MTD, 3% nodules, 2% polyp, 1% scarring, 1% lesion, 20% Parkinson's disease) and resulting dysphonia severity (0–51.3). The range of overall severity of dysphonia demonstrated a substantial overlap across groups; for instance, dysphonia severity ranged from 1.7 to 34.2, whereas adults with a voice disorder exhibited an overall severity of dysphonia ranging from 0.9 to 38.5. Speakers were therefore instructed to produce the RFF utterances across a range of vocal rates and amounts of vocal effort to simulate a range of laryngeal muscle tension levels. Yet the range of voice

sample characteristics captured in the current study was still limited, as all speakers were recorded in the same location (i.e., a sound-attenuated booth in the presence of constant noise from the endoscopic light source). As such, pitch strength categories were not incorporated in the current study. Instead, the methodology described by Lien (2015) was used to implement the physiologically relevant acoustic features in the development of the aRFF-APH algorithm.

In spite of the differences between the aRFF-AP and aRFF-APH algorithms, both resulted in a greater correspondence between acoustically and physiologically identified boundary cycles than did manual estimation. These results are surprising since manual estimation was considered the gold-standard technique for calculating RFF. Manual estimation serves as the gold standard for RFF estimates since trained technicians can exercise trial and error to identify the boundary cycle in difficult scenarios (e.g., poor recording environment and/or equipment, severe dysphonia) when boundary cycle masking, such as from concurrent aspiration and frication from articulation, is present. Yet the findings of the current study call into question whether manual RFF estimation should still be considered the gold-standard method. It is possible that the characteristics of the speaker database of the current study confounded this outcome, as all speakers were recorded in a sound-attenuated booth while undergoing an endoscopic examination. In particular, noise from the endoscopic light source could have masked the voice signals and/or speaker productions may have deviated from the norm due to the flexible nasendoscope. If so, manual RFF estimation techniques may not have been sensitive enough to isolate the physiological boundary cycle. Building on this theme, the

algorithms identify potential vocal cycles by leveraging a filtered version of the microphone signal. Specifically, the microphone signal is band-pass filtered using the estimated range of the speaker's $f_o$, which may reduce the effects of coarticulation from concurrent aspiration and frication during the production of the voiceless consonant. As such, the algorithms may not have been as affected by these recording conditions since the aRFF-AP algorithms were designed to account for such variations and the aRFF-APH algorithms were refined based on the physiologically determined vocal fold characteristics.

It is important to note that even though manual estimation resulted in the least number of correctly identified boundary cycles, most of these misclassifications occurred within two pitch periods of the true boundary cycle for both voicing offset and onset. These findings are similar to those comparing differences in boundary cycle selection between microphone- and accelerometer-derived RFF estimates using manual estimation techniques. Because a neck-surface accelerometer is able to capture the vibrations of the glottal source in the absence of vocal cycle masking due to frication and aspiration (as may occur during the production of an intervocalic fricative; Cheyne et al., 2003), the accelerometer signal is more sensitive in capturing the physiological vibrations of the vocal folds. Lien et al. (2015a) estimated that offset RFF values were extracted approximately two cycles closer to the vowel when using a microphone signal rather than an accelerometer signal. Onset RFF values, on the other hand, were computed less than one cycle away from the voiceless consonant when using a microphone signal relative to when using an accelerometer signal. The results of the current study support these

findings, wherein the majority of misclassifications occurred closer to the vowel for both voicing offset and onset when using manual RFF estimation.

Although the current study elucidates some factors that affect the acoustic outputs of the semi-automated RFF algorithm, the source of the discrepancy between acoustic and physiological boundary cycles using manual RFF estimation is still unclear. It is suspected that this discrepancy is the result of the algorithms leveraging a filtered version of the microphone signal to reduce the amplitude of vocal tract resonances, coarticulation due to concurrent frication and aspiration, and radiation of the lips. The algorithms use this filtered signal to identify potential vocal cycles. By only using the raw microphone signal to identify vocal cycles, the RFF values resulting from manual estimation may not reflect the true offset or onset of voicing as expected. Further investigation is necessary to examine this hypothesis, and should include an analysis of both laryngeal imaging and acoustics to comprehensively assess the relevance and validity of manual estimation as the gold-standard technique for calculating RFF. In doing so, laryngeal imaging would provide physiological confirmation of vocal fold vibrations that are indirectly captured via RFF. In addition to comparing manual and semi-automated boundary cycle selections, this investigation should aim to compare the boundary cycles obtained via manual RFF estimation when using each version of the acoustic signal. In the event that manual estimation is no longer considered as gold-standard RFF method, efforts should be made to develop new metrics of algorithmic performance, as current measures (e.g., root-mean-square error, mean bias error) are calculated in reference to RFF values obtained via manual estimation.

*Limitations and Future Directions*

Although the findings of the current study show promise for using RFF as a clinically relevant tool for assessing laryngeal muscle tension, steps must be undertaken to improve the clinical applicability of the aRFF-APH algorithm. As previously mentioned, the speaker database examined in the current study was limited in terms of voice sample characteristics (e.g., overall severity of dysphonia, recording conditions). Future work should therefore validate this algorithm in a larger set of speakers (using a training and test set) across a broad range of vocal function. In doing so, the aRFF-APH algorithm could be modified to include pitch strength categories to account for variations in voice sample characteristics.

In addition to algorithmic validation, it is worth pointing out that the current version of the aRFF-APH algorithm was only refined for use in microphone signals. Even though the majority of studies on RFF employed microphone signals, there has been increasing interest in using neck-surface vibrations generated during speech for ecological momentary assessment and ambulatory voice monitoring (e.g., Cheyne et al., 2003; Cortés et al., 2018; Fryd et al., 2016; Ghassemi et al., 2014; Hillman et al., 2006; Mehta et al., 2016; Mehta et al., 2015; Mehta et al., 2012a; Popolo et al., 2005; Švec, Titze, & Popolo, 2005; Van Stan et al., 2015a). Using an accelerometer, the physiological mechanisms of speech production can be non-invasively assessed in a way that minimizes the effects of supraglottic resonance, aspiration and frication due to coarticulation, and radiation of the lips. Moreover, accelerometers are less sensitive to the effects of background noise (Zanartu et al., 2009) and cannot be used to construct

intelligible speech (Cheyne et al., 2003). By capturing daily vocal behavior through a neck-surface accelerometer, vocal behaviors associated with excessive or imbalanced laryngeal muscle forces could be identified and monitored via RFF. An accelerometer-tuned RFF algorithm has been developed (Groll et al., 2020); however, future work should examine this algorithm to identify physiologically tuned features to identify the termination and initiation of vocal fold vibration. Doing so would further improve the clinical relevance of using RFF to assess and track laryngeal muscle tension.

### *Conclusions*

Although RFF has demonstrated marked potential for clinical implementation as an estimate of laryngeal muscle tension, the theoretical understanding of the physiological factors that influence the semi-automated RFF algorithm have largely remained unclear. The current study therefore examined the relationship between acoustic outputs from the algorithm and physiological vocal fold vibratory characteristics during voicing offsets and onsets. By enhancing the physiological relevance of the acoustic features used to estimate RFF, algorithmic accuracy increased with respect to identifying the true termination and initiation of vocal fold vibration. This accuracy was greater than that of the previous version of the RFF algorithm as well as the gold-standard, manual method for calculating RFF. These findings highlight improvements in the precision of using RFF to reflect the underlying physiological mechanisms for voicing offsets and onsets.

# CHAPTER 4. The Relationship between Vocal Fold Abductory Kinematics and Relative Fundamental Frequency: An Analysis across Young Adults, Older Adults, and Adults with Parkinson's Disease


## *Abstract*

**Purpose:** Relative fundamental frequency (RFF) is an acoustic measure that is thought to capture changes in laryngeal muscle tension and vocal fold abductory kinematics as a speaker devoices. Older adults typically exhibit lower RFF values than young adults, which has thus far been attributed to a prolonged abductory gesture for devoicing. Older adults with Parkinson's disease (PD) are reported to exhibit even lower RFF values than similar age controls, perhaps due to the interplay of a prolonged abductory gesture and increased levels of baseline laryngeal muscle tension. Despite these speculations, the contribution of vocal fold abduction to RFF has not yet been characterized. Thus, this study aimed to examine abductory patterns in young adult controls, older adult controls, and older adults with PD in order to elucidate the contribution of abduction to RFF during intervocalic voicing offsets.

**Methods:** Twenty-four individuals with Parkinson's disease ($M$ = 62.8 years, $SD$ = 9.6 years), twenty-four young adult controls ($M$ = 21.8 years, $SD$ = 3.4 years), and twenty-four older adult controls (age- and sex-matched to individuals with PD; $M$ = 63.1 years, $SD$ = 11.3 years) produced strings of the utterance, /ifi/ at their typical vocal pitch and loudness. Simultaneous recordings were made using a microphone and flexible nasendoscope. RFF was calculated from the acoustic signal, whereas the duration of the abduction gesture and glottic angle at voicing offset were identified through the

laryngoscopic images. Three separate analysis of variance models were constructed to examine differences in mean RFF at offset cycle 10, abduction duration, and glottic angle at voicing offset across speaker groups. An analysis of covariance model was then used to examine the relationship of RFF and abduction duration, glottic angle at voicing offset, speaker age, and speaker group.

**Results:** There were no statistically significant differences across groups for RFF at offset cycle 10 ($p = .084$, $\eta_p^2 = 0.07$), abduction duration ($p = .105$, $\eta_p^2 = 0.06$), or glottic angle at voicing offset ($p = .502$, $\eta_p^2 = 0.02$). However, speaker age ($p = .023$, $\eta_p^2 = 0.08$) and glottic angle at voicing offset ($p = .001$, $\eta_p^2 = 0.16$) were statistically significant predictors of RFF at offset cycle 10.

**Conclusions:** Vocal fold abductory patterns were not significantly different across young adult controls, older adult controls, and older adults with PD. However, speaker age and glottic angle prior to the termination of vocal fold vibration were found to be significantly related to RFF estimates obtained at acoustic voicing offset. The findings of this study indicate that RFF is, as hypothesized, related to abductory patterns during devoicing. Furthermore, speaker age is a predominating factor in the assessment of RFF, and should be considered in future studies aiming to assess differences in RFF within and across speakers.

### *Background*

Human speech production arises from the integration of aerodynamic, kinematic, and acoustic mechanisms. According to the classic source-filter theory of voice production, these mechanisms converge into a sound source and filtering process to drive speech production (Fant, 1960; Stevens, 2005). Sound sources are generated from airflow passing through narrow constrictions within the larynx, whereas the resonance characteristics of the vocal tract filter the sound source to produce speech sounds. Different sources are responsible for different speech sounds; for instance, the vibrating vocal folds serve as a sound source during the production of vowels, whereas airflow passing through oral articulatory constrictions (e.g., tongue elevated toward the hard palate) serves as the sound source for voiceless consonants. In the English language, the transition between a vowel and a voiceless consonant is marked by a change in source, specifically from the vibrating vocal folds to aspiration noise. This transition is associated with characteristic physiological patterns that are primarily attributed to laryngeal muscle tension and vocal fold kinematics (Löfqvist et al., 1989; Stepp et al., 2011d; Van Den Berg, 1958; Watson, 1998).

Laryngeal muscle tension is a crucial component in the termination of voicing that occurs when transitioning from a vowel to a voiceless consonant ("voicing offset"). This is because the regulation of laryngeal muscle tension is necessary to tense and abduct the vocal folds to cease vocal fold vibration (Boone et al., 2014, p. 42; Löfqvist et al., 1989). Numerous mechanisms have been posed to describe the physiological underpinnings of increased vocal fold tension during voicing offset; these include increased cricothyroid

activity (Stevens, 1977), increased thyroarytenoid activity (Hirano, 1974), and increased vocal fold stiffness from passive stretching that occurs as the result of changes in laryngeal height (Sonninen, Hurme, & Laukkanen, 1999; Stevens, 1977). Voice fundamental frequency ($f_o$) increases as a result of increased tension across the aforementioned mechanisms at voicing offset.

In addition to laryngeal muscle tension, vocal fold abductory kinematics may play a key role in enabling voicing offset. The posterior cricoarytenoid muscle is the sole intrinsic laryngeal muscle that contributes to vocal fold abduction. It supports the larynx during devoicing by acting as an antagonist to the cricothyroid and adductor muscles (i.e., thyroarytenoid, interarytenoid, lateral cricoarytenoid; Choi et al., 1993a; Faaborg-Andersen, 1957; Fujita et al., 1989; Hirano, 1988). Vocal fold abduction has been observed during vowels that precede voiceless consonants (Fukui & Hirose, 1983), which reduces the duration of vocal fold contact as the vocal folds continue to open (Rothenberg & Mahshie, 1988). Watson (1998) hypothesized that vocal fold abduction leads to lower $f_o$ values during voicing offset. The interplay of laryngeal muscle tension and vocal fold abduction are suspected to counteract each other during voicing offset to cease vocal fold vibration. These mechanisms may be captured by a non-invasive, objective measure called relative fundamental frequency (RFF).

RFF reflects short-term changes in $f_o$ during voicing offsets and onsets, and is typically extracted from the ten vocal cycles immediately preceding and following the voiceless consonant in a vowel–voiceless consonant–vowel production. The $f_o$ values of these 20 vocal cycles are then normalized to a relatively steady-state $f_o$ value of the

corresponding vowel (i.e., first vowel for voicing offset, second vowel for voicing onset) to enable comparisons of $f_o$ changes within and across speakers.

The pattern of RFF values during voicing offsets is thought to reflect the interplay of laryngeal muscle tension and vocal fold abduction that enable devoicing. In particular, RFF values of young adults typically exhibit a characteristically stable or slightly decreasing trend during voicing offset (Goberman et al., 2008; Watson, 1998). On the other hand, older adults typically demonstrate significantly lower offset RFF values (Stepp, 2013; Watson, 1998). The dissimilarity in offset RFF trends in these groups would suggest a possible difference in the mechanisms used for devoicing. As the mean $f_o$ value at voicing onset was found to match the highest mean $f_o$ value prior to voicing offset, Watson (1998) proposed that older adults may not be able to produce transient increases in vocal fold tension to assist in devoicing. Instead, it was suggested that older adults rely on a prolonged abductory gesture as their primary mechanism for devoicing, rather than a combination of tension and abduction. A prolonged abductory gesture implies that the gesture begins earlier during the preceding vowel and may include an increased glottic angle at the time of voicing offset (i.e., due to the vocal folds opening to cease vibration; Fukui et al., 1983). This mechanism was suggested as a byproduct of age-related vocal fold atrophy frequently observed in older adults, which is often characterized by changes in voice quality (e.g., breathiness, weakness, hoarseness, inability to sustain phonation; Takano et al., 2010), and atrophy of the vocalis muscle (Honjo & Isshiki, 1980; Rodeño, Sánchez-Fernández, & Rivera-Pomar, 1993).

The theoretical implications of a prolonged abductory gesture in older adults is

interesting to consider when comparing RFF values between adults with Parkinson's disease (PD) and age-matched controls (Goberman et al., 2008; Stepp, 2013). PD is a progressive neurodegenerative disease that typically develops in middle to late life (with incidence rates rising rapidly after 60 years of age; Van Den Eeden et al., 2003) and affects the central and peripheral nervous systems (Braak et al., 2003; Schapira et al., 2017). Thought to be a product of neural processes and morphological changes to the muscles, increased muscle tension is one of the hallmark motor symptoms exhibited in PD (Dietz et al., 1981; Edstrom, 1968; Mu et al., 2012; Rossi et al., 1996; Watts et al., 1986). The presence of tension in PD has been well-documented in the extremities (Berardelli et al., 1983; Cantello et al., 1991; Cantello et al., 1995; Edstrom, 1970; Meara et al., 1993; Prochazka et al., 1997; Robichaud et al., 2009; Rossi et al., 1996) axial muscles (e.g., hips; Anastasopoulos et al., 2009; Gurfinkel et al., 2006; Kroonenberg et al., 2006; Mak et al., 2007; Nagumo et al., 1993, 1996), and—more recently—in the laryngeal muscles. Specifically, excessive tension of the intrinsic laryngeal muscles has been reported in adults with PD when compared to age-matched controls (Zarzur et al., 2013; Zarzur et al., 2007). Prior work suggests that RFF is able to reflect this disparity in baseline tension, wherein adults with PD exhibit even lower RFF values compared to controls (Goberman et al., 2008; Stepp, 2013). Since voicing offset is thought to require the interplay of laryngeal muscle tension and vocal fold abduction to devoice, however, it is possible that the low RFF values observed in older adults with PD are due to an even more prolonged abductory gesture rather than from increased baseline levels of laryngeal muscle tension. Since the contribution of vocal fold abduction to RFF has not been

physiologically examined, it is thus unclear how tension and abduction play a role in RFF values in older adults with PD.

Although there is theoretical backing to support a prolonged abductory gesture in older adults and concurrently increased levels of baseline laryngeal muscle tension in PD, these conjectures have yet to be confirmed in relation to RFF. This is largely because RFF is estimated via an acoustic signal. Using a microphone to estimate RFF is useful for non-invasive clinical voice assessments; however, the acoustic signal indirectly reflects the glottal source. As such, vocal fold vibratory and abductory kinematics during devoicing cannot be characterized in relation to RFF when only examining the acoustic signal. Thus, it remains unclear (i) whether the observed discrepancies in RFF values between young and older adults truly correspond to a greater reliance of older adults on vocal fold abduction for devoicing, and further, (ii) whether the observed lower offset RFF values in PD reflect an increased reliance on abduction to cease voicing, increased levels of baseline laryngeal tension that arise with PD, or some other cause (e.g., laryngeal height).

### *Purpose of the Current Study*

RFF has been proposed as an acoustic estimate that reflects the degree of baseline laryngeal muscle tension. This measure shows promise in the clinical assessment and tracking of tension, as excessive and/or imbalanced laryngeal muscle forces have been implicated in a large proportion of voice disorders (Ramig et al., 1998). However, the specific contributions of the abductory gesture to RFF have not been physiologically assessed. As such, the goal of the current study was to examine the relationship between

vocal fold abductory kinematics and RFF. Three groups of speakers were assessed in the present study: young adult control speakers, older adult control speakers, and older adults with PD. These groups were chosen to explore hypothesized differences in laryngeal muscle tension and abductory mechanisms used to devoice.

To carry out this investigation, RFF measures were extracted during voicing offset and compared to time-aligned abductory measures obtained via high-speed videoendoscopy. These HSV measures included glottic angle at voicing offset and abduction duration. The relationships between RFF and these measures were used to characterize abductory patterns, as well as to elucidate the contribution of abductory characteristics to differences in RFF typically observed in older adult controls and older adults with PD when compared to young adult controls. As such, four hypotheses were proposed:

1. RFF values at voicing offset (offset cycle 10) will be significantly lower in older adult controls than in young adult controls due to a greater reliance on vocal fold abduction. Furthermore, RFF values at offset cycle 10 will be significantly lower in older adults with PD compared to age- and sex-matched controls due to increased baseline laryngeal muscle tension.
2. Older adult controls will exhibit significantly longer abduction durations and larger glottic angles at voicing offset compared to young adult controls.

3. Abduction duration and glottic angle at voicing offset will not significantly differ between older adults with PD and age-/sex-matched controls.
4. Abduction duration and glottic angle at voicing offset will be significantly, negatively related to RFF at voicing offset (i.e., larger abduction duration or glottic angle at voicing offset will be associated with lower RFF values at offset cycle 10).

## ***Methods***

### *Participants*

A total of 72 participants from the database described in *Chapter 3* were analyzed in the current study. This subset comprised individuals with PD ($N = 24$), age- and sex-matched control speakers ($N = 24$), and a set of sex-matched young control speakers ($N = 24$). All individuals with PD and age-matched controls were administered the Montreal Cognitive Assessment (MoCA) to determine cognitive status. An *a priori* cut-off of ≥21 was set to ensure all included participants had the capacity to consent to the study tasks (Dalrymple-Alford et al., 2010).

#### Speakers with Parkinson's Disease

A group of 24 individuals with idiopathic Parkinson's disease (6 female, 18 male) aged 43–75 years ($M = 62.8$ years, $SD = 9.6$ years) were enrolled in the study. This sex distribution is consistent with the higher incidence of PD in men compared to women (Gillies, Pienaar, Vohra, & Qamhawi, 2014; Van Den Eeden et al., 2003). All speakers were fluent in English, reported no history of hearing problems, and were diagnosed with

idiopathic Parkinson's disease by a neurologist. Individuals with PD were recorded while on their usual carbidopa/levodopa medication schedule to preserve typical vocal function. Individuals who used deep brain stimulation devices were requested to turn their device off for the duration of data collection to minimize the potential impacts of deep brain stimulation on laryngeal function.

**Table 4.1** shows demographic information for the 24 individuals with PD. A speech-language pathologist specializing in voice disorders assessed the overall severity of dysphonia[5] of each participant using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster et al., 2009), as described in detail in *Chapter 3*. The average overall severity of dysphonia (OS) score was 17.9/100 (*SD* = 11.8, *range* = 4.0–40.9). Additionally, the Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) was administered to each participant with Parkinson's disease to determine the extent of both motor and non-motor complications; each examination was administered and scored per protocol by a certified MDS-UPDRS administrator. The average severity of motor complications were moderate (*M* = 47.7, *SD* = 20.1) and ranged from mild to severe (*range* = 13–91; Martínez-Martín et al., 2015). The average Hoehn-Yahr score was 2.0 (*SD* = 1.1) and ranged from 0 (no disability) to 4 (severe disability; Goetz et al., 2004; Hoehn et al., 1967).

[5] Overall severity of dysphonia describes the "global, integrated impression of voice deviance" from normal, and is rated on a 100-millimeter visual analog scale out of 100 (wherein higher scores indicate a greater deviance from normal).

**Table 4.1. Demographic information of participants with disordered voices.**

| Participant | Sex | Age | CAPE-V OS | Parkinson's Disease Characteristics | | |
|---|---|---|---|---|---|---|
| | | | | Years Post-Dx | MDS-UPDRS-III | Hoehn-Yahr Scale |
| PD1 | M | 60 | 30.1 | 7 | 54 | 2 |
| PD2 | M | 49 | 5.8 | 7 | 47 | 1 |
| PD3 | F | 62 | 5.6 | 9 | 49 | 3 |
| PD4 | M | 45 | 10.4 | 10 | 51 | 2 |
| PD5 | F | 70 | 14.7 | 6 | 77 | 4 |
| PD6 | M | 50 | 7.1 | 0 | 17 | 0 |
| PD7 | M | 55 | 18.4 | 21 | 49 | 3 |
| PD8 | M | 62 | 10.0 | 3 | 50 | 2 |
| PD9 | F | 74 | 30.6 | 24 | 59 | 2 |
| PD10 | M | 73 | 33.6 | 9 | 19 | 1 |
| PD11 | M | 67 | 6.4 | 4 | 63 | 3 |
| PD12 | M | 67 | 19.4 | 2 | 38 | 2 |
| PD13 | M | 62 | 27.9 | 13 | 47 | 2 |
| PD14 | M | 59 | 4.0 | 2 | 23 | 2 |
| PD15 | M | 73 | 6.8 | 3 | 23 | 1 |
| PD16 | M | 68 | 5.0 | 6 | 38 | 2 |
| PD17 | M | 72 | 40.9 | 7 | 81 | 4 |
| PD18 | F | 73 | 33.3 | 8 | 52 | 2 |
| PD19 | M | 75 | 22.1 | 1.5 | 68 | 2 |
| PD20 | F | 65 | 15.4 | 10 | 48 | 3 |
| PD21 | M | 68 | 8.5 | 1 | 52 | 3 |
| PD22 | M | 43 | 35.8 | 5 | 91 | 3 |
| PD23 | M | 65 | 28.3 | 1 | 35 | 0 |
| PD24 | F | 51 | 9.7 | 5 | 13 | 0 |

*Note. Dx = Diagnosis, CAPE-V OS = Consensus of Auditory-Perceptual Evaluation of Voice, Overall Severity of Dysphonia, PD = Parkinson's disease, MDS-UPDRS-III = Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale: Part III, Motor Examination.*

## Control Speakers

A group of 48 individuals without voice disorders (12 female, 36 male) were recruited to participate in the study. Of the 48 participants, 24 speakers (6 female, 12 male; *M* = 63.1 years, *SD* = 11.3 years, *range* = 41–81 years) were recruited to serve as age[6]- and sex-matched controls to the group with Parkinson's disease; the remaining 24

[6] Control speakers were matched to those with Parkinson's disease within six years of age.

individuals were young adults (6 female, 12 male; *M* = 21.8 years, *SD* = 3.4 years, *range* = 18–31 years), of whom were sex-matched to the older controls and individuals with PD. All participants without voice disorders were fluent speakers of English, and had no history of speech, language, hearing, neurological, pulmonary or disorders. Participants had no trained singing experience beyond grade school in order to minimize variability in phonatory behaviors that may occur when differentiating between singers and non-singers (Stepp et al., 2011b). All were non-smokers and were screened by a certified voice-specializing speech-language pathologist for healthy vocal function via auditory-perceptual assessment (CAPE-V OS) and flexible nasendoscopic laryngeal imaging; the average OS score was 5.4 (*SD* = 1.2, *range* = 1.9–23.5) for young controls and 10.6 (*SD* = 7.7, *range* = 1.7–34.2) for older controls.

Hearing Status

Hearing screening data were collected for 67 of 72 participants. The five participants for which this data was not collected were young adult controls who reported no history of hearing disorders. Each of the remaining 19 young adult controls passed a hearing screening of pulsed pure tones (Burk et al., 2004) at frequencies of 125, 250, 500, 1000, 2000, and 4000 Hz under 25 dB HL in both ears (American Speech-Language-Hearing Association, 2005). All 24 older adult controls and 17 older adults with PD passed a hearing screening of pulsed pure tones (Burk et al., 2004) at frequencies of 125, 250, 500, and 1000 under 25 dB HL and at 2000 and 4000 Hz under 40 dB HL in at least one ear (Schow, 1991). One participant with PD (PD16 in **Table 4.1**) demonstrated a threshold of 45 dB HL at 2000 Hz. Two participants with PD (PD19, PD23) had a

threshold of 45 dB HL at 4000 Hz, and an additional two participants with PD (PD13, PD18) demonstrated a threshold of 50 dB HL at 4000 Hz. One participant with PD (PD17) had thresholds of 50 dB HL at 2000 Hz and 70 dB HL at 4000 Hz. Finally, one participant with PD (PD12) wore hearing aids during the course of the study, and demonstrated thresholds of 45 dB HL at 125 Hz and 30 dB HL at 1000 Hz.

*Recording Procedures*

All data analyzed in the current study were collected during the nasendoscopic examination described in detail in *Chapter 3*. In brief, participants were trained to produce iterations of the utterance, /ifi/, across three vocal rates (slow, regular, fast) and three levels of vocal effort (mild, moderate, maximum). Participants were then instrumented with a directional headset microphone (Shure SM35 XLR) and neck-surface accelerometer (BU series 21771 from Knowles Electronic, Itasca, IL). The microphone was placed 45° from the midline and 7 cm from the lips, and the accelerometer was positioned on the anterior neck, superior to the thyroid notch and inferior to the cricoid cartilage. These signals were pre-amplified (Xenyx Behringer 802 Preamplifier) and digitized at 30 kHz (National Instruments 6312 USB).

A flexible routine endoscope (Pentax, Model FNL-10RP3, 3.5-mm) was passed transnasally and into the hypopharynx for laryngeal visualization. A flexible slim endoscope (Pentax, Model FNL-7RP3, 2.4-mm) was used in the event that participant anatomy or comfort interfered with the acquisition process. Laryngeal images were recorded at 1 kHz via a camera (FASTCAM Mini AX100l; Model 540K-C-16GB; 256 × 256 pixels; 40-mm optical lens adapter) attached to the endoscope. A frame rate of 1 kHz

was used in this analysis to (i) track the fundamental frequency of vibration of the vocal folds, which is estimated to be 85–255 Hz during modal phonation in adults (Baken et al., 2000, p. 156), and (ii) capture gross abductory gestures, which occur within 104–227 ms (Dailey et al., 2005).

During the nasendoscopic examination, participants were instructed to produce eight /ifi/ utterances at each vocal rate and level of vocal effort, in the following order: slow rate, regular rate, fast rate, mild effort, moderate effort, maximum effort. This number of repetitions was selected based on the recording limitations of setup: the high-speed imaging, the synchronized microphone, accelerometer, and HSV recordings were restricted in duration to 7.940 seconds when the 3.5-mm endoscope was used and 8.734 seconds when the 2.4-mm endoscope was used. To account for trials in which productions at the end of the recording were incompletely captured or in cases where less than eight /ifi/ utterances were produced, each condition was repeated a minimum of two times (additional trials were recorded in the event that the vocal folds were not sufficiently captured). The length of this procedure lasted approximately 5–10 minutes.

Because the current study sought to isolate the relationship between RFF and vocal fold abduction during typical speech, only /ifi/ productions produced at a regular rate without intentional increases in vocal effort were retained for further analysis. Of the total productions, there were 200 instances in which 16 full /ifi/ productions were not captured for a speaker. This process resulted in total of 952 /ifi/ productions ([72 participants × 16 /ifi/ productions] – 200 incomplete or missing /ifi/ productions) across the three speaker cohorts.

## *Data Analysis*

### High-speed Video Image Processing

**Glottic Angle Waveform**

The methods used for HSV data extraction have been described in detail in *Chapter* 3, but will be summarized here in the context of the current study. The HSV images were first processed to determine /ifi/ usability. Trained technicians[7] manually inspected the video images comprising each /ifi/ production to determine whether the videos effectively captured the vocal folds during the transition into and out of the /f/. In the event that the vocal folds were obstructed (e.g., by the epiglottis) or not visible (e.g., due to poor image contrast) during the recording, the production was regarded as "unusable" and removed from further analysis. For usable videos, the trained technicians then ran an automated glottic angle extraction algorithm (Diaz-Cadiz et al., 2019) to track the glottic angle over time. If the vocal folds were not appropriately tracked, the technician could intervene by manually extracting vocal fold angles to inform the algorithm before running again. If errors still persisted following manual intervention, the trained technicians marked the /ifi/ production as unusable.

Nine trained technicians used the semi-automated MATLAB algorithm to extract

[7] Technicians were first trained in glottic angle identification at a conventional framerate of 30 frames per second (fps). The technicians were required to meet a training standard via two-way mixed-effects intraclass correlation coefficients for consistency of agreement [ICC(3,1]) ≥ .80 when marking glottic angles at 30 fps. The technicians were then trained to use a semi-automated glottic angle extraction algorithm developed in MATLAB (version 9.3; The MathWorks, Natick, MA) to extract glottic angles at 1000 fps. Similarly, the technicians were required to meet reliability standards of ICC(3,1) ≥ .80 compared to a gold-standard technician, described in Diaz-Cadiz et al. (2019). See *Chapter 3* for more details about this training scheme and the resulting technician reliabilities.

the glottic angle waveform for each /ifi/ production ($N = 952$). A single technician determined whether the /ifi/ production was usable and, if so, proceeded to extract the glottic angle waveform for the production. The technicians accepted the automated algorithmic results in 62.6% of productions ($N = 596$), and accepted the algorithmic results after performing manual extraction techniques on 20.8% of productions ($N = 198$). Of the remaining productions, 10.4% were considered unusable ($N = 99$) and a further 6.2% were rejected due to errors in algorithmic estimation ($N = 59$); an average of 2.2 /ifi/ productions ($SD = 3.3$) were classified as unusable or subject to algorithmic errors for each speaker. All productions that demonstrated problems in video usability or algorithmic processing were removed from subsequent analysis. As described in *Chapter 3*, algorithmic reliability was not assessed since prior work indicates that the algorithm yields good reliability ($ICC \geq .80$; Diaz-Cadiz et al., 2019). However, this initial data processing was then rechecked by a second trained technician. The additional analysis resulted in 794 usable /ifi/ productions for subsequent processing.

**Vocal Fold Abduction Time**

To assess vocal fold abductory kinematics, two metrics were extracted from the /ifi/ productions using methodology described in *Chapter 3*. Technicians were presented with a MATLAB (version 9.3; The MathWorks, Natick, MA) graphical user interface that showed time-aligned high-speed video frames, the microphone signal, the previously extracted glottic angle waveform, and the quick vibratory profile (QVP).

The QVP is a one-dimensional waveform that uses changes in light intensity of the video frame to estimate the vibratory motion of the vocal folds. Whereas the glottic

angle waveform is only sensitive to the vibration of the vocal folds, the QVP captures vocal fold vibration in addition to non-glottic activities such as camera or epiglottic motion (Ikuma et al., 2013). The QVP was therefore included as a supplement to the glottic angle waveform to assist technicians in discriminating the vibrating glottis in images of poor resolution, as well as to identify the time window containing the vowel to voiceless consonant transition. To calculate the QVP, the HSV frame was first centered over the glottis using methodology described in Diaz-Cadiz et al. (2019). Distinct from the typical QVP—which is calculated as the average of the minimum intensity of each row of the HSV frame—the QVP was computed here using a method that localizes the vibrating glottis (Ikuma et al., 2013). Specifically, changes in light intensity were examined in both the vertical and horizontal directions of the HSV frame; the average of the minimum pixel intensity per row (for vertical profile) or column (for horizontal profile) was then calculated. The horizontal and vertical profiles were then summed together to produce the QVP. From here, the QVP was high-pass filtered using a 7$^{th}$ order Butterworth filter using a cut-off frequency of 50 Hz to attenuate signal noise below a minimum $f_o$ of 50 Hz.

With this information, three technicians were instructed to use the time-aligned glottic angle waveform and QVP to identify two time points within the /ifi/ production. These time points were the start of abduction, described as the last full or maximum contact of the vocal folds during voicing offset, as well as the time of voicing offset, described as the termination of the last vibratory cycle before the voiceless consonant. In cases where the arytenoid cartilages blocked the view of the vocal folds (e.g., anterior-

posterior supraglottic constriction) during voicing offset, the start of abduction was marked as the time point in which the arytenoid cartilages began to move away from one another. The technicians then corroborated the selected time points via manual visualization of the raw HSV images (e.g., in the event that the glottic angle waveform failed to capture small glottic angle changes during the vibratory cycles). The microphone signal was presented to technicians in case the glottic angle waveform and QVP both failed to track the vibrations of the vocal folds (i.e., determined by comparing the waveforms against the raw HSV images). In such cases, technicians were able to mark the /ifi/ production to be rejected or re-processed using the aforementioned methodology. This analysis resulted in a total of 794 measures corresponding to the start of abduction and time voicing offset.

Technician intrarater reliability was assessed using the larger speaker database ($N$ = 122) in *Chapter 3* by reanalyzing 10% of participants in a separate sitting. Intrarater reliability was assessed in regard to extracting the start of abduction and time voicing offset via two-way mixed effects intraclass correlation coefficients (ICCs) for absolute agreement, producing an average intrarater reliability of .94 (*95% CI* = .88–1.0). Technician interrater reliability was assessed by instructing the three technicians to analyze the HSV images of the same participant. Interrater reliability was examined in regard to extracting the start of abduction and time voicing offset via two-way mixed-effects ICCs for consistency of agreement (single measures), resulting in an average interrater reliability of .83 (*95% CI* = .77–.89).

**Laryngeal Image-based Metrics of Vocal Fold Abduction**

To comprehensively investigate the relationship between vocal fold abduction and RFF, a series of HSV-derived measures were collected from the aforementioned analyses. The start of abduction and time of voicing offset were used to compute the duration of the abductory gesture ($T_{abd}$). In addition, the abductory gesture was also characterized by extracting the glottic angle at $t_{off}$ from the glottic angle waveform, called $\theta_{off}$. **Fig. 4.1** shows an example of these measures when extracted from the glottic angle waveform and QVP.

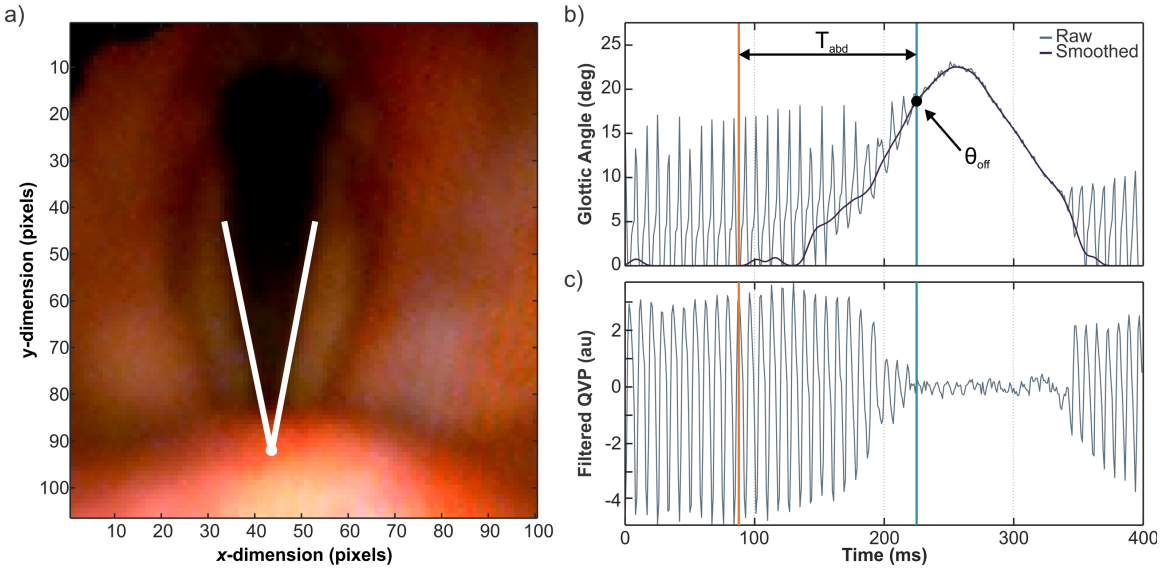


**Figure 4.1. (a) View of the vocal folds under flexible nasendoscopy, with the glottic angle marked from the anterior commissure to the vocal processes, (b) Raw glottic angle waveform (gray) with smoothed data overlay (purple), and (c) Filtered quick vibratory profile (QVP). Solid lines indicate the start of vocal fold abduction (orange) and time of voicing offset (blue). The start of abduction ($T_{abd}$) and glottic angle at voicing offset ($\theta_{off}$) are identified.**

Acoustic Signal Processing

Semi-automated RFF estimation was first performed on the microphone signals of the 794 /ifi/ productions with HSV data using the aRFF-AP algorithm in MATLAB (version 9.3). For this analysis, the aRFF-AP algorithm (developed in *Chapter 2*) was

used instead of the aRFF-APH algorithm (developed in *Chapter 3*), as this version was validated in a large speaker cohort ($N$ = 483) that spanned a range of dysphonia severity and recording conditions. As detailed in *Chapters 2* and *3*, the semi-automated RFF algorithms requires the user to confirm the location of the voiceless consonant, /f/, in each /ifi/ production; as such, incorrect locations were manually adjusted. RFF values were then calculated using the vocal cycles closest to the /f/. In order to examine the association of RFF with group and vocal fold abductory kinematics, RFF at voicing offset cycle 10 was retained for further analysis. Voicing offset instances that were rejected during algorithmic processing (e.g., due to voicing during the voiceless consonant, glottalization, or misarticulation) were removed from further analysis (215 /ifi/ productions).

**Statistical Analyses**

The analyses performed on the high-speed video images and microphone signals resulted in the following measures for each /i/-to-/f/ transition of 580 /ifi/ productions: (1) abduction duration, (2) glottic angle at voicing offset, and (3) RFF at voicing offset cycle 10. There was one older adult with PD for which fewer than two RFF productions were available for averaging (e.g., the algorithm could not identify potential vocal cycles, or resulting RFF values exhibited sharp transitions and were therefore rejected) and was subsequently removed from further processing. The result of this analysis yielded one abduction duration, one glottic angle at voicing offset, and one RFF value at voicing offset cycle 10 for each of the remaining 71 speakers.

The resulting measures were then evaluated with a series of statistical models to

determine the relationship between RFF values and vocal fold abductory kinematics. First, three separate analysis of variance (ANOVA) models were constructed to examine the effect of group (young controls, older controls, and individuals with PD) on voicing offset measures of RFF offset cycle 10, abduction duration, and glottic angle at voicing offset. In these models, each voicing offset measure was set as the response variable and group was set as a fixed factor. For these ANOVA models, significance was set *a priori* to $p < .05$, and partial eta squared ($\eta_p^2$) was calculated as an effect size. Additional *post hoc* analyses were conducted to examine trends in RFF at offset cycle 10, abduction duration, and glottic angle at voicing offset with respect to speaker age, sex, and MDS-UPDRS-III score (for older adults with PD only). Speaker age and MDS-UPDRS-III score were examined against the three voicing offset measures using Pearson's product-moment correlation coefficients. Sex was examined using a two-sample *t*-test.

An analysis of covariance (ANCOVA) model was then constructed to examine the effects of covariates of age, abduction duration, and glottic angle at voicing offset, as well as the fixed factor of group, on RFF offset cycle 10 (response variable). For this model, significance was set *a priori* to $p < .05$, and partial eta squared ($\eta_p^2$) was calculated as an effect size.

## *Results*

### *Relationship between Group and Measures of Voicing Offset*

**Table 4.2** shows the results for the models examining the effects of group on each of the three voicing offset measures. Group was not a significant factor in the model for RFF at offset cycle 10 ($p = .084$, $\eta_p^2 = .07$), abduction duration ($p = .105$, $\eta_p^2 = .06$), or

**Table 4.2. Results of the analysis of variance (ANOVA) models examining the effects of group on RFF at offset cycle 10, abduction duration, and glottal angle at voicing offset.**

| Model | Effect | *df* | *F* | *p* | $\eta_p^2$ | Effect Size Interpretation |
|---|---|---|---|---|---|---|
| RFF Offset Cycle 10 | Group | 2 | 2.57 | .084 | .07 | Small |
| Abduction Duration | Group | 2 | 2.33 | .105 | .06 | Small |
| Glottic Angle at Voicing Offset | Group | 9 | 0.70 | .502 | .02 | Small |

*Note. Effect size interpretations are based on criteria from Witte and Witte (2010).*

glottic angle at voicing offset ($p$ = .502, $\eta_p^2$ = .02).

Mean RFF values for offset cycle 10 were greatest in young controls ($M$ = -1.12 ST, $SD$ = 1.28 ST), followed by older controls ($M$ = -1.23 ST, $SD$ = 0.99 ST) then individuals with PD ($M$ = -1.81 ST, $SD$ = 1.17 ST). Mean abduction duration was longest for individuals with PD ($M$ = 65.1 ms, $SD$ = 17.8 ms), with young controls ($M$ = 57.1 ms, $SD$ = 15.2 ms) and older controls ($M$ = 54.2 ms, $SD$ = 20.1 ms) producing similar values. Finally, glottic angle at voicing offset was greatest in young controls ($M$ = 16.4 degrees, $SD$ = 4.1 degrees), followed by individuals with PD ($M$ = 15.4 degrees, $SD$ = 4.7 degrees) then older controls ($M$ = 15.2 degrees, $SD$ = 5.8 degrees).

Effects of Sex on Voicing Offset

**Fig. 4.2** shows the distribution of voicing offset measures based on speaker sex and group. The results of the two-sample $t$-test indicate that sex did not play a significant role in estimates of RFF at offset cycle 10, $t(31) = -0.35$, $p$ = .730; however, *post hoc* examination of trends in RFF at offset cycle 10 (see **Fig. 4.3a**) indicate that RFF was greatest in young adult women ($M$ = -0.75 ST, $SD$ = 1.34 ST), and lowest in older adult women with PD ($M$ = -2.19 ST, $SD$ = 0.97 ST). Young adult men ($M$ = -1.25 ST, $SD$ = 1.26 ST) exhibited lower RFF values than older adult men ($M$ = -1.16 ST, $SD$ = 1.12 ST), but greater RFF values than older adult men with PD ($M$ = -1.68 ST, $SD$ = 1.22 ST).

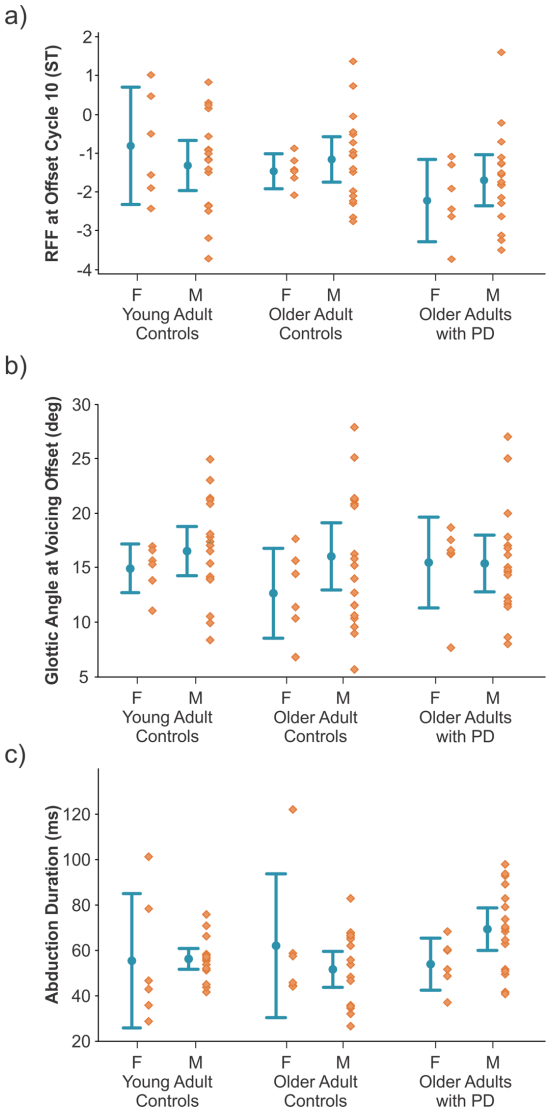


**Figure 4.2. Individual speaker (orange) and mean (blue) values for (a) relative fundamental frequency (RFF) at offset cycle 10, (b) glottic angle at voicing offset, and (c) abduction duration based on speaker sex and group. Error bars show 95% confidence intervals.**

Interestingly, older adult men demonstrated greater RFF values than older adult women (*M* = -1.46 ST, *SD* = 0.41 ST).

Young adult men exhibited the greatest glottic angle at voicing offset (*M* = 16.8 degrees, *SD* = 4.6 degrees), followed by older adult men (*M* = 16.0 degrees, *SD* = 6.2 degrees; see **Fig. 4.2b**). Older adult women showed the smallest glottic angles at voicing offset (*M* = 12.6 degrees, *SD* = 4.0 degrees). Angles were of similar magnitude between young adult women (*M* = 15.1 degrees, *SD* = 2.2 degrees), older adult women with PD (*M* = 15.4 degrees, *SD* = 4.0 degrees), and older adult men with PD (*M* = 15.4 degrees, *SD* = 5.1 degrees). As with RFF values at offset cycle 10, however, sex did not have a significant effect on glottic angle at voicing offset, $t(44) = -1.53$, $p = .134$.

As shown in **Fig. 4.2c**, older adult men with PD (*M* = 69.0 ms, *SD* = 18.3 ms)

demonstrated the longest duration of abduction, which was, on average, 15.1 ms longer than that of older adult women with PD (*M* = 53.9 ms, *SD* = 10.9 ms). Similarly, young adult men (*M* = 57.3 ms, *SD* = 9.0 ms) produced longer abduction durations than young adult women (*M* = 56.5 ms, *SD* = 28.1 ms). In contrast, older adult men (*M* = 51.6 ms, *SD* = 16.0 ms) exhibited a trend for shorter abduction durations than older adult women (*M* = 61.8 ms, *SD* = 30.1 ms). However, sex ultimately did not exhibit a significant effect on abduction duration, *t*(22) = -0.29, *p* = .772.

Effects of Age on Voicing Offset

**Fig. 4.3** describes trends in age across voicing offset measures of RFF at offset cycle 10, abduction duration, and glottic angle at voicing offset. Since sex was not a signficant factor in any of the measures of voicing offset, the effects of age were examined across males and females. Lines of best fit were not calculated for young adult controls due to

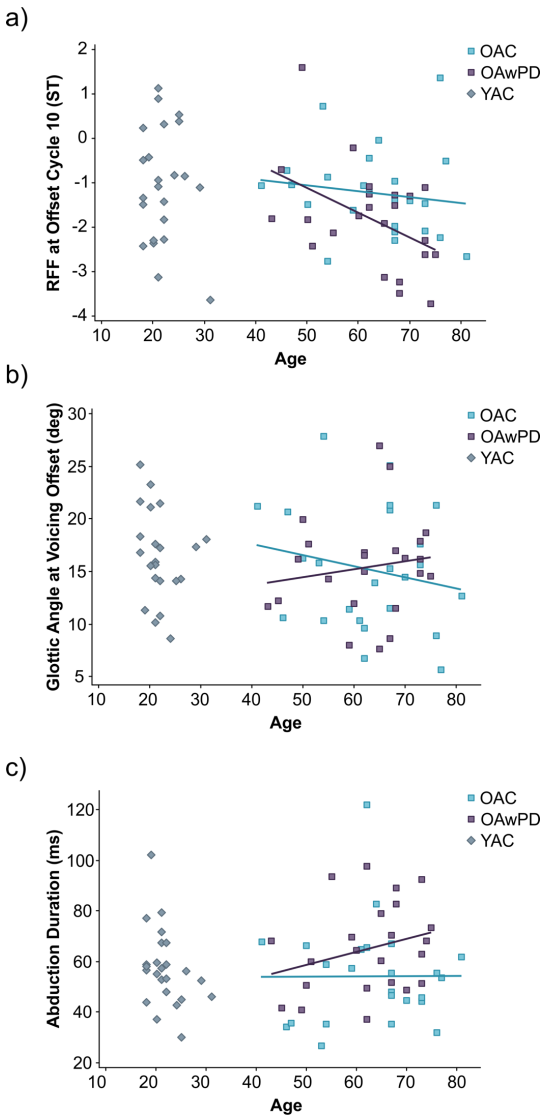


**Figure 4.3. Scatterplot of speaker values for (a) relative fundamental frequency (RFF) at offset cycle 10, (b) glottic angle at voicing offset, and (c) abduction duration based on speaker age and group. Older adult controls (OAC) are shown in light blue, younger adult controls (YAC) are shown in gray, and older adults with Parkinson's disease (OAwPD) are shown in purple. Lines of best fit are shown for OAC and OAwPD groups.**

the small range of ages within the group. The relationship between RFF values at offset cycle 10 and age resulted in a negligible[8] correlation coefficient of $r$ = -.15 ($p$ = .487) for older adult controls and a significant, low correlation coefficient of $r$ = -.46 ($p$ = .026) for older adults with PD (see **Fig. 4.3a**). The relationship between glottic angle at voicing offset and age resulted in negligible correlation coefficients of $r$ = -.20 ($p$ = .362) for older adult controls and $r$ = .15 ($p$ = .495) for older adults with PD. Finally, the relationship between abduction duration and age produced negligible correlation coefficents of $r$ = 0 ($p$ = .984) for older adult controls $r$ = .28 ($p$ = .192) for older adults with PD.

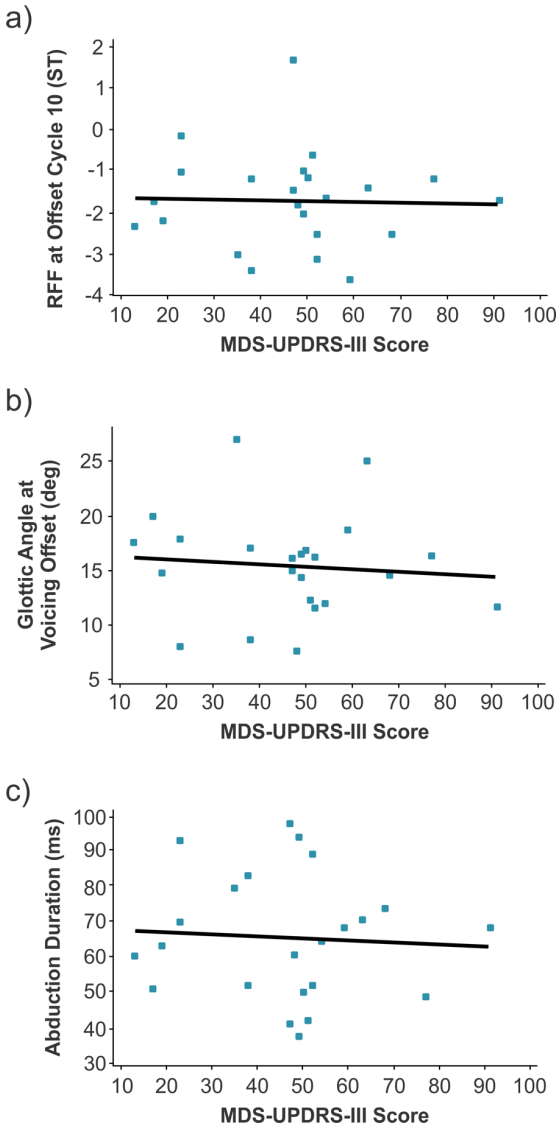


**Figure 4.4. Scatter plot of speaker values for (a) relative fundamental frequency (RFF) at offset cycle 10, (b) glottic angle at voicing offset, and (c) abduction duration relative to score on the Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale: Part III, Motor Examination (MDS-UPDRS-III) scale (for older adults with Parkinson's disease only).**

[8] Hinkle, Wiersma, and Jurs (2003) report that Pearson product-moment correlation coefficients ranging from 0 and .30 reflect a negligible correlation between variables, whereas values between .30 and .50 correspond to a low degree of correlation.

Effects of MDS-UPDRS-III Score on Voicing Offset

**Fig. 4.4** shows the relationship between voicing offset measures and the scores obtained by adults with PD from the motor examination section of the MDS-UPDRS (part III). The relationship between MDS-UPDRS-III score and each voicing offset measure was negligible, with all $r < .1$.

*Relationship between Vocal Fold Abduction and RFF*

**Table 4.3** summarizes the ANCOVA model examining the relationship between speaker age, glottic angle at voicing offset measures, abduction duration, and speaker group on RFF offset values. The model accounted for 26.3% of the variance in the data for RFF at offset cycle 10 (adjusted $R^2 = 20.6\%$). Whereas group and abduction duration did not exhibit a significant effect in the model for RFF at offset cycle 10, glottic angle at voicing offset ($p = .001$, $\eta_p^2 = 0.16$) produced a significant, medium effect on RFF at offset cycle 10 and age produced a significant, small effect ($p = .023$, $\eta_p^2 = 0.08$). The coefficient for glottic angle at voicing offset indicates that RFF at offset cycle 10 decreases by 0.09 ST per one degree increase in glottic angle at voicing offset. Similarly, the coefficient for abduction duration indicates that RFF at offset cycle 10 decreases by 0.04 ST per one year increase in speaker age (the other significant effect).

**Table 4.3. Results of the analysis of covariance model examining the effects of speaker age, abduction duration, glottic angle at voicing offset, and group on RFF offset cycle 10.**

| Effect | *df* | *Coef* | *F* | *p* | $\eta_p^2$ | Effect Size Interpretation |
|---|---|---|---|---|---|---|
| Speaker Age | 1 | -0.04 | 5.43 | .023 | .08 | Small |
| Glottic Angle at Voicing Offset | 1 | -0.09 | 13.08 | .001 | .16 | Medium |
| Abduction Duration | 1 | -9.73 | 1.82 | .182 | .03 | Small |
| Group | 2 | | 2.09 | .131 | .06 | Small |

*Note. Effect size interpretations are based on criteria from Witte et al. (2010).*

## *Discussion*

This study aimed to investigate the relationship between vocal fold abductory kinematics and RFF. Three distinct groups of speakers were enrolled to carry out this analysis: individuals with PD, older controls, and young controls. Simultaneous acoustic recordings and laryngeal images were captured via a microphone and flexible nasendoscope, respectively, as speakers produced the utterance, /ifi/, at their typical pitch and loudness. RFF was extracted from the acoustic signal, whereas glottic angle at voicing offset and abduction duration were computed from the laryngeal images. The relationships among these measures were used to characterize the abductory patterns in the three speaker groups, as well as to determine the contribution of the abductory gesture to measures of RFF.

In examining the role of vocal fold abduction in measures of RFF during intervocalic voicing offsets, it was determined that age ($\eta_p^2$ = .08) and glottic angle at voicing offset ($\eta_p^2$ = .16) were both significant predictors of RFF at offset cycle 10. However, group and abduction duration were not significant predictors in the model. The results of this analysis lend support to the hypothesis that glottic angle at voicing offset would be significantly, negatively related to RFF at offset cycle 10 (wherein larger glottic angles at voicing offset were related to lower RFF values). Yet these findings do not support the hypothesis that the other measure of vocal fold abduction—namely, abduction duration—would be significantly negatively related to RFF at offset cycle 10. As neither measure (glottic angle at voicing offset, abduction duration) significantly differed between young and older adults, the results of the current study do not support

the hypothesized role of vocal fold abduction for devoicing, wherein it was suspected that older adults exert a greater reliance on vocal fold abduction over laryngeal muscle tension for devoicing. Each voicing offset measure (i.e., RFF at voicing offset cycle 10, glottic angle at voicing offset, and abduction duration) is discussed relative to speaker factors of group, age, sex, and MDS-UPDRS-III score in detail below.

### *Physiologically Derived Measures of Vocal Fold Abduction*

Glottic angle at voicing offset and abduction duration were both measured relative to the physiological cessation of vocal fold vibration. It was hypothesized that both measures would be greater in older controls and individuals with PD as compared to young adult controls, and moreover, that these measures would not be statistically significantly different between older controls and individuals with PD. The results of each measure relative to these hypotheses are described in detail below.

#### Glottic Angle at Voicing Offset

Group was not a statistically significant factor in the model for glottic angle at voicing offset. Mean glottic angles at voicing offset were substantially similar across groups, ranging from a mean of 12.6–16.8 degrees. *Post hoc* examinations to understand potential driving factors in this analysis did not result in any noteworthy effects. In particular, the effect of sex on glottic angle was not statistically significant, whereas the relationship between glottic angle at voicing offset and both age and MDS-UPDRS-III score resulted in negligible correlations across group. Not only is it difficult to extract trends from these data because the means are similar across groups, but the range of angle magnitudes is fairly broad. In particular, the mean glottic angle at voicing offset was 15.6

degrees, but ranged from 5.6–27.8 degrees across speakers.

Although not captured in the same linguistic context, studies examining glottic angle magnitudes in the literature indicate similar findings. For instance, Dailey et al. (2005) saw that maximum abduction angles ranged from 31.0 to 77.0 degrees across 21 vocally healthy speakers during an /i-sniff/ task. Similarly, Iwahashi, Ogawa, Hosokawa, Kato, and Inohara (2016) observed an average maximum abduction angle of 59.6 degrees during a vowel phonation task. However, the observed range of maximum angles varied, with a standard deviation of 13.2 degrees. Taken together, these findings suggest that glottic angles are subject to a wide range of within-speaker variability. Although glottic angles at voicing offset may provide some insight into the abductory mechanisms used for devoicing during intervocalic voicing offsets, this measure is not able to distinguish across young adults, older adults, and adults with PD. As such, the results of the current study do not support the hypotheses that older controls exhibit larger glottic angles at voicing offset compared to young adult controls, or that the magnitude of these angles do not significantly differ from those with PD.

### Abduction Duration

The results surrounding abduction duration do not support the proposed hypotheses, as group was not a statistically significant factor in the model for abduction duration. Moreover, the observed trends in abduction duration values across group do not correspond with the proposed hypotheses, with older controls exhibiting shorter mean abduction durations ($M$ = 54.2 ms, $SD$ = 20.1 ms) than individuals with PD ($M$ = 65.1 ms, $SD$ = 17.8 ms) as well as young controls ($M$ = 57.1 ms, $SD$ = 15.2 ms).

The duration values in the current study are substantially higher than those reported in the literature. For instance, Patel et al. (2017) examined the relationship between the offset of the acoustic signal and post-phonatory oscillatory events. The authors found a mean abduction duration of 39.0 ms in females and 40.1 ms in males when comparing the complete cessation of the vocal folds to the time of first incomplete vocal fold closure. Unfortunately, it is difficult to compare the magnitudes obtained here to those obtained by Patel et al. (2017) due to differences in methodology. The authors examined 92 young adults using rigid laryngoscopic techniques, whereas the current study examined 72 adults (24 young controls, 24 older controls, 24 individuals with PD) using flexible laryngoscopy. As rigid laryngoscopy requires the tongue to be restricted during the endoscopic examination, participants were limited to producing repetitions of the syllable /hi/. It is possible that the differences in laryngoscope type (i.e., flexible versus rigid) affected vocal behaviors when producing the speech stimuli. Moreover, the mechanisms necessary to produce a /h/ may differ from those necessary to produce the /f/ of the /ifi/ productions examined in the current study.

Watson, Roark, and Baken (2012) investigated the relationship between acoustic and physiologic voicing offsets by comparing the acoustic signal to electroglottograph recordings of a sustained /ɑ/. The authors implemented electroglottography to track the contact of the vocal folds for comparison against the acoustic signal. The duration of the abductory gesture was identified as 20.0 ms from the time of acoustic voicing offset, which is approximately 15.6 ms lower when compared to older controls and 24.7 ms lower compared to young controls of the current study. The discrepancy in these values

may be the result of differences in methodology. In particular, Watson et al. (2012) examined abduction duration in a linguistically unconstrained context (i.e., sustained /ɑ/ vowel). The production of a sustained /ɑ/ will likely result in different abduction durations, as the /ifi/ productions examined in the current study require produce the intervocalic /f/. Additionally, the instrumentation and processing methods used to capture voicing offset differ between studies. Whereas the current study captured the initiation of abduction directly from images of the vibrating vocal folds (i.e., as the time of last full or maximal contact of the vocal folds), Watson et al. (2012) identified this time point from the cross-correlation of the amplitude of an electroglottograph signal that was band-pass filtered ±40% of the speaker's $f_o$. In combination with inherent differences in the definition of the abduction duration between studies, the electroglottograph signal is also known to suffer from artifacts caused by vertical movements of the larynx, irregular vocal fold vibratory motion, mucous on/around the vocal folds, as well as difficulties in obtaining sufficient waveforms in female speakers (e.g., due to vocal fold mass and amount of adipose tissue; Childers et al., 1990; Colton et al., 1990). Moreover, the termination of the abduction gesture was computed using acoustic voicing offset; however, the current study identified this time point as the cessation of vocal fold vibration from laryngoscopic images. As it is not uncommon for voicing offset to occur earlier in the acoustic signal than the true cessation of vocal fold vibrations (Patel et al., 2017), it is unsurprising that the absolute durations of the abductory gesture differ between studies.

It remains unclear why adults with PD exhibited a substantial (albeit, not

significant) increase in abduction duration compared to young and older controls. One mechanism that could explain this trend is a prolonged abductory gesture to counteract the increased baseline laryngeal muscle tension observed in PD (Zarzur et al., 2013; Zarzur et al., 2007). Adults with PD may not be able to further increase tension from baseline to control the rapid changes in vocal fold vibration at voicing offset. This would suggest that increased laryngeal muscle tension at baseline could reduce the speaker's ability to modulate the vocal folds to promote devoicing. Instead, exerting a greater reliance on vocal fold abduction via carrying out a prolonged abductory gesture may be necessary in order to cease vocal fold vibration during intervocalic voicing offsets. Thus, while these findings do not support the hypothesis that vocally healthy adults use prolonged abductory gesture for devoicing, it is feasible that older adults with PD require such a gesture to effectively cease vocal fold vibration.

### *RFF at Voicing Offset Cycle 10*

In the current study, RFF at voicing offset cycle 10 was originally hypothesized to be significantly associated with group. It was expected that older controls would exhibit smaller values than young adult controls due to an increased reliance on vocal fold abduction to enable devoicing. Furthermore, it was hypothesized that older adults with PD would demonstrate smaller values for RFF at offset cycle 10 due to an increased reliance on vocal fold abduction to enable devoicing as well as increased baseline laryngeal muscle tension.

The results of the current study do not support these hypotheses, as group was not a significant factor in the model for RFF at offset cycle 10. However, mean RFF trends at

offset cycle 10 corresponded to the hypotheses: RFF was greatest in young adult controls (*M* = -1.12 ST), followed by older adult controls (*M* = -1.23 ST), then older adults with PD (*M* = -1.81 ST). Samples from the young adult controls were slightly lower than mean RFF values for offset cycle 10 described in the literature, but were still within the range of RFF values reported in typical speakers for offset cycle 10 (Goberman et al., 2008; Robb et al., 2002; Stepp et al., 2010b; Stepp et al., 2012; Watson, 1998). In contrast, samples from the older adult controls were well within range of those reported in the literature. In older adult controls, mean RFF values for offset cycle 10 have been observed between -1.66 ST and -1.09 ST (Goberman et al., 2008; Stepp, 2013; Watson, 1998), corresponding well with the mean value of -1.23 ST obtained here. Likewise, the mean RFF value at offset cycle 10 observed in older adults with PD (*M* = -1.81 ST, *SD* = 1.17 ST) was similar to those reported in the literature for older adults with PD while on medication (as speakers were in the current study). Goberman et al. (2008) reported an average offset RFF value of approximately -2.20 ST at cycle 10 for older adults with PD while on medication and Stepp (2013) saw a mean offset RFF value of approximately -1.90 ST for older adults with PD while on medication.

The results of the current study suggest that age was a greater indicator of change in vocal fold abductory kinematics than the broad age groupings used to separate young adults from adults who were age- and sex-matched to speakers with PD. In particular, speaker groups (young adult controls, older adult controls, older adults with PD) were constructed to assess general trends in RFF in individuals who were suspected to use a modified abductory gesture to devoice (older adult controls, older adults with PD) as well

as individuals who exhibited increased levels of baseline laryngeal muscle tension (older adults with PD). However, the older adult control group (*SD* = 11.3 years) consisted of vocally healthy individuals as young as 41 years old, whereas the older adults with PD group (*SD* = 9.6 years) included speakers as young as 43 years old. As a result, the range of ages examined across these groups was extremely broad, and was met with an equally variable range of OS: the OS for older adult controls spanned 1.7 to 34.2, whereas that of the older adults with PD ranged from 4.0 to 40.9. Because age was a significant factor in the model for RFF, and because group was not, it is possible that age-specific changes to the laryngeal mechanism, such as presbylarynges (age-related vocal fold atrophy), may be a contributing factor to the differences in vocal fold abductory kinematics that are reflected in RFF. Unfortunately, there is no standardized mechanism for assessing vocal fold atrophy other than through a subjective examination of the vocal mechanism. Moreover, it is unclear whether suspected cases of vocal fold atrophy—including characteristics of vocal fold bowing, spindle-shaped glottal gap, prominent vocal processes, and thinning of the vocal fold mucosa (Angerstein, 2018; Bloch & Behrman, 2001; Isshiki, Shoji, Kojima, & Hirano, 1996; Omori et al., 1997; Pontes, Brasolotto, & Behlau, 2005; Rodeño et al., 1993; Takano et al., 2010)—are the result of morphologic and/or neuromuscular changes. As such, an investigation into objectively characterizing and quantifying the effects of age on the laryngeal mechanism is necessary. Elucidating these effects will provide insight into age-specific changes to laryngeal muscle tension and vocal fold abduction that are reflected in RFF.

*Limitations and Future Directions*

This study analyzed the contributions of vocal fold abduction to measures of RFF during intervocalic voicing offsets. Abductory patterns were characterized in three speaker groups, including young adult controls, older adult controls, and older adults with PD. Yet the identified relationships between vocal fold abduction and RFF may not be generalizable to speakers outside of these groups, as older adults with PD are merely a subset of individuals who exhibit excessive levels of intrinsic laryngeal muscle tension. For instance, speakers with vocal hyperfunction may exhibit excessive or imbalanced laryngeal muscle forces (Hillman et al., 1989). However, the manifestation of vocal hyperfunction is broad, wherein hyperfunctional vocal behaviors may occur in the presence or absence of organic pathology (e.g., vocal nodules), and may be the primary cause of a voice disorder or as a compensatory adaptation to glottal insufficiency. It is therefore unclear whether speakers that exhibit signs of vocal hyperfunction would demonstrate similar trends in vocal fold abductory kinematics as the older adults with PD examined here. Although this work attempts to elucidate the contribution of vocal fold abduction to RFF, future work should aim to expand upon the patterns described here. Characterizing abductory kinematics in speakers with other voice disorders associated with excessive tension may be a useful step toward isolating the differential contributions of tension and abduction in intervocalic voicing offsets.

Hearing screening data were collected to confirm that participants were able to hear experimenter instructions during the nasendoscopic procedures. However, these data were not collected for 5/72 participants (all young adult controls who reported no history

of hearing disorders), and moreover, not all participants passed at the standard hearing thresholds reported in the literature (see American Speech-Language-Hearing Association, 2005; Schow, 1991). As the goal of the current study was to directly examine the relationship between vocal fold abductory kinematics and RFF, hearing ability was not assessed. Although outside the scope of the current work, it is important to consider that hearing ability could be a contributing factor to the physiological mechanisms used for devoicing. Future work should therefore aim to comprehensively examine the effects of hearing ability on the known devoicing mechanisms of laryngeal muscle tension and vocal fold abduction, as well as on acoustic estimates of RFF.

Upon examining the relationship between speaker group and measures of voicing offset, it was determined that speaker group did not exhibit significant effects on RFF at offset cycle 10, glottic angle at voicing offset, or abduction duration. The results showed only small effect sizes ($\eta_p^2$ = .07, $\eta_p^2$ = .06, and $\eta_p^2$ = .02, respectively) between the voicing offset measures and speaker group, yielding non-significant $p$ values at the < .05 level. Based on these findings, a cohort of 447 participants (159 per speaker group) would be needed to report a significant small effect ($\eta_p^2$ = .02) with $p$ < .05 (G*Power v.3.1.9.2; Faul, Erdfelder, Buchner, & Lang, 2009). The results of this power analysis suggest that there may be differences in RFF at offset cycle 10, glottic angle at voicing offset, or abduction duration across speaker group; however, any significant differences in the voicing offset measures would be extremely small.

Prior work indicates that syllable stress is a contributing factor to measures of RFF (Park & Stepp, 2019); however, this variable was not controlled for in the current

study. Stressed syllables are generally produced using more vocal effort than unstressed syllables (Eriksson & Traunmüller, 2002), wherein unequal stress may substantially alter laryngeal muscle tension during recording. The current study instructed speakers to produce the maximum number of /ifi/ tokens possible within the limitations of the recording setup, resulting in an average of 7.8 (*SD* = 4.0, *range* = 2–19) /ifi/ tokens per speaker. Although speakers were trained to produce /ifi/ tokens with equal stress, syllable stress was not precisely monitored during the endoscopic procedure. Speaker RFF values for offset cycle 10 were then computed by averaging across the values obtained for individual productions in order to produce a more reliable estimate of RFF. Yet it is still possible that introducing first-syllable stress (i.e., /<u>i</u>fi/, with syllable stress denoted by the underline) altered laryngeal muscle tension during recording to, in turn, affect the $f_o$ contours of the captured /ifi/ productions. As such, future work should take care to instruct speakers to produce equal stress on the stimuli used to calculate RFF.

Finally, the recordings collected in the current study were acquired under flexible laryngoscopy. A numbing agent was not used during the scoping procedure so as not to affect laryngeal function (Dworkin et al., 2000b), yet the insertion of the flexible nasendoscope may have caused speakers to deviate from their typical vocal function by inducing stress and tension during recordings. For instance, a study by Hay, Oates, Giannini, Berkowitz, and Rotenberg (2009) described observable increases in general muscle tension (e.g., eyes shut tight, body stiffness, clenched jaw) as children underwent a flexible nasendoscopic procedure. If the insertion of the flexible nasendoscope were to increase laryngeal muscle tension during the recording procedure, then resulting RFF

values at offset cycle 10 would likely be lower than those typically reported. However, the values obtained across young adult controls, older adult controls, and older adults with PD were all well within the range of those reported in the literature. Thus, although possible, it is unlikely that the laryngoscopic procedure substantially affected the recordings collected in this study.

### *Conclusions*

Despite reasonable conjectures of a prolonged abductory gesture in older adults, vocal fold abductory patterns were not significantly different between young adult controls, older adult controls, and older adults with PD. However, measures of RFF at voicing offset were found to be related to speaker age and the glottic angle at voicing offset. In addition to corroborating abductory behaviors as a potential mechanism of RFF, these results further indicate that speaker age is an important factor to consider in the assessment of RFF. The findings from this study provide a framework for future investigations aimed at understanding the relationship between vocal fold abduction measures of RFF in disordered voices.

# CHAPTER 5. DISCUSSION

This work sought to improve the semi-automated relative fundamental frequency (RFF) algorithm for the objective quantification of laryngeal muscle tension, and then use the refined algorithm to determine the role of vocal fold abductory kinematics in estimates of RFF. The first study (*Chapter 2*) evaluated the effects of voice sample characteristics and $f_o$ estimation method on the correspondence between semi-automated and manual RFF estimates. The second study (*Chapter 3*) examined the relationship between acoustic features and vocal fold vibratory characteristics during voicing offset and onsets. The third study (*Chapter 4*) investigated the relationship between vocal fold abductory kinematics and RFF. The purpose of this final chapter is to link these studies together and provide recommendations to improve the future clinical applicability of RFF.

## *The Role of Laryngeal Muscle Tension in Voice Disorders*

Laryngeal muscle tension is a crucial component of voice production. The intrinsic laryngeal muscles are used to tense, abduct, and adduct the vocal folds for phonation, whereas the extrinsic laryngeal muscles stabilize, raise, and lower the larynx during speech and swallowing movements. Excessive or unbalanced laryngeal muscle forces may arise from overuse and/or misuse of the laryngeal mechanism in the absence of organic pathology (e.g., yelling), pathological changes to the vocal fold tissues (e.g., nodules), and neurological disorders affecting the laryngeal mechanism (e.g., Parkinson's disease; Boone et al., 2014; Ghassemi et al., 2014; Hillman et al., 1989). The development of these pathologies may cause functional changes in the tension of the

laryngeal mechanism in compensation for additional effort associated with achieving phonation.

It has been estimated that approximately 65% of individuals with voice disorders exhibit excessive laryngeal muscle tension (Ramig et al., 1998). Despite this prevalence, there currently exists no single, objective measure that is able to quantify the degree of tension present in the laryngeal mechanism. Clinical assessments typically include non-instrumental methods, such as acquiring a case history, patient-reported outcomes, auditory-perceptual judgments of voice, as well as carrying out manual palpations of the (para)laryngeal musculature (Morrison et al., 1986; Roy et al., 2013; Schwartz et al., 2009). Instrumental methods, on the other hand, comprise laryngeal visualization (e.g., flexible nasendoscopy), electroglottography, electromyography, accelerometry, and/or acoustic signal analysis techniques. Despite the availability of techniques for assessing extrinsic laryngeal muscle tension, many of these methods fall short in terms of validity, reliability, and/or specificity. As such, clinical assessments of laryngeal muscle tension remain heavily based on unreliable auditory-perceptual impressions of voice quality and manual palpations of the extrinsic laryngeal muscles. (Dejonckere et al., 2001; Maryn & Weenink, 2015; Stepp et al., 2011a)

### *Relative Fundamental Frequency as an Estimator of Laryngeal Muscle Tension*

Relative fundamental frequency (RFF) has received attention as a promising acoustic estimate for assessing and tracking the degree of baseline laryngeal muscle tension. RFF reflects short-term changes in instantaneous fundamental frequency ($f_o$) of voice during a vowel–voiced consonant–vowel (VCV) production. Voice $f_o$ relates to the

vibratory rate of the vocal folds, which is, in turn, associated with vocal fold length, mass, and tension (Van Den Berg, 1958). During a VCV production, the vocal folds vibrate to produce the vowel. However, the vocal folds must cease vibrating to transition into the voiceless consonant (voicing offset), then reinitiate vibration to transition into the second vowel (voicing onset). As the vocal folds stop and start vibrating, RFF is able to capture the changes in vibratory rate (i.e., $f_o$) that must occur to successfully produce the intended utterance. RFF is specifically calculated from the ten vocal fold vibratory cycles immediately before and after the voiceless consonant. The $f_o$ values of each of these cycles are normalized to a steady-state $f_o$ value obtained from the adjacent vowel. In doing so, RFF values may be compared within and across speakers.

Currently, the gold-standard method of calculating RFF is through tedious manual estimation techniques. A trained technician must visualize the acoustic waveform to extract the twenty vocal cycles of interest, wherein the majority of time is spent using trial and error to identify the vocal cycle closest to the voiceless consonant. A single reliable RFF estimate typically requires 20–40 minutes of time via manual estimation (Eadie et al., 2013; Lien et al., 2017).

### *Semi-automated RFF Estimation*

A semi-automated RFF algorithm ("aRFF") was developed to combat the time- and training-intensive nature of manual RFF estimation. The aRFF algorithm carries out rule-based signal processes techniques to identify the vocal cycles closest to the voiceless consonant in a VCV production, then uses these cycles to estimate RFF. This aRFF algorithm is advantageous over manual estimation since technicians do not need to be

extensively trained to use it, and moreover, the algorithm minimizes the need for manual intervention. Yet this method of calculating RFF is limited in that the accuracy of semi-automated RFF values varies across a wide range of voice signals (Lien et al., 2017). Thus, the current work sought to improve the accuracy and precision of semi-automated RFF estimates across a broad spectrum of vocal function. Detailed in *Chapter 2*, a new version of the semi-automated RFF algorithm ("aRFF-AP") was developed by implementing a new method of $f_o$ estimation as well as accounting for differences in voice sample characteristics (e.g., overall severity of dysphonia, signal acquisition quality) based on the acoustic measure, pitch strength (Camacho, 2012; Camacho et al., 2008; Kopf et al., 2017; Shrivastav et al., 2012). Algorithmic performance was compared between the aRFF and aRFF-AP algorithms, ultimately showing that errors related to the accuracy and precision of the algorithms (with respect to manual estimation) were reduced by 88.4% and 17.3%, respectively.

The development of the aRFF-AP algorithm was a crucial step toward applying RFF in the clinic. However, the results of this analysis elicited non-zero errors, suggesting that pitch strength categories alone were not sufficient to account for variations in voice samples and/or that manual estimation is not a true gold standard. The source of these potential issues likely involve the identification of the vocal cycle closest to the voiceless consonant: whereas manual estimation requires a trained technician to use trial and error to identify this boundary, the semi-automated RFF algorithm uses acoustic features. The process of manual RFF estimation is subjective in nature, such that the selected boundary between voiced and voiceless segments may not precisely

correspond to the true initiation or termination of voicing. The aRFF-AP algorithm, on the other hand, uses rule-based signal processing techniques to locate this boundary. Specifically, acoustic features are examined in time, and the pitch strength-tuned thresholds are used to assist the algorithm in choosing a potential location of the boundary between voiced and voiceless segments. Although these pitch strength-tuned thresholds were shown to improve the accuracy and precision of the algorithm relative to manual RFF estimation, it is possible that errors continue to occur because the acoustic features used to identify this boundary did not correspond with true initiation or termination of voicing. *Chapter 3* details the steps taken to examine these possibilities.

By examining simultaneous recordings made using a microphone and flexible nasendoscope, the initiation and termination of vocal fold vibration were related to acoustic features extracted from the acoustic signal. A new set of acoustic features that coincided well with voicing transitions were introduced into the semi-automated RFF algorithm, leading to an increased correspondence between the algorithmic and physiological boundary cycles. Not only was the algorithm developed from this work—"aRFF-APH"—significantly more likely to identify the true, physiological boundary cycle than aRFF-AP, but both versions of the semi-automated RFF algorithm were significantly more likely to identify this true boundary than manual estimation techniques. This progress may be due, in part, to the ability of the algorithms to leverage a filtered version of the microphone signal that amplifies the contribution of the speaker's $f_o$ and attenuates extraneous noise (e.g., from coarticulation due to concurrent aspiration and frication during the voiceless consonant).

Despite these promising results for the aRFF-APH algorithm, the aRFF-AP algorithm remains the gold-standard method for semi-automatically estimating RFF. This is because the overarching goals in refining the RFF algorithm were dissimilar between the two algorithm versions. As noted in *Chapter 2*, the aRFF-AP algorithm was developed to increase the *clinical applicability* of RFF, as the voice signals recorded in a clinic would presumably vary in terms of the severity of the speaker's dysphonia as well as the recording conditions when acquiring the voice signal. The acoustic measure, pitch strength, was therefore used to account for these variable sample characteristics by creating pitch strength-tuned categories for calculating RFF. This was possible since the speaker database used in *Chapter 2* incorporated a wide range of vocal function from 483 independent speakers: Not only were speakers recorded in a series of environments (waiting room, quiet room, sound-attenuated booth), but over 20 different primary voice complaints were included within the database. The resulting overall severity of dysphonia of the 483 speakers ranged from 0 to 100 in the study. Clear trends in pitch strength were identified when examining the acoustic features obtained from the microphone signal over time, such that algorithmic parameters could be tuned to estimate RFF based on a speaker's pitch strength.

On the other hand, the overarching goal of the work described in *Chapter 3* was to elucidate the *physiological relevance* of RFF. As previously mentioned, there is no objective, quantitative measure that is singlehandedly able to assess and track laryngeal muscle tension. Although RFF has been considered as a potential acoustic estimate of baseline laryngeal muscle tension, the relationship between RFF and vocal fold vibratory

offsets and onsets has not yet been characterized. This is because the primary means of calculating RFF is by using the microphone signal, which indirectly captures the glottal source. The speech acquired using a microphone may be affected by supraglottic resonance, effects of radiation from the lips, as well as boundary cycle masking due to coarticulation (Cheyne et al., 2003). It may not only be difficult to identify the boundary cycle because of these issues, but it is unclear if the selected boundary cycle truly represents the physiological termination or initiation of vocal fold vibration. The work described in *Chapter 3* thus sought to use these true time points to tune the current RFF algorithm to improve the precision with which the algorithm identifies voicing offsets and onsets. In doing so, an RFF algorithm tuned to vocal fold vibratory characteristics could then be used clinically to non-invasively, objectively, and quantitatively assess and track changes in laryngeal muscle tension. Future work should therefore validate the aRFF-APH algorithm using independent training and test sets constructed from a larger set of speakers across a broad range of vocal function.

### *The Relationship between Vocal Fold Abduction and RFF*

The steps taken in *Chapters 2* and *3* sought to improve the semi-automated RFF algorithm for use as a clinical estimation of baseline laryngeal muscle tension. However, there is evidence to suggest that RFF does not *only* reflect the degree of baseline laryngeal muscle tension. The characteristic pattern of RFF during voicing offset has been attributed to the interplay of laryngeal muscle tension and vocal fold kinematics (Löfqvist et al., 1989; Stepp et al., 2011d; Stevens, 1977; Watson, 1998). In particular, voicing offsets in young adults are characterized by a stable or slightly decreasing trend

in RFF (Goberman et al., 2008; Robb et al., 2002). These values may be due, in part, to transient increases in laryngeal muscle tension before, during, and after the production of the voiceless consonant to cease vocal fold vibration (Löfqvist et al., 1989; Stevens, 1977). It is also suspected that vocal fold abductory kinematics act in concert with elevated muscle tension to achieve devoicing during voicing offsets (Watson, 1998). However, the specific contribution of the abductory gesture is unclear.

Limitations in our understanding of the role of vocal fold abduction during intervocalic voicing offsets has led to difficulties in interpreting resulting RFF values. Specifically, older adults typically exhibit lower RFF values than young adults, which would suggest a difference in the mechanisms used for devoicing. Thus far, the dissimilarities in RFF values have been attributed to a prolonged abductory gesture for devoicing (Watson, 1998).Yet older adults with Parkinson's disease (PD)—a progressive, neurodegenerative disease with symptoms of increased intrinsic laryngeal muscle tension (Zarzur et al., 2013; Zarzur et al., 2007)—typically exhibit even lower RFF values than older adult controls. It is therefore unclear whether the observed discrepancies in RFF values between young and older adults is the result of a prolonged abductory gesture for devoicing, and further, whether the observed lower offset RFF values in PD reflect an increased reliance on abduction to cease voicing, increased levels of baseline laryngeal tension that arise with PD, or some other cause (e.g., laryngeal height).

Detailed in *Chapter 4*, it was found that vocal fold abductory patterns were not significantly different across young adult controls, older adult controls, and older adults with PD. These findings do not support the use of a prolonged abductory gesture by older

adults for devoicing. However, speaker age and glottic angle prior to the termination of vocal fold vibration were significantly related to RFF estimates obtained at acoustic voicing offset. The results of this study indicate that RFF is, as hypothesized, related to vocal fold abduction during voicing offsets. However, the identified relationships between vocal fold abduction and RFF must be extended to speakers outside of these groups, as older adults with PD are merely a subset of individuals who exhibit excessive levels of intrinsic laryngeal muscle tension. As such, future work should aim to expand upon the patterns described here by characterizing abductory kinematics in speakers with other voice disorders associated with excessive laryngeal muscle tension.

### *Conclusions*

The work detailed in this dissertation sought to (1) improve the accuracy and precision of the RFF algorithms for the objective quantification of laryngeal muscle tension, and (2) use the refined algorithm to determine the role of vocal fold abductory kinematics in estimates of RFF. Refining the method of $f_o$ estimation and accounting for variations in voice sample characteristics within the semi-automated RFF algorithm led to improved accuracy and precision relative gold-standard, manual RFF estimates. Incorporating acoustic features that captured the physiological termination and initiation of vocal fold vibration led to additional improvements in algorithmic accuracy. In examining the contribution of the vocal fold abduction to estimates of RFF, it was determined that abductory patterns play a significant role in resulting RFF measures at voicing offset. Taken together, these studies improved the clinical applicability of using RFF in conjunction with current clinical voice assessment techniques for assessing the

degree of baseline laryngeal muscle tension.

## BIBLIOGRAPHY

Abur, D., Lester-Smith, R.A., Daliri, A., Lupiani, A.A., Guenther, F.H., & Stepp, C.E. (2018). Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease. *PLoS One, 13*(1), e0191839.

Agarwal, M., Scherer, R.C., & Hollien, H. (2003). The false vocal folds: Shape and size in frontal view during phonation based on laminagraphic tracings. *Journal of Voice, 17*(2), 97-113.

Alharbi, G.G., Cannito, M.P., Buder, E.H., & Awan, S.N. (2019). Spectral/cepstral analyses of phonation in Parkinson's disease before and after voice treatment: A preliminary study. *Folia Phoniatrica et Logopaedica, 71*(5-6), 275-285.

Altman, K.W., Atkinson, C., & Lazarus, C. (2005). Current and emerging concepts in muscle tension dysphonia: A 30-month review. *Journal of Voice, 19*(2), 261-267.

American Speech-Language-Hearing Association. (2005). Guidelines for manual pure-tone threshold audiometry. Rockville, MD.

Anand, S., Kopf, L.M., Shrivastav, R., & Eddins, D.A. (2019a). Objective indices of perceived vocal strain. *Journal of Voice, 33*(6), 838-845.

Anand, S., Kopf, L.M., Shrivastav, R., & Eddins, D.A. (2019b). Using pitch height and pitch strength to characterize type 1, 2, and 3 voice signals. *Journal of Voice*, Advanced online publication. https://doi.org/10.1016/j.jvoice.2019.08.006

Anastasopoulos, D., Maurer, C., Nasios, G., & Mergner, T. (2009). Neck rigidity in Parkinson's disease patients is related to incomplete suppression of reflexive head stabilization. *Experimental Neurology, 217*(2), 336-346.

Andaloro, C., & La Mantia, I. (2019). Anatomy, Head and Neck, Larynx: Arytenoid Cartilage. In StatPearls [Internet]. Treasure Island, FL, USA: StatPearls Publishing. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK513252.

Angerstein, W. (2018). Vocal changes and laryngeal modifications in the elderly (presbyphonia and presbylarynx). *Laryngo-Rhino-Otologie, 97*(11), 772-776.

Angsuwarangsee, T., & Morrison, M. (2002). Extrinsic laryngeal muscular tension in patients with voice disorders. *Journal of Voice, 16*(3), 333-343.

Arnold, G.E., & Pinto, S. (1960). Ventricular dysphonia: New interpretation of an old observation. *The Laryngoscope, 70*, 1608-1627.

Aronson, A.E. (1990). *Clinical Voice Disorders: An Interdisciplinary Approach* (3 ed.). New York, NY, USA: Thieme Medical Publishers.

Aronson, A.E., & Bless, D.M. (2009). *Clinical Voice Disorders* (4 ed.). New York, NY, USA: Thieme Medical Publishers.

Askenfelt, A.G., & Hammarberg, B. (1986). Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures. *Journal of Speech, Language, and Hearing Research, 29*(1), 50-64.

Awan, S.N., Giovinco, A., & Owens, J. (2012). Effects of vocal intensity and vowel type on cepstral analysis of voice. *Journal of Voice, 26*(5), 670.e15-20.

Awan, S.N., & Roy, N. (2005). Acoustic prediction of voice type in women with functional dysphonia. *Journal of Voice, 19*(2), 268-282.

Awan, S.N., & Roy, N. (2006). Toward the development of an objective index of dysphonia severity: A four-factor acoustic model. *Clinical Linguistics & Phonetics, 20*(1), 35-49.

Awan, S.N., & Roy, N. (2009). Outcomes measurement in voice disorders: Application of an acoustic index of dysphonia severity. *Journal of Speech, Language, and Hearing Research, 52*(2), 482-499.

Awan, S.N., Roy, N., & Dromey, C. (2009). Estimating dysphonia severity in continuous speech: Application of a multi-parameter spectral/cepstral model. *Clinical Linguistics & Phonetics, 23*(11), 825-841.

Awan, S.N., Roy, N., Jette, M.E., Meltzner, G.S., & Hillman, R.E. (2010). Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clinical Linguistics & Phonetics, 24*(9), 742-758.

Azarov, E., Vashkevich, M., & Petrovsky, A. (2016, March). Instantaneous pitch estimation algorithm based on multirate sampling. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing,* (*ICASSP-2016)*, pp. 4970-74. Shanghai, China. https://doi.org/10.1109/ICASSP.2016.7472623

Azevedo, L.L., Cardoso, F., & Reis, C. (2003). Acoustic analysis of prosody in females with Parkinson's disease: Effect of L-dopa. *Arquivos de Neuro-Psiquiatria, 61*(4), 995-998.

Bainbridge, K.E., Roy, N., Losonczy, K.G., Hoffman, H.J., & Cohen, S.M. (2017). Voice disorders and associated risk markers among young adults in the united states. *The Laryngoscope, 127*(9), 2093-2099.

Baken, R.J., & Orlikoff, R.F. (2000). *Clinical Measurement of Speech and Voice*. San Diego, CA, USA: Singular Thomson Learning.

Baker, K.K., Ramig, L.O., Sapir, S., Luschei, E.S., & Smith, M.E. (2001). Control of vocal loudness in young and old adults. *Journal of Speech, Language, and Hearing Research, 44*(2), 297-305.

Baldner, E.B., Doll, E., & van Mersbergen, M.R. (2015). A review of measures of vocal effort with a preliminary study on the establishment of a vocal effort measure. *Journal of Voice, 29*(5), 530-541.

Bard, M.C., Slavit, D.H., McCaffrey, T.V., & Lipton, R.J. (1992). Noninvasive technique for estimating subglottic pressure and laryngeal efficiency. *Annals of Otology, Rhinology, and Laryngology, 101*(7), 578-582.

Bassich, C.J., & Ludlow, C.L. (1986). The use of perceptual methods by new clinicians for assessing voice quality. *Journal of Speech and Hearing Disorders, 51*(2), 125-133.

Behrman, A. (2005). Common practices of voice therapists in the evaluation of patients. *Journal of Voice, 19*(3), 454-469.

Behrman, A., Dahl, L.D., Abramson, A.L., & Schutte, H.K. (2003). Anterior-posterior and medial compression of the supraglottis: Signs of nonorganic dysphoniaor normal postures? *Journal of Voice, 17*(3), 403-410.

Belsky, M.A., Rothenberger, S.D., Gillespie, A.I., & Gartner-Schmidt, J.L. (2020). Do phonatory aerodynamic and acoustic measures in connected speech differ between vocally healthy adults and patients diagnosed with muscle tension dysphonia? *Journal of Voice*, Advanced online publication.

Berardelli, A., Sabra, A.F., & Hallett, M. (1983). Physiological mechanisms of rigidity in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry, 46*(1), 45–53.

Berardelli, I., Bloise, M.C., Bologna, M., Conte, A., Pompili, M., Lamis, D.A., Pasquini, M., & Fabbrini, G. (2018). Cognitive behavioral group therapy versus psychoeducational intervention in Parkinson's disease. *Neuropsychiatric Disease and Treatment, 14*, 399-405.

Berke, G.S., & Gerratt, B.R. (1993). Laryngeal biomechanics: An overview of mucosal wave mechanics. *Journal of Voice, 7*(2), 123-128.

Berry, D.A., Montequin, D.W., & Tayama, N. (2001). High-speed digital imaging of the medial surface of the vocal folds. *Journal of the Acoustical Society of America, 110*(5), 2539-2547.

Bhattacharyya, N. (2014). The prevalence of voice problems among adults in the united states. *The Laryngoscope, 124*(10), 2359-2362.

Bjorklund, S., & Sundberg, J. (2016). Relationship between subglottal pressure and sound pressure level in untrained voices. *Journal of Voice, 30*(1), 15-20.

Bloch, I., & Behrman, A. (2001). Quantitative analysis of videostroboscopic images in presbylarynges. *The Laryngoscope, 111*(11), 2022-2027.

Blumin, J.H., Pcolinsky, D.E., & Atkins, J.P. (2004). Laryngeal findings in advanced Parkinson's disease. *Annals of Otology, Rhinology, and Laryngology, 113*(4), 253-258.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9-10), 341-345.

Boone, D.R., McFarlane, S.C., Von Berg, S.L., & Zraick, R.I. (2014). *The Voice and Voice Therapy* (9 Ed.). Boston, MA, USA: Pearson.

Borg, G.A. (1982). Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise, 14*(5), 377-381.

Borrie, S.A., & Delfino, C.R. (2017). Conversational entrainment of vocal fry in young adult female american english speakers. *Journal of Voice, 31*(4), 513.e25-32.

Bottalico, P., Graetzer, S., & Hunter, E.J. (2016). Effects of speech style, room acoustics, and vocal fatigue on vocal effort. *Journal of the Acoustical Society of America, 139*(5), 2870-2870.

Boutsen, F., Park, E., Dvorak, J., & Cid, C. (2018). Prosodic improvement in persons with Parkinson disease receiving SPEAK OUT!® voice therapy. *Folia Phoniatrica et Logopaedica, 70*(2), 51-58.

Bowen, L.K., Hands, G.L., Pradhan, S., & Stepp, C.E. (2013). Effects of Parkinson's disease on fundamental frequency variability in running speech. *Journal of Medical Speech-Language Pathology, 21*(3), 235-244.

Braak, H., Del Tredici, K., Rub, U., de Vos, R.A., Jansen Steur, E.N., & Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging, 24*(2), 197-211.

Brandt, J.F., Ruder, K.F., & Shipp, T. (1969). Vocal loudness and effort in continuous speech. *Journal of the Acoustical Society of America, 46*(6B), 1543-1548.

Braunschweig, T., Flaschka, J., Schelhorn-Neise, P., & Döllinger, M. (2008). High-speed video analysis of the phonation onset, with an application to the diagnosis of functional dysphonias. *Medical Engineering & Physics, 30*(1), 59-66.

Broadfoot, C.K., Abur, D., Hoffmeister, J.D., Stepp, C.E., & Ciucci, M.R. (2019). Research-based updates in swallowing and communication dysfunction in Parkinson disease: Implications for evaluation and management. *Perspectives of the ASHA Special Interest Groups, 4*(5), 825-841.

Brockmann-Bauser, M., Bohlender, J.E., & Mehta, D.D. (2018). Acoustic perturbation measures improve with increasing vocal intensity in individuals with and without voice disorders. *Journal of Voice, 32*(2), 162-168.

Brockmann-Bauser, M., & Drinnan, M.J. (2011). Routine acoustic voice analysis: Time to think again? *Current Opinion in Otolaryngology & Head and Neck Surgery, 19*(3), 165-170.

Brockmann, M., Drinnan, M.J., Storck, C., & Carding, P.N. (2011). Reliable jitter and shimmer measurements in voice clinics: The relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task. *Journal of Voice, 25*(1), 44-53.

Broniatowski, M., Sonies, B.C., Rubin, J.S., Bradshaw, C.R., Spiegel, J.R., Bastian, R.W., & Kelly, J.H. (1999). Current evaluation and treatment of patients with swallowing disorders. *Otolaryngology–Head and Neck Surgery, 120*(4), 464-473.

Burk, M.H., & Wiley, T.L. (2004). Continuous versus pulsed tones in audiometry. *American Journal of Audiology, 13*(1), 54-61.

Cahill, L.M., Murdoch, B.E., Theodoros, D.G., Triggs, E.J., Charles, B.G., & Yao, A.A. (1998). Effect of oral levodopa treatment on articulatory function in Parkinson's disease: Preliminary results. *Motor Control, 2*(2), 161-172.

Camacho, A. (2007). *SWIPE: A sawtooth waveform inspired pitch estimator for speech and music.* (Doctoral dissertation, University of Florida, Gainesville, FL, USA), Retrieved from https://www.cise.ufl.edu/~acamacho/publications/dissertation.pdf

Camacho, A. (2012, Jul). On the use of auditory models' elements to enhance a sawtooth waveform inspired pitch estimator on telephone-quality signals. In *2012 11th International Conference on Information Science, Signal Processing and their Applications, (ISSPA-2012)*, pp. 1080-85. Montreal, QC, Canada. https://doi.org/10.1109/ISSPA.2012.6310450

Camacho, A., & Harris, J.G. (2008). A sawtooth waveform inspired pitch estimator for speech and music. *Journal of the Acoustical Society of America, 124*(3), 1638-1652.

Cannito, M.P., Suiter, D.M., Beverly, D., Chorna, L., Wolf, T., & Pfeiffer, R.M. (2012). Sentence intelligibility before and after voice treatment in speakers with idiopathic Parkinson's disease. *Journal of Voice, 26*(2), 214-219.

Cantello, R., Gianelli, M., Bettucci, D., Civardi, C., De Angelis, M.S., & Mutani, R. (1991). Parkinson's disease rigidity: Magnetic motor evoked potentials in a small hand muscle. *Neurology, 41*(9), 1449-1456.

Cantello, R., Gianelli, M., Civardi, C., & Mutani, R. (1995). Parkinson's disease rigidity: EMG in a small hand muscle at "rest". *Electroencephalography and Clinical Neurophysiology, 97*(5), 215-222.

Canter, G.J. (1965). Speech characteristics of patients with Parkinson's disease: II. Physiological support for speech. *Journal of Speech and Hearing Disorders, 30*(1), 44-49.

Carding, P.N., Horsley, I.A., & Docherty, G.J. (1999). A study of the effectiveness of voice therapy in the treatment of 45 patients with nonorganic dysphonia. *Journal of Voice, 13*(1), 72-104.

Carding, P.N., Steen, I.N., Webb, A., MacKenzie, K., Deary, I.J., & Wilson, J.A. (2004). The reliability and sensitivity to change of acoustic measures of voice quality. *Clinical Otolaryngology and Allied Sciences, 29*(5), 538-544.

Castillo-Guerra, E., & Ruiz, A. (2009). Automatic modeling of acoustic perception of breathiness in pathological voices. *IEEE Transactions on Biomedical Engineering, 56*(4), 932-940.

Cheyne, H.A., Hanson, H.M., Genereux, R.P., Stevens, K.N., & Hillman, R.E. (2003). Development and testing of a portable vocal accumulator. *Journal of Speech, Language, and Hearing Research, 46*(6), 1457-1467.

Chhetri, D.K., & Neubauer, J. (2015). Differential roles for the thyroarytenoid and lateral cricoarytenoid muscles in phonation. *The Laryngoscope, 125*(12), 2772-2777.

Chhetri, D.K., Neubauer, J., & Berry, D.A. (2012). Neuromuscular control of fundamental frequency and glottal posture at phonation onset. *Journal of the Acoustical Society of America, 131*(2), 1401-1412.

Chhetri, D.K., Neubauer, J., Sofer, E., & Berry, D.A. (2014). Influence and interactions of laryngeal adductors and cricothyroid muscles on fundamental frequency and

glottal posture control. *Journal of the Acoustical Society of America, 135*(4), 2052-2064.

Childers, D.G., Hicks, D.M., Moore, G.P., & Alsaka, Y.A. (1986). A model for vocal fold vibratory motion, contact area, and the electroglottogram. *Journal of the Acoustical Society of America, 80*(5), 1309-1320.

Childers, D.G., Hicks, D.M., Moore, G.P., Eskenazi, L., & Lalwani, A.L. (1990). Electroglottography and vocal fold physiology. *Journal of Speech, Language, and Hearing Research, 33*(2), 245-254.

Choi, H.S., Berke, G.S., Ye, M., & Kreiman, J. (1993a). Function of the posterior cricoarytenoid muscle in phonation: In vivo laryngeal model. *Otolaryngology–Head and Neck Surgery, 109*(6), 1043-1051.

Choi, H.S., Berke, G.S., Ye, M., & Kreiman, J. (1993b). Function of the thyroarytenoid muscle in a canine laryngeal model. *Annals of Otology, Rhinology, and Laryngology, 102*(10), 769-776.

Choi, H.S., Ye, M., & Berke, G.S. (1995). Function of the interarytenoid (IA) muscle in phonation: In vivo laryngeal model. *Yonsei Medical Journal, 36*(1), 58-67.

Chu, Y., & Kordower, J.H. (2007). Age-associated increases of α-synuclein in monkeys and humans are associated with nigrostriatal dopamine depletion: Is this the target for Parkinson's disease? *Neurobiology of Disease, 25*(1), 134-149.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ, USA: Erlbaum.

Coleman, L., Zakowski, M., Gold, J.A., & Ramanathan, S. (2013). Functional Anatomy of the Airway. In C.A. Hagberg (Ed.), *Benumof and Hagberg's Airway Management* (3 ed., pp. 3-20). Philadelphia, PA, USA: W.B. Saunders.

Coleman, R.F. (1988). Comparison of microphone and neck-mounted accelerometer monitoring of the performing voice. *Journal of Voice, 2*(3), 200-205.

Colton, R.H., Casper, J.K., & Leonard, R.J. (2011). *Understanding Voice Problem: A Physiological Perspective for Diagnosis and Treatment*. Philadelphia, PA, USA: Lippincott Williams & Wilkins.

Colton, R.H., & Conture, E.G. (1990). Problems and pitfalls of electroglottography. *Journal of Voice, 4*(1), 10-24.

Cortés, J.P., Espinoza, V.M., Ghassemi, M., Mehta, D.D., Van Stan, J.H., Hillman, R.E., Guttag, J.V., & Zañartu, M. (2018). Ambulatory assessment of phonotraumatic

vocal hyperfunction using glottal airflow measures estimated from neck-surface acceleration. *PLoS One, 13*(12), e0209017.

Cowan, J.P. (1993). *Handbook of Environmental Acoustics*. New York, NY, USA: Wiley.

Dabirmoghaddam, P., Aghajanzadeh, M., Erfanian, R., Aghazadeh, K., Sohrabpour, S., Firouzifar, M., Maroufizadeh, S., & Nikravesh, M. (2019). Comparative study of increased supraglottic activity in normal individuals and those with muscle tension dysphonia (MTD). *Journal of Voice*, Advanced online publication.

Dailey, S.H., Kobler, J.B., Hillman, R.E., Tangrom, K., Thananart, E., Mauri, M., & Zeitels, S.M. (2005). Endoscopic measurement of vocal fold movement during adduction and abduction. *The Laryngoscope, 115*(1), 178-183.

Dalrymple-Alford, J.C., MacAskill, M.R., Nakas, C.T., Livingston, L., Graham, C., Crucian, G.P., Melzer, T.R., Kirwan, J., Keenan, R., Wells, S., Porter, R.J., Watts, R., & Anderson, T.J. (2010). The MOCA. *Neurology, 75*(19), 1717.

Darley, F.L., Aronson, A.E., & Brown, J.R. (1969). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research, 12*(3), 462-496.

Dastolfo, C., Gartner-Schmidt, J., Yu, L., Carnes, O., & Gillespie, A.I. (2016). Aerodynamic outcomes of four common voice disorders: Moving toward disorder-specific assessment. *Journal of Voice, 30*(3), 301-307.

Dauer, W., & Przedborski, S. (2003). Parkinson's disease: Mechanisms and models. *Neuron, 39*(6), 889-909.

de Cheveigne, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America, 111*(4), 1917-30.

de Krom, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech, Language, and Hearing Research, 38*(4), 794-811.

De Letter, M., Santens, P., De Bodt, M., Van Maele, G., Van Borsel, J., & Boon, P. (2007). The effect of levodopa on respiration and word intelligibility in people with advanced Parkinson's disease. *Clinical Neurology and Neurosurgery, 109*(6), 495-500.

Dejonckere, P.H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., Van De Heyning, P., Remacle, M., & Woisard, V. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the

efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the committee on phoniatrics of the european laryngological society (ELS). *European Archives of Oto-Rhino-Laryngology, 258*(2), 77-82.

Dejonckere, P.H., Obbens, C., de Moor, G.M., & Wieneke, G.H. (1993). Perceptual evaluation of dysphonia: Reliability and relevance. *Folia Phoniatrica et Logopaedica, 45*(2), 76-83.

Dejonckere, P.H., Remacle, M., Fresnel-Elbaz, E., Woisard, V., Crevier-Buchman, L., & Millet, B. (1996). Differentiated perceptual evaluation of pathological voice quality: Reliability and correlations with acoustic measurements. *Revue de Laryngologie-Otologie-Rhinologie, 117*(3), 219-224.

Deliyski, D.D. (2010). Laryngeal High-Speed Videoendoscopy. In K.A. Kendall & R.J. Leonard (Eds.), *Laryngeal Evaluation: Indirect Laryngoscopy to High-speed Digital Imaging* (pp. 243-270). New York, NY, USA: Thieme Medical Publishers.

Deliyski, D.D., Petrushev, P.P., Bonilha, H.S., Gerlach, T.T., Martin-Harris, B., & Hillman, R.E. (2008). Clinical implementation of laryngeal high-speed videoendoscopy: Challenges and evolution. *Folia Phoniatrica et Logopaedica, 60*(1), 33-44.

Deliyski, D.D., Shaw, H.S., Evans, M.K., & Vesselinov, R. (2006). Regression tree approach to studying factors influencing acoustic voice analysis. *Folia Phoniatrica et Logopaedica, 58*(4), 274-288.

Desjardins, M., Halstead, L., Cooke, M., & Bonilha, H.S. (2017). A systematic review of voice therapy: What “effectiveness” really implies. *Journal of Voice, 31*(3), 392.e13-32.

Deuschl, G., Schade-Brittinger, C., Krack, P., Volkmann, J., Schäfer, H., Bötzel, K., Daniels, C., Deutschländer, A., Dillmann, U., Eisner, W., Gruber, D., Hamel, W., Herzog, J., Hilker, R., Klebe, S., Klo, M., Koy, J., Krause, M., Kupsch, A., Lorenz, D., Lorenzl, S., Mehdorn, H.M., Moringlane, J.R., Oertel, W., Pinsker, M.O., Reichmann, H., Reus, A., Schneider, G.-H., Schnitzler, A., Steude, U., Sturm, V., Timmermann, L., Tronnier, V., Trottenberg, T., Wojtecki, L., Wolf, E., Poewe, W., & Voges, J. (2006). A randomized trial of deep-brain stimulation for Parkinson's disease. *The New England Journal of Medicine, 355*(9), 896-908.

Diaz-Cadiz, M., McKenna, V.S., Vojtech, J.M., & Stepp, C.E. (2019). Adductory vocal fold kinematic trajectories during conventional versus high-speed videoendoscopy. *Journal of Speech, Language, and Hearing Research, 62*(6), 1685-1706.

Dietz, V., Quintern, J., & Berger, W. (1981). Electrophysiological studies of gait in spasticity and rigidity: Evidence that altered mechanical properties of muscle contribute to hypertonia. *Brain, 104*(3), 431-449.

Doellinger, M., Berry, D.A., & Berke, G.S. (2005). A quantitative study of the medial surface dynamics of an in vivo canine vocal fold during phonation. *The Laryngoscope, 115*(9), 1646-1654.

Döllinger, M., Kunduk, M., Kaltenbacher, M., Vondenhoff, S., Ziethe, A., Eysholdt, U., & Bohr, C. (2012). Analysis of vocal fold function from acoustic data simultaneously recorded with high-speed endoscopy. *Journal of Voice, 26*(6), 726-733.

Dong, E., Liu, G., Zhou, Y., & Cai, Y. (2002, Aug). Voice activity detection based on short-time energy and noise spectrum adaptation. In *6th International Conference on Signal Processing, (ICSP-2002)*, pp. 464-67. Beijing, China. https://doi.org/10.1109/ICOSP.2002.1181092

Dromey, C., Nissen, S.L., Roy, N., & Merrill, R.M. (2008). Articulatory changes following treatment of muscle tension dysphonia: Preliminary acoustic evidence. *Journal of Speech, Language, and Hearing Research, 51*(1), 196-208.

Dworkin-Valenti, J.P., Stachler, R.J., Stern, N., & Amjad, E.H. (2018). Pathophysiologic perspectives on muscle tension dysphonia. *Archives of Otolaryngology and Rhinology, 4*(1), 1-10.

Dworkin, J.P., Meleca, R.J., & Abkarian, G.G. (2000a). Muscle tension dysphonia. *Current Opinion in Otolaryngology & Head and Neck Surgery, 8*(3), 169–173.

Dworkin, J.P., Meleca, R.J., Simpson, M.L., & Garfield, I. (2000b). Use of topical lidocaine in the treatment of muscle tension dysphonia. *Journal of Voice, 14*(4), 567-574.

Eadie, T.L., & Baylor, C.R. (2006). The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *Journal of Voice, 20*(4), 527-544.

Eadie, T.L., & Doyle, P.C. (2005). Classification of dysphonic voice: Acoustic and auditory-perceptual measures. *Journal of Voice, 19*(1), 1-14.

Eadie, T.L., & Stepp, C.E. (2013). Acoustic correlate of vocal effort in spasmodic dysphonia. *Annals of Otology, Rhinology, and Laryngology, 122*(3), 169-176.

Eddins, D.A., Anand, S., Camacho, A., & Shrivastav, R. (2016). Modeling of breathy voice quality using pitch-strength estimates. *Journal of Voice, 30*(6), 774.e1-7.

Edstrom, L. (1968). Histochemical changes in upper motor lesions, Parkinsonism and disuse: Differential effect on white and red muscle fibres. *Experientia, 24*(9), 916-917.

Edstrom, L. (1970). Selective changes in the sizes of red and white muscle fibres in upper motor lesions and Parkinsonism. *Journal of the Neurological Sciences, 11*(6), 537-550.

Eller, R., Ginsburg, M., Lurie, D., Heman-Ackah, Y., Lyons, K., & Sataloff, R. (2008). Flexible laryngoscopy: A comparison of fiber optic and distal chip technologies. Part 1: Vocal fold masses. *Journal of Voice, 22*(6), 746-750.

Eriksson, A., & Traunmüller, H. (2002). Perception of vocal effort and distance from the speaker on the basis of vowel utterances. *Perception and Psychophysics, 64*(1), 131-139.

Eskenazi, L., Childers, D.G., & Hicks, D.M. (1990). Acoustic correlates of vocal quality. *Journal of Speech, Language, and Hearing Research, 33*(2), 298-306.

Espinoza, V.M., Zañartu, M., Van Stan, J.H., Mehta, D.D., & Hillman, R.E. (2017). Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 60*(8), 2159-2169.

Faaborg-Andersen, K. (1957). Electromyographic investigation of intrinsic laryngeal muscles in humans. *Acta Physiologica Scandinavica, 41*(140), 1-150.

Fahn, S. (1996). Is levodopa toxic? *Neurology, 47*(6), 184S-195S.

Fant, G. (1960). *Acoustic Theory of Speech Production* (2 ed.). The Hague, Netherlands: Mouton de Gruyter.

Fasano, A., Daniele, A., & Albanese, A. (2012). Treatment of motor and non-motor features of Parkinson's disease with deep brain stimulation. *The Lancet Neurology, 11*(5), 429-442.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160.

Ferrand, C.T. (2007). *Speech Science: An Integrated Approach to Theory and Clinical Practice*. Boston, MA, USA: Pearson/Allyn and Bacon.

Ferreira, J.J., Katzenschlager, R., Bloem, B.R., Bonuccelli, U., Burn, D., Deuschl, G., Dietrichs, E., Fabbrini, G., Friedman, A., Kanovsky, P., Kostic, V., Nieuwboer,

A., Odin, P., Poewe, W., Rascol, O., Sampaio, C., Schüpbach, M., Tolosa, E., Trenkwalder, C., Schapira, A., Berardelli, A., & Oertel, W.H. (2012). Summary of the recommendations of the EFNS/MDS-ES review on therapeutic management of Parkinson's disease. *European Journal of Neurology, 20*(1), 5-15.

Ferrer, C.A., Haderlein, T., Maryn, Y., de Bodt, M.S., & Nöth, E. (2018). Collinearity and sample coverage issues in the objective measurement of vocal quality: The case of roughness and breathiness. *Journal of Speech, Language, and Hearing Research, 61*(1), 1-24.

Fex, B., Fex, S., Shiromoto, O., & Hirano, M. (1994). Acoustic analysis of functional dysphonia: Before and after voice therapy (accent method). *Journal of Voice, 8*(2), 163-167.

Finck, C., & Lejeune, L. (2010). Structure and Oscillatory Function of the Vocal Folds. In S.M. Brudzynski (Ed.), *Handbook of Mammalian Vocalization: An Integrative Neuroscience Approach* (Vol. 19, pp. 427-438). London, UK: Academic Press.

Francis, D.O., Daniero, J.J., Hovis, K.L., Sathe, N., Jacobson, B., Penson, D.F., Feurer, I.D., & McPheeters, M.L. (2017). Voice-related patient-reported outcome measures: A systematic review of instrument development and validation. *Journal of Speech, Language, and Hearing Research, 60*(1), 62-88.

Friedman, A.D., Hillman, R.E., Landau-Zemer, T., Burns, J.A., & Zeitels, S.M. (2013). Voice outcomes for photoangiolytic KTP laser treatment of early glottic cancer. *Annals of Otology, Rhinology, and Laryngology, 122*(3), 151-158.

Fritts, L. (1994). Musical Instrument Samples. *The University of Iowa Electronic Music Studios.* Retrieved from http://theremin.music.uiowa.edu

Fryd, A.S., Van Stan, J.H., Hillman, R.E., & Mehta, D.D. (2016). Estimating subglottal pressure from neck-surface acceleration during normal voice production. *Journal of Speech, Language, and Hearing Research, 59*(6), 1335-1345.

Fujita, M., Ludlow, C.L., Woodson, G.E., & Naunton, R.F. (1989). A new surface electrode for recording from the posterior cricoarytenoid muscle. *The Laryngoscope, 99*(3), 316-320.

Fukui, N., & Hirose, H. (1983). Laryngeal adjustments in Danish voiceless obstruent production. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Copenhagen, 17*, 61-71.

Gallagher, D.A., & Schrag, A. (2012). Psychosis, apathy, depression and anxiety in Parkinson's disease. *Neurobiology of Disease, 46*(3), 581-589.

Gallena, S., Smith, P.J., Zeffiro, T., & Ludlow, C.L. (2001). Effects of levodopa on laryngeal muscle activity for voice onset and offset in Parkinson disease. *Journal of Speech, Language, and Hearing Research, 44*(6), 1284-1299.

Garaycochea, O., Navarrete, J.M.A., del Río, B., & Fernández, S. (2019). Muscle tension dysphonia: Which laryngoscopic features can we rely on for diagnosis? *Journal of Voice, 33*(5), 812.e15-18.

Gay, T., Hirose, H., Strome, M., & Sawashima, M. (1972). Electromyography of the intrinsic laryngeal muscles during phonation. *Annals of Otology, Rhinology, and Laryngology, 81*(3), 401-409.

Gelfer, M.P. (1995). Fundamental frequency, intensity, and vowel selection: Effects on measures of phonatory stability. *Journal of Speech, Language, and Hearing Research, 38*(6), 1189-1198.

Gervais-Bernard, H., Xie-Brustolin, J., Mertens, P., Polo, G., Klinger, H., Adamec, D., Broussolle, E., & Thobois, S. (2009). Bilateral subthalamic nucleus stimulation in advanced Parkinson's disease: Five year follow-up. *Journal of Neurology, 256*(2), 225-233.

Ghaemmaghami, H., Baker, B.J., Vogt, R.J., & Sridharan, S. (2010, Sept). Noise robust voice activity detection using features extracted from the time-domain autocorrelation function. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH-2010*, pp. 3118-3121. Makuhari, Chiba, Japan. https://eprints.qut.edu.au/40656/1/2011006688_H_Ghaemmaghami_ePrints.pdf

Ghassemi, M., Van Stan, J.H., Mehta, D.D., Zanartu, M., Cheyne, H.A., 2nd, Hillman, R.E., & Guttag, J.V. (2014). Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: Initial results for vocal fold nodules. *IEEE Transactions on Biomedical Engineering, 61*(6), 1668-1675.

Gillies, G.E., Pienaar, I.S., Vohra, S., & Qamhawi, Z. (2014). Sex differences in Parkinson's disease. *Frontiers in Neuroendocrinology, 35*(3), 370-384.

Goberman, A., Coelho, C., & Robb, M. (2002). Phonatory characteristics of Parkinsonian speech before and after morning medication: The on and off states. *Journal of Communication Disorders, 35*(3), 217-239.

Goberman, A.E., & Blomgren, M. (2008). Fundamental frequency change during offset and onset of voicing in individuals with Parkinson disease. *Journal of Voice, 22*(2), 178-191.

Goberman, A.M., & Blomgren, M. (2003). Parkinsonian speech disfluencies: Effects of L-dopa-related fluctuations. *Journal of Fluency Disorders, 28*(1), 55-70.

Goetz, C.G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G.T., Counsell, C., Giladi, N., Holloway, R.G., Moore, C.G., Wenning, G.K., Yahr, M.D., & Seidl, L. (2004). Movement Disorder Society task force report on the hoehn and yahr staging scale: Status and recommendations the Movement Disorder Society task force on rating scales for Parkinson's disease. *Movement Disorders, 19*(9), 1020-1028.

Goetz, C.G., Tilley, B.C., Shaftman, S.R., Stebbins, G.T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M.B., Dodel, R., Dubois, B., Holloway, R., Jankovic, J., Kulisevsky, J., Lang, A.E., Lees, A., Leurgans, S., LeWitt, P.A., Nyenhuis, D., Olanow, C.W., Rascol, O., Schrag, A., Teresi, J.A., van Hilten, J.J., & LaPelle, N. (2008). 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.

Gramming, P. (1991). Vocal loudness and frequency capabilities of the voice. *Journal of Voice, 5*(2), 144-157.

Groll, M.D., Vojtech, J.M., Hablani, S., Mehta, D.D., Buckley, D.P., Noordzij, J.P., & Stepp, C.E. (2020). Automated relative fundamental frequency algorithms for use with neck-surface accelerometer signals. *Journal of Voice*, Advanced online publication.

Gurfinkel, V., Cacciatore, T.W., Cordo, P., Horak, F., Nutt, J., & Skoss, R. (2006). Postural muscle tone in the body axis of healthy humans. *Journal of Neurophysiology, 96*(5), 2678-2687.

Guzman, M., Calvache, C., Romero, L., Munoz, D., Olavarria, C., Madrid, S., Leiva, M., Bortnem, C., & Pino, J. (2015). Do different semi-occluded voice exercises affect vocal fold adduction differently in subjects diagnosed with hyperfunctional dysphonia? *Folia Phoniatrica et Logopaedica, 67*(2), 68-75.

Guzman, M., Castro, C., Madrid, S., Olavarria, C., Leiva, M., Munoz, D., Jaramillo, E., & Laukkanen, A.M. (2016). Air pressure and contact quotient measures during different semioccluded postures in subjects with different voice conditions. *Journal of Voice, 30*(6), 759.e1-10.

Hammer, M.J., & Barlow, S.M. (2010). Laryngeal somatosensory deficits in Parkinson's disease: Implications for speech respiratory and phonatory control. *Experimental Brain Research, 201*(3), 401-409.

Hanson, D.G., Gerratt, B.R., & Ward, P.H. (1984). Cinegraphic observations of laryngeal function in Parkinson's disease. *The Laryngoscope, 94*(3), 348-353.

Hartnick, C.J., & Zeitels, S.M. (2005). Pediatric video laryngo-stroboscopy. *International Journal of Pediatric Otorhinolaryngology, 69*(2), 215-219.

Hast, M.H. (1966). Mechanical properties of the cricothyroid muscle. *The Laryngoscope, 75*, 537-548.

Hast, M.H. (1967a). Mechanical properties of the vocal fold muscles. *Practica Oto-Rhino-Laryngologica, 33*, 209-214.

Hast, M.H. (1967b). The respiratory muscle of the larynx. *Annals of Otology, Rhinology, and Laryngology, 76*(2), 489-497.

Hay, I., Oates, J., Giannini, A., Berkowitz, R., & Rotenberg, B. (2009). Pain Perception of Children Undergoing Nasendoscopy for Investigation of Voice and Resonance Disorders. *Journal of Voice, 23*(3), 380-388.

Heller Murray, E.S., Lien, Y.S., Van Stan, J.H., Mehta, D.D., Hillman, R.E., Pieter Noordzij, J., & Stepp, C.E. (2017). Relative fundamental frequency distinguishes between phonotraumatic and non-phonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 60*(6), 1507-1515.

Heman-Ackah, Y.D., Heuer, R.J., Michael, D.D., Ostrowski, R., Horman, M., Baroody, M.M., Hillenbrand, J., & Sataloff, R.T. (2003). Cepstral peak prominence: A more reliable measure of dysphonia. *Annals of Otology, Rhinology, and Laryngology, 112*(4), 324-333.

Heman-Ackah, Y.D., Michael, D.D., & Goding, G.S., Jr. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice, 16*(1), 20-27.

Heman-Ackah, Y.D., Sataloff, R.T., Laureyns, G., Lurie, D., Michael, D.D., Heuer, R., Rubin, A., Eller, R., Chandran, S., Abaza, M., Lyons, K., Divi, V., Lott, J., Johnson, J., & Hillenbrand, J. (2014). Quantifying the cepstral peak prominence, a measure of dysphonia. *Journal of Voice, 28*(6), 783-788.

Henderson, M.X., Trojanowski, J.Q., & Lee, V.M.Y. (2019). α-synuclein pathology in Parkinson's disease and related α-synucleinopathies. *Neuroscience Letters, 709*, 134316.

Herbst, C.T. (2019). Electroglottography – an update. *Journal of Voice, 34*(4), 503-526.

Hertegard, S., Gauffin, J., & Lindestad, P.A. (1995). A comparison of subglottal and intraoral pressure measurements during phonation. *Journal of Voice, 9*(2), 149-155.

Higgins, M.B., Chait, D.H., & Schulte, L. (1999). Phonatory air flow characteristics of adductor spasmodic dysphonia and muscle tension dysphonia. *Journal of Speech, Language, and Hearing Research, 42*(1), 101-111.

Hillenbrand, J., Cleveland, R.A., & Erickson, R.L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research, 37*(4), 769-778.

Hillenbrand, J., Getty, L.A., Clark, M.J., & Wheeler, K. (1995). Acoustic characteristics of american english vowels. *Journal of the Acoustical Society of America, 97*(5), 3099-3111.

Hillenbrand, J., & Houde, R.A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research, 39*(2), 311-321.

Hillman, R.E., Heaton, J.T., Masaki, A., Zeitels, S.M., & Cheyne, H.A. (2006). Ambulatory monitoring of disordered voices. *Annals of Otology, Rhinology, and Laryngology, 115*(11), 795-801.

Hillman, R.E., Holmberg, E.B., Perkell, J.S., Walsh, M., & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech and Hearing Research, 32*(2), 373-392.

Hillman, R.E., Montgomery, W.W., & Zeitels, S.M. (1997). Appropriate use of objective measures of vocal function in the multidisciplinary management of voice disorders. *Current Opinion in Otolaryngology & Head and Neck Surgery, 5*, 172-175.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (2003). *Applied Statistics for the Behavioral Sciences* (5 ed.). Boston, MA, USA: Houghton Mifflin.

Hirano, M. (1974). Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatrica et Logopaedica, 26*(2), 89-94.

Hirano, M. (1981). *Clinical Examination of Voice*. New York, NY, USA: Springer-Verlag.

Hirano, M. (1988). Vocal mechanisms in singing: Laryngological and phoniatric aspects. *Journal of Voice, 2*(1), 51-69.

Hirano, M., Kakita, Y., Ohmaru, K., & Kurita, S. (1982). Structure and Mechanical Properties of the Vocal Fold. In N.J. Lass (Ed.), *Speech and Language: Advances in Basic Research and Practice* (Vol. 7, pp. 271-297). Orlando, FL, USA: Academic Press.

Hirano, M., Kurita, S., Kiyokawa, K., & Sato, K. (1986). Posterior glottis. Morphological study in excised human larynges. *Annals of Otology, Rhinology, and Laryngology, 95*(6 Pt 1), 576-581.

Hirano, M., & Ohala, J. (1969a). Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. *Journal of Speech and Hearing Disorders, 12*(2), 362-373.

Hirano, M., Ohala, J., & Vennard, W. (1969b). The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation. *Journal of Speech and Hearing Research, 12*(3), 616-628.

Hixon, T.J., Weismer, G., & Hoit, J.G. (2018). *Preclinical Speech Science: Anatomy, Physiology, Acoustics, and Perception*. San Diego, CA, USA: Plural Publishing.

Hočevar-Boltežar, I., Janko, M., & Zargi, M. (1998). Role of surface EMG in diagnostics and treatment of muscle tension dysphonia. *Acta Oto-Laryngologica, 118*(5), 739-743.

Hoehn, M.M., & Yahr, M.D. (1967). Parkinsonism: Onset, progression and mortality. *Neurology, 17*(5), 427-442.

Hogikyan, N.D., & Sethuraman, G. (1999). Validation of an instrument to measure voice-related quality of life (V-RQOL). *Journal of Voice, 13*(4), 557-569.

Holmberg, E.B., Doyle, P., Perkell, J.S., Hammarberg, B., & Hillman, R.E. (2003). Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation in baseline and changes across voice therapy. *Journal of Voice, 17*(3), 269-282.

Holmberg, E.B., Hillman, R.E., Hammarberg, B., Sodersten, M., & Doyle, P. (2001). Efficacy of a behaviorally based voice therapy protocol for vocal nodules. *Journal of Voice, 15*(3), 395-412.

Holmes, R.J., Oates, J.M., Phyland, D.J., & Hughes, A.J. (2000). Voice characteristics in the progression of Parkinson's disease. *International Journal of Language & Communication Disorders, 35*(3), 407-418.

Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in f0 control. *Language and Speech, 42*(4), 401-411.

Hong, K.H., Kim, H.K., & Kim, Y.H. (2001). The role of the pars recta and pars oblique of cricothyroid muscle in speech production. *Journal of Voice, 15*(4), 512-518.

Hong, K.H., Ye, M., Kim, Y.M., Kevorkian, K.F., Kreiman, J., & Berke, G.S. (1998). Functional differences between the two bellies of the cricothyroid muscle. *Otolaryngology–Head and Neck Surgery, 118*(5), 714-722.

Honjo, I., & Isshiki, N. (1980). Laryngoscopic and voice characteristics of aged persons. *Archives of Otolaryngology, 106*(3), 149-150.

Hoodin, R.B., & Gilbert, H.R. (1989). Parkinsonian dysarthria: An aerodynamic and perceptual description of velopharyngeal closure for speech. *Folia Phoniatrica et Logopaedica, 41*(6), 249-258.

Hosokawa, K., Yoshida, M., Yoshii, T., Takenaka, Y., Hashimoto, M., Ogawa, M., & Inohara, H. (2012). Effectiveness of the computed analysis of electroglottographic signals in muscle tension dysphonia. *Folia Phoniatrica et Logopaedica, 64*(3), 145-150.

Houde, J.F., & Nagarajan, S.S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience, 5*, 82-82.

Hsiao, T.Y., Liu, C.M., Hsu, C.J., Lee, S.Y., & Lin, K.N. (2001). Vocal fold abnormalities in laryngeal tension-fatigue syndrome. *Journal of the Formosan Medical Association, 100*(12), 837-840.

Hülse, M. (1991). Zervikale dysphonie. *Folia Phoniatrica et Logopaedica, 43*(4), 181-196.

Hunter, E.J., Cantor-Cutiva, L.C., van Leer, E., van Mersbergen, M.R., Nanjundeswaran, C.D., Bottalico, P., Sandage, M.J., & Whitling, S. (2020). Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue. *Journal of Speech, Language, and Hearing Research, 63*(2), 509-532.

Ikuma, T., Kunduk, M., & McWhorter, A.J. (2013). Preprocessing techniques for high-speed videoendoscopy analysis. *Journal of Voice, 27*(4), 500-505.

Isshiki, N., Shoji, K., Kojima, H., & Hirano, S. (1996). Vocal fold atrophy and its surgical treatment. *Annals of Otology, Rhinology, and Laryngology, 105*(3), 182-188.

Iwahashi, T., Ogawa, M., Hosokawa, K., Kato, C., & Inohara, H. (2016). A detailed motion analysis of the angular velocity between the vocal folds during throat clearing using high-speed digital imaging. *Journal of Voice, 30*(6), 770.e1-8.

Jacob, S. (2007). *Human Anatomy: A Clinically-Orientated Approach*. London, UK: Elsevier-Health Sciences Division.

Jacobson, B.H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M.S., & Newman, C.W. (1997). The voice handicap index (VHI): Development and validation. *American Journal of Speech-Language Pathology, 6*(3), 66-70.

Jafari, N., Salehi, A., Izadi, F., Talebian Moghadam, S., Ebadi, A., Dabirmoghadam, P., Faham, M., & Shahbazi, M. (2017). Vocal function exercises for muscle tension dysphonia: Auditory-perceptual evaluation and self-assessment rating. *Journal of Voice, 31*(4), 506.e25-31.

Jafari, N., Salehi, A., Meerschman, I., Izadi, F., Ebadi, A., Talebian, S., Khoddami, S.M., Dabirmoghadam, P., Drinnan, M., Jordens, K., D'Haeseleer, E., & Van Lierde, K. (2020). A novel laryngeal palpatory scale (LPS) in patients with muscle tension dysphonia. *Journal of Voice, 34*(3), 488.e9-27.

Jalil, M., Butt, F.A., & Malik, A. (2013, May). Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals. In *2013 The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE-2013)*, pp. 208-212. Konya, Turkey. https://doi.org/10.1109/TAEECE.2013.6557272

Jankovic, J. (2008). Parkinson's disease: Clinical features and diagnosis. *Journal of Neurology, Neurosurgery, and Psychiatry, 79*(4), 368-376.

Jiang, J., Lin, E., Wang, J., & Hanson, D.G. (1999a). Glottographic measures before and after levodopa treatment in Parkinson's disease. *The Laryngoscope, 109*(8), 1287-1294.

Jiang, J., O'Mara, T., Chen, H.-J., Stern, J.I., Vlagos, D., & Hanson, D. (1999b). Aerodynamic measurements of patients with Parkinson's disease. *Journal of Voice, 13*(4), 583-591.

Jouvet, D., & Laprie, Y. (2017, Aug). Performance Analysis of Several Pitch Detection Algorithms on Simulated and Real Noisy Speech Data. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 1614-1618. Kos, Greece. https://doi.org/10.23919/EUSIPCO.2017.8081482

Karnell, M.P., Melton, S.D., Childes, J.M., Coleman, T.C., Dailey, S.A., & Hoffman, H.T. (2007). Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *Journal of Voice, 21*(5), 576-590.

Kawahara, H., de Cheveigné, A., Banno, H., Takahashi, T., & Irino, T. (2005). Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT. In *17th Annual Conference of the International Speech Communication Association (Interspeech 2016), Vols 1-5*, pp. 537-540. Lisbon, Portugal.

Kempster, G.B., Gerratt, B.R., Verdolini Abbott, K., Barkmeier-Kraemer, J.M., & Hillman, R.E. (2009). Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*(2), 124-132.

Kennard, E.J., Lieberman, J., Saaid, A., & Rolfe, K.J. (2015). A preliminary comparison of laryngeal manipulation and postural treatment on voice quality in a prospective randomized crossover study. *Journal of Voice, 29*(6), 751-754.

Kent, R.D., & Read, C. (2002). *The Acoustic Analysis of Speech*. Albany, NY, USA: Singular/Thomson Learning.

Khoddami, S.M., Ansari, N.N., Izadi, F., & Talebian Moghadam, S. (2013). The assessment methods of laryngeal muscle activity in muscle tension dysphonia: A review. *The Scientific World Journal, 2013*, 507397.

Khoddami, S.M., Ansari, N.N., & Jalaie, S. (2015). Review on laryngeal palpation methods in muscle tension dysphonia: Validity and reliability issues. *Journal of Voice, 29*(4), 459-468.

Khosla, S., Muruguppan, S., Gutmark, E., & Scherer, R. (2007). Vortical flow field during phonation in an excised canine larynx model. *Annals of Otology, Rhinology, and Laryngology, 116*(3), 217-228.

Klatt, D.H., & Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87*(2), 820-857.

Kleiner-Fisman, G., Fisman, D.N., Sime, E., Saint-Cyr, J.A., Lozano, A.M., & Lang, A.E. (2003). Long-term follow up of bilateral deep brain stimulation of the subthalamic nucleus in patients with advanced Parkinson disease. *Journal of Neurosurgery, 99*(3), 489-495.

Kooijman, P.G.C., de Jong, F.I.C.R.S., Oudes, M.J., Huinck, W., van Acht, H., & Graamans, K. (2005). Muscular tension and body posture in relation to voice handicap and voice quality in teachers with persistent voice complaints. *Folia Phoniatrica et Logopaedica, 57*(3), 134-147.

Kopf, L.M., Jackson-Menaldi, C., Rubin, A.D., Skeffington, J., Hunter, E.J., Skowronski, M.D., & Shrivastav, R. (2017). Pitch strength as an outcome measure for treatment of dysphonia. *Journal of Voice, 31*(6), 691-696.

Kotby, M.N., Shiromoto, O., & Hirano, M. (1993). The accent method of voice therapy: Effect of accentuations on FO, SPL, and airflow. *Journal of Voice, 7*(4), 319-325.

Koufman, J.A., & Blalock, P.D. (1991). Functional voice disorders. *Otolaryngologic Clinics of North America, 24*(5), 1059-1073.

Krausert, C.R., Olszewski, A.E., Taylor, L.N., McMurray, J.S., Dailey, S.H., & Jiang, J.J. (2011). Mucosal wave measurement and visualization techniques. *Journal of Voice, 25*(4), 395-405.

Kreiman, J., & Gerratt, B.R. (2000). Sources of listener disagreement in voice quality assessment. *Journal of the Acoustical Society of America, 108*(4), 1867-1876.

Kreiman, J., Gerratt, B.R., & Berke, G.S. (1994). The multidimensional nature of pathologic vocal quality. *Journal of the Acoustical Society of America, 96*(3), 1291-1302.

Kridgen, S. (2019). *Patient-reported events associated with the onset of phonotraumatic and nonphonotraumatic vocal hyperfunction.* (Master's thesis, MGH Institute of Health Professions, Boston, MA, USA), Retrieved from Proquest Digital Dissertations. (Publication No. 22618930)

Kroonenberg, P.M., Oort, F.J., Stebbins, G.T., Leurgans, S.E., Cubo, E., & Goetz, C.G. (2006). Motor function in Parkinson's disease and supranuclear palsy: Simultaneous factor analysis of a clinical scale in several populations. *BMC Medical Research Methodology, 6*, 26.

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. New York, NY, USA: Springer Science & Business Media.

Kunduk, M., Yan, Y., McWhorter, A.J., & Bless, D. (2006). Investigation of voice initiation and voice offset characteristics with high-speed digital imaging. *Logopedics Phoniatrics Vocology, 31*(3), 139-144.

Kuo, C., Holmberg, E.B., & Hillman, R.E. (1999, Mar). Discriminating speakers with vocal nodules using aerodynamic and acoustic features. In *1999 IEEE*

*International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99,* pp. 77-80. Phoenix, AZ, USA. https://doi.org/10.1109/ICASSP.1999.758066

Kutta, H., Steven, P., Kohla, G., Tillmann, B., & Paulsen, F. (2002). The human false vocal folds – an analysis of antimicrobial defense mechanisms. *Anatomy and Embryology, 205*(4), 315-323.

Kwan, L.C., & Whitehill, T.L. (2011). Perception of speech by individuals with Parkinson's disease: A review. *Parkinson's Disease, 2011*, 389767. 11 pages. https://doi.org/10.4061/2011/389767

Lagier, A., Vaugoyeau, M., Ghio, A., Legou, T., Giovanni, A., & Assaiante, C. (2010). Coordination between posture and phonation in vocal effort behavior. *Folia Phoniatrica et Logopaedica, 62*(4), 195-202.

Lamb, J.R., Schultz, S.A., Scholp, A.J., Wendel, E.R., & Jiang, J.J. (2020). Retest reliability for complete airway interruption systems of aerodynamic measurement. *Journal of Voice*, Advanced online publication.

Lametti, D.R., Nasir, S.M., & Ostry, D.J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience, 32*(27), 9351-9358.

Larson, C.R., Altman, K.W., Liu, H., & Hain, T.C. (2008). Interactions between auditory and somatosensory feedback for voice f0 control. *Experimental Brain Research, 187*(4), 613-621.

Leanderson, R., Meyerson, B.A., & Persson, A. (1971). Effect of L-dopa on speech in Parkinsonism: An EMG study of labial articulatory function. *Journal of Neurology, Neurosurgery, and Psychiatry, 34*(6), 679-681.

Liang, F.Y., Yang, J.S., Mei, X.S., Cai, Q., Guan, Z., Zhang, B.R., Wang, Y.J., Gong, J., Huang, X.M., Peng, J.R., & Zheng, Y.Q. (2014). The vocal aerodynamic change in female patients with muscular tension dysphonia after voice training. *Journal of Voice, 28*(3), 393.e7-10.

Lieberman, J. (1998). Principles and Techniques of Manual Therapy: Application In the Management of Dysphonia. In T. Harris, S. Harris, J. Rubin, & D. Howard (Eds.), *The Voice Clinical Handbook* (pp. 91-138). London, UK: Whurr Publishers.

Lien, Y.S. (2015). *Optimization and automation of relative fundamental frequency for objective assessment of vocal hyperfunction.* (Doctoral dissertation, Boston University, Boston, MA, USA), Retrieved from https://hdl.handle.net/2144/13645

Lien, Y.S., Calabrese, C.R., Michener, C.M., Heller Murray, E.S., Van Stan, J.H., Mehta, D.D., Hillman, R.E., Noordzij, J.P., & Stepp, C.E. (2015a). Voice relative fundamental frequency via neck-skin acceleration in individuals with voice disorders. *Journal of Speech, Language, and Hearing Research, 58*(5), 1482-1487.

Lien, Y.S., Gattuccio, C.I., & Stepp, C.E. (2014). Effects of phonetic context on relative fundamental frequency. *Journal of Speech, Language, and Hearing Research, 57*, 1259-1267.

Lien, Y.S., Heller Murray, E.S., Calabrese, C.R., Michener, C.M., Van Stan, J.H., Mehta, D.D., Hillman, R.E., Noordzij, J.P., & Stepp, C.E. (2017). Validation of an algorithm for semi-automated estimation of voice relative fundamental frequency. *Annals of Otology, Rhinology, and Laryngology, 126*(10), 712-716.

Lien, Y.S., Michener, C.M., Eadie, T.L., & Stepp, C.E. (2015b). Individual monitoring of vocal effort with relative fundamental frequency: Relationships with aerodynamics and listener perception. *Journal of Speech, Language, and Hearing Research, 58*(3), 566-575.

Limousin, P., Krack, P., Pollak, P., Benazzouz, A., Ardouin, C., Hoffmann, D., & Benabid, A.-L. (1998). Electrical stimulation of the subthalamic nucleus in advanced Parkinson's disease. *The New England Journal of Medicine, 339*(16), 1105-1111.

Lindestad, P.A., Blixt, V., Pahlberg-Olsson, J., & Hammarberg, B. (2004). Ventricular fold vibration in voice production: A high-speed imaging study with kymographic, acoustic and perceptual analyses of a voice patient and a vocally healthy subject. *Logopedics Phoniatrics Vocology, 29*(4), 162-170.

Lindestad, P.A., Fritzell, B., & Persson, A. (1991). Quantitative analysis of laryngeal EMG in normal subjects. *Acta Oto-Laryngologica, 111*(6), 1146-1152.

Lindestad, P.A., Sodersten, M., Merker, B., & Granqvist, S. (2001). Voice source characteristics in mongolian "throat singing" studied with high-speed imaging technique, acoustic spectra, and inverse filtering. *Journal of Voice, 15*(1), 78-85.

Liu, H., Behroozmand, R., Bove, M., & Larson, C.R. (2011). Laryngeal electromyographic responses to perturbations in voice pitch auditory feedback. *Journal of the Acoustical Society of America, 129*(6), 3946-3954.

Liu, H., Wang, E.Q., Metman, L.V., & Larson, C.R. (2012). Vocal responses to perturbations in voice auditory feedback in individuals with Parkinson's disease. *PLoS One, 7*(3), e33629.

Llewellyn-Thomas, H.A., Sutherland, H.J., Hogg, S.A., Ciampi, A., Harwood, A.R., Keane, T.J., Till, J.E., & Boyd, N.F. (1984). Linear analogue self-assessment of voice quality in laryngeal cancer. *Journal of Chronic Diseases, 37*(12), 917-924.

Löfqvist, A., Baer, T., McGarr, N.S., & Story, R.S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America, 85*(3), 1314-1321.

Löfqvist, A., Carlborg, B., & Kitzing, P. (1982). Initial validation of an indirect measure of subglottal pressure during vowels. *Journal of the Acoustical Society of America, 72*(2), 633-635.

Logemann, J.A., Fisher, H.B., Boshes, B., & Blonsky, E.R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *Journal of Speech and Hearing Disorders, 43*(1), 47-57.

Loucks, T.M., Poletto, C.J., Saxon, K.G., & Ludlow, C.L. (2005). Laryngeal muscle responses to mechanical displacement of the thyroid cartilage in humans. *Journal of Applied Physiology, 99*(3), 922-930.

Lowell, S.Y., Colton, R.H., Kelley, R.T., & Mizia, S.A. (2013). Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech. *Journal of Voice, 27*(4), 393-400.

Lowell, S.Y., Kelley, R.T., Awan, S.N., Colton, R.H., & Chan, N.H. (2012a). Spectral- and cepstral-based acoustic features of dysphonic, strained voice quality. *Annals of Otology, Rhinology, and Laryngology, 121*(8), 539-548.

Lowell, S.Y., Kelley, R.T., Colton, R.H., Smith, P.B., & Portnoy, J.E. (2012b). Position of the hyoid and larynx in people with muscle tension dysphonia. *The Laryngoscope, 122*(2), 370-377.

Lowell, S.Y., & Story, B.H. (2006). Simulated effects of cricothyroid and thyroarytenoid muscle activation on adult-male vocal fold vibration. *Journal of the Acoustical Society of America, 120*(1), 386-397.

Lowit, A., Dobinson, C., Timmins, C., Howell, P., & Kröger, B. (2010). The effectiveness of traditional methods and altered auditory feedback in improving speech rate and intelligibility in speakers with Parkinson's disease. *International Journal of Speech-Language Pathology, 12*(5), 426-436.

Mak, M.K., Wong, E.C., & Hui-Chan, C.W. (2007). Quantitative measurement of trunk rigidity in Parkinsonian patients. *Journal of Neurology, 254*(2), 202-209.

Marks, K.L., Lin, J.Z., Fox, A.B., Toles, L.E., & Mehta, D.D. (2019). Impact of nonmodal phonation on estimates of subglottal pressure from neck-surface

acceleration in healthy speakers. *Journal of Speech, Language, and Hearing Research, 62*(9), 3339-3358.

Martin, D., Fitch, J., & Wolfe, V. (1995). Pathologic voice type and the acoustic prediction of severity. *Journal of Speech, Language, and Hearing Research, 38*(4), 765-771.

Martínez-Martín, P., Rodríguez-Blázquez, C., Mario, A., Arakaki, T., Arillo, V.C., Chaná, P., Fernández, W., Garretto, N., Martínez-Castrillo, J.C., Rodríguez-Violante, M., Serrano-Dueñas, M., Ballesteros, D., Rojo-Abuin, J.M., Chaudhuri, K.R., & Merello, M. (2015). Parkinson's disease severity levels and MDS-Unified Parkinson's Disease Rating Scale. *Parkinsonism and Related Disorders, 21*(1), 50-54.

Martins, R.H., do Amaral, H.A., Tavares, E.L., Martins, M.G., Goncalves, T.M., & Dias, N.H. (2016). Voice disorders: Etiology and diagnosis. *Journal of Voice, 30*(6), 761.e1-9.

Maryn, Y., & Weenink, D. (2015). Objective dysphonia measures in the program praat: Smoothed cepstral peak prominence and acoustic voice quality index. *Journal of Voice, 29*(1), 35-43.

Mathieson, L. (2011). The evidence for laryngeal manual therapies in the treatment of muscle tension dysphonia. *Current Opinion in Otolaryngology & Head and Neck Surgery, 19*(3), 171-176.

Mathieson, L., Hirani, S.P., Epstein, R., Baken, R.J., Wood, G., & Rubin, J.S. (2009). Laryngeal manual therapy: A preliminary study to examine its treatment effects in the management of muscle tension dysphonia. *Journal of Voice, 23*(3), 353-366.

McCabe, D.J., & Titze, I.R. (2002). Chant therapy for treating vocal fatigue among public school teachers. *American Journal of Speech-Language Pathology, 11*(4), 356-369.

McKenna, V.S., Diaz-Cadiz, M.E., Shembel, A.C., Enos, N.E., & Stepp, C.E. (2018a). The relationship between physiological mechanisms and the self-perception of vocal effort. *Journal of Speech, Language, and Hearing Research, 62*(4), 815-834.

McKenna, V.S., Heller Murray, E.S., Lien, Y.S., & Stepp, C.E. (2016). The relationship between relative fundamental frequency and a kinematic estimate of laryngeal stiffness in healthy adults. *Journal of Speech, Language, and Hearing Research, 59*(6), 1283-1294.

McKenna, V.S., Llico, A.F., Mehta, D.D., Perkell, J.S., & Stepp, C.E. (2017). Magnitude of neck-surface vibration as an estimate of subglottal pressure during modulations of vocal effort and intensity in healthy speakers. *Journal of Speech, Language, and Hearing Research, 60*(12), 3404-3416.

McKenna, V.S., & Stepp, C.E. (2018b). The relationship between acoustical and perceptual measures of vocal effort. *Journal of the Acoustical Society of America, 144*(3), 1643-1658.

Meara, R.J., & Cody, F.W. (1993). Stretch reflexes of individual Parkinsonian patients studied during changes in clinical rigidity following medication. *Electroencephalography and Clinical Neurophysiology, 89*(4), 261-268.

Mehta, D. (2006). *Aspiration noise during phonation: Synthesis, analysis, and pitch-scale modification.* (Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA), Retrieved from https://scholar.harvard.edu/files/dmehta/files/mehtad_smthesis.pdf

Mehta, D.D., & Hillman, R.E. (2008). Voice assessment: Updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods. *Current Opinion in Otolaryngology & Head and Neck Surgery, 16*(3), 211-215.

Mehta, D.D., Van Stan, J.H., & Hillman, R.E. (2016). 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.

Mehta, D.D., Van Stan, J.H., Zañartu, M., Ghassemi, M., Guttag, J.V., Espinoza, V.M., Cortés, J.P., Cheyne, H.A., 2nd, & Hillman, R.E. (2015). Using ambulatory voice monitoring to investigate common voice disorders: Research update. *Frontiers in Bioengineering and Biotechnology, 3*, 155.

Mehta, D.D., Zanartu, M., Feng, S.W., Cheyne, H.A., 2nd, & Hillman, R.E. (2012a). Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform. *IEEE Transactions on Biomedical Engineering, 59*(11), 3090-3096.

Mehta, D.D., Zeitels, S.M., Burns, J.A., Friedman, A.D., Deliyski, D.D., & Hillman, R.E. (2012b). High-speed videoendoscopic analysis of relationships between cepstral-based acoustic measures and voice production mechanisms in patients undergoing phonomicrosurgery. *Annals of Otology, Rhinology, and Laryngology, 121*(5), 341-347.

Metter, E.J., & Hanson, W.R. (1986). Clinical and acoustical variability in hypokinetic dysarthria. *Journal of Communication Disorders, 19*(5), 347-366.

Mollaei, F., Shiller, D.M., Baum, S.R., & Gracco, V.L. (2016). Sensorimotor control of vocal pitch and formant frequencies in Parkinson's disease. *Brain Research, 1646*, 269-277.

Mollaei, F., Shiller, D.M., & Gracco, V.L. (2013). Sensorimotor adaptation of speech in Parkinson's disease. *Movement Disorders, 28*(12), 1668-1674.

Mooshammer, C. (2010). Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in german. *Journal of the Acoustical Society of America, 127*(2), 1047-1058.

Morrison, M. (1997). Pattern recognition in muscle misuse voice disorders: How i do it. *Journal of Voice, 11*(1), 108-114.

Morrison, M., Rammage, L., & Emami, A.J. (1999). The irritable larynx syndrome. *Journal of Voice, 13*(3), 447-455.

Morrison, M.D., Nichol, H., & Rammage, L.A. (1986). Diagnostic criteria in functional dysphonia. *The Laryngoscope, 96*(1), 1-8.

Morrison, M.D., & Rammage, L.A. (1993). Muscle misuse voice disorders: Description and classification. *Acta Oto-Laryngologica, 113*(3), 428-434.

Morrison, M.D., Rammage, L.A., Belisle, G.M., Pullan, C.B., & Nichol, H. (1983). Muscular tension dysphonia. *Journal of Otlaryngology, 12*(5), 302-306.

Mu, L., Chen, J., Sobotka, S., Nyirenda, T., Benson, B., Gupta, F., Sanders, I., Adler, C.H., Caviness, J.N., Shill, H.A., Sabbagh, M., Samanta, J.E., Sue, L.I., Beach, T.G., & Arizona Parkinson's Disease, C. (2015). Alpha-synuclein pathology in sensory nerve terminals of the upper aerodigestive tract of Parkinson's disease patients. *Dysphagia, 30*(4), 404-417.

Mu, L., Sobotka, S., Chen, J., Su, H., Sanders, I., Adler, C.H., Shill, H.A., Caviness, J.N., Samanta, J.E., Beach, T.G., & Arizona Parkinson's Disease, C. (2012). Altered pharyngeal muscles in Parkinson disease. *Journal of Neuropathology & Experimental Neurology, 71*(6), 520-530.

Murdoch, B.E., Manning, C.Y., Theodoros, D.G., & Thompson, E.C. (1997). Laryngeal and phonatory dysfunction in Parkinson's disease. *Clinical Linguistics & Phonetics, 11*(3), 245-266.

Nagumo, K., & Hirayama, K. (1993). A study on truncal rigidity in Parkinsonism—evaluation of diagnostic test and electrophysiological study. *Rinsho Shinkeigaku, 33*(1), 27-35.

Nagumo, K., & Hirayama, K. (1996). Axial (neck and trunk) rigidity in Parkinson's disease, striatonigral degeneration and progressive supranuclear palsy. *Rinsho Shinkeigaku, 36*(10), 1129-1135.

Nakano, K.K., Zubick, H., & Tyler, H.R. (1973). Speech defects of Parkinsonian patients: Effects of levodopa therapy on speech intelligibility. *Neurology, 23*(8), 865-870.

Nandhini, S., & Shenbagavalli, A. (2014, Mar). Voiced/Unvoiced Detection using Short Term Processing. In *IJCA Proceedings on International Conference on Innovations in Information, Embedded and Communication Systems ICIIECS-2014,* pp. 39-43. Coimbatore, India: International Journal of Computer Applications. https://www.ijcaonline.org/proceedings/iciiecs/number2/18661-1461

Nanjundeswaran, C., Jacobson, B.H., Gartner-Schmidt, J., & Verdolini Abbott, K. (2015). Vocal fatigue index (VFI): Development and validation. *Journal of Voice, 29*(4), 433-440.

Nash, E.A., & Ludlow, C.L. (1996). Laryngeal muscle activity during speech breaks in adductor spasmodic dysphonia. *The Laryngoscope, 106*(4), 484-489.

Nasri, S., Jasleen, J., Gerratt, B.R., Sercarz, J.A., Wenokur, R., & Berke, G.S. (1996). Ventricular dysphonia: A case of false vocal fold mucosal traveling wave. *American Journal of Otolaryngology, 17*(6), 427-431.

Neel, H.B., Harner, S.G., Benninger, M.S., Crumley, R.L., Ford, C.N., Gould, W.J., Hanson, D.G., Ossoff, R.H., & Sataloff, R.T. (1994). Evaluation and treatment of the unilateral paralyzed vocal fold. *Otolaryngology–Head and Neck Surgery, 111*(4), 497-508.

Nemr, K., Simões-Zenari, M., Cordeiro, G.F., Tsuji, D., Ogawa, A.I., Ubrig, M.T., & Menezes, M.H.M. (2012). GRBAS and CAPE-V scales: High reliability and consensus when applied at different times. *Journal of Voice, 26*(6), 812.e17-22.

Netsell, R., Lotz, W., & Shaughnessy, A.L. (1984). Laryngeal aerodynamics associated with selected voice disorders. *American Journal of Otolaryngology, 5*(6), 397-403.

Nguyen, D.D., & Kenny, D.T. (2009). Randomized controlled trial of vocal function exercises on muscle tension dysphonia in vietnamese female teachers. *Journal of Otolaryngology - Head & Neck Surgery, 38*(2), 261-278.

Núñez-Batalla, F., Díaz-Fresno, E., Álvarez-Fernández, A., Muñoz Cordero, G., & Llorente Pendás, J.L. (2017). Application of the acoustic voice quality index for

objective measurement of dysphonia severity. *Acta Otorrinolaringológica Española, 68*(4), 204-211.

Oates, J. (2009). Auditory-perceptual evaluation of disordered voice quality: Pros, cons and future directions. *Folia Phoniatrica et Logopaedica, 61*(1), 49-56.

Ogawa, M., Hosokawa, K., Yoshida, M., Iwahashi, T., Hashimoto, M., & Inohara, H. (2014). Immediate effects of humming on computed electroglottographic parameters in patients with muscle tension dysphonia. *Journal of Voice, 28*(6), 733-741.

Ogawa, M., & Inohara, H. (2018). Is voice therapy effective for the treatment of dysphonic patients with benign vocal fold lesions? *Auris Nasus Larynx, 45*(4), 661-666.

Ogawa, M., Yoshida, M., Watanabe, K., Yoshii, T., Sugiyama, Y., Sasaki, R., Watanabe, Y., & Kubo, T. (2005). Association between laryngeal findings and vocal qualities in muscle tension dysphonia with supraglottic contraction. *Nihon Jibiinkoka Gakkai Kaiho, 108*(7), 734-741.

Olanow, C.W., Agid, Y., Mizuno, Y., Albanese, A., Bonucelli, U., Damier, P., De Yebenes, J., Gershanik, O., Guttman, M., Grandas, F., Hallett, M., Hornykiewicz, O., Jenner, P., Katzenschlager, R., Langston, W.J., LeWitt, P., Melamed, E., Mena, M.A., Michel, P.P., Mytilineou, C., Obeso, J.A., Poewe, W., Quinn, N., Raisman-Vozari, R., Rajput, A.H., Rascol, O., Sampaio, C., & Stocchi, F. (2004). Levodopa in the treatment of Parkinson's disease: Current controversies. *Movement Disorders, 19*(9), 997-1005.

Omori, K., Slavit, D.H., Matos, C., Kojima, H., Kacker, A., & Blaugrund, S.M. (1997). Vocal fold atrophy: Quantitative glottic measurement and vocal function. *Annals of Otology, Rhinology, and Laryngology, 106*(7 Pt 1), 544-551.

Oren, L., Khosla, S., & Gutmark, E. (2014). Intraglottal geometry and velocity measurements in canine larynges. *Journal of the Acoustical Society of America, 135*(1), 380-388.

Park, Y., & Stepp, C.E. (2019). The effects of stress type, vowel identity, baseline f0, and loudness on the relative fundamental frequency of individuals with healthy voices. *Journal of Voice, 33*(5), 603-610.

Park, Y., Wang, F., Díaz-Cádiz, M.E., Vojtech, J.M., Groll, M.D., & Stepp, C.E. (Under Review). Vocal fold kinematics and relative fundamental frequency as a function of obstruent type and speaker age. *Journal of the Acoustical Society of America*.

Patel, R., Dailey, S., & Bless, D. (2008). Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders. *Annals of Otology, Rhinology, and Laryngology, 117*(6), 413-424.

Patel, R.R., Awan, S.N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Svec, J.G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American 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.

Patel, R.R., Forrest, K., & Hedges, D. (2017). Relationship between acoustic voice onset and offset and selected instances of oscillatory onset and offset in young healthy men and women. *Journal of Voice, 31*(3), 389.e9-89.e17.

Patel, R.R., Liu, L., Galatsanos, N., & Bless, D.M. (2011). Differential vibratory characteristics of adductor spasmodic dysphonia and muscle tension dysphonia on high-speed digital imaging. *Annals of Otology, Rhinology, and Laryngology, 120*(1), 21-32.

Pedrosa, V., Pontes, A., Pontes, P., Behlau, M., & Peccin, S.M. (2016). The effectiveness of the comprehensive voice rehabilitation program compared with the vocal function exercises method in behavioral dysphonia: A randomized clinical trial. *Journal of Voice, 30*(3), 377.e11-19.

Perez, K.S., Ramig, L.O., Smith, M.E., & Dromey, C. (1996). The Parkinson larynx: Tremor and videostroboscopic findings. *Journal of Voice, 10*(4), 354-361.

Perju-Dumbrava, L., Lau, K., Phyland, D., Papanikolaou, V., Finlay, P., Beare, R., Bardin, P., Stuckey, S., Kempster, P., & Thyagarajan, D. (2017). Arytenoid cartilage movements are hypokinetic in Parkinson's disease: A quantitative dynamic computerised tomographic study. *PLoS One, 12*(11), e0186611.

Peterson, E.A., Roy, N., Awan, S.N., Merrill, R.M., Banks, R., & Tanner, K. (2013). Toward validation of the cepstral spectral index of dysphonia (CSID) as an objective treatment outcomes measure. *Journal of Voice, 27*(4), 401-410.

Peterson, K.L., Verdolini-Marston, K., Barkmeier, J.M., & Hoffman, H.T. (1994). Comparison of aerodynamic and electroglottographic parameters in evaluating clinically relevant voicing patterns. *Annals of Otology, Rhinology, and Laryngology, 103*(5 Pt 1), 335-346.

Pinho, S.M.R., Pontes, P.A.L., Gadelha, M.E.C., & Biasi, N. (1999). Vestibular vocal fold behavior during phonation in unilateral vocal fold paralysis. *Journal of Voice, 13*(1), 36-42.

Plant, R.L., & Hillel, A.D. (1998). Direct measurement of subglottic pressure and laryngealresistance in normal subjects and in spasmodic dysphonia. *Journal of Voice, 12*(3), 300-314.

Plante, F., Meyer, G.F., & Ainsworth, W.A. (1995, Sept). A pitch extraction reference database. In *EUROSPEECH-1995*, pp. 837-840. Madrid, Spain. https://www.isca-speech.org/archive/eurospeech_1995/e95_0837.html

Poburka, B.J., Patel, R.R., & Bless, D.M. (2017). Voice-vibratory assessment with laryngeal imaging (VALI) form: Reliability of rating stroboscopy and high-speed videoendoscopy. *Journal of Voice, 31*(4), 513.e1-14.

Poletto, C.J., Verdun, L.P., Strominger, R., & Ludlow, C.L. (2004). Correspondence between laryngeal vocal fold movement and muscle activity during speech and nonspeech gestures. *Journal of Applied Physiology, 97*(3), 858-866.

Pontes, P., Brasolotto, A., & Behlau, M. (2005). Glottic characteristics and voice complaint in the elderly. *Journal of Voice, 19*(1), 84-94.

Popolo, P.S. (2017). Investigation of flexible high-speed video nasolaryngoscopy. *Journal of Voice, 32*(5), 529-537.

Popolo, P.S., Svec, J.G., & Titze, I.R. (2005). Adaptation of a pocket PC for use as a wearable voice dosimeter. *Journal of Speech, Language, and Hearing Research, 48*(4), 780-791.

Powell, M.E., Deliyski, D.D., Hillman, R.E., Zeitels, S.M., Burns, J.A., & Mehta, D.D. (2016). Comparison of videostroboscopy to stroboscopy derived from high-speed videoendoscopy for evaluating patients with vocal fold mass lesions. *American Journal of Speech-Language Pathology, 25*(4), 576-589.

Powell, M.E., Deliyski, D.D., Zeitels, S.M., Burns, J.A., Hillman, R.E., Gerlach, T.T., & Mehta, D.D. (2019). Efficacy of videostroboscopy and high-speed videoendoscopy to obtain functional outcomes from perioperative ratings in patients with vocal fold mass lesions. *Journal of Voice*, Advanced online publication.

Prochazka, A., Bennett, D.J., Stephens, M.J., Patrick, S.K., Sears-Duru, R., Roberts, T., & Jhamandas, J.H. (1997). Measurement of rigidity in Parkinson's disease. *Movement Disorders, 12*(1), 24-32.

Quatieri, T.F. (2008). *Discrete-Time Speech Signal Processing: Principles and Practice*. Saddle River, NJ, USA: Prentice Hall PTR.

Rabiner, L.R. (1977). Use of autocorrelation analysis for pitch detection. *IEEE Transactions on Signal Processing, 25*(1), 24-33.

Ramig, L.O., Fox, C., & Sapir, S. (2004). Parkinson's disease: Speech and voice disorders and their treatment with the Lee Silverman Voice Treatment. *Seminars in Speech and Language, 25*(2), 169-180.

Ramig, L.O., Fox, C., & Sapir, S. (2008). Speech treatment for Parkinson's disease. *Expert Review of Neurotherapeutics, 8*(2), 297-309.

Ramig, L.O., Sapir, S., Countryman, S., Pawlas, A.A., O'Brien, C., Hoehn, M., & Thompson, L.L. (2001). Intensive voice treatment (LSVT) for patients with Parkinson's disease: A 2 year follow up. *Journal of Neurology, Neurosurgery, and Psychiatry, 71*(4), 493-498.

Ramig, L.O., & Verdolini, K. (1998). Treatment efficacy: Voice disorders. *Journal of Speech, Language, and Hearing Research, 41*(1), S101-116.

Redenbaugh, M.A., & Reich, A.R. (1989). Surface EMG and related measures in normal and vocally hyperfunctional speakers. *Journal of Speech and Hearing Disorders, 54*(1), 68-73.

Reitermanova, Z. (2010, May). Data splitting. In *WDS'10 Proceedings of Contributed Papers: Part I – Mathematics and Computer Sciences*, pp. 31-36. Prague, Matfyzpress.

Richard, E.G., Robert, M.G., & William, W.M. (1999). Validation of a voice outcome survey for unilateral vocal cord paralysis. *Otolaryngology–Head and Neck Surgery, 120*(2), 153-158.

Robb, M.P., & Smith, A.B. (2002). Fundamental frequency onset and offset behavior: A comparative study of children and adults. *Journal of Speech, Language, and Hearing Research, 45*(3), 446-456.

Robbins, J.A., Logemann, J.A., & Kirshner, H.S. (1986). Swallowing and speech production in Parkinson's disease. *Annals of Neurology, 19*(3), 283-287.

Roberts, T.J., & Gabaldón, A.M. (2008). Interpreting muscle function from EMG: lessons learned from direct measurements of muscle force. *Integrative and Comparative Biology, 48*(2), 312-320.

Robichaud, J.A., Pfann, K.D., Leurgans, S., Vaillancourt, D.E., Comella, C.L., & Corcos, D.M. (2009). Variability of EMG patterns: A potential neurophysiological marker of Parkinson's disease? *Clinical Neurophysiology, 120*(2), 390-397.

Rodeño, M.T., Sánchez-Fernández, J.M., & Rivera-Pomar, J.M. (1993). Histochemical and morphometrical ageing changes in human vocal cord muscles. *Acta Oto-Laryngologica, 113*(3), 445-449.

Rosenthal, A.L., Lowell, S.Y., & Colton, R.H. (2014). Aerodynamic and acoustic features of vocal effort. *Journal of Voice, 28*(2), 144-153.

Rossi, B., Siciliano, G., Carboncini, M.C., Manca, M.L., Massetani, R., Viacava, P., & Muratorio, A. (1996). Muscle modifications in Parkinson's disease: Myoelectric manifestations. *Electroencephalography and Clinical Neurophysiology, 101*(3), 211-218.

Rothenberg, M., & Mahshie, J.J. (1988). Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech, Language, and Hearing Research, 31*(3), 338-351.

Roy, N. (2003). Functional dysphonia. *Current Opinion in Otolaryngology & Head and Neck Surgery, 11*(3), 144-148.

Roy, N. (2008). Assessment and treatment of musculoskeletal tension in hyperfunctional voice disorders. *International Journal of Speech-Language Pathology, 10*(4), 195-209.

Roy, N., Barkmeier-Kraemer, J., Eadie, T., Sivasankar, M.P., Mehta, D., Paul, D., & Hillman, R. (2013). Evidence-based clinical voice assessment: A systematic review. *American Journal of Speech-Language Pathology, 22*(2), 212-226.

Roy, N., Bless, D.M., Heisey, D., & Ford, C.N. (1997). Manual circumlaryngeal therapy for functional dysphonia: An evaluation of short- and long-term treatment outcomes. *Journal of Voice, 11*(3), 321-331.

Roy, N., Fetrow, R.A., Merrill, R.M., & Dromey, C. (2016). Exploring the clinical utility of relative fundamental frequency as an objective measure of vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 59*(5), 1002-1017.

Roy, N., Ford, C.N., & Bless, D.M. (1996). Muscle tension dysphonia and spasmodic dysphonia: The role of manual laryngeal tension reduction in diagnosis and management. *Annals of Otology, Rhinology, and Laryngology, 105*(11), 851-856.

Roy, N., Gray, S.D., Simon, M., Dove, H., Corbin-Lewis, K., & Stemple, J.C. (2001). An evaluation of the effects of two treatment approaches for teachers with voice disorders: A prospective randomized clinical trial. *Journal of Speech, Language, and Hearing Research, 44*(2), 286-296.

Roy, N., & Leeper, H.A. (1993). Effects of the manual laryngeal musculoskeletal tension reduction technique as a treatment for functional voice disorders: Perceptual and acoustic measures. *Journal of Voice, 7*(3), 242-249.

Roy, N., Merrill, R.M., Gray, S.D., & Smith, E.M. (2005). Voice disorders in the general population: Prevalence, risk factors, and occupational impact. *The Laryngoscope, 115*(11), 1988-1995.

Roy, N., Nissen, S.L., Dromey, C., & Sapir, S. (2009). Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy. *Journal of Communication Disorders, 42*(2), 124-135.

Roy, N., Weinrich, B., Gray, S.D., Tanner, K., Stemple, J.C., & Sapienza, C.M. (2003). Three treatments for teachers with voice disorders. *Journal of Speech, Language, and Hearing Research, 46*(3), 670-688.

Rubin, J.S., Blake, E., & Mathieson, L. (2007). Musculoskeletal patterns in patients with voice disorders. *Journal of Voice, 21*(4), 477-484.

Rubin, J.S., Lieberman, J., & Harris, T.M. (2000). Laryngeal manipulation. *Otolaryngologic Clinics of North America, 33*(5), 1017-1034.

Saffarian, A., Amiri Shavaki, Y., Shahidi, G.A., Hadavi, S., & Jafari, Z. (2019). Lee Silverman Voice Treatment (LSVT) mitigates voice difficulties in mild Parkinson's disease. *Medical Journal of the Islamic Republic of Iran, 33*, 5.

Salenius, S., Portin, K., Kajola, M., Salmelin, R., & Hari, R. (1997). Cortical control of human motoneuron firing during isometric contraction. *Journal of Neurophysiology, 77*(6), 3401-3405.

Samad, S.A., Hussain, A., & Fah, L.K. (2000, Sept). Pitch detection of speech signals using the cross-correlation technique. In *2000 TENCON Proceedings. Intelligent Systems and Technologies for the New Millennium*, pp. 283-286. Kuala Lumpur, Malaysia. https://doi.org/10.1109/TENCON.2000.893673

Samlan, R.A., Kunduk, M., Ikuma, T., Black, M., & Lane, C. (2018). Vocal fold vibration in older adults with and without age-related dysphonia. *American Journal of Speech-Language Pathology, 27*(3), 1039-1050.

Samlan, R.A., Story, B.H., & Bunton, K. (2013). Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling. *Journal of Speech, Language, and Hearing Research, 56*(4), 1209-1223.

Sapir, S., Ramig, L.O., & Fox, C.M. (2011). Intensive voice treatment in Parkinson's disease: Lee Silverman Voice Treatment. *Expert Review of Neurotherapeutics, 11*(6), 815-830.

Sataloff, R.T., Heman-Ackah, Y.D., & Hawkshaw, M.J. (2007). Clinical anatomy and physiology of the voice. *Otolaryngologic Clinics of North America, 40*(5), 909-929.

Schapira, A.H.V., Chaudhuri, K.R., & Jenner, P. (2017). Non-motor features of Parkinson disease. *Nature Reviews. Neuroscience, 18*(7), 435-450.

Scherer, R.C. (1991). *Aerodynamic Assessment In Voice Production*. Retrieved from http://www.ncvs.org/ProgressReports/NCVS%20Status%20&%20Progress%20Report%20Vol.%201,%20June%201991%20copy.pdf

Schow, R.L. (1991). Considerations in selecting and validating an adult/elderly hearing screening protocol. *Ear and Hearing, 12*(5), 337-348.

Schwartz, S.R., Cohen, S.M., Dailey, S.H., Rosenfeld, R.M., Deutsch, E.S., Gillespie, M.B., Granieri, E., Hapner, E.R., Kimball, C.E., Krouse, H.J., McMurray, J.S., Medina, S., O'Brien, K., Ouellette, D.R., Messinger-Rapport, B.J., Stachler, R.J., Strode, S., Thompson, D.M., Stemple, J.C., Willging, J.P., Cowley, T., McCoy, S., Bernad, P.G., & Patel, M.M. (2009). Clinical practice guideline: Hoarseness (dysphonia). *Otolaryngology–Head and Neck Surgery, 141*, 1-31.

Shahed, J., & Jankovic, J. (2007). Motor symptoms in Parkinson's disease. *Handbook of Clinical Neurology, 83*, 329-342.

Shiller, D.M., Laboissiere, R., & Ostry, D.J. (2002). Relationship between jaw stiffness and kinematic variability in speech. *Journal of Neurophysiology, 88*(5), 2329-2340.

Shim, H.-J., Jung, H., Koul, R., & Ko, D.-H. (2016). Spectral and cepstral based acoustic features of voices with muscle tension dysphonia. *Clinical Archives of Communication Disorders, 1*(1), 42-47.

Shrivastav, R., Eddins, D.A., & Anand, S. (2012). Pitch strength of normal and dysphonic voices. *Journal of the Acoustical Society of America, 131*(3), 2261-2269.

Shrivastav, R., & Sapienza, C.M. (2003). Objective measures of breathy voice quality obtained using an auditory model. *Journal of the Acoustical Society of America, 114*(4 Pt 1), 2217-2224.

Sidtis, J.J., Alken, A.G., Tagliati, M., Alterman, R., & Van Lancker Sidtis, D. (2016). Subthalamic stimulation reduces vowel space at the initiation of sustained production: Implications for articulatory motor control in Parkinson's disease. *Parkinson's Disease, 6*(2), 361-370.

Skodda, S., Grönheit, W., Mancinelli, N., & Schlegel, U. (2013). Progression of voice and speech impairment in the course of Parkinson's disease: A longitudinal study. *Parkinson's Disease, 2013*, 1-8.

Skodda, S., Grönheit, W., & Schlegel, U. (2012). Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease. *PLoS One, 7*(2), e32132.

Skodda, S., Visser, W., & Schlegel, U. (2010). Short- and long-term dopaminergic effects on dysarthria in early Parkinson's disease. *Journal of Neural Transmission, 117*(2), 197-205.

Solomon, N.P., & Hixon, T.J. (1993). Speech breathing in Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 36*(2), 294-310.

Sonninen, A., Hurme, P., & Laukkanen, A.M. (1999). The external frame function in the control of pitch, register, and singing mode: Radiographic observations of a female singer. *Journal of Voice, 13*(3), 319-340.

Spielman, J., Ramig, L.O., Mahler, L., Halpern, A., & Gavin, W.J. (2007). Effects of an extended version of the Lee Silverman Voice Treatment on voice and speech in Parkinson's disease. *American Journal of Speech-Language Pathology, 16*(2), 95-107.

Sprenger, F., & Poewe, W. (2013). Management of motor and non-motor symptoms in Parkinson's disease. *CNS Drugs, 27*(4), 259-272.

Stager, S.V., Bielamowicz, S.A., Regnell, J.R., Gupta, A., & Barkmeier, J.M. (2000). Supraglottic activity: Evidence of vocal hyperfunction or laryngeal articulation? *Journal of Speech, Language, and Hearing Research, 43*(1), 229-238.

Standring, S., Borley, N.R., & Gray, H. (2008). *Gray's Anatomy: The Anatomical Basis of Clinical Practice* (40 ed.). Edinburgh, Scotland: Churchill Livingstone/Elsevier.

Stebbins, G.T., & Goetz, C.G. (1998). Factor structure of the Unified Parkinson's Disease Rating Scale: Motor examination section. *Movement Disorders, 13*(4), 633-636.

Stelzig, Y., Hochhaus, W., Gall, V., & Henneberg, A. (1999). Kehlkopfbefunde bei patienten mit morbus Parkinson. *Laryngo-Rhino-Otologie, 78*(10), 544-551.

Stemple, J.C., Roy, N., & Klaben, B.K. (2018). *Clinical Voice Pathology: Theory and Management* (6 ed.). San Diego, CA, USA: Plural Publishing.

Stepp, C.E. (2013). Relative fundamental frequency during vocal onset and offset in older speakers with and without Parkinson's disease. *Journal of the Acoustical Society of America, 133*(3), 1637-1643.

Stepp, C.E., Heaton, J.T., Braden, M.N., Jette, M.E., Stadelman-Cohen, T.K., & Hillman, R.E. (2011a). Comparison of neck tension palpation rating systems with surface electromyographic and acoustic measures in vocal hyperfunction. *Journal of Voice, 25*(1), 67-75.

Stepp, C.E., Heaton, J.T., Jette, M.E., Burns, J.A., & Hillman, R.E. (2010a). Neck surface electromyography as a measure of vocal hyperfunction before and after injection laryngoplasty. *Annals of Otology, Rhinology, and Laryngology, 119*(9), 594-601.

Stepp, C.E., Heaton, J.T., Stadelman-Cohen, T.K., Braden, M.N., Jette, M.E., & Hillman, R.E. (2011b). Characteristics of phonatory function in singers and nonsingers with vocal fold nodules. *Journal of Voice, 25*(6), 714-724.

Stepp, C.E., Hillman, R.E., & Heaton, J.T. (2010b). The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research, 53*, 1220-1226.

Stepp, C.E., Hillman, R.E., & Heaton, J.T. (2010c). Use of neck strap muscle intermuscular coherence as an indicator of vocal hyperfunction. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 18*(3), 329-335.

Stepp, C.E., Hillman, R.E., & Heaton, J.T. (2010d). A virtual trajectory model predicts differences in vocal fold kinematics in individuals with vocal hyperfunction. *Journal of the Acoustical Society of America, 127*(5), 3166-3176.

Stepp, C.E., Hillman, R.E., & Heaton, J.T. (2011c). Modulation of neck intermuscular beta coherence during voice and speech production. *Journal of Speech, Language, and Hearing Research, 54*(3), 836-844.

Stepp, C.E., Merchant, G.R., Heaton, J.T., & Hillman, R.E. (2011d). Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 54*(5), 1260-1266.

Stepp, C.E., Sawin, D.E., & Eadie, T.L. (2012). The relationship between perception of vocal effort and relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research, 55*(6), 1887-1896.

Stevens, K.N. (1977). Physics of laryngeal behavior and larynx modes. *Phonetica, 34*(4), 264-279.

Stevens, K.N. (2005). The acoustic/articulatory interface. *Acoustical Science and Technology, 26*(5), 410-417.

Story, B.H. (2015). Mechanisms of Voice Production. In M. Redford (Ed.), *The Handbook of Speech Production* (pp. 34-58). West Sussex, UK: John Wiley and Sons.

Suárez-Quintanilla, J., Fernández Cabrera, A., & Sharma, S. (2019). Anatomy, Head and Neck, Larynx. In StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK538202.

Sukhostat, L., & Imamverdiyev, Y. (2015). A comparative analysis of pitch detection methods under the influence of different noise conditions. *Journal of Voice, 29*(4), 410-417.

Švec, J.G., & Granqvist, S. (2018). Tutorial and guidelines on measurement of sound pressure level in voice and speech. *Journal of Speech, Language, and Hearing Research, 61*(3), 441-461.

Švec, J.G., Titze, I.R., & Popolo, P.S. (2005). Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *Journal of the Acoustical Society of America, 117*(3 Pt 1), 1386-1394.

Svensson, P., Henningson, C., & Karlsson, S. (1993). Speech motor control in Parkinson's disease: A comparison between a clinical assessment protocol and a quantitative analysis of mandibular movements. *Folia Phoniatrica et Logopaedica, 45*(4), 157-164.

Swee, T.T., Salleh, S.H.S., & Jamaludin, M.R. et al. (2010). Speech pitch detection using short-time energy. In *International Conference on Computer and Communication Engineering (ICCCE'10)*, (pp. 1-6). Kuala Lumpur, Malaysia. https://doi.org/10.1109/ICCCE.2010.5556836

Szabo, A., Hammarberg, B., Hakansson, A., & Sodersten, M. (2001). A voice accumulator device: Evaluation based on studio and field recordings. *Logopedics Phoniatrics Vocology, 26*(3), 102-117.

Szkiełkowska, A., Krasnodębska, P., Miaśkiewicz, B., & Skarżyński, H. (2018). Electroglottography in the diagnosis of functional dysphonia. *European Archives of Oto-Rhino-Laryngology, 275*(10), 2523-2528.

Takano, S., Kimura, M., Nito, T., Imagawa, H., Sakakibara, K.-I., & Tayama, N. (2010). Clinical analysis of presbylarynx—vocal fold atrophy in elderly individuals. *Auris Nasus Larynx, 37*(4), 461-464.

Talkin, D. (1995). A Robust Algorithm for Pitch Tracking (RAPT). In W.B. Kleijn & K.K. Paliwal (Eds.), *Speech Coding and Synthesis* (pp. 495-518). New York, NY, USA: Elsevier Science.

Thomas, L.B., & Stemple, J. (2007). Voice therapy: Does science support the art? *Communicative Disorders Review, 1*, 49-77.

Titze, I.R. (1994). *Principles of Voice Production*. Englewood Cliffs, NJ, USA: Prentice Hall.

Titze, I.R. (1995). *Workshop on Acoustic Voice Analysis: Summary Statement*. Paper presented at the National Center for Voice and Speech, (pp. 1-36). Denver, CO, USA. http://www.ncvs.org/freebooks/WorkshopOnAcousticVoiceAnalysisProceedings_1995.pdf

Titze, I.R. (2006). Voice training and therapy with a semi-occluded vocal tract: Rationale and scientific underpinnings. *Journal of Speech, Language, and Hearing Research, 49*(2), 448-459.

Titze, I.R., Luschei, E.S., & Hirano, M. (1989). Role of the thyroarytenoid muscle in regulation of fundamental frequency. *Journal of Voice, 3*(3), 213-224.

Titze, I.R., Svec, J.G., & Popolo, P.S. (2003). Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues. *Journal of Speech, Language, and Hearing Research, 46*(4), 919-932.

Tiwari, M., & Tiwari, M. (2012). Voice - how humans communicate? *Journal of Natural Science, Biology, and Medicine, 3*(1), 3-11.

Tourville, J.A., & Guenther, F.H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26*(7), 952-981.

Tripoliti, E., Zrinzo, L., Martinez-Torres, I., Frost, E., Pinto, S., Foltynie, T., Holl, E., Petersen, E., Roughton, M., Hariz, M.I., & Limousin, P. (2011). Effects of subthalamic stimulation on speech of consecutive patients with Parkinson disease. *Neurology, 76*(1), 80-86.

Tsanas, A., Zañartu, M., Little, M.A., Fox, C., Ramig, L.O., & Clifford, G.D. (2014). Robust fundamental frequency estimation in sustained vowels: Detailed

algorithmic comparisons and information fusion with adaptive Kalman filtering. *Journal of the Acoustical Society of America, 135*(5), 2885-2901.

Tsuji, D.H., Hachiya, A., Dajer, M.E., Ishikawa, C.C., Takahashi, M.T., & Montagnoli, A.N. (2014). Improvement of vocal pathologies diagnosis using high-speed videolaryngoscopy. *International Archives of Otorhinolaryngology, 18*(3), 294-302.

Tykalová, T., Rusz, J., Čmejla, R., Klempíř, J., Růžičková, H., Roth, J., & Růžička, E. (2015). Effect of dopaminergic medication on speech dysfluency in Parkinson's disease: A longitudinal study. *Journal of Neural Transmission, 122*(8), 1135-1142.

Ueda, N., Oyama, M., Harvey, J.E., & Ogura, J.H. (1972). Influence of certain extrinsic laryngeal muscles on artificial voice production. *The Laryngoscope, 82*(3), 468-482.

Vahabzadeh-Hagh, A.M., Zhang, Z., & Chhetri, D.K. (2018). Hirano's cover-body model and its unique laryngeal postures revisited. *The Laryngoscope, 128*(6), 1412-1418.

Van Den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. *Journal of Speech, Language, and Hearing Research, 1*(3), 227-244.

Van Den Eeden, S.K., Tanner, C.M., Bernstein, A.L., Fross, R.D., Leimpeter, A., Bloch, D.A., & Nelson, L.M. (2003). Incidence of Parkinson's disease: Variation by age, gender, and race/ethnicity. *American Journal of Epidemiology, 157*(11), 1015-1022.

Van Houtte, E., Claeys, S., D'haeseleer, E., Wuyts, F., & Van Lierde, K. (2013). An examination of surface EMG for the assessment of muscle tension dysphonia. *Journal of Voice, 27*(2), 177-186.

van Leer, E., & Connor, N.P. (2015). Predicting and influencing voice therapy adherence using social-cognitive factors and mobile video. *American Journal of Speech-Language Pathology, 24*(2), 164-176.

Van Lierde, K.M., De Bodt, M., Dhaeseleer, E., Wuyts, F., & Claeys, S. (2010). The treatment of muscle tension dysphonia: A comparison of two treatment techniques by means of an objective multiparameter approach. *Journal of Voice, 24*(3), 294-301.

Van Stan, J.H., Mehta, D.D., Ortiz, A.J., Burns, J.A., Toles, L.E., Marks, K.L., Vangel, M., Hron, T., Zeitels, S., & Hillman, R.E. (2020). Differences in weeklong ambulatory vocal behavior between female patients with phonotraumatic lesions

and matched controls. *Journal of Speech, Language, and Hearing Research, 63*(2), 372-384.

Van Stan, J.H., Mehta, D.D., Zeitels, S.M., Burns, J.A., Barbu, A.M., & Hillman, R.E. (2015a). Average ambulatory measures of sound pressure level, fundamental frequency, and vocal dose do not differ between adult females with phonotraumatic lesions and matched control subjects. *Annals of Otology, Rhinology, and Laryngology, 124*(11), 864-874.

Van Stan, J.H., Roy, N., Awan, S., Stemple, J., & Hillman, R.E. (2015b). A taxonomy of voice therapy. *American Journal of Speech-Language Pathology, 24*(2), 101-215.

Vercueil, L., Linard, J.P., Wuyam, B., Pollak, P., & Benchetrit, G. (1999). Breathing pattern in patients with Parkinson's disease. *Respiration Physiology, 118*(2-3), 163-172.

Verdolini, K., Chan, R., Titze, I.R., Hess, M., & Bierhals, W. (1998). Correspondence of electroglottographic closed quotient to vocal fold impact stress in excised canine larynges. *Journal of Voice, 12*(4), 415-423.

Verdolini, K., & Ramig, L.O. (2001). Review: Occupational risks for voice problems. *Logopedics Phoniatrics Vocology, 26*(1), 37-46.

Vojtech, J.M., & Heller Murray, E.S. (2019a, Jan. 8, 2019). Tutorial for Manual Relative Fundamental Frequency (RFF) Estimation using Praat. Retrieved from https://sites.bu.edu/stepplab/research/rff/

Vojtech, J.M., Segina, R.K., Buckley, D.P., Kolin, K.R., Tardif, M.C., Noordzij, J.P., & Stepp, C.E. (2019b). Refining algorithmic estimation of relative fundamental frequency: Accounting for sample characteristics and fundamental frequency estimation method. *Journal of the Acoustical Society of America, 146*(5), 3184.

Von Doersten, P.G., Izdebski, K., Ross, J.C., & Cruz, R.M. (1992). Ventricular dysphonia: A profile of 40 cases. *The Laryngoscope, 102*(11), 1296-1301.

Watson, B.C. (1998). Fundamental frequency during phonetically governed devoicing in normal young and aged speakers. *Journal of the Acoustical Society of America, 103*(6), 3642-3647.

Watson, B.C., Roark, R.M., & Baken, R.J. (2012). Vocal release time: A quantification of vocal offset. *Journal of Voice, 26*(6), 682-687.

Watson, P.J., & Schlauch, R.S. (2008). The effect of fundamental frequency on the intelligibility of speech with flattened intonation contours. *American Journal of Speech-Language Pathology, 17*(4), 348-355.

Watts, C.R., & Awan, S.N. (2011). Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *Journal of Speech, Language, and Hearing Research, 54*(6), 1525-1537.

Watts, C.R., & Awan, S.N. (2015). An examination of variations in the cepstral spectral index of dysphonia across a single breath group in connected speech. *Journal of Voice, 29*(1), 26-34.

Watts, C.R., Hamilton, A., Toles, L., Childs, L., & Mau, T. (2019). Intervention outcomes of two treatments for muscle tension dysphonia: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research, 62*(2), 272-282.

Watts, R.L., Wiegner, A.W., & Young, R.R. (1986). Elastic properties of muscles measured at the elbow in man: II. Patients with Parkinsonian rigidity. *Journal of Neurology, Neurosurgery, and Psychiatry, 49*(10), 1177-1181.

Whitfield, J.A., & Goberman, A.M. (2014). Articulatory–acoustic vowel space: Application to clear speech in individuals with Parkinson's disease. *Journal of Communication Disorders, 51*, 19-28.

Wight, S., & Miller, N. (2015). Lee Silverman Voice Treatment for people with Parkinson's: Audit of outcomes in a routine clinic. *International Journal of Language & Communication Disorders, 50*(2), 215-225.

Witte, R.S., & Witte, J.S. (2010). *Statistics*. Hoboken, NJ, USA: Wiley.

Wolfe, V.I., Garvin, J.S., Bacon, M., & Waldrop, W. (1975). Speech changes in Parkinson's disease during treatment with L-dopa. *Journal of Communication Disorders, 8*(3), 271-279.

Woo, P. (2009). *Stroboscopy*. San Diego, CA, USA: Plural Publishing.

Woo, P., Colton, R., Casper, J., & Brewer, D. (1991). Diagnostic value of stroboscopic examination in hoarse patients. *Journal of Voice, 5*(3), 231-238.

Woźnicka, E., Niebudek-Bogusz, E., Morawska, J., Wiktorowicz, J., & Śliwińska-Kowalska, M. (2017). Laryngeal manual therapy palpatory evaluation scale: A preliminary study to examine its usefulness in diagnosis of occupational dysphonia. *Medycyna Pracy, 68*(2), 179-188.

Wuyts, F.L., De Bodt, M.S., & Van de Heyning, P.H. (1999). Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *Journal of Voice, 13*(4), 508-517.

Yin, J., & Zhang, Z. (2013). The influence of thyroarytenoid and cricothyroid muscle activation on vocal fold stiffness and eigenfrequencies. *Journal of the Acoustical Society of America, 133*(5), 2972-2983.

Yorkston, K.M., Beukelman, D.R., & Traynor, C. (1984). *Assessment of Intelligibility of Dysarthric Speech*. Austin, TX, USA: Pro-ed.

Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32-35.

Young, B., Matsuzaki, H., & Sasaki, C.T. (2015). Anatomy of the Larynx. In M.P. Fried & M. Tan (Eds.), Clinical Laryngology (2015 ed., pp. 1-8). Stuttgart, Germany: Georg Thieme Verlag.

Yuceturk, A.V., Yilmaz, H., Egrilmez, M., & Karaca, S. (2002). Voice analysis and videolaryngostroboscopy in patients with Parkinson's disease. *European Archives of Oto-Rhino-Laryngology, 259*(6), 290-293.

Zacharias, S.R.C., Deliyski, D.D., & Gerlach, T.T. (2018). Utility of laryngeal high-speed videoendoscopy in clinical voice assessment. *Journal of Voice, 32*(2), 216-220.

Zanartu, M., Galindo, G.E., Erath, B.D., Peterson, S.D., Wodicka, G.R., & Hillman, R.E. (2014). Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction. *Journal of the Acoustical Society of America, 136*(6), 3262.

Zanartu, M., Ho, J.C., Kraman, S.S., Pasterkamp, H., Huber, J.E., & Wodicka, G.R. (2009). Air-borne and tissue-borne sensitivities of bioacoustic sensors used on the skin surface. *IEEE Transactions on Biomedical Engineering, 56*(2), 443-451.

Zarzur, A.P., Duprat, A.C., Cataldo, B.O., Ciampi, D., & Fonoff, E. (2013). Laryngeal electromyography as a diagnostic tool for Parkinson's disease. *The Laryngoscope, 124*(3), 725-729.

Zarzur, A.P., Duprat, A.C., Shinzato, G., & Eckley, C.A. (2007). Laryngeal electromyography in adults with Parkinson's disease and voice complaints. *The Laryngoscope, 117*(5), 831-834.

Zeitels, S.M., Burns, J.A., Lopez-Guerra, G., Anderson, R.R., & Hillman, R.E. (2008). Photoangiolytic laser treatment of early glottic cancer: A new management strategy. *Annals of Otology, Rhinology, and Laryngology, 199*, 3-24.

Zhang, Z. (2016). Mechanics of human voice production and control. *Journal of the Acoustical Society of America, 140*(4), 2614-2614.

Zheng, Y.Q., Zhang, B.R., Su, W.Y., Gong, J., Yuan, M.Q., Ding, Y.L., & Rao, S.Q. (2012). Laryngeal aerodynamic analysis in assisting with the diagnosis of muscle tension dysphonia. *Journal of Voice, 26*(2), 177-181.

Ziegler, A., Verdolini Abbott, K., Johns, M., Klein, A., & Hapner, E.R. (2014). Preliminary data on two voice therapy interventions in the treatment of presbyphonia. *The Laryngoscope, 124*(8), 1869-1876.

Zraick, R.I., Kempster, G.B., Connor, N.P., Thibeault, S., Klaben, B.K., Bursac, Z., Thrush, C.R., & Glaze, L.E. (2011). Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *American Journal of Speech-Language Pathology, 20*(1), 14-22.

Zraick, R.I., Smith-Olinde, L., & Shotts, L.L. (2012). Adult normative data for the KayPENTAX Phonatory Aerodynamic System Model 6600. *Journal of Voice, 26*(2), 164-176.

Zwirner, P., & Barnes Gary, J. (1992). Vocal tract steadiness. *Journal of Speech, Language, and Hearing Research, 35*(4), 761-768.

Zwirner, P., Murry, T., & Woodson, G.E. (1991). Phonatory function of neurologically impaired patients. *Journal of Communication Disorders, 24*(4), 287-300.

# CURRICULUM VITAE