2019

Respiratory Muscle Strength Training to Improve Vocal Function of Patients with Presbyphonia

Maude Desjardins
Medical University of South Carolina

RESPIRATORY MUSCLE STRENGTH TRAINING TO IMPROVE VOCAL FUNCTION OF PATIENTS WITH PRESBYPHONIA

BY

Maude Desjardins

A dissertation submitted to the faculty of the Medical University of South Carolina in partial fulfillment of the requirement for the degree
Doctor of Philosophy
in the College of Health Professions

Dedications and Acknowledgement

This research was funded in part by a pilot grant from the Center on Aging of the Medical University of South Carolina, through a project entitled: “Respiratory Muscle Strength Training in Presbyphonia”.

I would like to individually thank each member of my committee for their valued support throughout my dissertation process. First and foremost, I thank my mentor, academic advisor, and Committee Chair Dr. Heather Bonilha, for her dedicated mentorship during the past four years. Dr. Bonilha has offered a guidance that allowed me to develop my critical thinking, my problem-solving abilities, my expertise in voice disorders, as well as my professional skills. She has provided the perfect balance between accompanying me in my academic and professional development and teaching me to become an independent researcher. Dr. Bonilha was always open-minded to my ideas and creativity, while at the same time ensuring that I was heading in the right direction. This made the whole process motivating, challenging, and rewarding, and at the same time I felt extremely well supported. I would like to thank Dr. Lucinda Halstead for having taken part in the project with passion and enthusiasm. She played a major role in recruiting the participants and the study couldn’t have been possible without her support. Dr. Halstead has shared her expertise in voice disorders with great generosity during my four years in the program, and I am extremely grateful to have studied and worked alongside such a passionate, experienced, and knowledgeable laryngologist. I would like to thank Dr. Annie Simpson for having devoted so much of her time in guiding me through the statistical analyses. Dr. Simpson made herself very available for accompanying me in the data analysis process and she helped me combine clinical thinking and statistical analyses. She was always willing to brainstorm and discuss the results to

help me “make the data talk”. I would like to thank Dr. Patrick Flume for having accepted to be part of this committee as the pulmonary expert. Despite being extremely busy, Dr. Flume provided clinical guidance regarding the spirometry results of the participants. He also met with me to discuss the results from the study regarding respiratory function and to guide me through their interpretation. I thank him for sharing his expertise, which was of great value for this project.

I would also like to acknowledge the work of Ms. Krysten Sears who provided standard of care voice therapy to the study participants. She fulfilled this role with passion, rigor, and positivity, and was always willing to help troubleshoot any unanticipated obstacle. She played an important role in the study, and I would like to thank her for that. I would also like to acknowledge her work, as well as the work of Mrs. Melissa Cooke, Ms. Angelina Schache, and Mrs. Julie Blair, as raters for the subjective voice measures.

I would like to give a special acknowledgement to Dr. Kit Simpson, who offered her precious advice and expertise regarding qualitative and exploratory analyses. I thank her for sharing her experience and wisdom with me, and for being of great help in a moment when I needed it the most.

Special thanks to my Montreal family, Louise, Robert and Audrey, and to my Charleston family, Corey, Patty, Janina, Melanie, Alison, Haydn, and Kate, for your love and support.

RESPIRATORY MUSCLE STRENGTH TRAINING TO IMPROVE VOCAL FUNCTION OF PATIENTS WITH PRESBYPHONIA

BY

Maude Desjardins

Approved by:

Chair, Project Committee	Dr. Heather Shaw Bonilha, PhD, CCC-SLP	7/24/19 Date
Member, Project Committee	Dr. Lucinda Halstead, MD	7/23/19 Date
Member, Project Committee	Dr. Annie Simpson, PhD	7/23/19 Date
Member, Project Committee	Dr. Patrick Flume, MD	7/22/19 Date
Dean, College of Health Professions	Dr. Zoher F. Kapasi, PT, Ph.D., MBA	7/25/19 Date

Abstract of Dissertation Presented to the
Doctor of Philosophy Program in Health and Rehabilitation Sciences
Medical University of South Carolina
In Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

RESPIRATORY MUSCLE STRENGTH TRAINING TO IMPROVE VOCAL FUNCTION OF PATIENTS WITH PRESBYPHONIA

By

Maude Desjardins

Chairperson: Dr. Heather Shaw Bonilha, PhD, CCC-SLP
Committee: Dr. Lucinda Halstead, MD
Dr. Annie Simpson, PhD
Dr. Patrick Flume, MD

Body of Abstract

Background: Presbyphonia is an age-related voice disorder characterized by vocal fold atrophy, and its effects on voice are potentially compounded by declines in respiratory function. We assessed: 1) the relationships between respiratory and voice function; 2) the effect of adding respiratory exercises to voice therapy; and 3) the impact of baseline respiratory function on the response to therapy in patients with presbyphonia.
Methods: Twenty-one participants underwent respiratory and voice assessments, from which relationships were drawn. Ten of these participants were blocked-randomized to receive either voice exercises only, or voice exercises combined with inspiratory muscle strength training or expiratory muscle strength training, for a duration of four weeks.
Results: FVC, FEV1, and MEP had an impact on phonation physiology through their effect on aerodynamic resistance and vocal fold pliability. Percent predicted values of FVC and FEV1 were strong predictors of perceived voice handicap. IMST induced the largest improvements in perceived handicap, and a lower baseline respiratory function was associated with a greater improvement, regardless of the intervention received.
Conclusion: Respiratory function impacts voice and the response to behavioral voice therapy. Adding IMST to voice exercises improves self-reported outcomes even in patients with a normal respiratory function.

Table of Content

List of Figures

List of Tables

Chapter 1: Introduction

Background and Significance

Voice Disorders

Phonation is the coordinated activity of the respiratory, laryngeal and neurological systems that creates a sound. When combined with the action of the articulatory system, phonation allows individuals to communicate through the production of words and sentences and is thereby at the center of most social interactions. As it is the case for all organs of the human body, those involved in voicing are subject to injury, disease or degeneration, which may result in a voice disorder, termed dysphonia. Voice disorders affect the quality, pitch and/or loudness of the voice and can hinder communication activities to various degrees depending on the severity of the disorder and the extent to which an individual relies on their voice.

Approximately one out of three individuals experience dysphonia throughout their life (Cohen, Kim, Roy, Asche, & Courey, 2012a, 2012b; Roy, Merrill, Gray, & Smith, 2005), with a point prevalence between 6.6% and 7.5% in the adult US population (Cohen, 2010; Roy et al., 2005). Based on these prevalence rates and on the adult population of the United States for the year of 2017 (244,340,854 individuals aged over 20 years old), between 16,126,496 and 18,325,564 Americans are currently experiencing dysphonia. With a direct health care cost for evaluation and management of voice disorders estimated at $577.18 per person over a period of one year

(Cohen et al., 2012a), the potential burden on the health care system ranges between 9 billion and 10.5 billion dollars. Cohen et al. (Cohen et al., 2012a) reported similar direct costs and observed that they are comparable to major chronic illnesses such as asthma, diabetes and chronic obstructive pulmonary disease (COPD). If not treated, voice disorders can impact quality of life by hindering work abilities and by leading to social isolation, depression and increased work absenteeism (Cohen et al., 2012b; Roy et al., 2005; Smith et al., 1996) Voice therapy with a speech language pathologist (SLP) is often the primary approach used with dysphonic patients. Depending on the etiology of the voice disorder, it can also be recommended along with surgery and/or pharmacological treatment (Cohen, Dinan, Kim, & Roy, 2015). Voice therapy aims at optimizing voice production and limiting maladaptive compensatory mechanisms. The SLP recommends and teaches exercises targeting the laryngeal, resonance and/or respiratory system. While the body of literature on voice therapy pertaining to the laryngeal and resonance systems is existing and growing (Desjardins, Halstead, Cooke, & Bonilha, 2017), there is a lack of evidence supporting the use of respiratory exercises with dysphonic patients.

Role of the Respiratory System in Voice Disorders

Pathophysiology

It is well established that the respiratory system plays a crucial role in the production of voice. The driving pressure arising from the lungs overcomes the resistance of the vocal folds and pulls them apart from each other and this sets off the vibratory process (Mathieson, 2006). The pressures generated by the lung-thorax unit are also involved in loudness and pitch adjustments. There are two ways of increasing pitch and loudness: 1) with fine-tuning of the tension, stiffness, length and mass (thickness) of the vocal folds (Zhang, 2016b) and 2) by altering the driving pressure generated by the lung-thorax unit (Zhang, 2016b). By increasing the air pressure coming

from the lungs and adjusting the tension of the vocal folds, the subglottal pressure is enhanced and this leads to a larger amplitude of vibration and a faster and stronger return to the initial position by the tensed vocal folds (Mathieson, 2006; McFarland, 2008). This also has the effect of augmenting the pitch of the sound because of the increased velocity of vibration. The respiratory muscles also play a role in airflow management by counteracting the recoil pressures of the lungs to slow the expiratory phase of speech, thereby allowing for the production of longer utterances and constant pressure.

While the normal physiology of respiration and phonation and the relationship between the two have been extensively described in the literature, there is a scarcity of data pertaining to the impact of a suboptimal respiratory system on voice production. Impaired respiratory function can be induced by compromised pulmonary function because of weak respiratory muscles or conditions such as asthma or COPD. An impaired respiratory function can influence spirometry outcomes, but the extent to which it may influence voice outcomes remains unclear. The respiratory and laryngeal systems are coupled and impairment in one of them may lead to dysfunction of the other one (Zhang, 2016c). If the respiratory system is not strong enough to generate sufficient air pressure, one has to rely mainly on laryngeal adjustments to build up substantial subglottic pressure and to increase loudness by augmenting the duration of contact and the medial compression of the vocal folds. Similarly, if the respiratory muscles are not strong enough to control the expiratory airflow during speech, the airflow conservation role relies solely on the vocal folds acting as a valve (Zhang, 2016c). Vaca et al. (Vaca, Mora, & Cobeta, 2015) found that elderly patients presenting with a laryngeal deficit (glottal gap) in addition to respiratory deficits (vital capacity or peak expiratory airflow below 80% of the reference value) had shorter phonation times, increased jitter, worse auditory perceptual ratings and higher Voice Handicap

Index scores than the groups with deficits in only the laryngeal or respiratory system. The impaired respiratory system could not compensate for the poor laryngeal valving function, thereby explaining the inferior voice outcomes in this subgroup.

Intervention

Exercises to increase breath support are commonly used by SLPs with a variety of voice patients, although common practices pertaining to breathing exercises appear to differ between clinics and clinicians (J. L. Gartner-Schmidt, Roth, Zullo, & Rosen, 2013; Sellars, Carding, Deary, MacKenzie, & Wilson, 2006). One possible reason for this variability is the paucity of evidence related to breathing exercises in the voice literature (Desjardins & Bonilha, 2019), despite the well documented relationship between the respiratory and laryngeal systems. Exercises recommended in certain voice therapy books include but are not limited to: diaphragmatic breathing, increasing extent of thoracic expansion, increasing period of rib elevation, and increasing period of expiratory airflow on phonemes /s, z, a, æ, i/(Boone, 2010; Mathieson, 2006). These exercises have seldom been tested individually in clinical trials with voice patients and their specific effects on respiratory and voice outcomes remains unclear (Desjardins & Bonilha, 2019). It is likely that most of these exercises are not intensive enough to induce changes in the respiratory system and impact voice parameters.

A more intensive intervention, respiratory muscle strength training (RMST) with a pressure threshold loading device, has received more attention in the literature. This type of intervention can be used to strengthen either the inspiratory muscles, the expiratory muscles, or both. The effect of expiratory muscle strength training (EMST) has been studied mostly on patients with important respiratory muscle weakness, as it is the case in patients with multiple sclerosis and Parkinson's disease (Chiara, Martin, & Sapienza, 2007; Silverman et al., 2006). Eight studies have assessed the effect of EMST on various populations, and improvements were found for at least

some participants in sound pressure level (SPL), subglottal pressure (Psub), words per minute, utterance length, coefficient of variation of fundamental frequency (F0), Voice Handicap Index 10 (VHI-10) scores, pitch range, and /s,z/ durations (Cerny, Panzarella, & Stathopoulos, 1997; Chiara et al., 2007; Darling-White & Huber, 2017; Johansson, 2012; Pereira, 2015; Ray, 2018; Tsai et al., 2016). Moreover, improved self-assessment scores, subglottal pressure at loud intensity, dynamic range, and vocal fold edges were found when combining voice therapy and EMST in professional voice users with laryngeal irritation, edema, and/or benign vocal fold lesions (Wingate, Brown, Shrivastav, Davenport, & Sapienza, 2007). Regarding inspiratory muscle strength training (IMST), most studies in the literature are case series targeting patients with upper airway obstruction diseases such as paradoxical vocal fold movement disorder, recurrent laryngeal papilloma and bilateral abductor vocal fold paralysis (S. E. Baker, Sapienza, & Collins, 2003; Mathers-Schmidt & Brilla, 2005; C. M. Sapienza, Brown, Martin, & Davenport, 1999). Only two studies have assessed the effect of IMST with regards to vocal quality; however, one of them did not assess the significance of pre-post intervention changes (Mueller, 2013). The other study, a single-subject cross-over trial, did not find improvement during the IMST phase in the two voice outcomes measured, SPL and pitch range (Ray, 2018).

Surprisingly, no studies on respiratory muscle training has been conducted on patients with an age-related voice disorder, in whom both the laryngeal and respiratory systems undergo significant changes affecting voice production.

Voice Disorders in the Elderly

Prevalence

Elderly people are at risk of experiencing reduced laryngeal and respiratory efficiency because of the changes that occur in the body with aging and that affect the systems involved in phonation. They represent an ideal subgroup of patients to study the impact of the respiratory system on voice outcomes because of the variability in the aging process and comorbid conditions impacting the laryngeal and respiratory systems. As a consequence, individuals are likely to present with various laryngeal and respiratory profiles, making it possible to assess their relative contribution on the resulting voice disorder.

The prevalence of dysphonia increases with age and peaks in individuals of 70 years and older (Cohen et al., 2012b). In a study using data from 54,600,465 individuals enrolled in a national claims database, people over 60 years old represented 27.8% of the total dysphonic population (Cohen et al., 2012b), corresponding to 149,270 individuals over a period of five years. The number of elderly people actually suffering from dysphonia is thought to be much higher. In fact, one out of five individuals over 65 years old experience difficulties with their voice (Golub, Chen, Otto, Hapner, & Johns, 2006), but less than a quarter of them consult a physician or a voice specialist (Turley & Cohen, 2009).

In a retrospective chart review of geriatric voice referrals at the Emory Voice Center, Davids et al. (Davids, Klein, & Johns, 2012) found that the most common diagnostic was vocal fold atrophy, accounting for a quarter of the voice diagnoses in this patient group. Another study using ICD-9-CM codes revealed that the most common laryngeal/voice disorder among older patients was "nonspecific hoarseness" (ICD-9-CM codes: 784.49, 784.42, 784.40, 784.41). Nonspecific causes of dysphonia were observed to increase in a relatively linear fashion with age, representing 49.24% of the diagnoses in the 65-69 bracket and 57.14% of the diagnoses in the >95 years old

bracket (Roy, Kim, Courey, & Cohen, 2015). The authors suggested that these nonspecific codes were likely to be used in cases of presbyphonia since there exist no code for this condition (Roy et al., 2015). Presbyphonia is a voice disorder that arises from age-related changes in the laryngeal, respiratory and neurological systems. It is often diagnosed based on the presence of atrophied and bowed vocal folds occasioning a glottal gap and prominent vocal processes. Compensatory involvement of the false vocal folds to aide in glottal closure may also be observed during videostroboscopy assessment. Patients with presbyphonia often complain of vocal fatigue, hoarseness, breathiness and reduced loudness. Although the etiology of presbyphonia stems from normal aging processes, elderly patients often present with comorbid conditions that can further impact voice production. Woo et al. found that more than 50% of their sample of 151 older voice patients had at least one systemic condition, the most common being pulmonary diseases (Woo, Casper, Colton, & Brewer, 1992). Moreover, Gregory et al. (Gregory, Chandran, Lurie, & Sataloff, 2012) found that 91% of their sample of 159 geriatric voice patients presented with laryngopharyngeal reflux (LPR).

Impacts of Presbyphonia

In elderly individuals, for whom health-related quality of life may already be declining due to various chronic conditions, the presence of a voice disorder further participates in the decline (Davids et al., 2012). Voice-related quality of life (V-RQOL) was measured in a sample of 107 individuals living in a senior independent living community in Atlanta, Georgia. Of those, more than 50% presented V-RQOL scores indicative of an altered qualify of life due to voice problems (Golub et al., 2006). Voice disorders in the elderly population can lead to social isolation and psychological repercussions such as increased anxiety and depression and reduced self-esteem (Davids et al., 2012; Gregory et al., 2012). Voice difficulties may prevent individuals from

communicating their needs to caregivers and induce frustration for the patient, caregiver and other communication partners (Davids et al., 2012). In the context where the individual and their partner are often experiencing hearing loss, the presence of a voice disorder may further hinder communication and negatively impact quality of life of both spouses (Cohen & Turley, 2009).

Why is it important to find a better intervention paradigm?

The number of patients seeking treatment for vocal fold atrophy has been gradually increasing in the past years and this trend is expected to be maintained in the future (Takano et al., 2010). The first reason for this upsurge is the rapid expansion of the elderly population in the United States. Individuals of 65 years and older actually represent 15.6% of the total national population and this percentage is anticipated to reach 20.6% in 2030, based on the predictions of the United States Census Bureau.

The second factor is the increase in labor force participation rate of elderly individuals. For the population 65 years and older, this rate has raised from 12.1% in 1990 to 16.1% in 2010, with a more significant increase for women (Kromer & Howard, 2013). Different reasons can explain the choice to continue working past the traditional working age. Motives may include financial responsibilities and the desire and ability to remain active longer due to enhanced medical technology and a resulting extended lifespan (Gregory et al., 2012). In a 7-year longitudinal study, Takano et al. (Takano et al., 2010) found that 33% of patients diagnosed with vocal fold atrophy were still active in the workforce and, therefore, were likely to rely on their voice for professional matters.

Problem Statement

Intervention: What We Know and What We Don't Know

The growth in the elderly population combined with their increased labor force participation rate is expected to increase the number of geriatric patients complaining of voice difficulties (Davids et al., 2012). This is likely to burden the health care system and contribute to the "2030 challenge" of ensuring necessary resources and effective services for the rising elderly population (Roy et al., 2015). Surgery is one treatment option for vocal fold atrophy, but is not always recommended for older people who are more at risk for complications. Moreover, a health care cost analysis of voice and laryngeal disorder management revealed that surgery represents 34.37% of the total procedure claims (Cohen et al., 2012a). In addition to being costly, surgery for vocal fold atrophy often yields suboptimal outcomes with a success rate that has been evaluated at 56% in elderly patients undergoing injection laryngoplasty, based on their voice handicap self-rated score (J. Gartner-Schmidt & Rosen, 2011). On the other hand, there is a lack of evidence supporting voice therapy for presbyphonia. Although some voice exercise programs present with growing evidence stemming from well-designed clinical trials, such as Vocal Function Exercises and Phonation-Resistance Training Exercise (PhoRTE; adapted from the Lee Silverman Voice Treatment approach) (Ziegler, 2014), discrepancies in the results, the lack of improvement in some of the outcome measures and the scarcity of direct assessment of the larynx through videostroboscopy warrant further studies to determine the most effective approach for presbyphonia. Because of their deficits in both the laryngeal and respiratory systems, it is our hypothesis that the most effective approach for these patients would be a combination of voice therapy and respiratory training.

Impairment-Specific Interventions

Even when belonging to the same voice disorder category, patients present with various profiles of laryngeal and respiratory mechanisms impairments. Consequently, there exist no specific treatment approach that would yield optimal results for all patients and this explains why many SLPs often combine techniques from different approaches to create a "hybrid therapy" (J. L. Gartner-Schmidt et al., 2013). This individualized treatment plan should be based on specific assessments and justified by evidence proving that the chosen technique will benefit the patient. In the literature, most studies have focused on testing intervention approaches for broad categories of voice disorders (J. L. Gartner-Schmidt et al., 2013) (e.g. LSVT for Parkinson's disease patients; resonant voice for muscle tension dysphonia patients, etc.) with little attention given to inter-individual variability. It is likely that the effectiveness of voice therapy would be enhanced if the treatment plan was tailored specifically for the patient's profile (Vaca et al., 2015).

Following this reasoning, it is essential to identify which patients would benefit the most from a respiratory training intervention. It is possible that patients with a well-preserved respiratory system might gain more from therapy time allocated exclusively to voice exercises, while patients with impairments in both systems may benefit from a combination of voice and respiratory exercises. This is particularly relevant to patients with presbyphonia because of the important variability in the aging process, resulting in various pathogenic profiles. In fact, Vaca et al. (Vaca et al., 2015) categorized 105 elderly participants (mean age of 75 years old) based on their deficits, which resulted into four laryngeal/respiratory profiles: glottal deficits only (21% of the participants), altered spirometry and no glottal gap (37%), altered spirometry and glottal gap (24%), and no deficit (18%).

To be even more specific, initial assessments should provide sufficient information to determine if a patient needing respiratory training should undergo EMST, IMST, or both. Most studies aiming at improving voice outcomes have focused on EMST and have generated conflicting results. Roy et al. (Roy et al., 2003) found that EMST resulted in increased maximum expiratory pressure (MEP) in teachers with voice disorders, but this increase was not translated into improved voice outcomes (self-perceived voice handicap or voice severity ratings). It is likely that these participants didn't have respiratory compromise and that voice improvements were not observed because of a ceiling effect of the treatment outcomes. It is also possible that some of these participants had weaknesses in their inspiratory muscles and were compensating with their larynx to lengthen the expiratory phase of their speech and subsequently their utterance length. In such cases, strengthening the expiratory muscles would not eradicate the hyperfunctional compensatory mechanism and IMST would be more appropriate. A similar dilemma applies to patients with presbyphonia: based on the patient's individual profile and functional deficits (e.g. decreased loudness, decreased utterance length, sensation of dyspnea during speech, etc.), what respiratory intervention would be the most likely to generate functional improvements? This question has not been answered to date in the voice literature.

Research Questions

The voids in the literature pertaining to the impact of respiratory impairments and respiratory training on voice outcomes, combined with the absence of an optimal treatment for presbyphonia and a growing population of treatment-seeking patients presenting with vocal fold atrophy justify the need for a study examining the effects of respiratory training on presbyphonic patients.

Three research questions arise:

1. Which outcome measures present with the strongest intra-subject reliability in a sample of patients with presbyphonia?

2. What is the respiratory function of patients with presbyphonia and how is it correlated with voice measures in this population?

3. a) What is the effect of RMST on respiratory and voice outcomes in patients with presbyphonia and b) how do baseline measures of respiratory function influence the response to the intervention?

To answer these questions, a clinical trial will be conducted on three groups of participants with presbyphonia: one group will receive voice therapy and IMST, one group will receive voice therapy and EMST, and one group will receive voice therapy only. The details of the methodology are described in Chapter 3 of this document.

The next chapter presents a literature review pertaining to the following aspects: normal voice anatomy and physiology; changes to voice anatomy and physiology occurring with aging; current intervention approaches for presbyphonia; suggested interventions (IMST and EMST); and outcome measures relevant to voice therapy research for patients with presbyphonia.

Chapter 2: Review of the Literature

Section 1: Normal Anatomy and Physiology of Voice Production

Voice production is the result of the coordinated activity of the respiratory, laryngeal, and neurological systems (Larson, 2017). These represent the three subsystems involved in voice production. Their anatomy and physiology are described in the following section.

Respiratory System

Lung-Thorax Unit

The lungs are located in the rib cage, which consists of 12 pairs of ribs each made of a bony part and a costal cartilage (J. Thomas, Fall 2016). The first pair is immobile and is attached to the manubrium (superior part of the sternum) anteriorly and to the vertebrae posteriorly. The following six pairs (the true ribs) are mobile; they also attach to the vertebrae posteriorly but anteriorly they are connected to the body of the sternum via synovial (gliding) joints that allow for anteroposterior expansion of the thorax (J. Thomas, Fall 2016). The eighth to twelfth pairs are the false ribs and are not directly connected to the sternum; the first three pairs are attached to adjacent ribs via cartilage and fibrous tissue and when they move they widen the thorax. The two last pairs, the floating ribs, are connected to the abdominal wall via fascia and therefore they

follow abdominal movements (Mathieson, 2006; Netter, 2014). The lungs are protected by the thorax and they are connected to the ribs by the pleura, which consists of two layers of tissue: the visceral pleura tightly covers the lungs and the parietal pleura is attached to the inner surface of the thorax and superior part of the diaphragm (Netter, 2014). Between the two membranes is a space called pleural cavity, which contains fluid that ensures lubrication and reduces friction during breathing (J. T. Hansen, 2014; Netter, 2014). The pressure of the pleural fluid is always negative, and this ensures the visceral and parietal pleurae to stay closely together, keeping the lungs from collapsing. The result is an elastic system called lung-thorax unit, with a resting point corresponding to a lung volume at which the forces of the lungs and the thorax balance each other (Mada, n.d.). The transpulmonary pressure (the difference between the alveolar pressure in the lungs and the pleural pressure) dicates the flow of air in and out of the lungs, allowing the lungs to follow the movements of the rib (Mada, n.d.).

Inspiration

Three principles from the physics of gases come into play during inspiration and exhalation: 1) when the volume of a space increases, the pressure of the gas inside decreases; when the volume decreases, the pressure of the gas inside increases (Boyle's law). 2) Gases always travel from a higher pressure area to a lower pressure area. 3) Pressures higher than atmospheric pressure are positive; pressures lower than atmospheric pressure are negative (McFarland, 2008). More specifically, inspiration occurs when the pressure inside the lung is negative, and exhalation occurs when the pressure inside the lungs is positive. The negative pressure is created by activating the inspiratory muscles, which leads to an increase in the thorax dimensions and consequently in lung volume. The volume change decreases the pressure inside the lungs, which becomes negative (because lower than the atmospheric pressure) and this allows for airflow to

enter. Inhalation is therefore considered an active phenomenon, because muscular activity is generated to overcome the resistance of the lung-thorax unit which is being pulled away from its resting point. The primary inspiratory muscle is the diaphragm, responsible for approximately 2/3 of the vital capacity (VC) (Williams, Bannister, Berry, & al., 1995). It has the shape of a dome and its fibers originate at the xiphoid process (lower part of the sternum), the costal cartilages of the six lower rib pairs, and the vertebrae L1-L3 (J. T. Hansen, 2014). They insert at the central tendon and pull on each other when the diaphragm contracts, which flattens it and moves it slightly forward (J. T. Hansen, 2014). This has the effect of increasing the vertical dimension of the thorax and expanding the lungs (Mathieson, 2006). The second most important muscles for inspiration are the external intercostal muscles, located between the ribs and attached to their lower borders. Because they insert on the upper border of the rib below their origin, the external intercostal muscle fibers have the effect of elevating the rib cage when they contract during inspiration (Netter, 2014), and this increases the anteroposterior and oblique dimensions of the thorax (Mathieson, 2006). Other accessory muscles participate in the inhalation process to support the activity of the diaphragm and the external intercostal muscles (table 1). In speech breathing, the inspirations are quicker than in rest breathing, to avoid long interruptions between utterances. For this reason, inhalations occur through the mouth instead of through the narrow nostril openings (Mathieson, 2006). Moreover, the inspiratory muscles are activated sooner than for rest breathing and there is a slight contraction of the abdominal muscles, which causes an elongation of the diaphragm fibers and enhances its contraction strength (McFarland, 2008).

Table 1. Accessory muscles of inspiration (Mathieson, 2006; Netter, 2014; J. Thomas, Fall 2016)

Muscle	Origin	Insertion	Action
Sternocleidomatoid	Sternal head: anterior surface of manubrium Clavicular head: upper surface of medial clavicle	Lateral surface of mastoid process; occipital bone	Supports the head Raises the sternum and the clavicles during the inspiration
Scalene (anterior)	Anterior tubercles of Transverse processes of C3–C6	Scalene tubercle on 1st rib	Elevates the 1st rib bends neck
Scalene (middle)	Posterior tubercles of transverse processes of C2–C7	Upper surface of 1st rib	Elevates 1st rib, bends neck
Scalene (posterior)	Posterior tubercles of transverse processes of C4–C6	Outer surface of 2nd rib	Elevates 2nd rib, bends neck
Levatores costarum	Transverse processes of C7-T11	Superior surface of the ribs	Stabilizes the spine and helps elevate the rib cage
Serratus posterior superior	Spinous processes C7-T3	Superior aspect of ribs 2-4	Elevates ribs

A maximal inspiration raises the lung volume to the Total Lung Capacity (TLC), and this upper limit is reached when the recoil force of the lung-thorax unit becomes so strong that it cannot be further overcome by the muscular strength of the inspiratory muscles. This is explained by the fact that, the further an elastic system is brought from its resting point, the stronger the pressure exerted to go back to that point. This force is called relaxation pressure and it is null at the resting end expiratory level (EEL)-or resting lung volume (figure 1), when the elastic recoil of the lungs is perfectly counterbalanced by the elastic recoil of the chest wall at the end of a

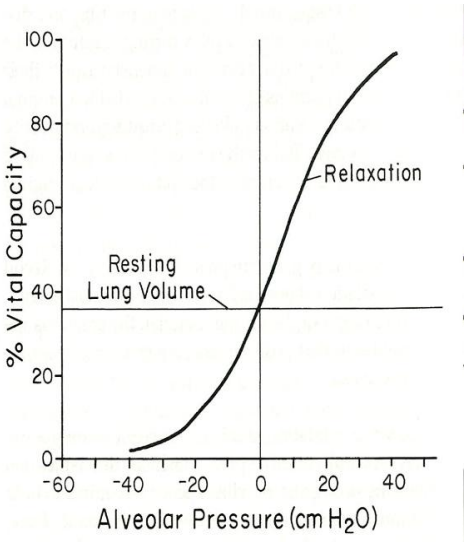


Figure 1. Relaxation Pressure Curve
http://www.easynotecards.com/notecard_set/17789

passive expiration, which corresponds to approximately 38% of the VC (McFarland, 2008). Below this point, the system will generate a negative pressure to trigger inspiration and return to its resting EEL at 38% of VC.

Expiration

Expiration is essential to phonation because it generates the airflow and pressure necessary to trigger vocal fold vibration and to generate adequate loudness and pitch. The pressure generated is a function of the trade-off between the natural pressures created by the volumes of air in the lungs and the muscle activity. For speech at a comfortable conversation level, lung volumes ranging from 35% to 60% of VC are required, which is more than during quiet breathing (Hixon, 1987). At 55% of VC, the alveolar pressure generated by the lung-thorax unit recoil forces is 6 cmH_20, which is ideal for speech (McFarland, 2008). An inhalation reaching a higher percentage of the VC will generate stronger positive relaxation pressures and will therefore create a subglottal pressure that is too elevated for conversational speech, leading to an increased intensity. In order to counterbalance the relaxation pressures and to reduce the subglottal pressure, a negative pressure has to be generated by voluntary muscular activity. Inspiratory muscles are therefore activated to help maintain an adequate and constant subglottal pressure throughout speech, without noticeable intensity variations (McFarland, 2008). Since the pulmonary volume is constantly changing during expiration, the muscles are continually working to maintain the subglottal pressure constant or to adjust it depending on the demand (McFarland, 2008). The end of a passive exhalation, when breathing at rest, occurs when the lung-thorax unit reaches equilibrium, which is when the elastic recoil of the lungs is perfectly counterbalanced by the elastic recoil of the chest wall at resting EEL. However, during speech, the expiration phase might continue beyond this point and into the expiratory reserve volume (ERV) in order to avoid

interruptions in a middle of a sentence. Once below the resting EEL, the expiratory muscles have to generate a positive pressure to counterbalance the relaxation pressures that have become negative (McFarland, 2008).

The primary expiratory muscles are the internal intercostal muscles. They lie deep to the external intercostal muscles, originate on the lower border of the ribs and insert on the costal cartilage of the ribs above their origin. This configuration allow them to close the intercostal spaces during forced exhalation (J. Thomas, Fall 2016). These muscles also fix the spaces between the ribs during inspiration and exhalation, preventing them from pushing out or drawing in (J. T. Hansen, 2014). Another important muscle group that helps to fix the intercostal spaces is the innermost intercostal muscles. They lie deep to the internal intercostal muscles and their fibers run in the same direction. They also originate from the lower border of the ribs, but they insert on the upper border of the ribs below their origin (Netter, 2014). The abdominal muscles play an important role in forced exhalation, but they are also active during the normal expiratory phase of conversational speech above resting EEL, with EMG amplitudes exceeding tidal breathing resting amplitudes (Hoit, Plassman, Lansing, & Hixon, 1988). The abdomen activation during speech at a conversational loudness level provides an opposing force against which the rib cage can rely to induce volume compressions and effectively transmit expiratory pressures to the upper respiratory tract, therefore improving its mechanical efficiency during voice production (Hoit et al., 1988; McFarland, 2008). It also allows for a passive expansion of the rib cage, which modifies the diaphragm configuration and helps control its ascension (McFarland, 2008). In their study using EMG, Hoit and colleagues (Hoit et al., 1988) found that the activation of the lateral abdominal muscles was higher than in the middle abdominal muscles during conversation speech, which led them to hypothesize that the external oblique abdominis, internal oblique abdominis, and/or transverse abdominis muscles were active whereas the rectus abdominis muscles, located

more towards the midline of the abdomen, was not. Moreover, they noted that the lower lateral sections were more active than the upper lateral section, a phenomenon that is also present during resting tidal breathing and that is explained by a response to the pressure of the abdominal content when in standing position (Hoit et al., 1988). As a matter of fact, the location and extent of abdominal muscle activity were shown to be influenced by factors such as body position, but also lung volume and task. For example, when higher expiratory muscular pressure was needed, such as when speaking at low lung volumes, middle EMG activity in addition to the lateral EMG activity were observed, suggesting an activity of all abdominal muscle groups (Hoit et al., 1988). Other accessory expiratory muscles include the transverse thoracis, the serratus posterior inferior, and the subcostal muscles (Mathieson, 2006; Netter, 2014; J. Thomas, Fall 2016). These muscles are mostly active during forced exhalation (Mathieson, 2006; Netter, 2014; J. Thomas, Fall 2016).

Co-activation of Inspiratory and Expiratory Muscles

Even though they have been categorized as "inspiratory muscles" and "expiratory muscles", these two groups of muscles each play a role during both the inspiratory and expiratory phases of speech production. As it was mentioned in the previous section, the inspiratory muscles (diaphragm and external intercostal muscles) are active during the expiratory phase of speech to counterbalance the relaxation pressures and maintain a constant subglottal pressure. This is accompanied by a co-contraction of the lateral abdominal wall which helps control the pressure that is directed to the upper respiratory tract (McFarland, 2008). This co-activation results in more rapid adjustments of the subglottal pressure and fundamental frequency, while avoiding undesirable pressure changes that could occur during speech (McFarland, 2008). This co-activation of inspiratory and expiratory muscles is also present during the inspiration phase of

speech breathing. During speech, the inspirations have to be quick and efficient in order to avoid unnecessary interruptions. The activation of the abdominal muscles in synchrony with the diaphragm and external intercostal muscles moves the abdominal wall inwards and induces more effective expansion of the rib cage by lengthening the diaphragm muscle fibers and increasing its force (McFarland, 2008).

Laryngeal System

The laryngeal and respiratory systems are closely related and a healthy phonation relies on the good coordination between the two. The vocal folds in adduction serve as a valve to modulate the expiratory airflow that is generated by the lower respiratory system (Mathieson, 2006).

The vocal folds are located in the larynx, which is a structure made of cartilages, ligaments, muscles, and membranes that extends from the base of the tongue to the trachea (from C3 to C6) (Mathieson, 2006; J. Thomas, Fall 2016). It lies anterior to the esophagus and is suspended in the neck by the hyoid bone. The larynx is part of the upper respiratory system, along with the oropharynx, the mouth, and the nasal cavity (Mathieson, 2006). The anatomy and physiology of the vocal folds as well as the function of the different muscles of the larynx will be described in the following sections as the phonation process is dissected.

Onset of Vibration

Even before the onset of phonation, a close relationship between the larynx and the lower respiratory system is witnessed: at the moment of inspiration, the vocal folds abduct rapidly during what has been named the "pre-phonatory inspiratory phase" (Wyke, 1983; cited in Mathieson, 2006). There is also a lowering of the larynx simultaneously with the descent of the diaphgragm, phenomenon known as the tracheal pull (Herbst, 2017). Abduction of the vocal folds

results from the contraction of the posterior cricoarytenoid (PCA) muscles, the only abductors of the vocal folds. The PCA muscles originate on the posterior surface of the cricoid cartilage and their fibers insert on the muscular process of the arytenoids (Netter, 2014). Their contraction pulls the arytenoids away from each other and rotates them in a way that pulls the vocal processes outwards, therefore abducting the vocal folds (Mathieson, 2006). After this short abduction phase, the arytenoids are brought together by the interarytenoid muscles (IA) (composed of a single transverse arytenoid muscle and a pair of oblique arytenoid muscles). This has for effect of bringing the vocal folds closer to the midline before they are adducted by the lateral cricoarytenoid (LCA) muscles (Mathieson, 2006). The LCA muscles originate on the lateral-superior edges of the cricoid cartilage and insert on the muscular processes of the arytenoid cartilages, which allows them to rotate the arytenoids anteriorly and to adduct the posterior portion of the vocal folds (Boone, 2010; Mathieson, 2006; Zhang, 2016b). To achieve a complete closure of the glottis, activation of the thyroarytenoid (TA) muscles is required. The TA muscles constitute the body of the vocal folds. Their fibers originate from the interior surface of the thyroid cartilage and insert on the anterolateral surface of the arytenoid cartilages (Yin & Zhang, 2014). Their contraction induces a bulging of the muscle as well as a rotation on the horizontal plane towards the midline, which allows for adduction of the anterior portion of the vocal folds (Chhetri, Neubauer, & Berry, 2012; Yin & Zhang, 2014) Once the vocal folds are adducted, subglottal pressure (generated by the alveolar pressure during the expiration phase of respiration) increases until it reaches the phonation threshold pressure (Farley & Barlow, 1994; cited in Mathieson, 2006). The phonation threshold pressure corresponds to the pressure necessary to overcome the intertie of the vocal folds and put them into movement. The vocal folds' resistance to the subglottal pressure depends on their size, viscoelastic properties, and tension (Titze, 1994; cited in Mathieson, 2006).

Self-Sustained Phonation

Once the subglottal pressure reaches the phonation threshold pressure, the vocal folds start to separate from their inferior border up to their superior border. Once they are peeled apart, a negative intraglottal pressure is induced in the glottis, created by the Bernouilli effect. The Bernouilli effect occurs when air moves from a large space to another through a constricted space (in this case the glottis), which leads to a high velocity and a low pressure at the site of constriction (Maran, 1988; cited in Mathieson, 2006). The negative pressure creates a suction on the vocal folds, which are pulled back towards the midline. This recoil to the midline is made possible due to the elastic properties of the vocal folds (Berke & Gerratt, 1993). This theory explains part of the vocal fold vibration but not its entirety. Newer models of phonation biomechanics suggest that the main condition necessary to transfer the energy from airflow to the vocal folds and induce vibration is the vertical phase difference engendered when the lower lips of the vocal folds shift laterally before the air reaches the upper lips and pulls them away from each other (Krausert et al., 2011). This phase difference creates a pressure asymmetry between the opening and closing phases of the vibration, which is necessary to achieve an energy transfer from the airflow to the vocal folds (Zhang, 2016b). The pressure changes in the subglottal and supraglottal vocal tracts, caused by the opening and closing of the vocal folds and resulting airflow variations, further participate in the push and pull action of the vocal folds during self-sustained vibration (Titze, 2001).

Vocal Folds

During the opening phase of the vibratory cycle, the subglottal pressure deforms the vocal folds' superficial layers, which results in the opening of the vocal folds from the lower lip to the upper lip (Krausert et al., 2011). In other words, it is the structural properties of the vocal folds, which have been described in the body-cover model, that allow the mucosal wave to take form and to spread vertically and medially, resulting in the transformation of air into sound (Krausert et al., 2011). The vocal fold is composed of the TA muscle and a mucous membrane. This mucous membrane is divided into the epithelium (covered by a thin layer of mucus) and the lamina propria, separated by the basement membrane (figure 2). The epithelium acts as a physical barrier to protect the lamina propria against injuries. It is composed of five to ten layers of stratified squamous cells, which are divided in the most basal layers and move to the suprabasal layers to replace old cells and ensure homeostasis (Levendoski, Leydon, & Thibeault, 2014). The epithelial cells are separated by different types of junctions, which play three important roles: they seal together adjacent cells, they stabilize the epithelium during vibration, and they provide pathways for communication between cells (Levendoski et al., 2014). The mucus barrier that lies on the epithelium also serves a protective function, as well as a lubricative one. Moreover, the viscosity of the mucus has an impact on the vocal fold vibration, therefore changes in its quantity and proprieties may affect phonation (Levendoski et al., 2014). The lamina propria is made of connective tissue and is divided into three sections: 1) the superficial layer (also called Reinke's space), 2) the intermediate layer, and 3) the deep layer (Mathieson, 2006). In the body-cover model, the structurally different parts of the vocal folds are reorganized into their functional roles: the epithelium and Reinke's space form the cover, the intermediate and deep layers of the lamina propria form the vocal ligament,

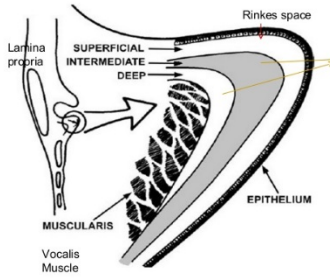


Figure 2. Vocal Fold Histology (Hirano, 1981)

and the TA muscle constitutes the body (Mathieson, 2006). The cover is loose and vibrates the most during phonation (Boone, 2010; Mathieson, 2006); deeper to the cover is the vocal ligament which is stiffer and provides a support to the vocal folds, and underneath it is the muscle (Boone, 2010).

The different roles played by the mucous layers are explained by their composition, more specifically by the components of their extracellular matrix (ECM). The ECM supports and surrounds the cells and is composed of different molecules that are produced by fibroblast cells (Thibeault, 2005) and released through exocytosis, resulting in fibers and ground substances (Thibeault, 2005). Fibers: the most omnipresent component of the ECM is collagen, present throughout the three layers of the lamina propria, but more significantly in the deep layer (Hammond, Gray, & Butler, 2000), which explains the increased stiffness of the vocal ligament compared with the cover of the vocal fold. Collagen represents approximately 43% of the total proteins in the lamina propria (Ishikawa & Thibeault, 2010) and is synthesized from procollagen, its precursor produced by the fibroblasts (J. K. Hansen & Thibeault, 2006; Ishikawa & Thibeault, 2010). A specific type of collagen fibers, reticular fibers, are also present in the ECM around the cells and are characterized by their small diameters (Ishikawa & Thibeault, 2010). The last type of fibers are elastic fibers, which are made out of elastin protein and are present in the superficial and deep layers of the vocal folds (J. K. Hansen & Thibeault, 2006). Elastin represents 8.5% of the total proteins in the lamina propria and provides the elasticity to the tissues (Ishikawa & Thibeault, 2010). Ground substances: the ground substance of the ECM is made of large interstitial molecules named glycosaminoglycans (GAGs), which main role is to preserve the viscoelastic characteristics of the tissue (Ishikawa & Thibeault, 2010). GAGs attract water molecules and therefore allow the tissue to remain hydrated. They also regulate the extracellular environment necessary for healthy cellular activity. Increases or decreases in the ground substance materials

have the potential to significantly affect the viscosity of the tissue (Ishikawa & Thibeault, 2010). Interstitial molecules include fibronectin, decorin, fibromodulin, cadherin, syndecan-1, syndecan-4, and hyaluronic acid (HA) (J. K. Hansen & Thibeault, 2006). HA has a substantial impact on the lamina propria's stiffness and viscosity. The stiffness of a tissue corresponds to its elastic restoring force in response to deformation (Friedrich et al., 2013; Zhang, 2016b) and is different from tension or stress, which correspond to the actual mechanic condition of the tissue (Zhang, 2016b). Viscosity is the resistance to deformation (Zhang, 2016b). Studies have shown that a reduction in HA leads to a decreased stiffness and an increased viscosity of the vocal folds and thus in suboptimal conditions for vibration including a higher phonation threshold pressure (Friedrich et al., 2013; J. K. Hansen & Thibeault, 2006). HA is also involved in scar formation and wound healing by playing a role in cell migration and proliferation (Bless & Welham, 2010; Branski, Verdolini, Sandulache, Rosen, & Hebda, 2006; Friedrich et al., 2013; J. K. Hansen & Thibeault, 2006). Its half-life is very short (0.5 to 4 days) and its concentration is higher in men's vocal folds (J. K. Hansen & Thibeault, 2006).

Although no laryngeal glands are found on the free edges of the vocal fold, they are numerous in the submucosa of the laryngeal saccule, which is the extension of the vestibule (space between the true and the false vocal folds). The thyroepiglottic muscle (part of the TA muscle) lies next to the saccule and separates it from the thyroid cartilage. When the muscle contracts, secretions from the laryngeal glands are being released on the vocal folds to protect them from dryness and friction during vibration. This lubrication is essential for a healthy phonation (Mathieson, 2006).

Laryngeal Adjustments

Once the vibration is initiated and maintained due to adequate viscoelastic properties of the vocal folds, different laryngeal adjustments can be performed to modify the sound, more specifically to adjust the fundamental frequency and loudness.

The frequency of the sound is determined by the time taken by the vocal folds to complete one vibratory cycle (its period) (Mathieson, 2006), and can be modified by changes in the vocal fold posturing and properties. More specifically, it can be increased by shortening length, increasing tension and stiffness, and reducing thickness (vibrating mass) of the vocal folds (Zhang, 2016b). The length of the vocal folds can be decreased by contracting the TA muscle and can be increased by activating the cricothyroid (CT) muscle (Zhang, 2016b). The CT muscle is made of two different bundles, a vertical one (pars recta) and an oblique one (pars oblique), which both originate from the anterolateral part of the cricoid cartilage and insert on the inferior border and inferior horn of the thyroid cartilage (Netter, 2014). Activation of both bundles results in an elongation of the vocal folds by bringing the two cartilages closer together through an anterior rotation about the cricothyroid joint and by inducing a slight backward movement of the cricoid cartilage and forward movement of the thyroid cartilage (Zhang, 2016b). Although contraction of the CT increases the length of the vocal folds, it also increases their tension and stiffness. In the case of the CT activation, the effect of stiffness and tension dominates the effect of the lengthening and thus results in an increased F0 (Zhang, 2016b). Length and tension are often associated, which could explain why an increased length of the vocal folds is often associated with an increased F0. Activation of the LCA, IA, and PCA muscles simultaneously to the CT contraction help stabilize the arytenoid cartilages and inhibits them from moving forward, further elongating and tensing the vocal folds for production of very high pitches (Zhang, 2016b). Elongation of the vocal folds through CT activation also decreases the thickness of the vibrating mass, which also has for effect

of increasing F0 (Zhang, 2016b). On the contrary, activation of the TA muscle increases the thickness of the vocal folds. Although an increased thickness in theory reduces F0, it also enhances the medial compression force, which leads to longer durations of vocal fold contact during vibration, which increases the stiffness and consequently F0. The role of the TA muscle in frequency modulation is thus very complex, as its activation can either increase or decrease F0, depending on the activation of the other laryngeal muscles, mostly the CT which is its antagonist (McFarland, 2008). Moreover, the TA contraction increases the tension of the body of the vocal folds but it decreases the tension in the cover layers by shortening the muscle but not the mucous membrane (Zhang, 2016b). If there is co-activation of the CT and the TA, the tension of the vocal folds will be increased, without an increase in the length (McFarland, 2008).

Different strategies can be used to increase the amplitude of the vocal fold vibration, and therefore the intensity of the sound. At the level of the larynx, increasing the duration of contact and the medial compression of the vocal folds will have the effect of boosting the harmonics close to the first formant and therefore will increase the intensity of the sound (Zhang, 2016b). This enhanced approximation of the vocal folds can be done by activating the LCA and IA muscles (McFarland, 2008) or by augmenting the thickness/contact surface and medial compression of the vocal folds by contracting the TA (Zhang, 2016b). However, increases in intensity usually don't merely rely on laryngeal adjustments, but are the result of simultaneous increases in subglottal pressure and laryngeal resistance (Stathopoulos & Sapienza, 1993), as explained in the following section.

Co-activation of Respiratory and Laryngeal Systems

The main strategy to enhance the intensity of the voice is to increase the subglottal pressure, which induces a larger amplitude of vibration and a faster and stronger return to the initial

position by the tensed vocal folds (Mathieson, 2006; McFarland, 2008). This also has the effect of augmenting the pitch of the sound because of the relationship between frequency and tension of the vocal folds, as described above. At high intensities, laryngeal adjustments play a minor role in controlling the intensity of the sound (Zhang, 2016b). However, some adjustments need to be made by the laryngeal muscles in order to increase the resistance of the vocal folds to the high subglottal pressure. Without a sufficient resistance, the expiratory pressure generated in the pulmonary alveoli will quickly overcome the phonation threshold pressure without allowing for subglottal pressure to build up under the vocal folds (McFarland, 2008). Vocal fold resistance is a function of the vocal folds' adduction and medial compression, which relies on the activation of the LCA, IA, and TA muscles (Zhang, 2016b). Moreover, increasing glottal resistance proportionally to subglottal pressures avoids changes in mean airflow rates and allows for a longer and more controlled expiration phase during speech (Zhang, 2016c). The airflow conservation role of the glottis is important to reduce the effort of the respiratory muscles (Zhang, 2016c). In fact, it has been reported that a weakness in one system (i.e. laryngeal or respiratory) may lead to hyperfunction in the other system in an attempt to compensate (Zhang, 2016c).

Source-Filter Model (Resonance System)

So far, it has been presented that loudness can be altered by the respiratory and laryngeal systems. However, there is another mean through which loudness can be increased and reduced, and it involves the articulatory system (supraglottic vocal tract) (Zhang, 2016b). The supraglottic tract comprises the structures above the larynx and includes the different parts of the pharynx (hypopharynx, oropharynx, and nasopharynx) as well as the oral and nasal cavities (Mathieson, 2006). The three sections of the pharynx are bordered by the superior, middle and inferior

pharyngeal constrictor muscles, which modify the circumference of the pharynx (Mathieson, 2006).

Modifications in the shape of the vocal tract not only affect loudness of the sound, but also its pitch and quality. Although humans cannot change the length of their trachea and vocal tract, they can adjust the height of their larynx (Titze, 2008), the shape of their epilarynx (including the aryepiglottic folds and the superior portion of the epiglottis) and the shape of their vowels to optimize the acoustic properties of the vocal tract (Titze & Worley, 2009). The position of the larynx is altered by the extrinsic laryngeal muscles, which can be classified into two categories: suprahyoid and infrahyoid. The suprahyoid muscles originate from above the larynx and insert on the hyoid bone (Mathieson, 2006). When they contract they elevate the larynx (digastric, stylohyoid, mylohyoid and geniohyoid muscles) (Mathieson, 2006; Netter, 2014). The infrahyoid muscles originate from below the larynx and insert on the hyoid bone or thyroid cartilage (in the case of the sternothyroid) (Mathieson, 2006). When they contract, they pull the larynx downwards (sternohyoid, sternothyroid, omohyoid and thyrohyoid) (Mathieson, 2006; Netter, 2014).

Acoustically, an interaction between the source (vocal folds) and the filter (vocal tract) results in a good tuning between the harmonics that are being produced at the source and the formants (range of frequencies enhanced by the vocal tract resonance) (Titze & Worley, 2009). Because the source-filter model is nonlinear, certain vowels and voice qualities are more effective at certain pitches and specific vocal tract shapes are more effective with particular vocal fold configurations. A megaphone shape (narrow epilarynx and wide mouth) such as in the vowel /a/ or during a shout, interacts well with a high glottal resistance and therefore with a pressed glottal adduction (Titze & Worley, 2009). On the other hand, an inverted megaphone (wide epilarynx and narrow mouth) or neutral vocal tract shape has a more effective interaction with a low glottal

resistance/moderate vocal fold adduction (Titze & Worley, 2009). Speakers can therefore adjust their vocal tract shape and vocal fold adduction for an ideal source-filter interaction. The supraglottic tract configuration can also affect the pitch of the voice without a change in the fundamental frequency, the implication of which being that F0 is not always a precise predictor of the pitch (Mathieson, 2006). As a matter of fact, the sound that is perceived by the listener is not the same as the one produced at the source (the glottal signal), but is the result of the sound after it has been filtered by the supraglottic tract (Mathieson, 2006). The glottal signal is made of a series of harmonics and depending on the shape of the vocal tract and the different articulators, some of them will resonate more than others. These zones of frequencies enhanced by the vocal tract resonance are called formants (Titze & Worley, 2009). This allows for the production of the different vowels and gives the "color" to the voice that is unique to every speaker (Mathieson, 2006).

Hormonal Influence on the Larynx

The larynx is a hormone-sensitive organ (D'Haeseleer, Depypere, Claeys, Van Borsel, & Van Lierde, 2009). It is affected in a different way by each class of sexual hormone, and these influences have been thoroughly described by Abitbol et al. (Abitbol, Abitbol, & Abitbol, 1999). While estrogens have a proliferative effect on the laryngeal mucosa and reduces the desquamation process in the superficial layers of the lamina propria, progesterone exerts the opposite effect and hastens desquamation of the superficial layers (Abitbol et al., 1999). Estrogen and progesterone also play contrasting roles with regards to glandular secretions and capillary permeability, which tend to be increased by estrogens and decreased by progesterone. As a result, progesterone has a dehydrating effect on the vocal fold mucosa. Androgens are also associated with dehydration of

the mucosa, as well as with its atrophy. Moreover, androgens are thought to have an impact on skeletal muscles, causing hypertrophy and a decrease in fat cells (Abitbol et al., 1999). Considering these effects, the levels and ratios of the different hormones play a significant role on the maintenance of optimal conditions for voice production, which may be disrupted during life events such as puberty and menopause (Awan, 2006).

Neurological System

Central Nervous System

Phonation is the result of the coordinate activity of laryngeal, respiratory, and articulatory muscles, with motor fibers originating from different locations in the central nervous system. The intrinsic laryngeal muscles are controlled through the vagus nerve by motoneurones located in the nucleus ambiguus of the medulla, while the motoneurones controlling the external muscles lie in the 2nd cervical segment and send signals through the ansa cervicalis (Jurgens, 2009). The respiratory muscles are operated by neurones located in the ventral horn of the thoracic and upper lumbar spinal cord and the articulatory muscles are regulated by many different groups of motoneurones, including those from the facial nucleus and the trigeminal motor nucleus (Jurgens, 2009). Since there is no evidence of a direct connection between these groups of motoneurones (Jurgens, 2009; Thoms & Jurgens, 1987), how is it possible that the result of their co-activation is this perfectly well coordinated activity that is phonation? The explanation

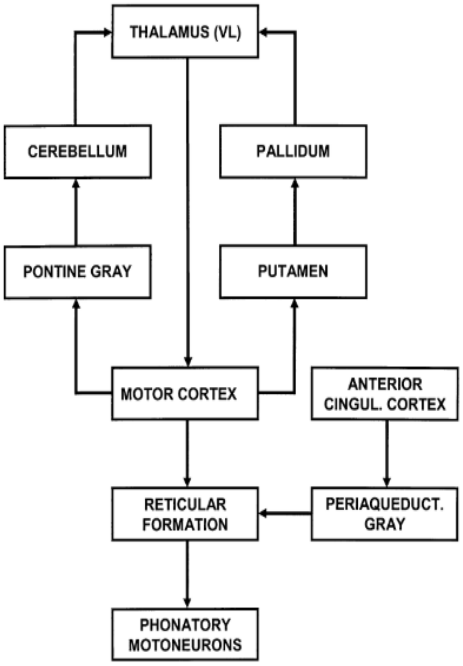


Figure 3. Neural Control of Phonation (Jurgens, 2009)

lies in the existence of a region in the brain that receives input from all the motoneurones involved in phonation and that coordinates their activity (figure 3). This region is located in the reticular formation of the lower brainstem and plays a major role in the two pathways involved in vocalization, as described by Jurgens in his model on the neural control of phonation: 1) the readiness to vocalize and 2) the motor patterning of vocalizations (Jurgens, 2009).

1) Readiness to Vocalize

The voluntary control of phonation starts in the cortex of the medial frontal lobe, in a region called anterior cingular cortex (ACC) (Jurgens, 2009; Jurgens & von Cramon, 1982; Rubens, 1975). The confirmation of the ACC's role in voluntary phonation has been provided by the observation of communication behaviors of patients with a lesion in this part of the brain. These patients could still vocalize, but they wouldn't initiate phonation unless a question was directly asked to them (Jurgens, 2009; Jurgens & von Cramon, 1982; Rubens, 1975). The ACC sends projection to the periaqueductal gray (PAG), in the caudal half of the midbrain (upper part of the brainstem). The PAG has been identified as an important area for voice production not only in humans, but in the large family of mammalians (Davis, Zhang, Winkworth, & Bandler, 1996). It is responsible for involuntary and emotional sounds in humans, such as a pain cries (Jurgens, 2009). It is also in charge of the initiation and intensity of the sounds (Jurgens, 2009). The PAG projects its fibers to the reticular formation of the lower brainstem (Mantyh, 1983) (figure 3).

2) Motor Control

The second neural pathway of voluntary vocalization starts in the precentral gyrus of the primary motor cortex of the right and left brain hemispheres (Jurgens, 2009; Mathieson, 2006). The cortical homunculus, developed by Dr. Wilder Penfield (Penfield & Boldrey, 1937), is a map of the different body part representations within the motor cortex (figure 4). Brown and colleagues (Brown, Ngan, & Liotti, 2008) identified a definite region within the primary motor cortex that

specifically controls the adduction/abduction as well as the tension/relaxation of the vocal folds, which is located next to the lip region that was identified by Dr. Penfield (figure 5). Moreover, Ramsay and his colleagues (Ramsay et al., 1993) found an overlapping of the regions of the primary motor cortex that are activated during phonation with those activated during controlled exhalation. Loucks and his colleagues (Loucks, Poletto, Simonyan, Reynolds, & Ludlow, 2007) had similar findings and concluded to a "common volitional sensorimotor system" for phonation and voluntary respiration, which might help facilitate the coordination of these systems during phonation activities (Brown et al., 2008). Quiet breathing is a reflex activity and is coordinated in the respiratory control center (West, 1976, cited in Mathieson, 2006) in the reticular formation of the brain stem, more specifically in the pons and medulla (Mathieson, 2006). However, during speech the breathing activity becomes voluntary and is regulated to some extent by the cortex (Mathieson, 2006), as suggested by the results from Loucks and Ramsay and their colleagues (Loucks et al., 2007; Ramsay et al., 1993). This voluntary control is gradually developed from a young age when the child starts to vocalize and eventually starts talking. When individuals reach adulthood, speech breathing is so well integrated that it has become automatic (Mathieson, 2006). Literature regarding the predominance of a certain hemisphere for vocalization control is conflicting. Some studies found a left hemisphere predominance (Loucks et al., 2007; Ramsay et al., 1993), whereas others found a right hemisphere predominance (Riecker, Ackermann, Wildgruber, Dogil, & Grodd, 2000) and some concluded to a symmetrical pattern of activation (Ozdemir, Norton, & Schlaug, 2006). This

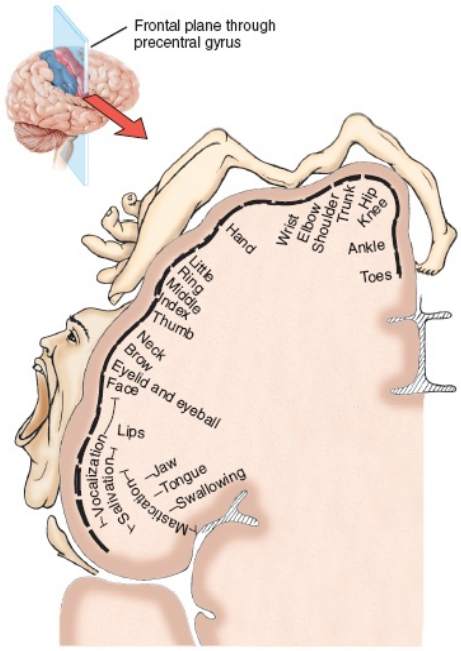


Figure 4. Cortical Homunculus (http://keywordsuggest.org/gallery/682317.html)

could be explained by differences in the tasks administered, generating diverse cognitive and motor responses (Loucks et al., 2007).

Following Jurgens' model (Jurgens, 2009), fibers from the motor cortex project to the reticular formation of the lower brainstem, where the two pathways of readiness to vocalize and motor patterning converge and allow for phonation. As shown in figure 4, the motor cortex also receives information from other brain regions (Jurgens, 2009). The motor commands are sent from the motor cortex to the cerebellum and the pallidum via the pontine gray and putamen respectively, and are pre-processed before being sent back to the motor cortex through the ventrolateral thalamus (Jurgens, 2009).

Fibers from the laryngeal region of the motor cortex reach the medulla in the brainstem by descending in the corticobulbar pathway (group of neurons that travel from the motor cortex to the medulla). With the corticospinal tract, the corticobulbar tract is part of the pyramidal system, which refers to the upper motor neurons. Once the corticobulbar tract reaches the medulla, part of its fibers decussate to the contralateral side while the rest of the fibers remain in the ipsilateral side (Mathieson, 2006). Fibers sending motor signals to the different parts of the larynx therefore synapse with either the contralateral or ipsilateral nuclei of the 10th cranial nerve (vagus nerve), located in the nucleus ambiguus of the medulla (J. T. Hansen, 2014). The fibers stemming from the vagal nuclei are part of the peripheral nervous system and form the lower motor neuron pathway (Mathieson, 2006).

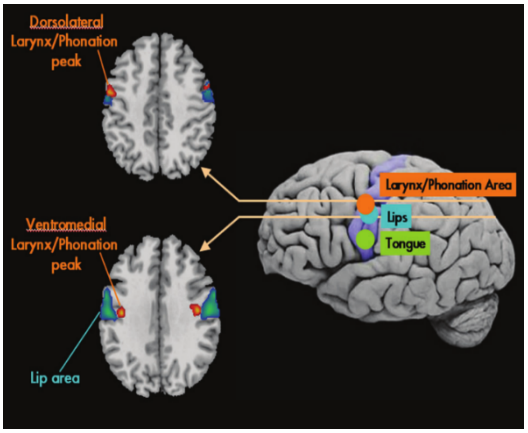


Figure 5. Larynx/Phonation Area in the Motor Cortex (Brown, Ngan, & Liotti, 2008)

Peripheral Nervous System

The efferent fibers of the vagal nerve descend down the neck (within the carotid sheath) through a hole in the skull called the jugular foramen. Just below the jugular foramen, a part of the vagus nerve branches into the superior laryngeal nerve (SLN), which itself branches into an internal and an external branch (J. Thomas, Fall 2016). The external SLN provides motor innervation to only one muscle of the larynx, the cricothyroid muscle. The internal SLN contains sensory and parasympathetic fibers and innervates the tissues of the larynx above the vocal folds (Mathieson, 2006). On both the right and the left sides, the main trunks of the vagus nerve maintain their course downward. On the right side, it branches into the right recurrent laryngeal nerve (RLN) once it reaches the subclavian artery (J. Thomas, Fall 2016) (figure 6). The right RLN curves under the right subclavian artery and ascend to reach the larynx by traveling between the trachea and the esophagus (Mathieson, 2006). On the left side, the vagus nerve branches into the left RLN at the level of the aortic arch (figure 6). Because of its location, and because of its extensive pathway under the arch of the aorta and upwards, the left RLN is more susceptible to be affected by pressure from intrathoracic masses and by iatrogenic injuries (Mathieson, 2006). Once they have reached the larynx, the right and left RLN split into an anterior and a posterior branch before they enter the larynx and supply motor innervation to all the intrinsic muscles of the larynx, with the exception of the cricothyroid muscle, which is innervated by the external branch of the SLN as abovementioned.

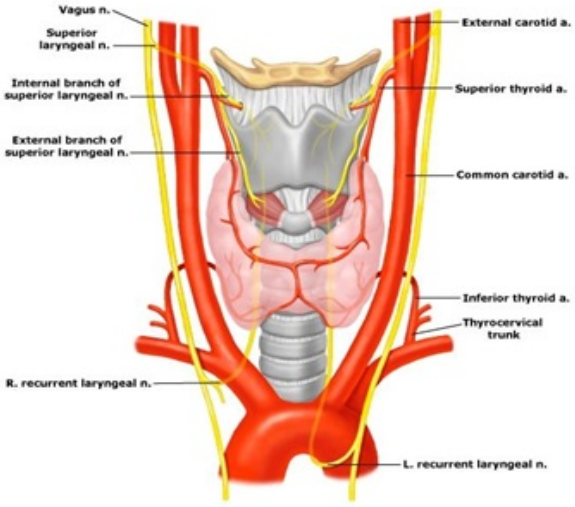


Figure 6. Laryngeal Innervation (http://www.forwardthinkingchiro.com/blog/)

The peripheral nerves such as the SLN and the RLN are responsible for transmitting the information between the central nervous system and the muscles or organs. The contact between the motor neuron and the muscle fiber is called the neuromuscular junction (J. Thomas, Fall 2016). When the neurons synapse, through the action of neurotransmitters, the resting membrane potential is disrupted and this leads to the creation of an action potential (J. Thomas, Fall 2016). The action potential is an impulse that travels in the neuron's axon, which is covered with segments of myelin sheath made of Schwann cells (J. Thomas, Fall 2016). The myelin provides the insulation necessary for an effective impulse transmission: the action potential jumps along the Nodes of Ranvier, the depolarized regions between the myelin sheaths (J. Thomas, Fall 2016). The result of this "saltatory conduction" is a conservation of energy and an enhanced conduction velocity (J. Thomas, Fall 2016).

Feedback System

Most peripheral nerves are mixed, which means that they contain both efferent and afferent fibers. While efferent fibers send commands from the central nervous system to the muscles, afferent fibers send information from the muscles to the central nervous system. This feedback allows the brain to constantly adjust the motor commands to obtain the desired movements. For the intrinsic laryngeal muscles, this is made possible by the mechanoreceptors contained in the mucosa of the larynx, which are sensitive to stretch, touch and pressure and therefore respond to vibration and airflow (Hammer & Krueger, 2014). These mechanoreceptors travel via the afferent fibers of the vagus nerve and reach the lateral part of the medulla, from which projections are sent to the primary somatosensory cortex (Hammer & Krueger, 2014). Along with somatosensory feedback, auditory feedback plays a primordial role in phonation. An important brain region for auditory feedback is the superior temporal gyrus (STG). As it was mentioned in

the previous section, the primary motor cortex regions activated by phonation are similar to those activated by controlled exhalation because of their close relationship. However, a stronger response is found in the auditory region, more precisely the superior temporal gyrus (STG), during phonation when compared with controlled exhalation (Loucks et al., 2007). This confirms the involvement in STG in the phonation feedforward system (Larson, Altman, Liu, & Hain, 2008).

Section 2: Age-Related Changes in Voice Anatomy and Physiology

This section describes the anatomical and physiological changes in the different systems involved in voice production: the respiratory, laryngeal, and neurological system.

The impact of these changes on the different voice outcomes will be discussed in the last section, along with the associated outcome measures that are relevant for assessing pre- post-treatment changes in patients with presbyphonia.

Changes in the Respiratory System

Respiratory Muscle Strength

Respiratory muscles are mainly composed of skeletal muscles, which are subject to muscle fiber atrophy with aging. In this process named sarcopenia the proportion of type II fibers (fast-twitch fibers) is decreased and this leads to a reduction in muscle mass, which in turn induces a decrease in muscle strength and power (Kim & Sapienza, 2005). The extent and progression of the sarcopenia process varies among the population and depends on different factors (such as nutrition, physical activity, neuromuscular, molecular and hormonal status, age, body mass index) (Melton, Khosla, & Riggs, 2000; Reychler, Delacroix, Dresse, Pieters, & Liistro, 2016), but can result in a 50% decrease in total muscle mass by age 80 (Lexell, Taylor, & Sjostrom, 1988). The physiological process that leads to muscle atrophy involves the neurological changes that take place with aging (described in a subsequent section), but can also result from a direct decrease in muscle protein synthesis (Tolep & Kelsen, 1993). More specifically, the cross sectional areas of intercostal muscles decrease with age, with an onset around 50 years old and a greater reduction in internal intercostal muscles (expiratory muscles) than in external intercostal muscles (Lalley, 2013). Atrophy of the diaphragm has also been noted, with a decrease in fast twitch (type II) fibers

(Lalley, 2013). Structural changes in respiratory muscles are reflected in declined maximum expiratory pressure (MEP) and maximum inspiratory pressure (MIP) (Lalley, 2013).

Lungs and Chest Wall Compliance

Respiratory function is not only affected by alterations in muscle strength. Structural changes occurring in the lungs and in the thorax greatly affect respiratory function in elderly individuals. With age, changes with regards to the number and organization of the elastic fibers surrounding the alveolar ducts take place in the connective tissue of the lungs. In addition to inducing the collapsing of small airways and increasing resistance in the airway at low long volume (Enright, Kronmal, Manolio, Schenker, & Hyatt, 1994), these changes increase the compliance ("change in volume per unit change in pressure"(Mada, n.d.)) of the lungs, which become more distensible and lose some of their elastic recoil (Janssens, Pache, & Nicod, 1999).

The aging effect on lung elasticity has been noted to be more marked in men, who lose more elastic recoil than women over time. Because men's elastic recoil are higher than women at a young age, the recoil forces of both sexes becomes similar with aging (Bode, Dosman, Martin, Ghezzo, & Macklem, 1976). While the lung compliance is increased with aging, the chest wall compliance is decreased due to a narrowing of the intervertebral disk spaces and a calcification of the articulations between the ribs and the spine (Janssens et al., 1999). Moreover, the shape of the thorax is subjected to changes caused by age-related osteoporosis, in some cases leading to a kyphotic spine and impacting the curvature and functioning of the diaphragm (Janssens et al., 1999). The structural alterations in the thorax increase the stiffness (resistance to deformation) of the chest wall and explain its reduced compliance and an increased elastic recoil (Janssens et al., 1999). The resulting static pressure curve of the lung-thorax unit in older people when compared to younger people is depicted in figure 7.

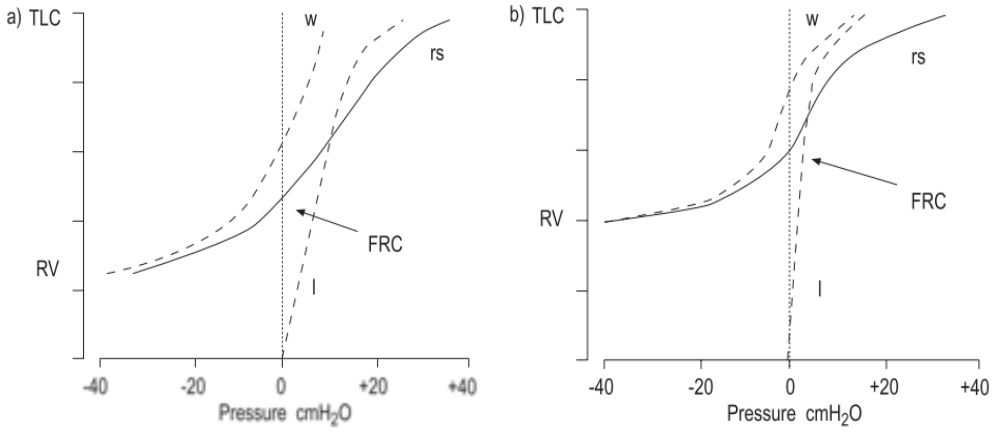


Figure 7. Static pressure curves of the lungs (l), chest wall (w) and respiratory system (rs) in a a) 20-yr-old and b) 60-yr-old. RV: residual volume; FRC: functional residual capacity; TLC: total lung volume. (Janssens et al., 1999; Adapted from Turner et al.,

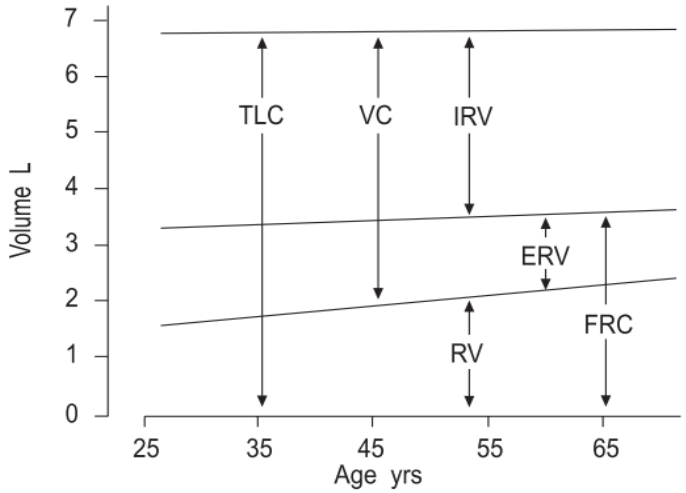


Figure 8. Evolution of lung volumes with ageing. TLC: total lung capacity; VC: vital capacity; IRV: inspiratory reserve volume; ERV: expiratory reserve volume; FRC: functional residual capacity; RV: residual volume. (Janssens et al., 1999; adapted from CRAPO et al., 1982)

Changes in the lungs and chest walls as well as changes in respiratory muscle strength have an impact on pulmonary function (Lalley, 2013). Firstly, they lead to a greater residual volume in older individuals (Janssens et al., 1999). Janssens et al. mention that residual volume increases by approximately 50% between the ages of 20 and 70 years old, and this leads to a reduced vital capacity in these individuals (figure 8). The functional residual capacity (FRC) is increased with age, which means that older individuals have a smaller functional reserve and that they breathe at higher lung volumes than younger people (Janssens et al., 1999; Kim & Sapienza, 2005; Tolep & Kelsen, 1993). In spirometry testing, changes in the respiratory system translate into a reduced forced vital capacity (FVC) and a reduced force expiratory volume (FEV_1) (Janssens et al., 1999; Kim & Sapienza, 2005). However, the total lung capacity is not affected by aging (Janssens et al., 1999).

Impact on Phonation Physiology

Combined, the aforementioned changes related to the respiratory system have a substantial impact on voice production. Confronted with an altered respiratory mechanism, elderly people use compensatory mechanisms for voice production. While young adults initiate speech at mid-lung volume (35-60% of their vital capacity) (Hixon, 1973), older adults use a higher lung volume and the difference is more important in men (Huber & Spruill, 2008). By filling their lungs at mid-volume, adults are relying on the natural recoil pressure of the system to produce the required air pressure for voice production without having to use much muscle activity (Hixon, 1973). However, because of the loss of elasticity in the lungs, the recoil pressure is not as effective in older individuals and this explains why they need to initiate speech at higher lung volumes (Huber, 2008). Another way of increasing the subglottal pressure would be to use the expiratory muscles. However, this is not the mechanism that is primarily adapted by elderly individuals when speaking

at a comfortable loudness, for three reasons: 1) with age, the decline in expiratory muscle strength is more important than the decline in inspiratory muscle strength, as may be reflected by values of MEP and MIP. Starting at 65 years old, there is a decline of 2-3 cmH_20/year in MEP and of 1 cmH_20/year in MIP (Enright, Kronmal, Manolio, et al., 1994). Since the inspiratory muscles are better preserved, using them to increase lung volume prior to initiating speech might be a more effective mechanism for elderly individuals than to rely on expiratory muscles (Huber, 2008). 2) Considering that older people have a reduced functional reserve, they are susceptible of reaching their expiratory reserve volume (ERV) more rapidly than younger adults (Huber, 2008; Tolep & Kelsen, 1993). Once ERV is reached, one has to generate pressure using the expiratory muscles in order to overcome the negative pressure created by the lung-thorax unit (Huber, 2008). Therefore, by using a higher lung volume the chances of speaking below the resting point are reduced. 3) The third reason is related to the changes in the vocal folds and the resulting poor laryngeal valving(Hoit & Hixon, 1987). Because of the reduced laryngeal resistance, more air is likely to escape during the closed phase of the vocal folds vibration and consequently a larger volume of air is spent during speech (Hoit & Hixon, 1987; Huber & Spruill, 2008). By using a higher lung volume, individuals can speak for longer durations before taking a pause for a breathing (Huber, 2008). However, despite the use of this strategy, it has been found that older adults produce shorter utterances when compare to a younger group, regardless of the voice loudness (Huber, 2008). Moreover, Huber and Spruill (Huber & Spruill, 2008) found that, when talking loud over background noise, older individuals tend to use their abdomen significantly more than when talking at a comfortable level, whereas younger people tend to rely more on passive recoil forces to increase loudness(Huber & Spruill, 2008). The authors explain this result with the hypothesis that, since older adults are already initiating phonation at high lung volumes for comfortable sound level, increasing the volume even more might be too demanding for the inspiratory muscle

strength available (Huber, 2008). Nonetheless, by using their expiratory muscles to increase loudness, elderly individuals need to take more pauses for breath intake because they are expending air at an increased flow rate (Huber & Spruill, 2008).

In summary, changes in the vocal folds affect laryngeal resistance and create a need for an increased subglottal pressure. However, changes in respiratory muscle strength and in the lung-thorax unit compliance reduce the ability of older individuals to generate and maintain a sufficient subglottal pressure (Huber & Spruill, 2008). The consequences of these changes are more obvious in challenging situations in which more muscular force is needed, i.e. during production of longer utterances and during loud speech (Huber, 2008).

Changes in the Laryngeal System

Some of the key clinical findings of presbyphonia are a bowing of the vocal folds, prominence of the vocal processes, and the presence of a glottal gap. This section will discuss the etiology of these observed laryngeal changes as well as other age-related alternations that occur in the laryngeal system.

Epithelium and Lamina Propria

Numerous changes in the mucous membranes of the vocal folds (epithelium and lamina propria) have been reported in older individuals, affecting both the structure and function of these tissues. Comparisons between old and young vocal folds revealed increased cell desquamation and reduced quantity of cells (2-3 cells instead of 5-7 cells of thickness) in the epithelium of aged vocal folds (Goncalves, Dos Santos, Pessin, & Martins, 2016; Hammond et al., 2000). Moreover, in some of the elderly specimens, cell junctions in the epithelium were characterized by deep sulci, which were not present in younger vocal folds (Goncalves et al., 2016). Age-related changes have also been found in the lamina propria, with an impact on the shape and viscoelastic properties of the

vocal fold (Hammond et al., 2000). In the superficial layer of the lamina propria, collagen fibers are more abundant and distributed irregularly and elastic fibers are reduced in number and disorganized (M. Hirano, 1974; Martins et al., 2015). In the deep layer, the collagen network tends to be denser than in younger vocal folds (Goncalves et al., 2016). Changes in the aged larynx should not only be considered from a quantitative standpoint, but also from a qualitative one: the collagen fibers, although increased in number, are not fulfilling their role as well as in younger vocal folds, therefore leading to a decreased tissue resistance (Hammond et al., 2000). Moreover, their augmented presence does not reflect an increased collagen production, but rather a decrease in the turnover process (Hammond et al., 2000): because of an impaired expression in the enzyme collagenase, the mature collagen fibers do not degrade and instead become distorted and rearrange in clusters and disorganized layouts, thereby affecting the flexibility of the tissue (Goncalves et al., 2016). Regarding elastic fibers, they have been found to undergo significant morphologic and metabolic changes with aging: they vary in size, contain less microfibrils and their surface is less uniform (K. Sato & Hirano, 1997). They authors conclude that these changes are important enough to play a role in the perceptual changes associated with the aging voice (K. Sato & Hirano, 1997).

Although the structural changes in the fibroblast cells of aged larynges have received little attention, some authors have stressed their potential role in the protein alterations in the elderly vocal folds. Hirano et al. (M. Hirano, Sato, & Nakashima, 2000) found that fibroblasts in the macula flava (MF) of the vocal folds were less active in the geriatric group when compared to the adult group and that the fibroblasts in Reinke space (RS) were more active in the geriatric group, as shown by the degree of development of their rough endoplasmic reticulum (rER) and Golgi apparatus (GA). These results are important given that fibroblasts in MF are responsible for producing collagen and elastic fibers for the vocal ligament, whereas those in RS produce

intercellular material such as fibronectin and HA and participate in wound repair. As a matter of fact, the concentration of HA has been found to be reduced in older vocal folds (S. Hirano, Tateya, Kishimoto, Kanemaru, & Ito, 2012; K. Sato, Hirano, & Nakashima, 2002). According to Hirano et al. and to Roberts et al. (Roberts, Morton, & Al-Ali, 2011), the reduced activity of the fibroblasts in the aged vocal folds are partly responsible for the changes observed in the lamina propria and epithelium of elderly people.

Laryngeal Muscles

Age-related atrophy of the TA muscle and consequential reduction in force, speed of contraction and endurance (McMullen & Andrade, 2009) have been reported by many authors, but the findings regarding patterns of fiber loss vary across studies. The speed of contraction of a muscle fiber depends on the myosin heavy chain (MHC) isoform, which is the main protein in skeletal muscles (Suzuki et al., 2002). Five MCH isoforms have been identified in animal and human intrinsic laryngeal muscles: I (slow), IIA, IIB, IIX and IID (fast to different degrees) and IIL (super-fast) (Suzuki et al., 2002). One study found a loss of type I fibers (Kersing & Jennekens, 2004), while another study reported only a loss of type II fibers (Malmgren, Fisher, Bookman, & Uno, 1999) and a third study reported a loss of both types of fibers in aged TAs (T. Sato & Tauchi, 1982). Despite the reduced number of muscle fibers, a recent experiment by Ziade and colleagues (Ziade, Semaan, Ghulmiyyah, Kasti, & Hamdan, 2016) using magnetic resonance imaging (MRI) found no significant differences in dimensions (length, width, height) and volume between the TA muscles of aged (65 and older) and younger larynges (Ziade et al., 2016). The unchanged volume and dimensions could be explained by a compensation hypertrophy in the remaining muscle fibers (Malmgren et al., 1999). Other explanations could lie in the alterations occurring in the connective tissue of the TA muscle: an increase in ragged red fibers (resulting from the increased

mitochondrial activity) (Kersing & Jennekens, 2004) and in collagen fibers (Kersing & Jennekens, 2004; Rodeno, Sanchez-Fernandez, & Rivera-Pomar, 1993) could partly account for the unchanged muscle dimensions and volume despite the loss of muscle fibers.

The degeneration process is not limited to the TA muscle but also affect the other intrinsic laryngeal muscles as confirmed by observed changes in proportions and diameters of type I and type II fibers and reduced EMG amplitudes (K. K. Baker, Ramig, Sapir, Luschei, & Smith, 2001) in other laryngeal muscles. Interestingly, Suzuki et al.(Suzuki et al., 2002) found alterations in MHC isoform profiles only in adductory muscles and not in the PCA, suggesting that the function of respiration might be better preserve than those of phonation and airway protection (Suzuki et al., 2002). In the TA and the LCA, a decrease in type IIB isoforms was found along with an increase in a slower fiber type, IIX. In the LCA muscle, an increase in the type IIA isoform was noted. The authors concluded to a replacement of fast contracting fibers to slower contracting fibers, a phenomenon similarly observed in denervated muscles (Suzuki et al., 2002).

One of the underlying mechanism that could be partly responsible for the loss of certain muscle fibers is the decrease in blood flow (hypoperfusion) that has been observed in the intrinsic laryngeal muscles. A study by Lyon & Malmgren (Lyon & Malmgren, 2010) showed a significant blood flow reduction of 60% in the TA muscles, of 50% in the CT muscles and of 42% in the PCA muscles in older rat larynges versus younger ones. In humans, while no change in capillary length have been found with aging, alterations related to the arrangement of the capillary contacts have been noted in intrinsic laryngeal muscles and are likely to impact blood flow (Lyon, Steer, & Malmgren, 2007). Decreased blood flow has a direct effect on muscular fatigue resistance as it affects the oxygen and nutrient supply to the muscle and reduces cellular waste elimination, leading to an altered muscle function (Lyon & Malmgren, 2010; L. B. Thomas, Harrison, & Stemple, 2008). Moreover, chronic hypoperfusion can induce an enhanced production of reactive oxygen

species (ROS) by the mitochondria and cause damage to the cells, thereby inducing a loss of muscle fibers (Lyon & Malmgren, 2010).

Cartilages, Bones and Joints

Mineralization (calcification) and ossification processes are observed in the cartilages of the larynx with aging as a response to mechanical stress induced by the laryngeal muscles on the tissues over time (von Glass & Pesch, 1983). These changes are observed only in the hyaline cartilages (thyroid, cricoid and arytenoid cartilages) since ossification does not occur in elastic cartilages (epiglottis, corniculate and cuneiform cartilages) (Turkmen et al., 2012). The process starts as soon as during the second decade of life, with an increased progress in the fifth decade for the thyroid cartilage and in the third decade for the cricoid cartilage (Turkmen et al., 2012). Along with the progressive ossification of the arytenoid and cricoid cartilages, changes in the cricoarytenoid joints have been reported and affect the range of movement of the articulation (Casiano, Ruiz, & Goldstein, 1994). These include a disorganization of collagen fibers, irregularity and thinning of the articular surface, and fibrosis of the periarticular region (Casiano et al., 1994; Ramig et al., 2001). Tissue changes in the insertion zones of the vocal ligament have also been observed in the anterior commissure, where fibrocartilage is slowly replaced by bone with advancing age (Paulsen, Kimpel, Lockemann, & Tillmann, 2000). The stiffening caused by ossification of the cricoarytenoid joint and vocal ligament insertion zone is thought to limit the extent of approximation of the vocal folds, further hindering their adduction in older speakers (Paulsen et al., 2000; Ramig et al., 2001).

The ossification of the laryngeal cartilages has an impact on the resonance of the sound, which is further altered by the lowering of the larynx and by the changes in the length and volume of the vocal tract (Linville & Fisher, 1985; Xue & Hao, 2003). A large part of the dimension changes in the

vocal tract is thought to be attributable to the increment in oral cavity volume and these alterations impact the formant frequencies of the different vowels (Xue & Hao, 2003). More research is needed regarding the effects of craniofacial skeleton enlargement and facial muscle atrophy on voice and more specifically on resonance (Xue & Hao, 2003).

Mucosal Glands

The concentration of laryngeal glands decreases with aging as the glandular tissues are either atrophied or replaced by adipose tissue (Tomita, Nakashima, Maeda, Umeno, & Sato, 2006). In addition, serous and mucous glands undergo age-related morphologic changes affecting the cytoplasm of the cells and these changes impact the quantity and properties of the secretions, which become thicker (Johns, Arviso, & Ramadan, 2011; K. Sato & Hirano, 1998). These alterations hinder the lubrication process of the vocal folds and this is thought to play a role in impeding vocal function (K. Sato & Hirano, 1998).

Impact on Phonation Physiology

The abovementioned physiological changes in the larynx have significant repercussion on vocal function. Because of the changes occurring in the epithelium, lamina propria and TA muscle, the vocal folds become thinner and weaker (Ximenes Filho, Tsuji, do Nascimento, & Sennes, 2003). This vocal fold atrophy leads to an incomplete glottal closure and a shorter closed phase during vibration, particularly in men. In fact, age-related changes in the vocal fold tissue and laryngeal cartilages have been found to be more extensive in men than women, with 67% of older men presenting with vocal fold atrophy versus 26% of older women in a study by Honjo and Isshiki (Honjo & Isshiki, 1980). The glottal gap caused by the atrophy is further increased by changes in the joints and insertion zones of the vocal ligaments as well as in the other adductory muscles. To

compensate for this lack of closure, speakers tend to increase muscular adduction force (Higgins & Saxman, 1991) which can involve medialization of the false vocal folds towards the midline. The result of this compression is the need for a greater subglottal pressure to overcome the phonation threshold pressure and initiate or maintain vocal fold vibration (Higgins & Saxman, 1991). The need for a greater subglottal pressure is further induced by the changes in the structural properties of the vocal fold epithelium and lamina propria and laryngeal mucosal glands. By increasing the stiffness of the vocal folds and reducing its lubrication, these alterations hinder the vertical and medial propagation of the mucosal wave and the vertical height difference indispensable for a healthy vibration (Zhang, 2016b).

Even when compensatory mechanisms are used to promote vocal fold adduction, an increased airflow and a shorter closed phase often persist in older subjects and are indicative of the persistence of poor laryngeal valving (Higgins & Saxman, 1991; Hoit & Hixon, 1987). This adds on to the limitations of the respiratory muscles and participates in decreasing utterance length and vocal quality in this population.

Changes in the Neurological System

Central Nervous System and Feedback

Although it has been suggested that the CNS loses fine motor control with aging and that this may affect voice production (Johns et al., 2011), research that correlates age-related changes in upper neurons with vocal function are sparse. Liu et al. (Liu, Chen, Jones, Huang, & Liu, 2011) examined the effect of age of the pitch-shift reflex (Larson, 1998) and found that the magnitude of the vocal response to a shift in the artificially fed back F0 changed as a function of age. The magnitude of the response increases up to the age of 51-60 years old, after which it starts to gradually decrease (Liu et al., 2011). Statistical analyses confirmed that the changes in voice F0 observed with aging

were not correlated with the alterations in pitch-shift responses. The authors explain their results by a decrease vocal motor control induced by the physiological changes in the laryngeal muscles, but also by alterations in the CNS. In fact, aging has been associated with a decrease in the number of synapses and a subsequent reduction in grey matter volume (Sowell et al., 2003). The decrease in inhibitory synapses may reduce the capacity of the cortical and subcortical systems to inhibit the vocal response to a F0 shift, therefore resulting in altered response magnitudes (Amenedo & Diaz, 1998; Liu et al., 2011). Elderly people also present with different neural representations of sounds than younger individuals, as demonstrated by distinctive responses in the temporal lobes (Bellis, Nicol, & Kraus, 2000). The effect that these changes, as well as hearing impairment, could have on voice production has not been studied yet.

Changes in proprioception related to the laryngeal respiratory system are also mentioned in the literature. In fact, older people have a reduced ability to integrate information provided by sensors via the afferent somatosensory route (Janssens et al., 1999). Regarding the respiratory system, they are also thought to rely more on feedback from the respiratory muscles in opposition to younger people who rely more on lung proprioception to gauge lung volume for phonation.

Peripheral Nervous System

As it has been previously detailed, aging brings about remodeling in the respiratory and laryngeal muscles, including changes in the proportion of slow and fast fiber types. Part of the reason for this remodeling lies in the interdependent relationship between the muscles and the motor neurons. With age, morphological changes occur in the superior and recurrent laryngeal nerves including a reduction in Schwann cells and myelinated fibers as well as a loss of large axons and a decrease in their diameter (Nakai, Goto, Moriyama, Shiraishi, & Nonaka, 2000; Tiago, Pontes, & Brasil Ode, 2008). Mortelliti et al. found a 31% decrease in myelinated nerve fibers in the superior

laryngeal nerves from the older group when compared with the younger group (Mortelliti, Malmgren, & Gacek, 1990). Similar findings have been reported affecting the myelinated fibers of the phrenic nerve (innervating the diaphragm) and resulting in reduced action potential amplitudes of this important inspiratory muscle (Imai et al., 2005; cited in Laller 2013) (Lalley, 2013). As a result of these changes, the nerves lose conduction velocity. Since the nerves are responsible for transmitting the neural input to the muscle and are therefore essential for the muscles' survival, this loss in conduction effectiveness leads to partial denervation of the muscle fibers and explains manifestations such as fiber loss and muscle atrophy (L. B. Thomas et al., 2008).

Neuromuscular Junction

As a result of the spontaneous axonal degeneration described above, changes happen at the level of the neuromuscular junction (Perie, St Guily, Callard, & Sebille, 1997), which are similar to alterations observed in denervated muscles (Connor, Suzuki, Lee, Sewall, & Heisey, 2002). Firstly, the percentage of motor fibers that are innervated through multiple branches of a same axon is significantly reduced in intrinsic laryngeal muscles (TA, PCA, IA, CT) (Perie et al., 1997). Secondly, the mean lengths of the neuromuscular junctions are decreased with age. These changes in innervation reduce the ability to sustain synaptic transmission and therefore participate in the sarcopenia process of the laryngeal muscles (Connor et al., 2002). In order to limit the impact on muscle function, a reorganization of the neuromuscular junctions takes place, during which the muscle fibers that were denervated are re-innervated by sprouts of axons belonging to adjacent fibers, otherwise known as the denervation/re-innervation process (Perie et al., 1997).

A similar phenomenon is observed in respiratory muscles innervation. The loss of motor neurons in the spinal cord leads to denervation of type II fibers in the respiratory muscles, which are then

re-innervated by adjacent slow-twitch motor neurons (type I) through axonal branching (Tolep & Kelsen, 1993). This process results in alterations in respiratory muscle structure and function. Moreover, changes in the neuromuscular junctions at the level of the diaphragm and of the intercostal muscles have been observed in aged rats and in humans, including a decrease in acetylcholine receptors and acetylcholine release, as well as a reduce number of end plates (Smith et al., 1990; Jan and Van Remmen, 2011; cited in (Lalley, 2013)).

Coordination Between Systems

Changes in the respiratory, laryngeal and neurological systems result in a decreased sensorimotor function, affecting the coordination between the subsystems. In fact, studies on animals have shown alterations in the temporal relationship between laryngeal and respiratory movements during quiet breathing, with older animals often initiating inspiration during the closed phase of the vocal folds (Nagai, 2005; cited in Johns et al., 2011)(Johns et al., 2011). Regarding phonation in humans, the lack of coordination between the laryngeal and respiratory systems adds up to the decreased rapidity of movements and can translate into timing deficiencies (Mathieson, 2006). These include the production of sounds on voiceless phonemes and on pauses during speech (the vocal fold abduction is too slow) as well as in increased voice-onset time (the vocal fold adduction is not well coordinated with the exhalation) (Mathieson, 2006).

Changes in the Endocrinal System

While the fundamental frequency of men tends to increase with aging, that of women tends to decrease. Many authors have explained this contrast by hormonal changes that occur with menopause in women around 50 years old, bringing about various biological changes throughout

the body, including the larynx (D'Haeseleer et al., 2009). With menopause, the ratio of estrogen-progesterone to androgens is altered and results in an enhanced influence of androgens on the organs (D'Haeseleer et al., 2009). The increased ratio of testosterone to estrogens has also been reported as having an effect on voice, possibly involved in the decrease in F0 (Lindholm, Vilkman, Raudaskoski, Suvanto-Luukkonen, & Kauppila, 1997). Visual examinations of the larynx in menopaused women revealed edema of the free edges of the vocal folds, a loss of the white appearance of the mucosa, and the presence of microvarices (Abitbol et al., 1999; Schneider, van Trotsenburg, Hanke, Bigenzahn, & Huber, 2004). The edema of the vocal fold causes an increase in the vibrating mass and a subsequent decrease in F0 and a reduced vocal range (Abitbol et al., 1999; Schneider et al., 2004). Reduced flexibility and stability of the voice have also been reported by menopaused singers, along with a loss of the higher register (Boulet & Oddens, 1996).

Even though the abovementioned changes in the larynx appearance are thought to be the result of hormonal changes in menopaused women, it remains challenging to differentiate some of the menopausal symptoms from other age-related changes in the larynx (D'Haeseleer et al., 2009). Moreover, there is a lack of scientific studies to support the mechanism through which hormonal changes may induce laryngeal manifestations. In fact, studies have reported conflicting results regarding the mediation process of hormonal influence and the presence of sex hormone receptors in the larynx (D'Haeseleer et al., 2009; L. B. Thomas et al., 2008). For these reasons, some authors have suggested that menopause might affect the larynx indirectly. One example of that is the high prevalence of gastroesophageal reflux symptoms that are associated with menopause and that are known to induce edema of the vocal folds (D'Haeseleer et al., 2009).

Section 3: Current Treatment Approaches for Presbyphonia

Behavioral Therapy

Different exercise programs have been applied to the treatment of presbyphonia, but the biological rationale for this type of intervention has not been thoroughly studied (Johnson, Ciucci, & Connor, 2013), often because of feasibility issues (Johnson et al., 2013). Although strengthening exercises have been shown to reverse the sarcopenia process in limb muscles, their physiological effect on laryngeal age-related alterations still remains unclear (L. B. Thomas et al., 2008). To address this question, Johnson and colleagues (Johnson et al., 2013) used an animal model to study the impact of exercise on neuromuscular plasticity. The results suggest that vocal training can mitigate some of the age-related differences in the neuromuscular junction of the TA muscle by reducing motor endplate dispersion in the muscle. This is likely to improve synaptic transmission, which is dependent on the synapse spatial arrangement, and therefore to increase muscle strength and allow for a reliable muscle contraction (Johnson et al., 2013).

In the current intervention paradigm, the first line of treatment for presbyphonia is voice therapy with a speech and language pathologist before suggesting a medical or surgical approach to improve glottal closure (Lu, Presley, & Lammers, 2013). However, the literature supporting behavioral approaches for presbyphonic patients is still sparse. Few research papers have been published in relation to behavioral voice therapy for the aging voice, and most of them have a low level of evidence (J. M. Oates, 2014). Nonetheless, these studies have generated data on the effect of different exercise protocols on the presbylarynx and have painted the picture of treatment approaches that are currently used in clinics by SLPs.

Vocal Function Exercises

Stemple's program of Vocal Function Exercises (VFE) is the treatment approach that has received most attention in presybyphonia studies to date. Its impact on the aging voice was examined in seven studies (E. E. Berg, Hapner, Klein, & Johns, 2008; Gorman, Weinrich, Lee, & Stemple, 2008; Kaneko et al., 2015; Sauder, Roy, Tanner, Houtz, & Smith, 2010; Tanner, Sauder, Thibeault, Dromey, & Smith, 2010; Tay, Phyland, & Oates, 2012; Ziegler, 2014), although sometimes combined with other approaches such as resonant voice therapy (E. E. Berg et al., 2008) or unsuccessful injection laryngoplasty (Tanner et al., 2010).

The program consists of a series of four specific exercises that were designed to improve the strength, endurance, flexibility and stability of the laryngeal and respiratory systems and to build up the balance between muscular effort and airflow (Stemple, Lee, D'Amico, & Pickup, 1994; Tay et al., 2012). The exercises are as followed: 1) sustain the vowel /i/ on the musical note F (above middle C for women; below middle C for men) for as long as possible; 2) glide from the lowest note to the highest note on the word "knoll"; 3) glide from the highest note to the lowest note of the word "knoll"; 4) sustain the notes C-D-E-F-G (starting from middle C for women and an octave below middle C for men) on the word "oll", for as long as possible. The placement of the voice is forward-focused for all the exercises (the rationale of a forward-focused voice will be detailed in the section on resonant voice). The protocol is usually practiced once a week with a SLP and the participants complete daily practices of five minutes twice daily for the duration of the therapy, which varies from four (Ziegler, 2014) to 12 (Gorman et al., 2008) weeks in the literature.

The VFE program is thought to induce a positive effect on the vocal fold vibration by promoting its regularity (Stemple et al., 1994), which in turns decreases acoustic measures of jitter and shimmer as well as the perceptual correlate of roughness (Tay et al., 2012). As a matter of fact,

roughness and jitter have been associated with the periodicity of vibration (Gorham-Rowan & Laures-Gore, 2006). This reduced roughness as well as decreased acoustic measures of perturbation have been observed in a study by Tay et al., who examined the effects of six weeks of VFE on a group of 22 aging choral singers randomized to a treatment or no treatment group (Tay et al., 2012). In this study, significant improvements in maximum phonation time (MPT) have also been observed in the VFE group but not in the control group. The authors explain these improvements by the enhanced coordination between the phonatory and respiratory systems, as well as by the increased strength and endurance of the laryngeal and respiratory muscles. The investigators didn't find improvements in auditory perception of breathiness and strain and they explain that finding by a lack of sensitivity of the auditory-perceptual instrument (the Perceptual Voice Profile (J. Oates & Russel, 1997)) when used with non-impaired voices, as it was the case in those healthy choir singers. On the other hand, Sauder and colleagues (Sauder et al., 2010) applied six weeks of VFE on nine participants with diagnosed presbyphonia and found significant reductions in breathiness and strain during reading tasks as judged by blinded listeners. In this study, the exercises also led to a lessened score on the Voice Handicap Index (VHI) and decreased self-assessed severity and phonatory efforts but no significant difference was found in MPT and in acoustic and stroboscopy measures (Sauder et al., 2010). Even so, the authors conclude that the improvements in perceptual measures for functional reading tasks and the reductions in perceived handicap, severity and vocal effort are sufficient to demonstrate the clinical validity of VFE for patients with presbyphonia (Sauder et al., 2010). They support that the absence of change in glottal gap, which is one of the hallmarks of presbyphonia, might be due to a lack of sensitivity of the laryngeal imaging technique used (videostroboscopy). They also suggest the alternative hypothesis that functional and perceptual improvements might have occurred even in the absence of an improvement in glottal closure because of combined alterations in the three

subsystems of phonation leading to enhanced vocal efficiency (Sauder et al., 2010). Gorman and his colleagues (Gorman et al., 2008) found statistically significant improvements in MPT and in some aerodynamic measures (including subglottal pressure and measures of glottal airflow) after a 12-week VFE intervention in 19 elderly male participants. The authors hypothesize that the exercises may have helped compensate for the decreased neural control of the laryngeal muscles by permitting more motor unit recruitment. This may have increased the muscle function in the larynx and helped counteract the age-induced atrophy, therefore leading to an improved medial compression of the vocal folds (Gorman et al., 2008). Even though no visual examination of the larynx was conducted as part of the outcome measures, the authors mention that the significantly improved aerodynamic measures relate to glottal closure, since an enhanced vocal fold adduction would lead to a better airflow management (Gorman et al., 2008).

Only two study found improvement in glottal closure. One of them is a case study of two twin participants (Tanner et al., 2010), therefore providing low evidence that VFE lead to an improved glottal closure in patients with presbyphonia. The other one is a retrospective study with historical controls and revealed improvements in GRBAS, MPT, jitter (but not shimmer), normalized mucosal wave amplitude, normalized glottal gap, and VHI-10 (Kaneko et al., 2015). No improvements were found in shimmer, intensity, mean flow rate and bowing index. It is possible that vocal fold atrophy is resistant to intervention, because it is not only originating from the muscle atrophy, but also from changes in the connective tissue of the lamina propria and those might not be reversible with exercise, thus explaining the lack of change in bowing index (Kaneko et al., 2015). Even though the exercises might be effective in improving the function of the TA muscle and therefore the vibratory status (Kaneko et al., 2015), in some cases this might not be sufficient to close the gap observed in this population. This further stresses the importance of addressing the three systems involved in phonation and not only vocal fold adduction. The lack of

improvement on certain outcome measures could be partly explained by the fact that VFE don't directly address the strengthening of the respiratory system and therefore overlook a key component of presbyphonia. However, statements regarding the mechanisms of action of VFE in patients with presbyphonia remain assumptions since the underlying physiological effects of the treatment have not been confirmed (Sauder et al., 2010). The available data stems from assessment of the clinical manifestations of the voice disorder before and after the intervention. Some limitations that could explain the discrepancies in the results of the studies include: differences in treatment durations, inclusion and exclusion criteria chosen to recruit participants, small sample sizes, and sensitivity and reliability issues of the different assessment instruments (see section on outcome measures). It is also noteworthy that, of the abovementioned studies, only two (Tay et al., 2012; Ziegler, 2014) included a concurrent control group. One of them, by Ziegler and colleagues (Ziegler, 2014), is the only one to have directly compared four weeks of VFE with another intervention approach as well as with a no intervention control group, in prospective randomized controlled trial (level II evidence) (table 2). They found that VFE leads to

Table 2. Levels of Evidence from the National Health and Medical Research Council (NHMRC) (https://www.nhmrc.gov.au/_files_nhmrc/file/guidelines/developers/nhmrc_levels_grades_evidence_120423.pdf)

Level	Intervention [1]	Diagnostic accuracy [2]	Prognosis	Aetiology [3]	Screening Intervention
I [4]	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies
II	A randomised controlled trial	A study of test accuracy with: an independent, blinded comparison with a valid reference standard,[5] among consecutive persons with a defined clinical presentation[6]	A prospective cohort study[7]	A prospective cohort study	A randomised controlled trial
III-1	A pseudorandomised controlled trial (i.e. alternate allocation or some other method)	A study of test accuracy with: an independent, blinded comparison with a valid reference standard,[5] among non-consecutive persons with a defined clinical presentation[6]	All or none[8]	All or none[8]	A pseudorandomised controlled trial (i.e. alternate allocation or some other method)
III-2	A comparative study with concurrent controls: ▪ Non-randomised, experimental trial[9] ▪ Cohort study ▪ Case-control study ▪ Interrupted time series with a control group	A comparison with reference standard that does not meet the criteria required for Level II and III-1 evidence	Analysis of prognostic factors amongst persons in a single arm of a randomised controlled trial	A retrospective cohort study	A comparative study with concurrent controls: ▪ Non-randomised, experimental trial ▪ Cohort study ▪ Case-control study
III-3	A comparative study without concurrent controls: ▪ Historical control study ▪ Two or more single arm study[10] ▪ Interrupted time series without a parallel control group	Diagnostic case-control study[6]	A retrospective cohort study	A case-control study	A comparative study without concurrent controls: ▪ Historical control study ▪ Two or more single arm study
IV	Case series with either post-test or pre-test/post-test outcomes	Study of diagnostic yield (no reference standard)[11]	Case series, or cohort study of persons at different stages of disease	A cross-sectional study or case series	Case series

positive changes in voice-related quality of life (V-RQOL) but not in perceived phonatory effort contrarily to the other intervention, phonation-resistance training exercise (PhoRTE). The details of the PhoRTE approach, as well as possible reasons that could explain its slight superiority over VFE, are presented in the following section.

PhoRTE and LSVT Interventions

The PhoRTE program, adapted from the Lee Silverman Voice Treatment approach, consists of four exercises: 1) loud maximum sustained phonation on the vowel /a/; 2) loud ascending and descending pitch glides over the entire pitch range on the vowel /a/; 3) functional phrases using a loud and high voice; and 4) phrases from exercise #3 in a loud and low voice.

The rationale behind this approach, which was designed specifically for patients with presbyphonia based on the LSVT program, is that an effective intervention should target the underlying causes of the disorder and thus in the case of presbyphonic patients should load both the laryngeal and the respiratory systems (Ziegler, 2014). The authors of the PhoRTE intervention support that this approach allows to target the phonatory biomechanics directly but also indirectly, via improved respiratory biomechanics (see figure 9). The load on the two targeted systems is achieved by asking the participants to produce a loud/strong voice. They are encouraged to use a low abdominal breathing pattern, however the only feedback offered during the execution of the exercises is to maintain a "strong voice". More precisely, the targeted intensity is 80-90 dB and this goal is monitored via a microphone placed in front of the participant (Ziegler, 2014) (it is not clear from the article if the participant has access to the visual feedback of the produced and expected sound pressure levels). Producing a loud voice is a high intensity task as it requires sustained effort from the laryngeal and respiratory muscles. This effort, or

overload, is what is thought to induce the neuromuscular changes necessary to improve strength and to reduce the perceived effort during voice production (Ziegler, 2014).

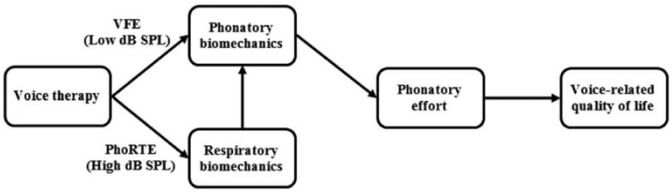


Figure 9. Model linking voice therapy to changes in phonatory and respiratory biomechanics, phonatory effort, and voice-related quality of life (Ziegler, 2014).

Reduced phonatory effort can also be explained by the source-filter interaction (Titze, 2008). When performing the exercises on the vowel /a/, participants have a wide-open mouth and a narrowed pharynx, therefore making the form of megaphone with their vocal tract. The megaphone shape has the effect of raising the frequency of the first formant and amplifying the fundamental and second harmonic of the sound at the source (Ziegler, 2014). It is also an optimal configuration to promote vocal fold adduction and to transfer the sound from the source to the lips, thus providing maximal intensity (Ziegler, 2014). This source-filter interaction might explain the significant improvements in perceived phonatory effort in the PhoRTE group and the lack of improvement in the VFE group. The VFE are produced on the rounded vowel /o/, stimulating an inverted megaphone shape in the vocal tract (a wide pharynx and a narrowing at the lips), which is thought by the authors to decrease vocal fold adduction (Ziegler, 2014). In the PhoRTE group, the assumed increased vocal fold adduction and resulting reduction in glottal gap could have lowered the respiratory effort necessary to maintain a sufficient subglottal pressure and this may have led to a reduced perceived phonation effort. The intervention also led to significant increases in voice-related quality of life, although this result was found in both the VFE and the PhoRTE groups (Ziegler, 2014).

The PhoRTE intervention is based on the same principles as the LSVT. LSVT was initially intended for patients with Parkinson disease (PD) with the goal of improving their respiratory drive and vocal fold adduction and consequently increase their voice loudness and quality (Lu et al., 2013).

The program, which can only be conducted by a LSVT-certified SLP, consists of 16 60-minute sessions over a four-week period. The first half of each session is allocated to exercises designed to increase pitch range and maximum phonation time (maximum F0 range and sustained vowels at loud intensity (Ramig et al., 2001)), followed by the production of functional sentences at loud intensity (Lu et al., 2013). During the second half of the sessions, participants practice their loud and strong voice in different speech tasks including reading, question answering, word production, and conversation (Lu et al., 2013; Ramig et al., 2001). Participants are also given exercises to practice at home to promote carryover of the treatment effects (Lu et al., 2013; Ramig et al., 2001). The LSVT protocol was conceived following motor learning and skill acquisition principles: it is intensive, it involves numerous repetitions (minimum of 12 to 15), it requires high effort from the participants and the instructions are simple and targeted ("loud voice")(Ramig et al., 2001). The rigorous vocal exercises combined with a focus on pitch and intensity are expected to minimize vocal fold bowing and promote vocal fold adduction(Lu et al., 2013).

In patients with PD, LSVT has been shown to increase subglottal pressure, SPL, and F0 variations and to improve glottic closure and functional speech production (Dromey, Ramig, & Johnson, 1995; Ramig, Countryman, O'Brien, Hoehn, & Thompson, 1996; Ramig, Countryman, Thompson, & Horii, 1995; Ramig & Dromey, 1996). Deeming that these improvements would be beneficial for patients with presbyphonia, Ramig and his colleagues(Ramig et al., 2001) were the first ones to apply LSVT to this patient population, in 2001. Their study included three participants and demonstrated an increased SPL in the three of them, an improved voice quality in two of them, and a better vocal fold adduction in one of them. All patients reported a positive change in voice production. Another case series study evaluated the effect of LSVT on two patients with presbyphonia and included quantitative measurements of glottal gap size (GGS) (Lu et al., 2013). The GGS was significantly reduced in both participants, who also demonstrated a higher

frequency of complete closure per vibratory cycle. Significant improvements in SPL and pitch range were also found in both subjects, and MPT improved in one subject. Although the subjects didn't experience any laryngeal trauma during the four-week program, the level of supraglottic activity remained the same following treatment and no follow-up assessment beyond two weeks after the end of the program was conducted to ensure that the level of hyperfunction would not increase over time. As a matter of fact, the shortcoming in the PhoRTE and LSVT approaches is the likeliness that participants will rely on compensatory strategies to produce high intensity utterances, both at the laryngeal and respiratory levels. At the laryngeal level, medial compression of the false vocal folds could be enhanced to promote more adduction force of the vocal folds, along with a general increase in muscle tension. At the respiratory level, participants might rapidly reach their expiratory reserve and this could also promote a strained phonation (hyperfunctional laryngeal configuration). Since very limited feedback is provided in those intervention methods, participants do not have tools to modify their laryngeal and respiratory behaviors. The impact of these compensatory mechanisms might have detrimental effects on the voice in the long term, even though a higher vocal loudness is achieved at a short-term level. The long-term effects of these therapies have not been studied on patients with presbyphonia.

An aspect of the LSVT intervention that could represent an impediment to its implementation is the intensity component. It might be unrealistic for elderly patients to attend four therapies per week for four weeks, because of possible mobility or geographical barriers (Lu et al., 2013). Various alternatives have been studied with PD patients, such as an online delivery program (Constantinescu, 2011), an assistive computer system (Halpern, 2012), or an extended version of the program (16 sessions over two months)(Spielman, Ramig, Mahler, Halpern, & Gavin, 2007). Although these options have been proved to be non-inferior to the original LSVT protocol, they have not yet been tested with the presbyphonic population.

Studies on LSVT for patients with presbyphonia are limited to case series and therefore have a low level of evidence (IV). The study on PhoRTE is a well-designed randomized controlled trial, associated with a high level of evidence (II). However, the study by Ziegler and colleagues (Ziegler, 2014) is the only one to have tested these exercises on patients with presbyphonia and contains limitations such as small sample size and limited set of outcome measures. More studies that would evaluate the effects of PhoRTE on the aged larynx are warranted.

Resonant Voice and Semi-Occluded Vocal Tract Exercises

The vocal tract configuration used in the VFE are thought to promote a semi-occluded vocal tract and a resonant voice, which help recalibrate the vocal fold adduction while promoting vocal efficiency (Gorman et al., 2008; Titze, 2006). In the case of patients with a glottal gap and possible false vocal fold hyperfunction, this would have this effect of recalibrating a better adduction while reducing the false vocal folds involvement. For this reason, resonant voice therapy and semi-occluded vocal tract exercises are both employed by SLP with patients with presbyphonia, although no study to date has specifically examined the effect of these treatment approaches on elderly patients (J. Gartner-Schmidt & Rosen, 2011; Mau, Jacobson, & Garrett, 2010).

Semi-occluded vocal tract exercises (SOVTE) include tasks such as lip trills and tongue trills, flow resistance straws or tubes and nasal consonants. The principles underlying the use of these exercises stem from the source-filter model of voice production (Titze & Worley, 2009), which supports that the vocal tract plays an active role in converting aerodynamic energy into acoustic energy. The semi-occlusion in the vocal tract enhances the interaction between the source (vibrating vocal folds) and the filter (the supraglottal tract) and this allows for an optimal vocal efficiency and economy (Titze, 2006). SOVTE increase the inertance of the vocal tract ("acoustic property of an air mass [usually a column of air in a tube] being accelerated or decelerated by

pressure")(Titze, 2001) (p. 520). Back pressures are generated in the vocal tract which, when reaching the glottis, create a slightly abducted/adducted vocal fold arrangement, a reduced vibration amplitude and a decreased phonation threshold pressure (Titze, 2006). The vocal folds, instead of being tightly adducted at the vocal processes through the action of the LCA, are uniformly adducted/abducted over their whole length through more TA involvement (Titze, 2006). In summary, SOVTE amplify the sound by optimizing the source-filter interaction instead of relying on an increased vibration amplitude, therefore protecting the vocal folds from collision impact (Titze, 2006).

The most effective semi-occluded configuration for vocal economy and efficiency has been found to be a narrow epilarynx and a wide mouth (in this case the semi-occlusion is located in the back of the vocal tract, in the epilarynx) (Titze, 2006). However, exercises with a wide epilarynx and a narrow mouth opening (such as in lip trills, tongue trills, straw phonation, humming, and VFE) are a safer start for voice therapy because the inverted megaphone shape generates lower vibration amplitudes at the level of the vocal folds and lower acoustic pressures (Titze, 2006). Moreover, it allows for easier control and more sensation of the occlusion, which is located at the front of the vocal tract (lips) (Titze, 2006). The pressure created behind the occlusion is translated into vibrations that allow the speaker to associate a healthy vocal fold configuration with proprioceptive sensations (Titze, 2006). Nevertheless, since the inverted megaphone shape doesn't produce the highest vocal loudness, the SLP should eventually bring the patient to a more open mouth configuration, in which case the epilarynx might naturally narrow in order to preserve the vocal efficiency gained through the SOVTE (Titze, 2006). This is particularly relevant for patients with presbyphonia since one of the main therapeutic goal with this population is to increase voice intensity. Relying on the source-filter interaction to do so is a promising alternative

for them because it allows for an increased intensity without solely depending on an increased vibration amplitude, which would inevitably tax the respiratory system.

Once an efficient vocal production is reached, the resulting intensified acoustic pressures in the oral cavity lead to another kinesthetic sensation: tissue vibrations in the face (alveolar ridge and maxillary bones)(Titze, 2006; Verdolini-Marston, Burke, Lessac, Glaze, & Caldwell, 1995). This proprioceptive feedback is the basis of resonant voice therapy, in which the patient learns to recognize these sensations and use them to maintain an adequate vocal fold configuration (Titze, 2006; Verdolini, Druker, Palmer, & Samawi, 1998). Resonant voice therefore results from the same acoustic phenomenon than in SOVTE and has the same goal: to maximize vocal production while decreasing the collision impact on the vocal folds by inducing a slightly adducted/abducted configuration (Verdolini-Marston et al., 1995; E. M. Yiu, Lo, & Barrett, 2017). Exercises are graded in a hierarchy from sounds (usually starting with humming) to words, sentences, reading, and lastly natural conversation (E. M. Yiu et al., 2017). Specific voice therapy programs based on resonance voice have been developed, such as Verdonlini's Lessac-Madsen Resonant Voice Therapy (LMRVT) and Stemple's resonance therapy program (E. M. Yiu et al., 2017). Resonant voice therapy has been studied mostly on patients with muscle tension dysphonia/hyperfunctional dysphonia (Ogawa et al., 2014; Ogawa et al., 2013; Edwin ML Yiu & Ho, 2002) and on professional voice users (Barrichelo & Behlau, 2007; Roy et al., 2003; Verdolini et al., 1998). A review of nine resonant voice studies (Barrichelo & Behlau, 2007; F. C. Chen, Ma, & Yiu, 2014; S. H. Chen, Hsiao, Hsiao, Chung, & Chiang, 2007; Ogawa et al., 2014; Ogawa et al., 2013; Roy et al., 2003; Verdolini et al., 1998; Verdolini-Marston et al., 1995; Edwin ML Yiu & Ho, 2002) revealed positive outcomes in perceptual voice quality, acoustic and aerodynamic measures, self-reported measures and phonatory effort measures following two to nine therapy sessions. One study found improvements in vocal fold closure after eight weeks of therapy (S. H.

Chen et al., 2007) and one study found significant reduction in medial and antero-posterior compression immediately following humming exercises (Ogawa et al., 2013).

While these results seem promising for patients with presbyphonia, who often present with glottal insufficiency and compensating hyperfunction, it is unclear what the impact of resonant voice and SOVTE would be on laryngeal features for this patient population. As a matter of fact, the level of evidence supporting the use of SOVTE and resonant voice therapy with patients with presbyphonia is very low. Only two retrospective studies (J. Gartner-Schmidt & Rosen, 2011; Mau et al., 2010) have reported treatment outcomes following these approaches. Although a retrospective cohort study would usually be associated with a level of evidence of III-2 (table 2), in both cases the voice therapy was described as an amalgam of direct techniques tailored to the patients and included LMRVT, flow phonation, SOVTE, resonant voice, and/or LSVT, making it impossible to isolate the treatment approach and to compare it to another treatment or no treatment.

(Stretch) and Flow Phonation

Flow phonation was also included in the direct voice technique approaches used by SLPs in the two retrospective reviews on treatment for presbyphonia (J. Gartner-Schmidt & Rosen, 2011; Mau et al., 2010). This approach, also known as "stretch and flow" (S_nF), was first described by Stone and Casteel (1982; cited in Watts et al., 2015)(C. R. Watts et al., 2015) and since then has been commonly used by SLPs to treat patients with hyperfunctional dysphonia (C. R. Watts et al., 2015). The approach consists of a hierarchy of vocal tasks that aim to develop a balance between the respiratory, phonatory and resonatory systems by producing an effortless voice. The program starts with a prolonged exhaled, followed by slow whispering, slow breathy voice, breathy speech at a faster rate and lastly production of a normal voice with minimal effort (C. R. Watts et al.,

2015). The general level of evidence to support S_nF was very low until recently, when Watts and colleagues (C. R. H. Watts, Amy; Toles, Laura; Childs, Lesley; Mau, Ted, 2015) conducted a randomized controlled trial on patients with muscle tension dysphonia. However, no study has examined the effect of this treatment approach on patients with presbyphonia or any other type of hypofunctional voice disorder, and the rationale for using flow phonation with patients with glottal insufficiency is unclear. The results of Watts' clinical trial revealed significant improvements in VHI, MPT and cepstral peak prominence (CPP) measures, indicating that flow phonation is effective in improving hyperfunctional voices (C. R. H. Watts, Amy; Toles, Laura; Childs, Lesley; Mau, Ted, 2015). This could be beneficial for patients with presbyphonia who present with secondary hyperfunction; however it might also be counterproductive regarding glottal insufficiency. Clinical trials assessing the impact of S_nF on patients with presbyphonia would be needed to answer these questions.

Symptomatic Approaches and Breathing Exercises

The treatment approaches described so far are categorized as "physiological approaches" because they are designed to modify physiology and aim at a more efficient voice production. They don't target a specific voice symptom but rather phonation as a whole. Some SLPs choose to follow these programs, while others use an amalgam of different techniques depending on the symptoms of the patient, thus following a "symptomatic approach". Voice techniques suggested in speech therapy books for patients with presbyphonia include the following:

Glottal attacks: using words and sentences starting with a vowel and use a gentle glottal attack to promote vocal fold adduction. This strategy targets the laryngeal features of glottal gap and vocal fold atrophy/bowing (Mathieson, 2006).

Vocal glides and arpeggios: glides and arpeggios executed on vowels are suggested to increase vocal pitch range, which is often decreased in older people. The glides are usually started at mid-range and gradually extending to higher and lower pitches (Mathieson, 2006). Vocal glides are similar to the second and third exercises of the VFE program and target muscle flexibility and contractile function.

Focused voice: Focused voice is another term for resonant voice and is often suggested as a voice technique in speech therapy book, for various voice disorders including presbyphonia (Mathieson, 2006).

Phrases of increasing loudness: some books suggest to gradually increase the patient's vocal loudness using short sentences (Mathieson, 2006). The issue with this strategy is that the patients might not have the muscular/pulmonary resources to increase loudness and might not succeed or might produce an unhealthy phonation. It is therefore important that the patient be encouraged to integrate other voice technique while increasing vocal loudness, such as a focused voice.

Glottal fry: glottal fry is usually employed with patients who have hyperfunctional voices although it has been suggested for patients with presbyphonia, who may present with compensatory hyperfunction. The rationale is that during glottal fry, the vocal folds are very relaxed and the airflow and subglottal pressure are reduced (Boone, 2010).

Auditory and visual feedback and masking: auditory feedback plays an important role in speech therapy. It can be provided by the therapist (modeling), or it can consist of a playback from the patient's voice. For patients who have an impaired auditory feedback, as it may be the case in older patients, other strategies are also suggested. Visual feedback can help the patient reach a targeted F0, loudness or air volumes in the lungs (Boone, 2010). This may be relevant for presbyphonic patients because older individuals also tend to have a decreased ability to integrate

information from proprioceptive sensors in the respiratory system (Janssens et al., 1999). Lastly, masking is proposed for patients with an impaired auditory feedback. The rationale behind this strategy is that a compromised auditory monitoring may lead to a suboptimal voice production and that masking it with white noise could result in clearer voice production (Boone, 2010). While this is appealing for patients with presbyphonia, there is no rationale to date that support the use of this technique with this population.

Counseling on vocal hygiene: vocal hygiene counseling (indirect voice therapy) is recommended in mostly all books of speech therapy, for all types of voice disorder. While it is important to limit vocal abuses and stay hydrated, indirect voice therapy should always be combined with direct voice therapy, because there is no evidence that vocal hygiene by itself leads to improvements in voice outcomes (Ruotsalainen, 2007).

Breathing exercises: Presbyphonic patients have less air volume and less respiratory muscle strength available for speech production because of the age-related changes in the respiratory system. Therefore, breathing exercises that aim to lengthen maximum phonation time are frequently used by SLP when working with patients with presbyphonia. The most common techniques, and those described (very briefly) in speech therapy books are: diaphragmatic breathing, increasing extent of thoracic expansion, increasing period of rib elevation, increasing period of expiratory airflow on phonemes /s, z, a, æ, i/. Emphasis is placed on feeling the abdomen and ribcage expand and retract with the hands and on maintaining a straight posture (Boone, 2010). The concern with these exercises is that they do not load the respiratory system and therefore it is likely that they won't result in increased respiratory muscle strength. Simply modifying the breathing pattern by encouraging abdominal breathing will not result in an enhanced support for phonation if the chest wall is rapidly collapsing due to inspiratory muscle

weakness. Moreover, there is very low evidence in the literature supporting the use of these techniques to improve respiratory and/or voice outcomes (Desjardins & Bonilha, 2019).

This lack of evidence is an issue with most of the abovementioned symptomatic approaches, with the exception of the focused voice techniques, which has been extensively studied. When employed in voice studies, these approaches are usually combined with various other techniques, which makes it challenging to draw causal relationships between the treatment and the outcomes. A good example of this issue is the Cognitive Vocal Program (CVP) (Nemr et al., 2014). The CVP is a therapy program that was conceived for patients with a presbylarynx. It consists of stretching, breathing, and voice techniques, including ascending and descending glissandos, repetitions of syllables with /p/ and /b/ and humming. Breathing exercises involved "costodiaphragmatic breathing pattern" with prolongation of air on the sounds /f, v, s, z, i, e/. The authors reported improvements in loudness, jitter, HNR, vocal quality, MPT, F0, glottal closure, and mucosal wave. However, the study comprised only three participants and did not include a control group nor statistical analyses, in addition to testing the efficacy of numerous techniques simultaneously.

Neuromuscular electrical stimulation

One study evaluated the impact of neuromuscular electric stimulation (NMES) on bowed vocal folds (Lagorio, Carnaby-Mann, & Crary, 2010). The hypothesis was that the positive effects of voice therapy would be enhanced by the simultaneous administration of NMES. To test this hypothesis, Lagorio and colleagues (Lagorio et al., 2010) conducted a prospective case series with seven participants exhibiting glottal insufficiency caused by bilateral vocal fold bowing. The NMES was delivered through a VitalSim NMES device (Empi Corporation, St Paul, Minnesota), which releases a biphasic pulsed current through electrodes that are strategically placed on the neck of

the participants. In Lagorio's study, one pair of electrodes was placed above and below the cricothyroid membrane to target the cricothyroid muscle. Another pair of electrodes was placed inferiorly to the horns of the hyoid bone, targeting the SLNs. The intensity of the current was regulated during the initial stroboscopy evaluation: the electrical stimulation was gradually increased until the vocal folds reached an improved glottal closure. The vocal exercises, a hierarchy of different vocalizations, were produced while the NMES was being administered. Treatment outcomes were assessed after three weeks of treatment, five days a week for one hour. Results indicate increased MPT and glottal closure as well as decreased supraglottic compression and improvements in VHI scores (Lagorio et al., 2010). These improvements were maintained up to three months following the end of the treatment. Although these results may seem promising, this study holds the significant limitation of not containing a voice exercises- only control group. For this reason, the individual contribution of NMES is yet to be known and the underlying mechanism that led to the observed improvements remains unknown. The authors hypothesized that the increased glottal closure may have been stimulated by an enhanced contraction of the CT muscle leading to more elongation and tension in the vocal folds (Lagorio et al., 2010).

A clinical trial compared voice therapy alone with voice therapy combined with NMES in patients with unilateral vocal fold paralysis (Ptok & Strack, 2008). The results showed a stronger improvement in vocal fold vibration irregularity in the combined group, but no significant improvement in MPT for either group. Moreover, the vocal exercises that were used in the study are not described, which further hinders the interpretation of the results. In summary, there is not sufficient evidence to support the addition of NMES to voice exercises in patients with age-related vocal fold atrophy and more research is needed before suggesting a change in clinical practice.

Surgical Options

Surgery is rarely the first line option for patients with presbyphonia. Their enhanced vulnerability to stressors due to a decline in physiologic reserves (frailty) (Johns et al., 2011) makes them more at risk for postoperative complications and longer hospital lengths of stay (Johns et al., 2011). However, when voice therapy does not yield satisfying results or if the patient presents with severe vocal fold atrophy and glottal insufficiency, surgical options are considered, with the two main options being injection augmentation and type I thyroplasty (Mau et al., 2010). A retrospective cohort review by Gartner-Schmidt and Rosen (J. Gartner-Schmidt & Rosen, 2011) was conducted on 275 patients with a primary diagnosis of vocal fold atrophy who presented at the University of Pittsburgh's Voice Center. Eighty-six of them had follow-up information in their file and of those, 15% had surgery alone while 44% had voice therapy alone, and 9% had surgery after unsuccessful voice therapy. In another retrospective cohort study, by Mau and colleagues (Mau et al., 2010), 67 patients with presbyphonia were examined. Of those, 48 received voice therapy only, six received voice therapy combined with injection augmentation, nine received injection only, and four received a thyroplasty type I intervention (with or without an injection augmentation).

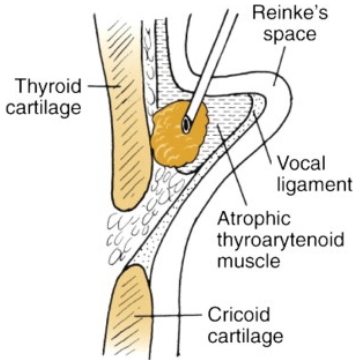


Figure 10. Injection Augmentation. From Fakhry C, Flint PW, Cummings CW. In: Cummings otolaryngology. 5th edition.

Injection Augmentation (Injection Laryngoplasty)

During injection augmentation, a biomaterial is injected with a needle deep to the lamina propria into the paraglottic space (figure 10) to medialize the atrophied vocal fold (Johns et al., 2011; Seino & Allen, 2014). The procedure can be done in the operating room, but office-based procedures in which the patient is awake and under

topical anesthesia are more common (Johns et al., 2011). Because they are often performed percutaneously, office-based injection augmentations present some risks for elderly patients who are likely to be on anticoagulation medication (Johns et al., 2011). Various biomaterials can be used and have been tested in the literature, with the most common ones being Cymetra (collagen-based), autologous fat, calcium hydroxylapatite (CaHA) microspheres (Radiesse Voice) or CaHA gel (Radiess Gel) (Cantillo-Banos, Jurado-Ramos, Gutierrez-Jodas, & Ariza-Vargas, 2013). These biomaterials are gradually degraded and therefore their effects are only temporary (Seino & Allen, 2014). CaHA injections have been found to have the longest duration with follow-up studies indicating an average benefit period of at least one year with improved MPT, perceptual ratings of voice quality and closed quotients, and no major complications (Carroll & Rosen, 2011; Kwon, An, Ahn, Kim, & Sung, 2010; Rosen et al., 2009). In the retrospective chart review by Gartner-Schmidt and Rosen (J. Gartner-Schmidt & Rosen, 2011), including injections of Cymetra and CaHA, success of treatment was defined by a change of at least five points on a total of 40 points on the VHI-10, a cut-off chosen by the authors based on clinical experience. Based on this criterion, the rate of success was 56% for patients who received injection augmentation only and 17% for patients who received injection augmentation following voice therapy. These low success rates highlight the limitations of surgery management for patients with presbyphonia and reveal that reducing the glottal gap by augmenting the bulk of the TA muscle may not be sufficient to improve functional voice outcomes, contrarily to in patients with vocal fold paralysis who demonstrated greater improvements (J. Gartner-Schmidt & Rosen, 2011; Seino & Allen, 2014). This can be explained by the fact that the injection augmentation does not restore the altered microarchitecture of the lamina propria, affected by age-related alterations (J. Gartner-Schmidt & Rosen, 2011). Therefore, even though glottal closure is improved by the intervention with subsequent increase in the closed phase of vibration and in loudness, the flaccidity of the vocal

folds may remain (J. Gartner-Schmidt & Rosen, 2011). The lack of tone and the difference of tone between the two vocal folds have a negative impact on vibration regularity, and thus perceived roughness is likely to persist even after the injection (J. Gartner-Schmidt & Rosen, 2011). The patients' expectations with regards to the effects of surgery on their voice might explain the lack of improvement in VHI scores.

Moreover, it is noteworthy that biomaterials such as Radiesse can lead to undesirable results if the amount of injected material and location of the injection are not adequate. An over-injection or superficial injection would increase the rigidity of the tissue and hinder the mucosal wave and vibratory properties of the vocal folds (Seino & Allen, 2014). On the other hand, insufficient injection would not sufficiently improve glottal closure and in this case a complementary type I thyroplasty might be necessary (Seino & Allen, 2014), which will be described in the following section.

Type I Thyroplasty (Medialization Thyroplasty)

While injection augmentations are mostly used for mild glottal gaps, type I thyroplasties have been recommended for patients with severe glottal insufficiency (gap>3 mm, VHI≥90) (Cantillo-Banos et al., 2013; Seino & Allen, 2014). However, some authors suggest that both methods are equally valid in mild and moderate cases and that clear criteria to determine one or the other technique are not provided in the literature (Cantillo-Banos et al., 2013). Type I thyroplasty, also designated as laryngeal framework surgery or medialization thyroplasty, is a permanent solution to glottal insufficiency because the vocal folds are medialized using implants that are not intended to be absorbed, with materials such as silastic (Montgomery Thyroplasty Implant system), Gore-Tex ribbon, hydroxyapatite, or titanium (TVFMI) (Johns et al., 2011; Seino & Allen, 2014). Through an opening in the neck, a small window is removed from the lamina of the thyroid cartilage and

the implant is inserted in the paraglottal space to medialize the vocal fold (Johns et al., 2011; Seino & Allen, 2014). The procedure can be done while the patient is awake and phonating, which helps indicate to the laryngologist when an adequate glottal closure is achieved (Johns et al., 2011). Type I thyroplasty has been found to be an effective intervention for patients with presbyphonia (Netterville, Stone, Luken, Civantos, & Ossoff, 1993; Shah, Deal, & Buckmire, 2013), yielding improvements in perceptual ratings of voice quality and voice-related quality of life (Shah et al., 2013). However, in a retrospective chart review by Shah et al. (Shah et al., 2013), improvements in glottal function index (GFI), a self-reported measure specifically for glottal dysfunction (Bach, Belafsky, Wasylik, Postma, & Koufman, 2005), were not statistically significant for patients with vocal fold atrophy and were in fact smaller than for the other groups of participants (vocal fold paresis and scarring). Similar results were found by Buckmire and colleagues (Buckmire, Bryson, & Patel, 2011). In this retrospective study, significant changes were found for voice-related quality of life, GRBAS composite score, and GFI after an average of almost eight months post-intervention. However, sub-analyses revealed that the subgroup of patients with vocal fold atrophy showed the least improvement when compared to patients with vocal fold paralysis, hypomobility, or scarring (Buckmire et al., 2011). Lu et al. (Lu, Casiano, Lundy, & Xue, 1998) compared the effect of thyroplasty between patients with vocal fold bowing or scarring and patients with unilateral vocal fold paralysis (UVFP). The results reveal better outcomes in the group with UVFP, once again indicating that the evidence for medialization thyroplasty for vocal fold paralysis cannot necessarily be generalized to presbyphonic patients.

Because the implants are not being absorbed by the tissue, there is a risk of inflammation and foreign body reaction (Zeitels, Mauri, & Dailey, 2003). Moreover, just like it is the case with injection augmentations, there is a risk of under-correction which would result in persistence of

the glottal insufficiency, and a risk of over-correction, which would affect the vibration of the vocal folds (Johns et al., 2011) .

Most of the articles on the current surgical options for presbyphonia are retrospective chart reviews, without concurrent controls (level III-3) or without controls (level IV). Moreover, the main inclusion criteria is usually often vocal fold atrophy or glottal insufficiency, which can include other diagnoses such as vocal fold paralysis. Therefore, although the literature reveals positive voice outcomes that justify the use of these approaches in clinics, the overall level of evidence for patients with age-related vocal fold atrophy remains moderate.

Novel Approaches

The current surgical approaches constitute mechanical compensatory strategies to the glottal insufficiency in presbylarynges but they don't address the underlying physiological causes of the gap and therefore only offer a partial solution. This may explain why the improvements in voice outcomes and related quality of life are lesser in presbyphonic patients when compared with other subgroups of patients with glottal insufficiency such as those with vocal fold paralysis (Seino & Allen, 2014). Novel approaches that directly address the cellular changes in the lamina propria are currently being developed in animal models and translated into early phase clinical trials (Ohno & Hirano, 2014). These approaches stem from both regenerative medicine and tissue engineering and are based on the premise that age-related manifestations in the larynx (e.g. bowing, glottal incompetence and reduced mucosal wave amplitude) are mainly caused by the imbalance between synthesis and degradation of the ECM components such as hyaluronic acid (HA) and collagen (Ding & Gray, 2001). As it was described in a previous section, HA synthesis is decreased in aged vocal folds and collagen deposition tends to increase significantly (Ohno & Hirano, 2014). HA is indispensable to the maintenance of optimal viscoelastic properties for vocal

fold vibration. Collagen is important because it provides structural support and malleable strength to the tissue, but a disruption in its synthesis and degradation cycle leads to an excessive accumulation in the lamina propria, which reduces the mucosal wave amplitude and therefore affects the vibration of the vocal folds (Ohno & Hirano, 2014). Researchers gained interest in molecules that could potentially restore the synthesis and degradation process of these ECM components: growth factors. Growth factors are peptides that present with various biological functions involved in cellular growth, proliferation and migration (Ohno & Hirano, 2014).

Basic Fibroblast Growth Factor

Basic Fibroblast Growth Factor (bFGF) has been found to promote the growth of fibroblasts and to stimulate HA production in aged vocal folds during in vitro experiments (S. Hirano, Bless, del Rio, Connor, & Ford, 2004). Studies on rats also revealed increased HA production in aged vocal folds (S. Hirano et al., 2005; Ohno et al., 2009b). These promising results of *in vitro* studies and animal models led to *in vivo* experiments on humans, with the first case study published in 2008 (S. Hirano, Kishimoto, Suehiro, Kanemaru, & Ito, 2008). In 2012, Hirano and his colleagues (S. Hirano et al., 2012) conducted a clinical trial on 10 patients with vocal fold atrophy due to aging. The participants' vocal folds were transorally injected with bFGF dissolved in saline via a curved needle. Seven days after the intervention, results indicated complete glottal closure and improved mucosal wave in the participants. Moreover, improvements in MPT, mean flow rate and acoustic parameters (jitter, shimmer, noise-to-harmonic ratio) were also reported and lasted for at least one year after treatment (S. Hirano et al., 2012). This study by Hirano, which contained no control group, is the only clinical trial of a bFGF intervention to have been conducted on patients with presbyphonia. The current evidence for bFGF intervention for vocal fold atrophy is therefore limited to a level IV. However, more research on bFGF is being conducted and yielding

encouraging results, mostly on vocal fold scarring. One of the shortcomings of bFGF is its rapid absorption in vitro (Hiwatashi et al., 2017) leading to a short working time and in some cases the need for multiple injections (Kobayashi et al., 2017). Most recent research has focused on the development of a drug delivery system that would allow for gradual release of bFGF into the vocal folds. Studies testing collagen-gelatin sponges (Hiwatashi et al., 2017) as well as gelatin hydrogels (Kobayashi et al., 2017) to deliver the growth factor into scarred vocal folds have been conducted on animals (in rat and canine models). Results indicate higher mucosal wave and vibratory amplitudes in the groups using a sustained delivery system, therefore indicating that a solution to the rapid bFGF absorption in vivo is currently coming to fruition (Hiwatashi et al., 2017; Kobayashi et al., 2017). However, studies are still largely at the basic science stage and concentrating on scarred vocal folds. There is no evidence to date that these gradual release strategies would yield significant improvements in patients with vocal atrophy associated with presbyphonia.

Hepatocyte Growth Factor

Hepatocyte growth factor (HGF) is another molecule to have received attention in the voice field, because of its antifibrotic properties (Matsumoto & Nakamura, 2001). As a matter of fact, HGF strongly induces the expression of matrix metalloproteinases (MMPs), which play an important role in the degradation of various ECM components including collagen. It was suggested that the decreased procollagen expression and increased collagen deposition in aged vocal folds are associated with a downregulation in MMPs expression (Ding & Gray, 2001) and that HGF could have the potential of restoring the balance in collagen synthesis and degradation (Ohno et al., 2009a). This hypothesis was tested by Ohno and colleagues (Ohno et al., 2009a), who administered HGF in aged rat vocal folds. They found that the gene expression of MMPs (MMP-2 and MMP-9) was significantly increased in the HGF-treated group when compared to the sham-

treatment group (Ohno et al., 2009a). The expression of hyaluronan synthase-3 was also significantly increased, with a resulting increase in hyaluronan density in the lamina propria (Ohno et al., 2009a). Basic science studies are still ongoing in animal models and human vocal folds *in vitro* to find the optimal dosage, administration route and drug delivery system for HGF interventions (Ohno & Hirano, 2014). Moreover, clinical trials are ongoing for vocal fold scarring and sulcus (Ohno & Hirano, 2014). However, there are no clinical trial available in the literature to support the injection of HGF in patients with presbyphonia.

Section 4: Possible Future Treatment for Presbyphonia - Respiratory Muscle Strength training

Clinical Feasibility

The body of literature on respiratory muscle strength training (RMST) has been growing in the last years. The objective of RMST is to strengthen either the inspiratory muscles or the expiratory muscles, which can be done via different modalities: voluntary isocapnic hyperpnea, flow resistive loading, incentive spirometry, or pressure threshold loading (McConnell & Romer, 2004). The first method, in which the resistance depends on the user's airflow rate (C. M. Sapienza, 2008), is physically demanding and time consuming, as it requires the maintenance of a high level of ventilation (up to 50 breaths per minute) for 30 minutes (McConnell & Romer, 2004), and is therefore not adequate for the elderly population. Flow resistive loading, in which the individual breathes through an orifice of variable diameter, is also dependent on the airflow and therefore this variable needs to be controlled, which adds complexity to the intervention. Incentive spirometers are also influenced by airflow rate, in addition to offering a low resistance (C. M. Sapienza, 2008). Pressure threshold loading consists of using a mouthpiece with a spring-loaded valve (C. M. Sapienza et al., 1999). The valve blocks the airflow until the threshold pressure is reached. Once the threshold is met, the valve allows the airflow to pass through as long as a sufficient pressure is maintained. With a threshold valve, the resistance to inspiration or expiration is independent of the airflow produced by the participants (McConnell & Romer, 2004), which makes it more reliable than the other methods. It is also portable and easy to use, cost effective (McConnell & Romer, 2004), and can be easily performed with a limited amount of

instruction (C. M. Sapienza et al., 1999), which makes it clinically feasible and ideal for the elderly population.

Mechanism of Action

Respiratory muscles, similarly to other striated skeletal muscles, respond to conditioning and can be strengthened with exercises that involve overloading (S. E. Baker, Sapienza, Martin, et al., 2003; C. M. Sapienza et al., 1999). As recommended by the American College of Sports Medicine (cited in Baker et al., 2003) (S. E. Baker, Sapienza, Martin, et al., 2003) the optimal effect on strengthening is obtained with a near maximal load, which is defined by intensity, duration, and frequency of training. When the load is sufficient to activate the neuromuscular system beyond its normal level of activity, the system is forced to adapt and this results in changes in muscle function (Daniilidou, 2007). In order for the benefits to be optimal, the exercise has to induce a functional reorganization in the cortex (increased neuronal excitability and synaptic response), a phenomenon that has been associated with specific skilled-tasks (C. M. Sapienza & Wheeler, 2006). Although RMST does not involve phonation, Sapienza argues that since it requires the integration and coordination of many muscle groups, it is thought to be complex and specific enough to lead to neuroplasticity and threshold changes affecting respiratory function (C. M. Sapienza & Wheeler, 2006). Nonetheless, the carryover of the effects of RMST to specific voicing tasks remains a crucial aspect of the therapy to reach the best possible voice and functional outcomes.

Previous Studies and Impact on Voice

Inspiratory muscle training with a pressure threshold device has been studied mostly in cases of airway obstruction disorders, such as COPD (Lotters, van Tol, Kwakkel, & Gosselink, 2002), asthma (Silva et al., 2013), and upper airway diseases including congenital juvenile laryngeal papilloma (scarring and laryngeal web) (case study)(C. M. Sapienza et al., 1999), exercise-induced vocal fold dysfunction(Mathers-Schmidt & Brilla, 2005; Sandnes et al., 2013), as well as congenital and acquired bilateral abductor vocal fold paralysis (case studies)(S. E. Baker, Sapienza, & Collins, 2003; S. E. Baker, Sapienza, Martin, et al., 2003). The results from the IMST literature are promising, as they confirm that the intervention induces morphologic changes in the respiratory muscles and leads to improvements in clinical measures. In a study by Ramirez-Sarmiento et al. (Ramirez-Sarmiento et al., 2002), a group of patients with COPD underwent five weeks of IMST and a control group received a sham IMST treatment. Following the intervention, biopsies were taken from the external intercostal muscles and the results revealed a 38% increase in the proportion of type I fibers and a 21% increase in the size of type II fibers (Ramirez-Sarmiento et al., 2002). Clinically, these changes were translated into improvements in strength and endurance of the inspiratory muscles (Ramirez-Sarmiento et al., 2002). A study by Souza et al.(Souza et al., 2014) explored the effects of eight weeks of IMST on women aged 60 to 80 years old in a well-designed double-blinded randomized controlled trial. The results revealed significant improvements in MIP and MEP as well as in diaphragm thickness (when contracted) and mobility. These changes were not found in the sham control group. The authors suggest that the improvements detected in respiratory function might have been caused by a combination of muscle hypertrophy (as shown by enhanced diaphragm thickness) and neuromuscular adaptations induced by the training (Souza et al., 2014). The literature regarding changes in pulmonary function following inspiratory training is conflicting, as some studies have reported

improvements in pulmonary function measures (forced vital capacity, forced expiratory volume, vital capacity) and others did not (Mills, Johnson, Barnett, Smith, & Sharpe, 2015; Reyes, Cruickshank, Nosaka, & Ziman, 2015). This could be explained by the different patient populations and differences in training protocol.

Surprisingly, the impact of IMST on voice outcomes has not been studied in patients with respiratory muscle weakness as it is the case with presbyphonia. Studies on elderly individuals and on patients with Parkinson's disease and multiple sclerosis, who also present with suboptimal respiratory function, have either used IMST to improve ventilatory functions (mostly to reduce the sensation of dyspnea)(Kim & Sapienza, 2005) or have focused on **expiratory muscle strength training** (C. M. Sapienza, 2008) to improve nonventilatory functions (coughing, swallowing, and speaking). Eight studies have assessed the effect of EMST on voice outcomes in various patient populations and found improvements in some outcome measures (SPL, Psub, words per minute, utterance length, coefficient of variation of fundamental frequency (F0), VHI-10 scores, pitch range, and /s,z/ durations), while some outcomes remained unchanged (Cerny et al., 1997; Chiara et al., 2007; Darling-White & Huber, 2017; Johansson, 2012; Pereira, 2015; Ray, 2018; Tsai et al., 2016). It is noteworthy that in those studies, the EMST protocol was applied by itself and was not combined with voice exercises, which may have hindered the generalization to voice production. One study combined voice exercises with EMST during a 5-week intervention protocol and found significant improvements in MEP as well as in subglottal pressure (produced at loud intensity), Voice Handicap Index and Voice Rating Scare scores, and dynamic range. The improvements were greater than those obtained with voice therapy alone (Wingate et al., 2007).

The suggested rationale for choosing EMST to improve voice outcomes is that since elastic recoil force of the lungs is reduced, active expiratory muscles should be activated to compress the chest wall and therefore the lungs to a smaller volume and generate the required airway pressure for

the task (Kim & Sapienza, 2005; Silverman et al., 2006). There are two possible pathways to increase respiratory driving pressure: by recruiting the expiratory muscles, as suggested in previous research, or by intensifying the passive recoil pressure by increasing the inspiratory lung volume. During normal voice production, respiratory pressure is mostly generated from the lungs' passive recoil forces, with occasional activation of expiratory muscles in cases of very loud or long duration speech (C. M. Sapienza & Wheeler, 2006). The activation of inspiratory muscles during the active expiration phase results in a better control of the airflow, which allows for longer and louder speech without using the expiratory reserve volume and without activating expiratory muscles. The strategy of initiating phonation at higher lung volumes is already naturally adopted by older individuals, given that the inspiratory pressure is better preserved than the expiratory pressure (Hoit & Hixon, 1987; Huber & Spruill, 2008). It is also thought to be an optimal strategy because at lower lung volumes, speakers have to use more expiratory effort to maintain an adequate subglottal pressure while at the same time physiological expiratory pressures decrease (Zhang, 2016c). This is counterproductive and therefore it is recommended to maintain the lung volume well above the residual volume (Zhang, 2016c).

Patients with presbyphonia need even more strengthening of their inspiratory muscles to compensate for their impaired laryngeal valving that restrains them from using glottal resistance to decrease the rate of lung volume decline and conserve airflow. Moreover, since it has been found that older adults produce shorter utterances when compare to a younger group (Huber, 2008), inspiratory muscle training would allow for a better control of the expiration during speech which would lead to longer utterances and less pauses. Frequent breath pauses reduce the linguistic cues in the message and affect the listener's experience, who tend to perceive the speaker as less competent (Huber, 2008). Despite these arguments, IMST might not be the best approach for all patients with presbyphonia. Some of individuals may have a well-preserved MIP

and may already be initiating phonation at very high lung volume. Because of their reduced ribcage compliance and lung recoil forces, some patients may still have difficulties generating adequate subglottal pressure for phonation and the early activation of expiratory muscles might be an effective strategy for them. Because of a lack of research, it is not known which approach, IMST or EMST, would yield the best possible outcomes for patients with an age-related voice disorder.

Future Studies

Both IMST and EMST represents a promising rehabilitation paradigm for patients with presbyphonia who present with rises in expiratory reserve volume (Gibson, Pride, O'Cain, & Quagliato, 1976) in addition to reduced respiratory muscle strength. Respiratory training is not intended to replace voice therapy, but rather to complement it in order to generate the best possible outcomes. Current standard of care consists of voice therapy, with or without breathing exercises. The hypothesis is that current breathing exercises and voice therapy programs are not specific and intensive (in terms of loading) enough to induce the neuromuscular and hypertrophic changes necessary for improving respiratory outcomes. IMST and EMST have been proven to be effective in engendering those changes in elderly individuals (Kim, Davenport, & Sapienza, 2009; Souza et al., 2014), leading to significant improvements in clinical measures of respiratory pressures. However, this is not enough to support their use with presbyphonic patients, because the impact of these changes on voice outcomes has not yet been established.

In order to demonstrate that RMST should be integrated to current standard of care, a study should be conducted comparing the voice and respiratory outcomes of patients receiving voice therapy and RMST with those of patients receiving voice therapy only. This will allow for the testing of two hypotheses: 1) that respiratory training combined with voice therapy leads to significantly greater respiratory improvements than voice therapy only 2) that the respiratory

improvements induced by RMST result in significantly greater improvements in voice outcomes than those obtained solely with voice therapy.

Section 5: Clinical Manifestations of Presbyphonia and Outcome Measures for Presbyphonia Research

It is important that the outcome measures used to assess changes following treatment be strongly associated with the vocal manifestations of the voice disorder. These manifestations, such as breathiness, decreased loudness and incomplete glottal closure, are by themselves manifestations of the underlying physiological processes of aging including atrophy of the TA and respiratory muscles and degeneration of the lamina propria of the vocal folds. Therefore, to assess if the treatment was specific enough to target the relevant physiological mechanisms of presbyphonia, outcome measures should be chosen carefully. Moreover, while each outcome measure has its own weight/significance, the importance of a multidimensional assessment of the treatment outcomes should be taken into consideration to draw a complete representation of the voice and its changes following treatment and to allow for future comparison with other relevant studies.

Laryngeal Features

Laryngeal endoscopy with stroboscopy (LES) is used in voice clinics to diagnose patients, plan therapeutic goals and assess changes following treatment and has been described as the most valuable method for evaluating laryngeal function in presbyphonic patients(Bloch & Behrman, 2001; Vaca, Cobeta, Mora, & Reyes, 2017). In research, LES is used to choose participants based on specific inclusion criteria, such as vocal fold bowing, presence of a glottal gap and/or vocal fold processes prominence in the case of presbyphonia. Visual examination of the larynx would allow for direct observation of laryngeal changes following treatment, however LES is rarely used as an

outcome measure in the studies of voice therapy for presbyphonia. While many authors hypothesized that their intervention yielded changes in vocal fold vibration and vocal fold adduction, in many cases these assumptions were not verified. Only few studies, mostly case series, have confirmed an improved glottal closure after behavioral voice therapy (Lu et al., 2013; Tanner et al., 2010).

Although glottal insufficiency is thought to be one of the main pathogenic features of the presbylarynx (Pontes, Yamasaki, & Behlau, 2006; Vaca et al., 2017), additional age-related changes in laryngeal function and morphology are also at the basis of changes in elderly voices and should therefore be considered as relevant/clinically meaningful outcome measures. An observational study by Pontes et al. (Pontes et al., 2006) revealed that elderly larynges differ significantly from young ones on different parameters including: **vocal fold edges** (bowing), **glottal gap configuration**, **supraglottic hyperfunction**, and **phase and amplitude symmetry of vocal fold vibration.** Although the difference was not always significant, aged vocal folds have also been found to exhibit a decreased **regularity in vocal fold vibration**, an increased **vibration amplitude**, a decreased **mucosal wave symmetry (Pontes et al., 2006)** and a dominant **open phase (Murty, Carding, & Kelly, 1991)**. While correlated, these parameters provide complementary information on the laryngeal function and should therefore all be taken into consideration when assessing the impact of an intervention on the presbylarynx.

Bowing of the vocal folds, thought to be caused by atrophy of the TA muscle and changes in the ECM components of the lamina propria, has been related to glottal insufficiency (Bloch & Behrman, 2001; Pontes et al., 2006) and provides a high sensitivity for the diagnosis of an incomplete glottal closure. However, some patients will exhibit vocal fold bowing during rest breathing but will achieve a complete glottal closure during phonation by medially compressing the vocal folds with the ventricular folds (Vaca et al., 2017). A particular attention should be given

to the configuration/location of the glottal gap as a stroboscopic parameter. In fact, no significant difference is found in glottal gap occurrence between young old subjects whereas a significant difference has been found regarding the shape of the glottal gap, with a spindle shaped gap more frequently associated with aged larynges and a triangular posterior chink more frequently found in young larynges (particularly in women)(Linville, 1996; Pontes et al., 2006). Vocal fold bowing and glottal gap are also related to the parameter of phase closure which has been found to be predominantly open in patients with presbyphonia, but which could also be influenced by the degree of supraglottic hyperfunction.

A literature review on the use of LES as an outcome measure in voice treatment studies has examined the sensitivity to change of different laryngeal parameters. The results revealed that some of the abovementioned parameters present a good sensitivity to change: phase symmetry showed a rate of improvement of 85% and location/shape of glottal gap had a rate of improvement of 80% (Bonilha H, 2017). This signifies that these parameters were sensitive enough to detect a statistically significant change following the administered treatment. The features of vocal fold edges and of supraglottic hyperfunction were not assessed often enough regarding their statistical significance in order to provide data on their sensitivity to change. However, other parameters that demonstrated a good sensitivity to change and that are relevant for presbyphonia research include **regularity of vibration** (88%), **amplitude of vibration** (83%), **mucosal wave quality** (83%), **mucosal wave amplitude** (75%), and **magnitude of glottal gap** (72%) (Bonilha H, 2017).

No study to date has established the magnitude of change in these parameters that would correlate with a functionally meaningful improvement for presbyphonic patients. Such thresholds are challenging to establish because they might differ from one patient to another but also because few research studies to date have employed standardized rating forms for stroboscopic

assessment of laryngeal anatomy and physiology (Bonilha H, 2017). The lack of a widely used standardized form for stroboscopy ratings of laryngeal parameters, combined with the use of ambiguous terms to name these parameters and the use of various scales to measure them, has led to intra- and inter-rater reliability issues in LES assessment. Parameters that have been identified to be the less reliable are those involving a movement pattern (functional/temporally-dependent), such as amplitude of vibration, mucosal wave, phase symmetry, and regularity of vibration(Bonilha, Focht, & Martin-Harris, 2015; Poburka & Bless, 1998). The implementation of a standardized rating form such as the Vibratory Assessment With Laryngeal Imaging (VALI)(Poburka, Patel, & Bless, 2016), with definitions, graphics, and rating instructions for each parameter as well as video samples for training has been shown to improve the reliability of these parameters and should therefore be used for voice research.

Aerodynamic Measures and Phonation Duration Measures

Aerodynamic measures and phonation duration measures are the result of laryngeal function, pulmonary function and respiratory strength (Vaca et al., 2017). They can also be influenced by the resonance system because of the acoustic pressures created by the source-filter interaction. Because they encompass several components of voice production, it is challenging to make assumption on physiological mechanisms underlying aerodynamic measures, unless other outcome measures are being used to individually assess laryngeal function and respiratory function. However, aerodynamic measures have the advantage of providing information on the combined activity of these systems and therefore offer functional and clinically meaningful data. This is particularly pertinent in the case of presbyphonia because of its known alterations in both phonatory and respiratory mechanisms. Aerodynamic and phonation duration measures that are relevant for presbyphonia research and that have been shown to demonstrate good outcome

measures properties include **laryngeal resistance (RLaw)**, **mean airflow**, estimated **subglottal pressure** (**Psub)**, and **maximum phonation time (MPT).**

Mean airflow rate has been shown to increase in older subjects, which can be explained by an incomplete glottal closure (Higgins & Saxman, 1991). Higgins found the subglottal pressure to be increased in older men when compared to younger men and explains this result by the increased stiffness of the vocal folds, which may require a higher Psub (Higgins & Saxman, 1991). These findings are different from those of Baker et al.(K. K. Baker et al., 2001) who found similar absolute and normalized Psub values for older and younger subjects across three loudness conditions (soft, comfortable and loud). The variability in age-related changes and in compensation mechanisms could partly account for these discrepancies: some patients compensate for their glottal insufficiency by increasing vocal fold adduction while others may increase respiratory effort (K. K. Baker et al., 2001). Moreover, variations in pulmonary and respiratory function can influence the available subglottal pressure. Laryngeal resistance is computed by dividing Psub by mean airflow and provides information of the valving function of the vocal folds. RLaw has been shown to be reduced in older individuals and this is thought to be correlated with their reduced adductory function (Melcon, Hoit, & Hixon, 1989). MPT is influenced by both the laryngeal and respiratory function, and has consistently been found to be reduced with aging (J. Gartner-Schmidt & Rosen, 2011). MPT is rapid and easy to administer and is often used as an outcome measure in voice treatment studies (Desjardins et al., 2017).

Authors have examined the expected variations in aerodynamic measures in healthy subjects and have found that some parameters present with a good intra-subject reliability when assessed on different days or at different moments within one day. **Laryngeal resistance**, **mean airflow**, estimated **Psub**, and **maximum phonation time (MPT)** showed no significant difference when assessed on the same speaker at different time points, demonstrating a good reliability (Garrison,

2009; L. Lee, J. C. Stemple, & M. Kizer, 1999; Linda Lee, Joseph C Stemple, & Marla Kizer, 1999; Leeper & Graves, 1984). It was also found that the standard deviations were smaller and the reliability was better when the parameters of intensity and fundamental were controlled (Garrison, 2009; Linda Lee et al., 1999; Leeper & Graves, 1984). Awan et al. (Awan, Novaleski, & Yingling, 2013) calculated intraclass correlation coefficients (ICCs) for various aerodynamic measures to gather information on relative reliability, which is the extent to which subjects maintain their respective ranks in the sample (Weir, 2005) . The study revealed strong ICCs for MPT, Psub, and laryngeal resistance. For these three outcome measures, ICCs were greater than 0.75, which has been associated with excellent reproducibility(Marx, Menezes, Horovitz, Jones, & Warren, 2003). The ICC for mean airflow was strong for males but weaker in females (ICC=0.67), which remains acceptable for test-retest reliability (Awan et al., 2013). In this study, the intensity and fundamental frequency were not controlled. The authors state that, even though stronger test-retest reliability values may be obtained when controlling for these variables, it may negatively impact the validity of the measurements by masking the true phonatory behaviors (Awan et al., 2013; Higgins, Netsell, & Schulte, 1994). For this reason, they suggest that the measurements be taken while allowing the subjects to use comfortable loudness and pitch.

A systematic review on voice therapy effectiveness studies revealed that this category of outcome measures (including MPT, mean airflow, laryngeal resistance and Psub) were used in a third of the 15 reviewed studies and that they improved significantly in half of those (Desjardins et al., 2017). The authors mention that a lack of improvement in these measures should not necessarily be interpreted as an absence of treatment effect. Because these measures are a result of many physiologic components (including respiratory muscle activity, integrity of vocal fold tissue, glottal competence and phonatory style (Desjardins et al., 2017; Zraick, Smith-Olinde, & Shotts, 2012)), it is possible that an improvement in one of these components was present but not sufficient to

impact the resulting aerodynamic measure. This reveals the lower sensitivity to change of aerodynamic measures, but it also highlights their clinical meaningfulness as functional measures since they integrate many aspects of voice production. It is hard to determine if the 50% rate of significant improvement found in the systematic review (Desjardins et al., 2017) was also caused by a high inter-subject variability. In discussions on this topic, Higgins and his colleagues (Higgins et al., 1994) suggest that their high inter-subject variability does not discredit aerodynamic measures as outcome measures for voice research. They calculated coefficients of variation in 21 subjects and concluded that a change of more than 15% in subglottal pressure (Psub) and of more than 25% in mean phonatory airflow should represent a true physiological change, not caused by intra-subject variability. The use of cut-off values such as those suggested by Higgins et al. should greatly help with the interpretation of aerodynamic measures. Whether or not this change correlates with clinically meaningful voice remains to be determined.

Laryngeal resistance, Psub and mean airflow can be measured with the voice efficiency protocol of the KayPENTAX PAS 6600. Although it is possible that a certain amount of measurement error be attributable to the system or to the user's oversight, the PAS system has been shown to have a strong calibration accuracy and a good validity of the built-in flowhead and pressure transducer (Garrison, 2009; Zraick et al., 2012).

Auditory-Perceptual Judgement of Voice Quality

Auditory-perceptual judgements of voice quality are a common part of voice assessment both in the clinical and research fields. They are the auditory perceptual consequence of the underlying voice physiology and therefore are affected by the age-related changes in the respiratory, laryngeal and neurological systems. A study on 72 elderly subjects revealed that the most common vocal alterations observed in this population are: general dysphonia grade, followed by

roughness, then breathiness(Pessin, Tavares, Gramuglia, de Carvalho, & Martins, 2017). The same study revealed that vocal alterations of different severities were present in 90% of subjects aged 60-75 years old and in 96% of subjects over 75 years old. Other auditory perceptual traits have been reported in the literature as being characteristic of presbyphonia, including altered pitch (lower in women and higher in men), decreased pitch range, increased strain and unsteadiness/instability, occasional vocal tremor and decreased loudness(Lu et al., 2013; Tay et al., 2012).

Two main standardized assessment forms are widely used in the voice field to rate perceptual judgments of voice quality: the Grade Roughness Breathiness Asthenia Strain (GRBAS, or GRBASI-including the parameter of instability) and the **Consensus Auditory Perceptual Evaluation of Voice (CAPE-V-including overall severity, breathiness, roughness, strain, pitch and loudness)(Nemr et al., 2012)**. The CAPE-V seems like the most appropriate outcome measure instrument for presbyphonia research because it includes the parameters of pitch and loudness, which are typically affected in this patient population. The CAPE-V instrument also includes a classification of the resonance parameter and allows for the description additional features (Nemr et al., 2012). As study comparing the reliability of the GRBAS and the CAPE-V showed that both instruments present with strong ICCs for intra-judge reliability, ranging from 0.923 to 0.985 (Nemr et al., 2012). Strong correlations were also found for inter-judge reliability for both instruments (Nemr et al., 2012). The CAPE-V and GRBAS have also been found to be highly correlated regarding the general severity grade parameters, thus supporting their concurrent validity (Nemr et al., 2012). The CAPE-V's concurrent validity was also confirmed in a study that compared raters' judgments of overall severity (using the CAPE-V scale) to a consensus of severity judgments that had been a priori established on 59 voices (22 normal and 37 disordered) (Zraick et al., 2011).

The CAPE-V has been found to be sensitive for detecting even subtle changes in voice quality, more so than the GRBAS (Nemr et al., 2012; Wuyts, De Bodt, & Van de Heyning, 1999). This could be explained by the different types of scales that are used in the two instruments. While the GRBAS employs an ordinal scale that offers only three choices to the rater (mild, moderate, severe), the CAPE-V uses a visual analog scale that allows for more nuances and might be helpful for rating small changes in voice(Nemr et al., 2012). In a literature review on voice therapy effectiveness studies, all studies that used the overall severity parameter from the CAPE-V, except one, found a statistically significant improvement following behavioral voice therapy. This confirms the strong sensitivity to change of the CAPE-V and its relevance as an outcome measure, particularly in presbyphonia research since changes following treatment have been shown to be minimal and/or harder to detect for this patient population (Gillespie, Dastolfo, Magid, & Gartner-Schmidt, 2014).

Acoustic Measures

Although perceptual judgments of voice quality present with good reliability profiles, they remain subjective measures and should be complemented by objective auditory measures in the context of multidimensional voice assessments. Quantifying the severity and characteristics of a disordered voice offers a precise way to assess treatment effectiveness and patient improvement following an intervention. Acoustic assessments are therefore recommended by ASHA in their most recent guidelines for instrumental assessment of voice (Shaheen Awan, 2005), as they are the correlates of the perceptual parameters of pitch, loudness and voice quality. The recommended parameters include:

1) Measures of sound level in decibels (dB):

a. **Habitual sound pressure level** (SPL): mean SPL during connected speech, which correlated with vocal loudness.

b. **Maximal and minimal SPL**: softest and loudest sustained sounds that can be produced by the subject.

2) Measures of vocal frequency in Hertz (Hz):

a. **Mean fundamental frequency** (F0): the number of vibratory cycles per second(Pessin et al., 2017); correlates with the vocal pitch.

b. **F0 standard deviation**: provides information on the stability of the F0.

c. **Minimal and maximal F0**: these values can be used to compute the F0 range, expressed in semitones.

3) Measures of vocal quality:

a. **Jitter** (%): cycle-to-cycle perturbation of the period.

b. **Pitch perturbation quotient (PPQ)**

c. **Shimmer** (%): cycle-to-cycle perturbation of the period amplitude.

d. **Amplitude perturbation quotient (APQ)**

e. **NHR** (dB): measures the degree of noise (aperiodicity) in the voice signal.

f. **Voice turbulence index (VTI)**

g. **Soft phonation index (SPI)**

h. **Smoothed cepstral peak prominence** (CPP) (dB): measure of voice quality derived from the cepstrum. It is the ratio of the observed amplitude of the cepstral peak to the expected amplitude based on linear regression (Peterson et al., 2013). A periodic signal will display a more prominent cepstral peak when compared to an aperiodic signal because the fundamental frequency and the harmonics are well defines(Peterson et al., 2013).

With computer software for cepstral measure analyses becoming more accessible, ASHA advocates for the inclusion of a cepstral measure such as CPPS in the acoustic analysis of voice quality. They justify this recommendation with the growing body of evidence that claims that cepstral measures (frequency-based) are more appropriate than traditional acoustic measures (time-based) for analyzing connected speech, which makes them more ecologically valid (Maryn, Roy, De Bodt, Van Cauwenberge, & Corthals, 2009). In addition, time-based measures rely on the periodic signal and thus their accuracy for severely dysphonic voices is not optimal (Gillespie et al., 2014). On the other hand, frequency-based measures examine the dominance of the F0 over superfluous noise in the signal and do not depend on the periodicity of the waveform (Gillespie et al., 2014). CPP has been shown to be reliable in men and moderately reliable in women when tested on vocally healthy volunteers (Leong et al., 2013).

Despite these considerations, traditionally-used time-based measures such as F0, jitter, shimmer and HNR provide clinicians and researchers with valuable information regarding the impact of a treatment on specific acoustic characteristics. Moreover, these measures are sensitive enough to differentiate voices of young and elderly subjects and therefore to depict the acoustic consequences of the age-related physiological changes occurring in the larynx. Acoustic analyses in older subjects have consistently revealed a lower F0 in women and a higher F0 in men (Pessin et al., 2017). While the higher F0 in men can be explained by a reduced mass of the vocal folds caused by the muscle atrophy, the lower F0 found in women is thought to be related to post-menopausal hormonal changes (Awan, 2006). Other acoustic characteristics have been observed in elderly subjects, including increased in amplitude variations (shimmer) and reduced HNR (Deliyski, 2001; Ferrand, 2002; Gorham-Rowan & Laures-Gore, 2006). Observations regarding

frequency variations values (such as jitter) for presbyphonic voices are conflicting in the current literature (Tay et al., 2012), and this might be related to the data acquisition procedures.

As a matter of fact, attention should be given to the context in which acoustic measures are being recorded in order to ensure their validity and reliability and therefore their ability to detect changes following a treatment. Acoustic measures have indeed been criticized by some authors on the basis of a lack of reliability and sensitivity to change (Carding, Wilson, MacKenzie, & Deary, 2009; Leong et al., 2013). However, in the abovementioned literature review on voice therapy effectiveness (Desjardins et al., 2017), acoustic measurements showed significant improvements between 85% and 90% of the time and yet most of the parameters used were time-based (cepstral measures were included in only one of the 15 reviewed articles). In their guidelines, ASHA has provided detailed information on how to conduct acoustic recordings and analyses in order to limit measurement error. The strict following of these guidelines should greatly improve the reliability of acoustic measures and consequently enhance their value as relevant outcome measures for voice treatment studies.

Self-Assessment Measures

Self-assessment measures represent a unique and indispensable category of outcome measures because they assess the functional impact of the voice disorder on the patient's life and their perception and symptoms related to the problem. Self-reported measures don't necessarily correlate with other outcome measures and this is not because of a lack of validity: it is because they assess a different construct, namely the concept of handicap. The impact of a voice disorder on the patient is of course influenced by the voice disorder itself (and therefore the stroboscopic, aerodynamic, auditory perceptual and acoustic findings), but other factors come into play. These factors include the level of activity and participation of the patient and the degree to which their

vocal demands can be met despite the voice disorder. This type of information is particularly relevant for patients with presbyphonia who might present with diverse profiles of activity and participation. This could partly explain why, even though age-related changes in the subsystems related to voice production are universal, only a certain proportion of elderly people are considered to have a voice disorder (although this number is also affected by the number of people who seek treatment and by the inter-subject variability in the physiological aging process). The **Voice Handicap Index** (VHI) is one of the most broadly used self-reported questionnaire in the voice field. It evaluates the degree of voice-related handicap experienced by the patients in their day to day life. The questionnaire contains 30 items, which are divided into three categories: functional, emotional and physical (Jacobson et al., 1997). For each item, the degree of perceived handicap is rated on a 5-point scale, from 0 (no problem) to 4 (always a problem), for a total of 120 points(Jacobson et al., 1997). The test has a strong test-retest reliability for the three subcategories as well as for the total score, and a strong internal consistency (Jacobson et al., 1997). Moreover, the correlation between the patients' perception of their voice disorder severity and the scores on the VHI were moderately strong in the validation study, indicating an adequate construct validity (Jacobson et al., 1997). The original version is lengthy and this could represent an issue for the older population. For this reason, the use of the validated abridged version, the **VHI-10**, seems more appropriate for presbyphonia research. This version contains 10 questions which were specifically chosen based on their clinical relevance, through expert consensus but also through item analysis by comparing a dysphonic and a control group as well as a pre- and post-treatment conditions (Rosen, Lee, Osborne, Zullo, & Murry, 2004). The concurrent validity of the VHI-10 was confirmed by comparing its results with those obtained with the original VHI, which yielded a correlation exceeding 0.90. The results suggest that the VHI-10 is valid and might even represent a more robust and sensitive instrument than the VHI to assess voice-related

handicap in patients with voice disorders in addition to being faster to administer (Rosen et al., 2004). In order to facilitate the interpretation of the results in terms of clinical meaningfulness, Arrfa et al.(Arffa, Krishna, Gartner-Schmidt, & Rosen, 2012) calculated normative values for the VHI-10 and came to the conclusion that a score above 11 is indicative of a voice-related handicap. Although the VHI is useful in assessing the level of handicap caused by the voice disorder and the level of improvement following treatment, it is not specific to the manifestations of presbyphonia. The **Glottal Function Index** (GFI) (Bach et al., 2005) was conceived specifically to assess the symptoms related to glottal insufficiency and therefore applies directly to presbyphonic patients, considering that the presence of a glottal gap is one of the main hallmarks of this voice disorder. The advantage of a disease specific impairment instrument such as the GFI is that it provides information on the level of handicap and symptoms directly associated with the condition and evaluates how the treatment was successful in targeting its underlying mechanisms (Bach et al., 2005). Because it is targeted, the GFI is very short and can be easily administered at each therapy session, allowing for interim data collection. The instrument comprises four items, each rated on an ordinal scale from 0 (no problem) to 5 (severe problem), for a total of 20. The GFI presents with an excellent test-retest reliability and with good criterion and construct validity (Bach et al., 2005). Its criterion validity and sensitivity to change were confirmed based on its consistent and significant decrease in scores following established surgical treatment for glottal insufficiency. Its construct validity was confirmed by a good correlation with the VHI results. A GFI score higher than 4 (mean + 2SD) was determined as the cut-off value to differentiate normal and abnormal voices (Bach et al., 2005).

Perceived Phonatory Effort (PPE) is another easy-to-administer outcome measure relevant for patients with presbyphonia. It was used in a clinical trial for presbyphonic patients by Ziegler et al. (Ziegler, 2014), in which participants were asked to rate their perceived effort during phonation

on a scale where 100 represents comfortable effort, 200 represents twice as much as comfortable, 50 represents half the comfortable effort, etc. (Ziegler, 2014) This outcome measure detected a significant reduction in perceived effort in one of the intervention groups.

Laryngopharyngeal reflux (LPR) is a widespread chronic condition that can have a significant impact on the larynx and the voice, including edema of the vocal folds and perceived hoarseness and vocal fatigue (Belafsky, Postma, & Koufman, 2001, 2002). Most voice studies on treatment effectiveness don't incorporate the presence of LPR in their exclusion criteria because of its scope (it could be present in up to 50% of voice patients (Koufman, Amin, & Panetti, 2000)) and also because many people are affected by reflux without being aware of it. Including subjects with LPR in voice studies enhances the generalizability of the results to a wider patient population. However, it remains important to document the presence and symptoms of reflux as it may impact the response to voice therapy. The **Reflux Symptom Index** (RSI) (Belafsky et al., 2002) is a self-reported measure that is commonly administered as part of voice evaluations to assess the symptoms related to LPR. It contains nine items and the authors state that it can be administered in less than one minute (Belafsky et al., 2002). Each item is score on an ordinal scale with values from 0 (no problem) to 5 (severe problem), for a total of 45 points. Similarly to the GFI, the RSI has been compared to the VHI to confirm its construct validity, which has been shown to be good. Its criterion-based validity is also good since the RSI can correctly detect improvements of LPR symptoms following treatment(Belafsky et al., 2002). The instrument has also be proven to be reliable(Belafsky et al., 2002). Lastly, the cut-off value to differentiate a LPR related problematic is a RSI score higher than 13. The authors mention that a certain amount of reflux can occur in most people and thus this cut-off value helps interpret the results in a clinically meaningful way (Belafsky et al., 2002).

Lastly, the **Communicative Participation Item Bank** (CPIB) (Baylor et al., 2013) is a 10-item instrument designed to assess the construct of communicative participation across various communication disorders and situations. The items are score from 0 (the condition interferes very much with the situation) to 3 (the condition does not interfere at all). The total score ranges from 0 to 30, with a higher score being more favorable.

Pulmonary Function and Respiratory Muscle Strength Measures

In the context of presbyphonia research, it is clinically relevant to determine if patients who present with voice complaints also have reduced pulmonary function and/or decreased respiratory muscle strength when compared to norms for healthy elderly individuals. This data could help understand why some elderly subjects present with poorer voice outcomes than others and could partly explain why some patients who undergo voice therapy don't improve as much as others. It could also provide data on the effects of respiratory and voice therapies on these measures and the resulting impact on voice outcomes.

The two main pulmonary function parameters obtained with spirometry are **forced vital capacity (FVC)** and **force expiratory volume in one second (FEV1)**. FVC a measure of the lungs' vital capacity performed during a complete and forced exhalation from full inspiration to the end of the expiratory reserve volume (Miller et al., 2005). FEV1 is the volume forcefully exhaled during the first second of the FVC task (Miller et al., 2005). Both tasks are measured in liters at body temperature and ambient pressure saturated with water vapour (BTPS) and can be altered in both obstructive and non-obstructive lung disorders (Pellegrino et al., 2005) as well as in normal aging. As a matter of fact, both FVC and FEV1 have been shown to decrease with age in a linear fashion (Loth et al., 2013). When tested on sample of 3528 subjects, those spirometry measure have been found to have an excellent reliability (Detels, Coulson, Tashkin, & Rokaw, 1975). Moreover, FEV1

is thought to be the most repeatable measure of pulmonary function (Pellegrino et al., 2005). The sensitivity to change of these measures following IMST is unclear from the data in the literature and this could partly be explained by the differences in patient populations. Moreover, the lack of change following a treatment with patients who have a rapidly progressive disease should be interpreted carefully, as it does not imply that the treatment was ineffective or that the measure has a poor sensitivity to change, but rather that the intervention may be maintaining the patient at a certain level and avoiding decline (Pellegrino et al., 2005).

Measures of vital capacity do not only reflect the integrity of the lungs since they are also influenced by the strength of the respiratory muscles. In fact, a low vital capacity or an increased residual volume can either be a sign of disease of the lungs or they can be indicative of respiratory muscle weakness(Evans & Whitelaw, 2009). Weak inspiratory muscles won't allow to expand the lungs to their full capacity, whereas weak expiratory muscles won't allow to compress the lungs completely to a normal residual volume(Evans & Whitelaw, 2009). Many conditions, such as aging, comprise changes in both lung structure and respiratory muscle strength. Complementing spirometry measures with outcome measures that provide direct information on the inspiratory and expiratory muscle strength will offer a more precise picture of the respiratory system status before and after the intervention.

To directly measure the number and the size of the respiratory muscle fibers before and after an intervention would be invasive and wouldn't provide clinically meaningful information. This is why indexes such as **Maximum Inspiratory Pressure (MIP)** and **Maximum Expiratory Pressure (MEP)** are useful, because they indirectly measure respiratory muscle strength (Kim & Sapienza, 2005; C. M. Sapienza et al., 1999) by appraising the pressures at the mouth (Romer & McConnell, 2004). This can be done by using a hand-held pressure meter or a manometer. Values obtained with a hand-held device have been shown to be reliable and valid when compared to those obtained

with a manometer (Hamnegard et al., 1994; Romer & McConnell, 2004). A reliability study by Romer et al.(Romer & McConnell, 2004) revealed lowest intra-subject variability than previously published findings, with excellent agreement ratios for MIP, MEP and FVC. The authors explain these results by the fact that the subjects in their experiment were thoroughly familiarised with the procedure a priori and that the measurements were all taken by the same investigator, therefore reducing secondary variance. Reducing bias is also likely to increase sensitivity to change of these measures, which is already strong. In fact, many studies testing inspiratory and expiratory muscle strength training have included MIP and MEP in their outcome measures and have obtained statistically significant differences (Souza et al., 2014).

In summary, presbyphonia is a complex and multi-variable phenomenon. Its features and their underlying physiological basis need to be thoroughly understood to develop the best possible intervention and to adequately measure the efficacy of the newly developed treatment paradigm.

Chapter 3: Methodology

Research Questions

1. Which outcome measures present with the strongest intra-subject reliability in a sample of patients with presbyphonia?

2. What is the respiratory function of patients with presbyphonia and how is it correlated with voice measures in this population?

3. a) What is the effect of RMST on respiratory and voice outcomes in patients with presbyphonia and b) how do baseline measures of respiratory function influence the response to the intervention?

Specific Aims

Specific Aim 1: To determine which outcomes present with the strongest intra-subject reliability in a sample of patients with presbyphonia. Hypothesis: Respiratory muscle strength measure (MIP and MEP), cepstral peak prominence smooth (CPPS), mean F0, mean SPL, and self-reported measures will be the most reliable outcomes.

Specific Aim 2: To assess respiratory function and its correlation with voice outcomes in patients with presbyphonia. Hypothesis 1: Patients with presbyphonia will present with a reduced respiratory function when compared to predicted values.

Hypothesis 2: Participants with a lower respiratory function will present with poorer voice outcomes than participants with a higher respiratory function when controlling for baseline atrophy.

Specific Aim 3: To acquire pilot data on the effects of RMST on respiratory and voice outcomes in patients with presbyphonia.

Aim 3a: To determine what effect sizes are to be expected within and between participants following a RMST intervention. Hypothesis 1: Small effect sizes will be found within participants for respiratory muscle strength measures (MIP and MEP), smoothed cepstral peak prominence (CPPS), self-reported outcomes, and aerodynamic measures. Hypothesis 2: The RMST groups (IMST and/or EMST) will demonstrate greater improvement in respiratory outcomes, and in voice-related outcomes than the VFE-only group.

Aim 3b [exploratory]: To determine how baseline measures of respiratory function and atrophy influence the effects of RMST. Hypothesis 1: IMST will have maximal effect on patients with decreased maximum inspiratory pressure (MIP). Hypothesis 2: EMST will have maximal effect on patients with decreased maximum expiratory pressure (MEP). Hypothesis 3: Participants with more severe atrophy will be less responsive to non-surgical approaches, RMST and VFE, than participants with milder atrophy.

This is an initial feasibility and mechanistic study that has the potential to reveal that RMST (IMST/EMST) can significantly improve outcomes from voice therapy by acting in synergy with VFE. Currently, much time is spent in voice therapy on subthreshold 'breathing exercises' that lack evidence. This project will provide guidance for evidence-based and theory driven approaches to leverage respiratory function to improve voice therapy. Conversely, negative findings from this

study would indicate that respiratory function is ineffective and not a viable target for therapy in this patient population.

Overview

After being enrolled in the study (inclusion and exclusion criteria are described below), treatment-seeking patients with a diagnosis of presbyphonia underwent baseline assessments including: voice (videostroboscopy, acoustics, aerodynamics, self-assessments) and respiratory (pulmonary function and respiratory muscle strength) measures. Participants were then blocked-randomized (based on their respiratory muscle strength) to one of three groups using a 3-parallel arm design: IMST+VFE (I+VFE), EMST+VFE (E+VFE) and VFE (interventions described below). Participants attended four 1-hour therapy sessions with a SLP and a research team member. Each session consisted of 15 minutes of pre-session measures; followed by 10 minutes of IMST, EMST or VFE for groups I+VFE, E+VFE and VFE, respectively; then 20 minutes of voice exercises for all three groups and; lastly 15 minutes of patient education and post-session measures. Participants were asked to practice their respective exercises for 10 minutes, twice daily, seven days a week during the duration of the study. A final assessment session, after the 4^{th} therapy session, mirrored the baseline measures. Outcomes that rely on subjective judgment (visual ratings of videostroboscopy and auditory-perceptual judgments of voice quality) were rated by trained judges (SLPs) blinded to the group assignment. [**Note**: Long-term follow-up was not conducted. This study's focus is on the mechanistic effects of the intervention; future studies, given positive outcomes from this study, will explore the effectiveness and long-term effects.]

Study Design

Aims 1 and 2

This is a prospective cross-sectional study (Aim 2), with a repeated measures baseline assessment (Aim 1). Participants were assessed on various voice and respiratory measures, with the objective of assessing potential relationships between respiratory function and voice. Patients with various levels of respiratory function were included in the study. A subset of participants underwent two baseline assessment sessions within two months of the first session, in order to assess intra-subject reliability of the measures. Specific details regarding eligibility criteria, assessment tasks and outcomes measures are provided in subsequent sections.

Aim 3

This is a prospective stratified randomized 3-arm parallel pilot study. Participants were divided into four strata based on baseline respiratory muscle strength measures: 1) within normal limits MIP and MEP; 2) preserved MIP and decreased MEP; 3) preserved MEP and decreased MIP; 4) decreased MIP and MEP. Participants were block-randomized into three groups (allocation ratio of 1:1:1): IMST and VFE (I+VFE); EMST and VFE (E+VFE); voice therapy only (VFE).
Stratification was used to ensure a balanced representation of different respiratory profiles between intervention groups.

Participants

The study was approved by the Institutional Review Board of the Medical University of South Carolina. Participants were recruited from the treatment-seeking population at the Evelyn Trammell Institute for Voice and Swallowing (division of the Department of Otolaryngology-Head & Neck Surgery) - Medical University of South Carolina (MUSC). As part of usual care, patients

underwent a videostroboscopic assessment conducted by the laryngologist. During the 10-month recruitment period, all consecutive patients who were diagnosed with presbyphonia underwent a brief screening interview by the laryngologist, after which qualifying patients were offered participation in the study. A member of the study team met with interested participants and obtained formal written consent. Table 1 presents the inclusion and exclusion criteria for Aims 1 and 2 of the study. Table 2 presents *additional* exclusion criteria that were applied for Aim 3.

Table 3. Inclusion and Exclusion Criteria for Aims 1 and 2

Inclusion Criteria	**Exclusion Criteria**
Is 50 years or older	Presents with a vocal fold pathology other than presbyphonia
Has a diagnosis of presbyphonia given by the laryngologist and consistent with the hallmarks of the disorder (vocal fold bowing, glottal gap during phonation, and/or prominence of the vocal processes indicative of vocal fold atrophy)	Has a known neurologic or progressive neuromuscular disease
	Has dysarthria
	Has any other condition judged by the physician as being unsuitable for spirometry or muscle pressure testing
	Is unable to give informed consent

Table 4. Additional Exclusion Criteria for Aim 3

Exclusion Criteria
Has received voice therapy in the past year
Has a medical condition that could be aggravated by the experimental intervention, or any condition judged by the physician as being unsuitable for RMST.
Has a language disorder
Has a hearing loss that is not adequately managed
Has a cognitive disorder that might affect treatment compliance

Respiratory Assessments and Outcome Measures

Pulmonary Function

A computer-based spirometer was used (Koko®Sx 1000, NSPIRE HEALTH), and pulmonary function testing was conducted following the American Thoracic Society and European Respiratory Society recommendations(Miller et al., 2005). Prior to the task, participants were instructed to put on a nose clip and to made sure that their lips were sealed around the mouthpiece connected to the pneumotach. They were then asked to breathe normally in the spirometer for about four breaths before taking a full inspiration, blowing out with force during at least 6 seconds, and taking another full inspiration. Those maneuvers were repeated at least three times, until the two best measurements of FVC and FEV1 differed by no less than 0.15L, to ensure acceptable repeatability (Miller et al., 2005). The study team member responsible for conducting spirometry testing received training in the MUSC pulmonary department.

In addition to FVC and FEV1, the FEV1/FVC ratio was also collected from spirometry testing. Both raw and predicted normal values were reported. Predicted normal values were generated by the spirometer software based on the participants' age, sex, and height.

Maximum Respiratory Strength

1) Maximum Inspiratory Pressure (MIP): participants were instructed to exhale slowly and completely, seal lips firmly around the mouthpiece, and then breathe in forcefully in the respiratory pressure meter (MicroRPM, MD Spiro/Micro Direct). The maneuver was repeated approximately four times, with the goal of matching the two highest trials within 10 cmH_2O)(Sachs, Enright, Hinckley Stukovsky, Jiang, & Barr, 2009).

2) Maximum Expiratory Pressure (MEP): participants were instructed to take a maximal inspiration, seal lips firmly around the mouthpiece, and then breathe out forcefully in the

respiratory pressure meter (MicroRPM, MD Spiro/Micro Direct). The maneuver was repeated approximately four times, with the goal of matching the two highest trials within 10 cmH_2O)(Sachs et al., 2009).

The lower limit of normal for MIP and MEP were calculated using reference equations provided by Enright et al.(Enright, Kronmal, Manolio, et al., 1994).

Voice Assessments and Outcome Measures

Laryngeal Features

Videostroboscopy was used to visualize the vocal folds, using either a transnasal flexible endoscope or a rigid endoscope via the mouth, depending on the tolerability of the subject (KayPENTAX, model 9400; Olympus, VISERA Elite). Transnasal endoscopy: prior to the examination, topical lidocaine (numbing medication) and neosynepherine (vasoconstrictive medication) in a spray were administered if needed in the nasal cavities. Rigid endoscopy: prior to the examination, topical lidocaine was administered in the back of the mouth if needed to inhibit gagging. Participants were asked to sustain phonation on the vowel /i/ at a comfortable pitch and loudness.

Images obtained from videostroboscopy were rated through consensus by two SLPs specialized in voice disorders, using the Voice-Vibratory Assessment with Laryngeal Imaging (VALI) rating form (Poburka et al., 2016). This standardized assessment form was developed with the objective of improving the validity and reliability of visual examinations from videostroboscopy and high-speed videoendoscopy (HSV). The form contains definitions, graphics, rating instructions for each

of the 10 laryngeal parameter: 1) amplitude (magnitude of lateral excursion of the vocal folds), 2) mucosal wave (magnitude of movement of the mucosa), 3) nonvibratory portion, 4) supraglottal activity (antero-posterior and medio-lateral), 5) phase symmetry, 6) regularity (periodicity of vibration), 7) vertical level of approximation, 8) free-edge contour (smoothness and straightness), 9) glottal closure (type), and 10) phase closure. See Figure 11 for the complete assessment form and rating scales. Prior to rating the images, the judges underwent the VALI video training suggested by the authors of the form (Poburka et al., 2016). The video training presents each of the laryngeal parameters and how to assess them with the form. Training has been shown to improve the reliability of ratings, especially for temporally dependent ones such as phase closure and regularity of vibration (Poburka et al., 2016)

Videostroboscopy images were also used to measure bowing index (BI), using the *Fiji Is Just ImageJ* software for biological image analysis (Schindelin et al., 2012) developed by Wayne Rasband (U.S. National Institutes of Health, Bethesda, Maryland, USA 1997-2008). A still image of the open glottis was captured from the videostroboscopy assessment and entered into the software. The image was taken immediately prior to or following a sustained /i/ at comfortable pitch and loudness, during quiet breathing, with the vocal folds remaining in a resting position for at least 166.66 milliseconds (5 frames at a rate of 30 frames per second).

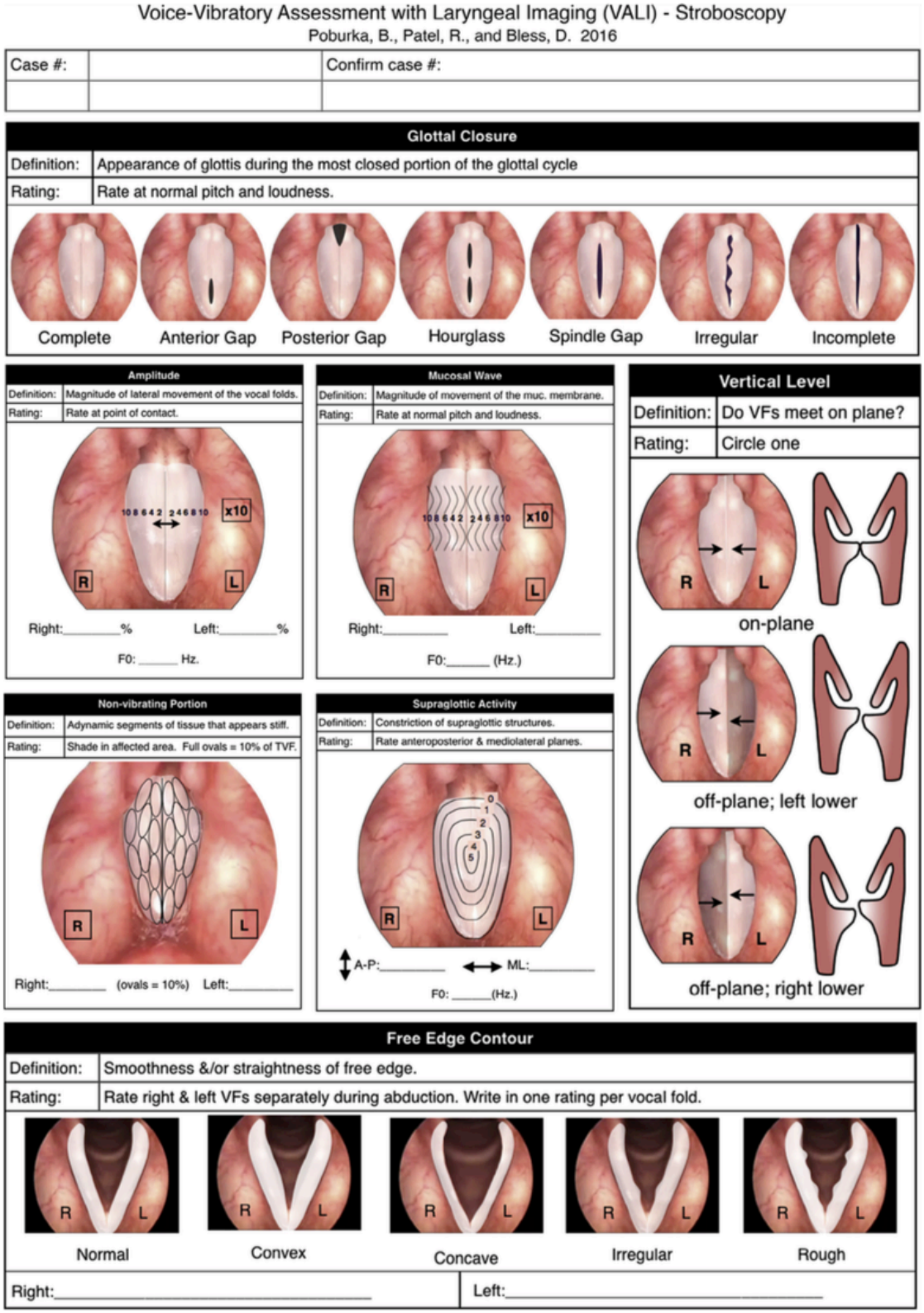
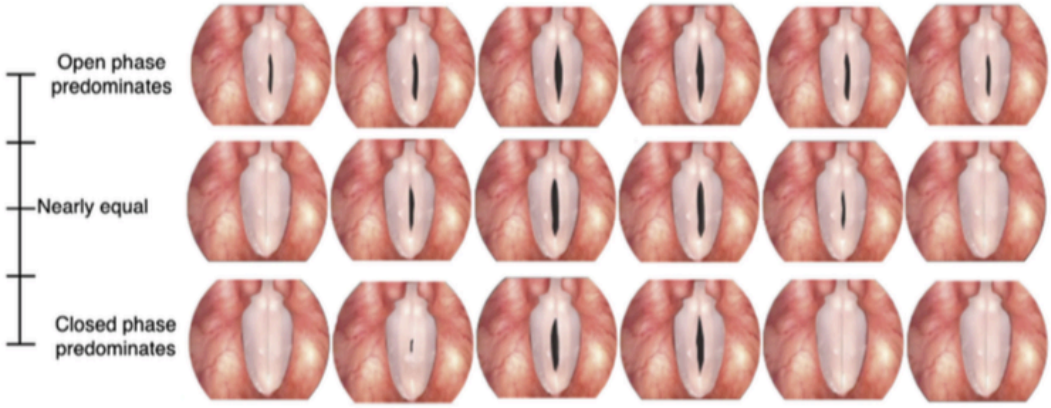
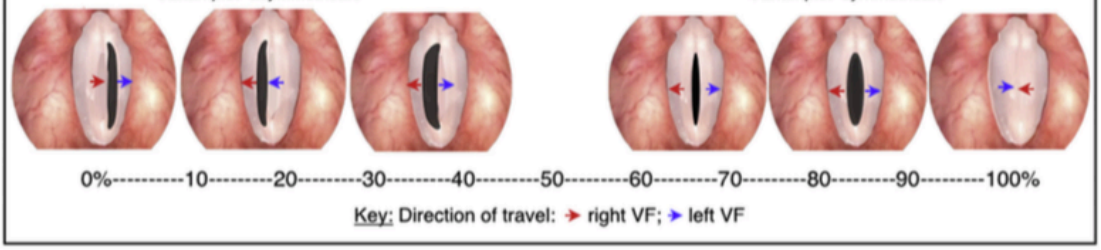


Figure 11. VALi Rating Form for Stroboscopy. Voice-Vibratory Assessment with Laryngeal Imaging (VALI) - Stroboscopy (Poburka, B., Patel, R., and Bless, D. 2016).

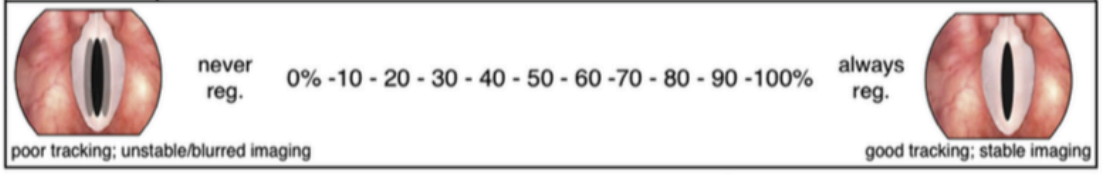
Phase Closure	
Definition:	The relative durations of appearance of consecutive glottal cycles.
Rating:	Circle a hash mark on the continuum below. Rate at point of contact.



Phase Symmetry	
Definition:	The degree to which the VFs move as mirror-opposite images of each other (180° phase difference).
Rating:	Select the % of exam time that vibration is symmetrical.



Regularity	
Definition:	Consistency of cycles.
Rating:	Circle % of time that vibration is regular



Nonvibratory Observations	
Definition:	Report observations that were not noted elsewhere (e.g.,structural differences, erythema, varices, mucous).

Figure 11. (continued)

BI was calculated following the methodology suggested by Omori et al.(Omori et al., 1997), which was replicated by Kaneko et al.(Kaneko et al., 2015). The length of the membranous vocal fold (L) was measured from the anterior commissure to vocal process by using the straight-line tool in the software (see figure 11). Using the same straight-line tool, the maximum distance (d) between the edge of the vocal fold and the line linking the anterior commissure to the tip of the vocal process was measured. BI was computed with the following formula: BI = d L x 100, and the sum of the BI for both vocal folds was used for analyses (Kaneko et al., 2015; Omori et al., 1997). The BI measurements were performed by two SLPs specialized in voice disorders, and the measures were averaged.

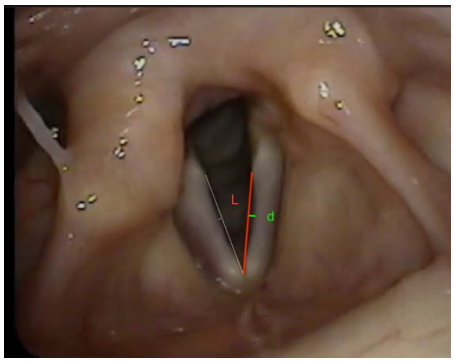


Figure 12. Measurement of Bowing Index

Acoustic Assessments and Maximum Phonation Time

Acoustic data was recorded using a hand-held microphone held at a distance of 5.5 inches from the patient's mouth and at a 45˚ angle. Recordings were made using the *Multi-Dimensional Voice Program Advanced and the Real-Time Pitch* program from the *Computerized Speech Laboratory* (CSL) (KayPENTAX, Montvale, NJ). Specific tasks and measurements recorded for each task are described in table 5.

Table 5. Tasks and Outcome Measures for Acoustic Analysis and MPT

Tasks	Outcome Measures
Sustained /a/ and /i/ at comfortable pitch and loudness for approximately 5 seconds	Jitter percent, shimmer percent, noise-to-harmonic ratio (NHR), soft phonation index (SPI), voice turbulence index (VTI), amplitude perturbation quotient (APQ), pitch perturbation quotient (PPQ), and standard deviation of F0 (STD), smoothed cepstral peak prominence measure (CPPS)
Sustain /a/ at comfortable pitch and loudness for as long as possible	Maximum phonation time (MPT)
Glide from the lowest pitch to the highest pitch on /a/	Minimum fundamental frequency (F0) Maximum F0
Glide from the softest voice to the loudest voice on /a/	Maximum sound pressure level (SPL)
Read the 6 sentences from the CAPE-V	Mean F0 during reading Mean SPL during reading CPPS during reading
Natural speech for approximately one minute	Mean F0 during conversation Mean SPL during conversation

In order to obtain smoothed cepstral peak prominence (CPPS), recordings of sustained /a/ at comfortable pitch and loudness and the three middle sentences from the CAPE-V (21 syllables) were imported into the software Pratt (version 6.0.46) and were analyzed using the methodology recommended by Maryn and Weenink (Maryn & Weenink, 2015) and reported in Phadke et al. (Phadke et al., 2018).

Auditory-Perceptual Judgments of Voice Quality

Voice recordings were also used to obtain perceptual ratings of voice quality, using the CAPE-V. Voice samples from the following tasks were submitted to raters: /a/ and /i/ sustained at comfortable pitch and loudness, readings from the six CAPE-V sentences, and sample of

spontaneous speech. The following parameters were assessed using a visual analog scale of 100 mm: overall severity, roughness, breathiness, strain, pitch, and loudness.

Aerodynamic Assessments

The Phonatory Aerodynamic System (KayPENTAX, model 6600) was used to collect measures of airflow and pressure. The participants wore a face mask over their nose and mouth. A small tube connected to a pressure transducer was passed through a hole in the mask and inserted between the lips. Subjects were asked to produce short utterances each comprised of five /pi/ syllables at a rate approximating 1.5 - 2 syllables per second (/pi-pi-pi-pi-pi/). Each string of five syllables was produced on one breath/exhalation. The task was repeated three times at comfortable pitch and loudness. The following measures were recorded: 1) mean airflow during voicing; 2) mean peak air pressure (P_s); and 3) aerodynamic resistance.

Self-Assessments

Four questionnaires were filled out by the subjects to assess the impact of their voice disorder on their daily activities: 1) The Voice Handicap Index 10 (VHI-10), which contains 10 clinically relevant questions chosen through expert consensus to capture the degree of perceived handicap related to voice(Rosen et al., 2004). Each item is rated on a 5-point scale, from 0 (no problem) to 4 (always a problem)(Jacobson et al., 1997). A score above 11 is indicative of a voice-related handicap (Arffa et al., 2012). 2) The GFI(Bach et al., 2005), a disease-specific impairment instrument meant to assess symptoms related to glottal insufficiency. The questionnaire contains four items rated on an ordinal scale from 0 (no problem) to 5 (severe problem), with a maximal score of 20 indicative of the highest severity. A GFI score higher than 4 is considered the cut-off value differentiating normal and abnormal voices (Bach et al., 2005). 3) The RSI(Belafsky et al., 2002), a validated self-

administered questionnaire to assess symptoms related to LPR. It contains nine items scored on an ordinal scale with values from 0 (no problem) to 5 (severe problem), for a total of 45 points. The cut-off value to differentiate a LPR related problematic is a RSI score higher than 13. 4) The Communicative Participation Item Bank (CPIB) (Baylor et al., 2013), a 10-item instrument designed to assess the construct of communicative participation across various communication disorders and situations. The items are score from 0 (the condition interferes very much with the situation) to 3 (the condition does not interfere at all). The total score ranges from 0 to 30, with a higher score being more favorable.

Lastly, subjects were asked to rate their perceived phonatory effort (PPE) on a direct magnitude estimation scale (Stevens, 1975) following the speaking and voicing tasks recorded with the CSL. They were instructed that 100 represents the most comfortable effort, comparable to before they experienced any difficulty with their voice. If their current effort was twice as much, they would say 200. If it was three times as much, they would say 300, and so on.

Assessment Timepoints

Participants included exclusively in aim 2 underwent only one assessment session (baseline), and did not receive any intervention. A subset of participants underwent two baseline sessions to assess intra-subject reliability (Aim 1).

Participants included in aim 3 underwent two main assessment sessions: at pre-treatment (baseline) and post-treatment (after the fourth treatment session). In addition, interim assessments were conducted before and after each session.

Main assessment sessions included all measures listed above: pulmonary function tests, respiratory muscle strength tests, laryngeal imaging, acoustic assessments, auditory-perceptual judgments of voice quality, aerodynamic assessments, and patient self-assessments (Appendix I).

Interim assessment sessions included: acoustic assessments and respiratory muscle strength tests.

The two following sections apply only to aim 3

Randomization

Following baseline assessments, participants were divided into four strata based on baseline respiratory muscle strength measures: 1) within normal limits MIP and MEP; 2) preserved MIP and decreased MEP; 3) preserved MEP and decreased MIP; 4) decreased MIP and MEP. The lower limit of normal (LLN) for MIP and MEP were determined using Enright et al.'s reference equations (Enright, Kronmal, Manolio, et al., 1994). Table 6 provides the reference equations predicting MIP and MEP based on gender, weight, and age, as well as the number to be subtracted from the result to determine LLN. Using a random number algorithm, participants were block-randomized into three groups (allocation ratio of 1:1:1): IMST and VFE (I+VFE); EMST and VFE (E+VFE); voice therapy only (VFE). The goal of the stratification was to ensure a balanced representation of different respiratory profiles between intervention groups.

The interventions are described in the following sections.

Table 6. Reference Equations for Respiratory Muscle Strength(Enright, Kronmal, Manolio, et al., 1994)

	Linear Regression Equations	LLN
Women	MIP = (0.133*Weight) - (0.805*Age) + 96 MEP = (0.344*Weight) – (2.12*Age) + 219	-32 -52
Men	MIP = (0.131*Weight) – (1.27*Age) + 153 MEP = (0.250*Weight) – (2.95*Age) + 347	-41 -71

Abbreviations: LLN: lower limit of normal; MIP: maximum inspiratory pressure; MEP: maximum expiratory pressure

Interventions

Participants attended four 1-hour therapy sessions with a SLP and a research team member. Each session consisted of 15 minutes of pre-session measures; followed by 10 minutes of IMST, EMST or VFE for groups I+VFE, E+VFE and VFE, respectively; then 20 minutes of voice exercises for all three groups and; lastly 15 minutes of patient education and post-session measures. Participants were asked to practice their respective exercises for 10 minutes, twice daily, seven days a week during the duration of the study.

Inspiratory Muscle Strength Training (IMST)

IMST was conducted using an inspiratory pressure threshold trainer (Philips Respironics® Threshold IMT or POWERbreathe® Medic Plus, for patients with a MIP of 55 cmH_20 and over), which consists of a mouthpiece with a spring-loaded valve (C. M. Sapienza, 2008; C. M. Sapienza et al., 1999). The valve blocks the airflow until the threshold pressure is achieved and allows airflow as long as the sufficient pressure is maintained. The threshold was set at 75% of the participant's initial MIP, as reported in the literature with elderly participants (C. H. Huang, Yang, Wu, & Lee, 2011). However, the load was lowered if the participant was unable to perform the exercise regimen. The load was adjusted weekly by the SLP based on the participant's MIP. Daily practices consisted of 5 sets of 5 breaths with the device with a 1-minute break between sets, repeated twice daily (Kim et al., 2009; Mills et al., 2015; Souza et al., 2014).

Expiratory Muscle Strength Training (EMST)

EMST was conducted using an expiratory pressure threshold trainer (EMST150^{TM}) which consists of a mouthpiece with a spring-loaded valve. The valve blocks the airflow until the threshold pressure is produced (set at 75% of the participant's initial MEP, per literature with similar age

groups (Pitts et al., 2009; C. Sapienza, Troche, Pitts, & Davenport, 2011; Troche et al., 2010)), and allows the airflow as long as the sufficient pressure is maintained. The load was lowered if the participant was not able to perform the exercise regimen. Participants were instructed to exhale forcefully into the mouthpiece after a full inspiration. The load was adjusted weekly by the SLP to maintain a threshold at 75% of the participant's MEP. Daily practices consisted of 5 sets of 5 breaths with the device with a 1-minute break between sets, repeated twice daily (Pitts et al., 2009; C. Sapienza et al., 2011; Troche et al., 2010).

Voice therapy

Participants were instructed to follow the four steps of the VFE protocol, developed by Stemple (Stemple, 2005) and commonly used by SLPs with patients with presbyphonia. The protocol was slightly adapted and the following exercises were used: (a) sustain "mi" on the musical note F (above middle C for women; below middle C for men) for as long as possible. (b) Glide from the lowest note to the highest note on the word "whoop". (c) Glide from the highest note to the lowest note of the word "boom". (d) Sustain the notes C-D-E-F-G (starting from middle C for women and an octave below middle C for men) on the word "moo", for as long as possible. Each note was repeated until the participants found the right placement (forward-focused voice), as judged by the SLP. Daily practices, 2x/day, consisted of two repetitions of the VFE protocol for the I+VFE and E+VFE groups and four repetitions of the VFE protocol for the VFE group. Recordings of the voice exercises or a link to an online keyboard were provided to facilitate home practice, depending on the preference of the participant.

Home Practice

Participants were instructed to practice their respective exercises twice daily, and to log their practice in a compliance journal (days of training, perceived effects of treatment or other comments). In addition to in-person therapy sessions, weekly phone calls to the participants were given to clarify aspects of the training and encourage treatment adherence.

Statistical Approaches and Data Analysis

All statistical analyses were conducted using SAS statistical software (version 9.4, released 2016, SAS Institute Inc., Cary, N.C., USA) or IBM SPSS Statistics for Windows (version 24, released 2016, IBM Corp., Armonk, N.Y., USA).

Measures of Reliability

Intraclass Correlation Coefficients (ICCs)

ICCs are indexes of reliability that provide information on the ratio of true variance to total variance (true variance plus error)(Shrout & Fleiss, 1979; Weir, 2005). More specifically, the Shrout and Fleiss $ICC_{3,1}$ was computed to assess intra-subject reliability of each outcome (Shrout & Fleiss, 1979). This model is based on a two-way mixed model ANOVA, in which the effect of the trial is fixed and therefore is considered to be the same for all participants (same therapists administering the task, and same equipment). The true score variance, since it is unknown, is estimated from the between-subjects variance resulting from the ANOVA's mean square values (Weir, 2005). A disadvantage from this model is that it only accounts for random error, and not for systematic error. Therefore, if all subjects vary consistently across trials, the resulting ICC may be strong despite a significant trial effect (Weir, 2005). This may cause an issue when using the

outcome measures to assess pre- to post-treatment effect, and therefore it is recommended that an ANOVA be conducted in addition to the ICC, to assess potential systematic trial effect (Weir, 2005).

ICCs can have a value between 0 (or lower) and 1, for which 0 indicates a concordance that could have occurred by chance, whereas 1 indicates perfect reliability (100% of the variance is explained by the true variance)(Chinn, 1991; Weir, 2005). The ICC values can be interpreted following guidelines established in the literature and applied in various studies: values of less than 0.40 are considered indicative of poor reliability; values equal or greater than 0.40 and equal or smaller than 0.75 indicate fair to good reliability, and values greater than 0.75 indicate excellent reliability (Awan et al., 2013; Rosner, 2005). Importantly, outcome measures presenting with an ICC of less than 0.60 were excluded from further analyses, as they were deemed not reliable enough to accurately inform on pre- to post-therapy changes (Awan et al., 2013; Chinn, 1991).

Standard Error of Measurement (SEM)

SEM represents the random error, or "noise", that is expected when one subject is tested more than once for the same outcome measure (Hopkins, 2000). Contrarily to ICC, SEM is not affected by the between-subject variability and represents an absolute measure of reliability (Weir, 2005). SEM theoretically corresponds to the within-subject standard deviation (Hopkins, 2000); however, since this exact value cannot be computed because there is usually a small number of trials for each participant, it is estimated by using the ICC value in the following formula: $SEM=SD\sqrt{1-ICC}$, where SD corresponds to the between-subject standard deviation across all trials (Hopkins, 2000; Weir, 2005). SEM can also be determined by computing the square root of the mean square error term in the ANOVA output ($\sqrt{M_{se}}$)(Hopkins, 2000; Weir, 2005). In fact, SEM is

equivalent to the SD of the residuals (random error) for each subjects (Hopkins, 2000). The former calculation method was used in the present study.

Coefficient of variation (CV)

CV is the ratio of the SEM to the mean multiplied by 100 (Awan et al., 2013; Higgins et al., 1994; Hopkins, 2000; Weir, 2005). Therefore, it relates the estimated within-subjects spread of measurements to the mean and provides a relative magnitude of the SEM (Awan et al., 2013). This measure facilitates comparisons across outcome measures and across samples, because it is unit-less and independent of the size of the value. In fact, for most measurements, SEM increases as the value of the outcome increases, which makes comparisons of reliability difficult across different population samples (Hopkins, 2000; Nevill & Atkinson, 1997).

Minimum difference (MD) or Minimum Detectable Change (MDC)

MD provides an indication of how much difference is needed for two measurements to be considered different, and not only due to random error in measurement (Weir, 2005). In other words, it is the least amount of change that needs to occur between two trials in order to be ensured that there was an intervention effect. This measure is therefore very relevant in interpreting the results of an intervention study. It is calculated with the following formula: MD= SEM x 1.96 x $\sqrt{2}$ (Weir, 2005).

Data analysis: Aim 1

For Aim 1, the methodology described by Awan et al. was followed (Awan, 2006). Firstly, the distribution of the data was verified for each outcome with the Shapiro-Wilk test, using the SAS statistical software. For each outcome presenting with a non-normal distribution, a linear

regression analysis was conducted between the two trials. Residuals that were more than three SD from the mean were identified, and the corresponding outliers were removed. The Shapiro-Wilk test was then conducted again on the data to test for normality. Distributions that remained non-normal even after the removal of outliers were log10 transformed. A two-way ANOVA was then conducted with one between-subject factor (gender) and one within-subject factor (time). Time effect was assessed to identify any systematic variability between the two asessements, which is not taken into account in the $ICC_{3,1}$ calculation (Shrout & Fleiss, 1979). If a gender effect was found, further analyses were conducted separately for males and females to avoid bias from heteroscedasticity.

Data was then imported in IBM SPSS Statistics for Windows for computation of ICCs. To be consistent with the $ICC_{3,1}$ as described by Shrout and Fleiss (Shrout & Fleiss, 1979), *two-way mixed* model and *consistency* were chosen from the scroll-down menus. The *single measures* values were reported. Outcome measures presenting with an ICC of 0.60 or more were deemed to present with sufficient intra-subject reliability (Awan et al., 2013; Chinn, 1991).

SEM was calculated with the formula: $SEM=SD\sqrt{1-ICC}$, in which SD represents the pooled standard deviation for both trials. CV was calculated by dividing SEM by the grand mean of all trials, and by multiplying the result by 100. MD was computed with the formula $MD= SEM \times 1.96 \times \sqrt{2}$. SEM and MD were not reported for outcomes with non-normal distributions, because log transformations of these values are not valid measures (Awan et al., 2013; Bland & Altman, 1996). However, it was shown that the SEM of the log transformed value was approximately equal to the CV of the original outcome, so this association was used to provide CV of log transformed variables ($CV=(\ln(10)SEM)\times100$) (Euser, Dekker, & le Cessie, 2008). CV can only be computed for variables with an absolute 0 and therefore was not provided for PPE, which was rated on a magnitude scale with no absolute 0 (Euser et al., 2008).

Data Analysis: Aim 2

Descriptive statistics including mean, standard deviation, median, and interquartile range, were computed on each respiratory and voice outcome, for the total sample and for males and females separately. Normality of the distributions was assessed with the Shapiro-Wilk test (alpha level set at 0.05). A Mann-Whitney U test was conducted to assess any significant difference between genders (alpha level set at 0.05). Even when the data was normal, the non-parametric test was chosen considering the small sample size. Missing data from the self-assessment questionnaires were imputed using linear regression on the other questions from the same questionnaire.

In SAS, voice and respiratory outcomes were standardized (mean=0, SD=1) and then added +5 to eliminate negative values. An exploratory factor analysis was conducted on the standardized respiratory and voice measures to extract the latent constructs of the sets of outcome measures. The cut-off value of the loading score for inclusion in the factor was set at 0.71 because of the small sample size (Comrey & Lee, 2013). A weighted variable was then created for each factor, by multiplying the standardized values of its variables by their associated loading score. In total, two respiratory factors and eight voice factors were created. A cluster analysis was then conducted and included all the respiratory and voice factors, as well as age and gender.

Spearman correlations were conducted between all respiratory and voice factors and between age and all factors. Linear regressions were then computed between respiratory variables and the voice factors that were found to have a significant ($p \leq 0.05$) or marginally significant ($p \leq 0.1$) correlation with one of the respiratory factors (we considered statistically significant and marginally significant values to be relevant because this is a pilot study. We erred on the side of smaller type II versus type I error). Model fit was verified by visual examination of the residual

plots, and all assumptions for linear regression were met. We examined if adding age/gender increased the model's explanatory ability by looking at their respective p-values as well as the adjusted R-squared for the overall model. Age and gender were kept in the model only if they improved its explanatory ability.

Data Analysis: Aim 3a

Comparisons of the three intervention groups at baseline were conducted using a one-way ANOVA for continuous data in SPSS. The assumption of homogeneity of variances was verified for each variable using Levene's test. If the assumption was not met, result from the Welch test was reported. For ordinal data, groups were compared using the Kruskal-Wallis analysis of variance by ranks.

For single-subject analyses, standard mean difference (SMD) were computed for each variable. To compute SMD, the numerator is usually calculated by averaging the baseline data and the intervention data (or 3 baseline points and the last 3 intervention points, to obtain SMD_3). The former is then subtracted from the latter to obtain the numerator. The denominator is the pooled standard deviation from the population sample, at baseline (Gierut, Morrisette, & Dickinson, 2015). Since only two baseline data points (or one for some participants) were obtained in this study, the two baseline data points were averaged, as well as the two last intervention points, to obtain SMD_2. Magnitude of the effect sizes were based on Cohen's (1988) and Sawilowsky's (2009) interpretation guidelines: small (0.20), medium (0.50), large (0.80), very large (1.20), and huge (2.0).

For group analyses, a two-way mixed ANOVA was computed in SPSS for each continuous variable with time as a within-subject factor and intervention group as a between-subject factor. Tukey test was used for post-hoc group comparisons. Once again we considered statistically significant (0.05) and marginally significant (0.1) values to be relevant because this is a pilot study. We erred on the side of smaller type II versus type I error.

Pre-post effect sizes within and between the intervention groups were computed using Cohen's d. The standard deviation used in the calculation of the effect sizes was that of the dependent variable at pre-treatment for the total sample (N=10). Because of the very small sample size of each of the intervention group, the standard deviation of the total sample was more likely to approach the real standard deviation of that measure for this particular population, thus reducing the risk of over or underestimating the true effect size (Carlson & Schmidt, 1999). Effect sizes between groups were adjusted considering the baseline means for each group (Carlson & Schmidt, 1999; Durlak, 2009). The formula used was the following: $d_b=(E_2-E_1)-(C_2-C_1)/S_{E1C1}$ where E is the mean of the experimental group (IMST or EMST) and C is the mean of the control group (VFE); 2 represents post-treatment and 1 pre-treatment; and S represents the standard deviation of the population sample at pre-treatment (experimental and control groups pooled)(Carlson & Schmidt, 1999). Effect sizes within each group were calculated using the following formula: $d_w=G_2-G_1/ S_{E1C1}$, where G_1 is the mean for the dependent variable at pre-treatment for this group, G_2 is the mean at post-treatment, and S_{E1C1} is the pooled standard deviation of all the groups (experimental and control groups) at pre-treatment.

Lastly, Spearman correlations were computed in SAS, between the pre-post difference in respiratory measures and voice measures.

Data Analysis: Aim 3b

Multiple linear regressions were conducted in SAS to assess the relationships between the difference in VHI-10 score and potential predictors and covariates. The standardized estimate parameter (β) was computed to facilitate comparison across predictors, which have different units. The β represents the change in the outcome, in terms of number of SDs, that occurs for a change of one SD in the independent variable (Nathans, Oswald, & Nimon, 2012).

A univariate ANCOVA was then conducted in SPSS with the change in VHI-10 as the dependent variable, the intervention group as the independent variables, and baseline respiratory function (using the strongest predictor) as the covariate. The results were compared to those of a univariate ANOVA (without controlling for baseline respiratory function). For both the ANOVA and the ANCOVA, pairwise comparisons between intervention groups were computed, with Bonferroni adjustment for multiple comparisons.

Chapter 4: Results

Results Aim 1

To determine the measurements that were most appropriate for investigating the relationship between respiration and voice, we assessed the stability of common measurements using intra-subject reliability in participants that had two baseline assessments.

Participant Characteristics

Twelve participants (7 females, 5 males) underwent two baseline assessments for analysis of intra-subject reliability. Table 7 presents the characteristics of these participants.

Table 7. Participant characteristics for test-retest reliability assessment

SUBJECT	AGE	GENDER	SMOKING STATUS	PROFESSIONAL VOICE USE	REFLUX SYMPTOM INDEX	DAYS BETWEEN TRIALS
1	75	Female	Quit	Yes	13	16
2	79	Female	Quit	No	19	15
3	66	Female	Never	No	36	14
4	59	Male	Never	No	18	8
5	56	Male	Never	No	31	28
6	67	Female	Quit	No	20	23
7	75	Male	Quit	Yes	15	15
9	58	Male	Never	Yes	17	5
12	65	Female	Quit	Yes	25	62
13	79	Female	Never	No	21	49
16	68	Female	Quit	Yes	18	15
17	74	Male	Quit	Yes	6	5
MEAN (SD) OR FREQUENCY	68.42 (8.05)	Female: 7 Male: 5	Quit: 7 Never: 5	Yes: 6 No: 6	19.92 (7.93)	21.25 (17.53)

Rater Reliability

Intra-rater reliability was calculated for the CAPE-V, as well as for each of its individual parameters. Reliability for the CAPE-V was excellent, with an overall ICC of 0.938. Intra-rater reliability for each of the parameters were: 0.958 for overall severity; 0.895 for roughness; 0.976 for breathiness; 0.544 for strain; 0.719 for pitch; and 0.873 for loudness. All parameters presented with an excellent intra-rater reliability except for the parameters of strain and pitch.

Time and Gender Effects

Table 8 shows the results of the mixed ANOVAs for effects of time and gender. Four variables showed a significant time effect, indicating a potential systematic error: SPL reading ($F_{1,10}$=6.735; p=0.027); F0 reading ($F_{1,10}$=6.546; p=0.028); F0 speech ($F_{1,10}$=7.461; p=0.021); and SPI /i/ ($F_{1,10}$=13.274; p=0.007). For the four variables, the mean values decreased significantly from the first to the second trial.

Seventeen variables demonstrated a significant gender effect and/or a significant time by gender interaction effect (see Table 8). Gender effects were found as expected for all measures involving F0 (F0 reading F0 speech, minimum F0, and maximum F0), and for measures of respiratory muscle strength (MIP and MEP). In addition, significant gender effects were found for jitter /a/, shimmer /a/, PPQ /a/, STD /a/, jitter /i/, PPQ /i/, and for the overall and breathiness parameters of the CAPE-V. Significant time by gender interactions were found for STD /a/, PPE, and for the roughness and loudness parameters of the CAPE-V.

For all variables with significant gender or interaction effects, separate reliability analyses were conducted for males and females in order to avoid bias caused by the gender differences.

Table 8. Two-Way Mixed ANOVA Results for within-subject effects (time) and between-subject effects (gender)

	Time	**Time*gender**	**Gender**
Variable	F (p-value)	F (p-value)	F(p-value)
SPL reading	6.735 (0.027)*	0.648 (0.440)	2.02 (0.186)
SPL speech	3.396 (0.095)	0.385 (0.549)	0.500 (0.496)
Max SPL	0.610 (0.453)	0.187 (0.675)	0.752 (0.406)
F0 reading	6.546 (0.028)*	2.465 (0.147)	30.817 (0.000)*
F0 speech	7.461 (0.021)*	2.217 (0.167)	34.013 (0.000)*
Max F0	0.360 (0.562)	0.183 (0.678)	8.547 (0.015)*
Min F0	0.126 (0.730)	0.034 (0.857)	22.271 (0.001)*
CPPS sustained /a/	0.002 (0.966)	0.545 (0.477)	0.776 (0.399)
CPPS reading	0.181 (0.680)	0.007 (0.937)	0.981 (0.345)
Jitter % /a/	0.012 (0.916)	0.845 (0.382)	20.534 (0.001)*
Shimmer % /a/ (log10)	0.082 (0.781)	1.321 (0.280)	7.780 (0.021)*
NHR /a/ (log10)	0.188 (0.675)	0.11 (0.919)	2.137 (0.178)
SPI /a/	2.10 (0.168)	2.782 (0.126)	1.347 (0.273)
VTI /a/	0.038 (0.851)	0.182 (0.680)	0.045 (0.836)
APQ /a/ (log10)	0.702 (0.424)	0.497 (0.499)	3.921 (0.079)
PPQ /a/	0.073 (0.793)	0.995 (0.345)	18.745 (0.002)*
STD /a/ (log10)	0.203 (0.663)	5.773 (0.040)*	18.199 (0.002)*
Jitter % /i/	0.002 (0.970)	0.268 (0.618)	7.904 (0.023)*
Shimmer % /i/ (log10)	0.044 (0.839)	1.202 (0.301)	0.846 (0.382)
NHR /i/ (log10)	0.133 (0.724)	0.491 (0.501)	0.364 (0.561)
SPI /i/	13.274 (0.007)*	2.543 (0.149)	0.058 (0.816)
VTI /i/	0.405 (0.540)	1.582 (0.240)	3.514 (0.094)
APQ /i/ (log10)	0.244 (0.633)	0.596 (0.460)	0.011 (0.918)
PPQ /i/	0.001 (0.972)	0.154 (0.705)	8.346 (0.020)*
STD /i/	1.439 (0.265)	2.033 (0.192)	4.627 (0.064)
PPE (log10)	0.201 (0.663)	8.802 (0.014)*	0.198 (0.666)
MPT	0.580 (0.464)	3.478 (0.092)	0.278 (0.610)
MIP	0.388 (0.547)	2.222 (0.167)	6.945 (0.025)*
MEP	0.506 (0.493)	0.549 (0.476)	17.242 (0.002)*
Overall	2.699 (0.131)	0.450 (0.518)	7.876 (0.019)*
Roughness	4.270 (0.066)	5.846 (0.036)*	4.178 (.068)
Breathiness	0.227 (0.644)	0.227 (0.644)	12.835 (0.005)*
Strain (log10)	1.095 (0.320)	1.039 (0.332)	1.420 (0.261)
Pitch[a]	1.255 (0.289)	1.255 (0.289)	3.775 (0.081)
Loudness[a]	3.216 (0.103)	5.247 (0.045)*	2.262 (0.164)

*Indicates significance at a 0.05 alpha level.

[a]Levene's test of equality of variances was significant

SPL=sound pressure level; F0=fundamental frequency; CPPS=smoothed cepstral peak prominence; NHR= noise-to-harmonic ratio; SPI=soft phonation index; VTI=voice turbulence index; APQ=amplitude perturbation quotient; PPQ=pitch perturbation quotient; STD=standard deviation of F0; PPE=perceived phonatory effort; MPT=maximum phonation time; MIP=maximum inspiratory pressure; MEP=maximum expiratory pressure

Intraclass Correlation Coefficients

Table 9 displays the two trials' means and SDs (or medians and interquartile ranges for log-transformed or non-normal outcomes), as well as the ICCs for each variable. ICCs varied between -0.866 and 0.970. The following variables were found to have a poor reliability (ICC$<$0.40): minimum and maximum F0 (in females), jitter /a/, jitter /i/ (in females), shimmer /a/ (in males), shimmer /i/, SPI /a/, VTI /a/, VTI /i/, PPQ /a/, PPQ /i/ (in females), STD /a/ (in males), STD /i/, NHR /i/, APQ /i/, PPE, MPT, and roughness (in females). The following variables showed a fair to good reliability (0.40$\leq$ICC$\leq$0.75): maximum SPL, minimum F0 (males), CPPS /a/, CPPS reading, shimmer /a/ (females), APQ /a/, STD /a/ (females), SPI /i/, MEP (males). The following variables demonstrated an excellent reliability (ICC$>$0.75): SPL reading, SPL speech, F0 reading, F0 speech, maximum F0 (males), NHR /a/, jitter /i/ (males), PPQ /i/ (males), MIP, MEP (females), breathiness (in females), strain, pitch (in females), and loudness (in males).

Variables with an ICC of 0.60 or greater were considered to present with sufficient intra-subject reliability to inform on pre- to post-treatment changes. Variables with an ICC greater or equal to 0.60 (for both females and males when analyses were conducted separately), were: SPL reading (ICC=0.807), SPL speech (ICC=0.850), F0 reading (ICC=0.806 for females and ICC=0.853 for males), F0 speech (ICC=0.881 for females and ICC=0.957 for males), CPPS /a/ (ICC=0.666), CPPS reading (ICC=0.715), NHR /a/ (ICC=0.782), APQ /a/ (ICC=0.642), SPI /i/ (ICC=0.643), MIP (ICC=0.804 for females and ICC=0.941 for males), MEP (ICC=0.853 for females and ICC=0.746 for males), overall perceptual judgment of voice quality (ICC=0.805 for females and ICC=0.858 for males).

Table 10 displays the ICCs and other measures of reliability (SEMs, MDs, and CVs) for these variables. However, SPL reading, F0 reading, F0 speech, and SPI /i/ were removed because they

presented with a significant time effect indicative of systematic error. The pitch parameter from the CAPE-V was also removed because it presented with zero variance for males.

Table 9. Intraclass Correlation Coefficients

Variable	Test Mean (SD)	Re-test Mean (SD)	$ICC_{3,1}$
SPL reading	76.24 (5.80)	73.71 (5.32)	0.807
SPL speech	70.75 (5.40)	72.53 (5.69)	0.850
Max SPL	96.89 (7.18)	98.52 (6.92)	0.429
F0 reading			0.938
Females	194.92 (23.35)	191.32 (25.85)	0.806
Males	137.82 (12.62)	122.75 (9.14)	0.853
F0 speech			0.970
Females	209.67 (30.14)	195.56 (28.01)	0.881
Males	121.74 (17.01)	117.59 (16.88)	0.957
Max F0			0.677
Females	745.92 (139.37)	755.26 (195.39)	0.116
Males	454.01 (181.77)	509.94 (216.33)	0.876
Min F0			0.203
Females	148.55 (40.76)	145.96 (33.02)	-0.553
Males	102.17 (26.28)	93.95 (11.96)	0.589
CPPS /a/	22.13 (3.02)	22.03 (2.08)	0.666
CPPS reading	16.02 (1.36)	16.13 (0.81)	0.715
Jitter % /a/			0.707
Females	2.63 (0.83)	2.93 (1.06)	0.362
Males	1.06 (0.38)	0.82 (0.23)	-0.449
Shimmer/a/			0.579
Females	6.14 (3.05)	6.95 (2.20)	0.738
Males	3.99 (0.85)	3.59 (0.74)	-0.866
NHR /a/	Median: 0.15 IQ range: 0.06	Median: 0.16 IQ range: 0.39	0.782
SPI /a/	22.46 (12.46)	16.67 (7.16)	0.293
VTI /a/	0.04 (0.02)	0.04 (0.01)	0.169
APQ /a/	Median: 3.75 IQ range: 1.74	Median: 3.97 IQ range: 2.16	0.642
PPQ /a/			0.693
Females	1.56 (0.47)	1.79 (0.72)	0.363
Males	0.62 (0.22)	0.48 (0.12)	-0.467
STD /a/			0.549
Females	Median: 6.93 IQ range: 3.60	Median: 19.14 IQ range: 20.57	0.469
Males	Median: 2.60 IQ range: 1.43	Median: 1.97 IQ range: 0.21	0.066
Jitter % /i/			0.213
Females	1.78 (0.29)	1.98 (1.23)	-0.216
Males	0.95 (0.61)	0.74 (0.31)	0.885

Shimmer % /i/	Median: 2.71 IQ range: 2.51	Median: 2.74 IQ range: 1.79	0.018
NHR /i/	Median: 0.13 IQ range: 0.04	Median: 0.13 IQ range: 0.02	-0.475
SPI /i/	15.53 (6.4)	10.01 (3.73)	0.643
VTI /i/	0.04 (0.02)	0.05 (0.01)	-0.045
APQ /i/	Median: 2.36 IQ range: 1.18	Median: 2.41 IQ range: 1.54	0.033
PPQ /i/			0.196
Females	1.04 (0.17)	1.15 (0.73)	-0.240
Males	0.52 (0.29)	0.44 (0.19)	0.970
STD /i/	4.29 (1.83)	10.18 (14.10)	0.040
PPE	Median: 135 IQ range: 100	Median: 150 IQ range: 160	0.162
MPT	17.48 (5.76)	18.60 (8.45)	0.165
MIP			0.890
Females	75.57 (26.30)	83.86 (25.02)	0.804
Males	117.80 (24.35)	114.40 (20.86)	0.941
MEP			0.909
Females	96.14 (18.05)	96.00 (18.56)	0.853
Males	149.60 (24.28)	156.60 (38.67)	0.746
Overall			0.892
Females	56.00 (18.15)	58.86 (17.45)	0.805
Males	27.60 (16.59)	34.40 (13.90)	0.858
Roughness			0.686
females	42.71 (15.55)	60.57 (13.08)	0.355
Males	32.40 (24.07)	31 (20.52)	0.937
Breathiness			0.760
females	51.00 (25.99)	45.29 (24.05)	0.454
males	9.20 (13.03)	9.20 (13. 99)	0.945
Strain	Median: 0.90 IQ range: 0.84	Median: 0.93 IQ range: 0.59	0.493
Pitch			0.784
Females	Median: 4.00 IQ range: 11	Median: 4.00 IQ range: 6.00	0.741
Males	Median: 0 IQ: 0	Median: 0 IQ: 0	No variance
Loudness			0.764
females	Median: 11 IQ range: 38	Median: 0 IQ range: 11	0.811
males	Median: 0 IQ range: 0	Median: 0 IQ range: 0	0.471

Abbrevations: SD: standard devation; ICC: intraclass correlation coefficient; SPL: sound pressure level; F0: fundamental frequency; CPPS: smoothed cepstral peak prominence; NHR: noise-to-harmonic ratio; SPI: soft phonation index; VTI: voice turbulence index; APQ: amplitude perturbation quotient; PPQ: pitch perturbation quotient; STD: standard deviation of F0; PPE: perceived phonatory effort; MPT: maximum phonation time; MIP: maximum inspiratory pressure; MEP: maximum expiratory pressure

Table 10. Measures of Reliability

Variable	Test Mean (SD)	Re-test Mean (SD)	$ICC_{3,1}$	SD (pooled)	SEM	CV (%)	MD
SPL speech	70.75 (5.40)	72.53 (5.69)	0.850	5.54	2.15	3.00	5.95
CPPS /a/	22.13 (3.02)	22.03 (2.08)	0.666	2.55	1.47	6.67	4.08
CPPS reading	16.02 (1.36)	16.13 (0.81)	0.715	1.08	0.58	3.60	1.60
NHR /a/	Median: 0.15 IQ range: 0.06	Median: 0.16 IQ range: 0.39	0.782	N/A	N/A	7.00	N/A
APQ /a/	Median: 3.75 IQ range: 1.74	Median: 3.97 IQ range: 2.16	0.642	N/A	N/A	9.70	N/A
MIP			0.890				
Females	75.57 (26.30)	83.86 (25.02)	0.804	25.66	11.36	14.25	31.49
Males	117.80 (24.35)	114.40 (20.86)	0.941	22.60	5.49	4.73	15.22
MEP			0.909				
Females	96.14 (18.05)	96.00 (18.56)	0.853	18.30	7.02	7.30	19.45
Males	149.60 (24.28)	156.60 (38.67)	0.746	31.47	15.86	10.36	43.97
Overall severity			0.892				
Females	56.00 (18.15)	58.86 (17.45)	0.805	17.80	7.86	13.69	21.79
Males	27.60 (16.59)	34.40 (13.90)	0.858	15.25	5.74	18.53	15.92

SD=standard deviation; ICC=intraclass correlation coefficient; SEM=standard error of measurement; CV=coefficient of variation; MD=minimum difference; SPL=sound pressure level; CPPS=smoothed cepstral peak prominence; NHR=noise-to-harmonic ratio; APQ=amplitude perturbation quotient; MIP=maximum inspiratory pressure; MEP=maximum expiratory pressure

Other Measures of Reliability (SEMs, MDs, and CVs)

Among the variables presenting with an ICC greater or equal to 0.60, SPL speech had the smallest CV (CV=3.00%), and overall severity in males had the largest CV (CV=18.53%). The CV for MIP was much smaller in males (CV=4.73%) than in females (CV=14.25%). This also translated into a greater MD needed in females in order to be confident that a true difference occurred (MD= 31.49 cmH_20 in females; MD=15.22 cmH_20 in males). The opposite was found regarding MEP, for which higher CVs and MDs were found in males (CV=10.36%; MD=43.97 cmH_20) than in females (CV=7.30%; MD=19.45 cmH_20). CPPS /a/, CPPS reading, NHR /a/, and APQ /a/ were the only voice quality

measures that presented with adequate reliability based on the 0.60 cut-off for ICC. CPPS /a/ showed a CV of 6.67 and a MD of 4.08dB. CPPS during reading showed a higher reliability than CPPS /a/, as well as a smaller CV (CV=3.60%) and MD (MD=1.60 dB). NHR /a/ and APQ /a/ presented with CVs of 7% and 9.7%, respectively. SEMs and MDs were not computed for these variables because they were log transformed. Auditory-perceptual judgment of overall severity was the only parameter of the CAPE-V to present with an adequate ICC for both females and males. Overall severity as well as standard deviations were greater in females than in males. The SEM, CV, and MD were also greater in females when compared to males (Table 10).

Results: Aim 2

To address Aim 2, we first assessed the descriptive statistics for the various measurements. We then used exploratory factor analysis to extract the main constructs measured by the respiratory and voice outcomes, and cluster analysis to identify groups of participants with similar profiles based on these constructs. Lastly, we used correlations and linear regression models to specify the relationships between respiratory and voice constructs. All measurements were taken from the participants' baseline assessements.

Participant Characteristics

Twenty-one participants met inclusion criteria for this aim. Age was normally distributed among the participants, and the mean age was 72 years old (range: 56 to 91). Ten participants were females and eleven participants were males. The median age for both females and males were 74 years old. Twelve participants had never smoked and nine participants were former smokers. Nine participants were active or retired professional voice users. Table 11 displays the characteristics of the participants. The majority of participants were Caucasian, one was Asian.

Table 11. Participant Characteristics

CHARA	AGE	GENDER	SMOKING STATUS	PROFESSIONAL VOICE USE	RACE	REFLUX SYMPTOM INDEX
1	75	Female	Quit	Yes	Caucasian	13
2	79	Female	Quit	No	Caucasian	19
3	66	Female	Never	No	Caucasian	36
4	59	Male	Never	No	Asian	18
5	56	Male	Never	No	Caucasian	31
6	67	Female	Quit	No	Caucasian	20
7	75	Male	Quit	Yes	Caucasian	15
8	91	Female	Never	No	Caucasian	5
9	58	Male	Never	Yes	Caucasian	17
10	74	Male	Quit	No	Caucasian	16
11	80	Female	Never	No	Caucasian	8
12	65	Female	Quit	Yes	Caucasian	25
13	79	Female	Never	No	Caucasian	21
14	73	Female	Never	No	Caucasian	25
16	68	Female	Quit	Yes	Caucasian	18
17	74	Male	Quit	Yes	Caucasian	6
18	82	Male	Never	Yes	Caucasian	9
19	75	Male	Quit	No	Caucasian	22
20	64	Male	Never	No	Caucasian	7
21	65	Male	Never	Yes	Caucasian	14
22	81	Male	Never	Yes	Caucasian	17
MEAN (SD) OR FREQUENCY	71.71 (8.94)	Male:11 Female:10	Never:12 Quit: 9	Yes:9 No: 12	Caucasian:20 Asian:1	17.24 (8.00)

Respiratory Measures

Table 12 presents the group results for pulmonary function and respiratory muscle strength. Appendix II presents individual results for these respiratory outcomes.

All respiratory measures presented with a normal distribution across the sample, with the exception of the FEV1/FVC ratio (raw and predicted percent values). As expected, the raw values of FVC, FEV1, and MEP were significantly different between males and females (p=0.013, p=0.020, and p=0.003, respectively). Although it did not reach significance, a considerable difference in means and medians for MIP was also found between genders, with males having a larger MIP than females.

Males presented with a mean FVC of 3.80 liters (L) (SD=1.15), and females with a mean FVC of 2.56 L (SD=0.43). The mean FEV1 was 2.77 L (SD=0.94) in males and 1.87 L (SD=0.32) in females. The FEV1/FVC ratio had a median of 0.73 (IQ range=0.06) across the total sample, and a median percent predicted value of 98% (IQ range=9). The mean percent predicted value of FVC was 89.81% (SD=23.81) and the mean percent predicted value of FEV1 was 88.43% (SD=26.38), for the total sample. One third of the participants (seven participants) presented with FVC% and FEV1% values below 80%. One participant presented with a percent predicted FEV1/FVC below 80%.

Males had a mean MEP of 141.18 cmH_20 (SD=34.89), while females had a mean MEP of 92.90 cmH_20 (SD=15.78). As expected, MIP was lower than MEP in both genders, with a mean of 93.73 cmH_20 (SD=36.95) in males and 70.50 cmH_20 (SD=23.58) in females. Nine percent of the males (one participant) and 0% of the females had a MIP below the lower limit of normal (LLN) expected

for their age, gender, and weight. Eighteen percent of the males (two participants) and 20% of the females (two participants) fell below the LLN for MEP.

Table 12. Descriptive Statistics for Pulmonary Function and Respiratory Muscle Strength Measures

Respiratory Parameter	ALL		Males		Females		p-value
	Mean (SD)	Median (IQ)	Mean (SD)	Median (IQ)	Mean (SD)	Median (IQ)	
FVC (L)	3.21 (1.07)	2.97 (1.55)	3.80 (1.15)	3.83 (1.49)	2.56 (0.43)	2.45 (0.69)	**0.013***
FVC %	89.81 (23.81)	85.00 (32.00)	87.73 (24.08)	85.00 (23.00)	92.10 (24.56)	89.50 (45.00)	1.000
FEV1 (L)	2.34 (0.84)	2.06 (1.17)	2.77 (0.94)	2.84 (1.06)	1.87 (0.32)	1.85 (0.39)	**0.020***
FEV1 %	88.43 (26.38)	88.00 (31.00)	87.00 (27.46)	88.00 (30.00)	90.00 (26.52)	85.00 (38.00)	1.000
FEV1/FVC	0.73 (0.06)	0.73 (0.06)	0.72 (0.08)	0.73 (0.06)	0.73 (0.05)	0.73 (0.06)	0.756
FEV1/FVC %	97.38 (8.15)	98.00 (9.00)	97.55 (9.77)	100.00 (6.00)	97.20 (6.43)	95.50 (8.00)	0.426
MIP (cmH_20)	82.67 (32.77)	80.00 (54.00)	93.73 (36.95)	107.00 (59.00)	70.50 (23.58)	64.00 (25.00)	0.085
MEP (cmH_20)	118.19 (36.48)	114.00 (58.00)	141.18 (34.89)	147.00 (57.00)	92.90 (15.78)	89.50 (15.00)	**0.003***

*Mann-Whitney U test significant at an alpha level of 0.05
SD=standard deviation; IQ=interquartile range; FVC=forced vital capacity; FEV1=forced expiratory volume in one second; MIP=maximal inspiratory pressure; MEP=maximal expiratory pressure

Laryngeal Features

Laryngeal features were rated using the VALI form. Three participants (males) had a complete glottal closure, three (males) had an anterior gap, three (2 females, 1 male) had a posterior gap, one (female) had an hourglass gap, and eight (5 females, 3 males) had a spindle gap. Regarding vertical level of approximation, nine participants (3 females, 6 males) had vocal folds on-plane and nine (5 females, 4 males) had vocal folds off-plane. Twenty participants (10 females, 10 males) had a concave free edge contour of the right vocal fold, while one (male) had a convex free edge. Sixteen participants (8 females, 8 males) had a concave free edge of the left vocal fold, while three

(1 female, 2 males) had a convex edge and one (male) had a normal edge. Only one participant (male) had non-vibrating segments of tissue on their vocal folds; however, the small size of the area was not deforming the vocal fold and was not clinically significant, which is why the participant could still be included in the study. Lastly, all participants had a predominantly open phase.

Amplitude of vibration, mucosal wave, supraglottic activity (anteroposterior and mediolateral), phase symmetry, and regularity were not normally distributed across the sample. Means and medians and measures of variability are presented in Table 13. Results were not different across males and females, as indicated by the results from the Mann-Whitney U test.

Table 13 also displays the results for bowing index, which were not significantly different between males and females. Bowing index was normally distributed across the sample and the mean was 8.80, with a standard deviation of 4.19.

Table 13. Descriptive Statistics for Laryngeal Features

Laryngeal Feature	All		Males		Females		p-value
	Mean (SD)	Median (IQ)	Mean (SD)	Median (IQ)	Mean (SD)	Median (IQ)	
Amplitude of vibration (right)	57.50 (12.38)	60.00 (10.00)	62.00 (11.35)	60.00 (0)	50.00 (10.95)	50.00 (20.00)	0.118
Amplitude of vibration (left)	56.67 (17.15)	60.00 (40.00)	58.00 (14.76)	60.00 (20.00)	55.00 (20.70)	40.00 (40.00)	0.633
Amplitude of mucosal wave (right)	57.50 (20.49)	60.00 (40.00)	64.00 (15.78)	60.00 (20.00)	46.67 (24.22)	50.00 (40.00)	0.181
Amplitude of mucosal wave (left)	54.71 (26.01)	60.00 (40.00)	62.00 (22.01)	70.00 (40.00)	44.29 (29.36)	40.00 (60.00)	0.230
Supraglottic activity (anteroposterior)	1.52 (0.98)	2.00 (1.00)	1.64 (0.92)	2.00 (1.00)	1.40 (1.07)	2.00 (2.00)	0.705
Supraglottic activity (mediolateral)	2.67 (1.24)	3.00 (3.00)	2.55 (1.13)	3.00 (2.00)	2.80 (1.40)	3.5 (3.00)	0.512
Phase symmetry	66.11 (27.47)	75.00 (40.00)	70.00 (28.28)	85.00 (40.00)	61.25 (27.48)	70.00 (35.00)	0.360
Regularity of vibration	69.47 (30.64)	90.00 (40.00)	74.55 (28.41)	90.00 (30.00)	62.50 (34.12)	80.00 (60.00)	0.545
Bowing Index	8.80 (4.19)	8.28 (3.99)	10.07 (5.10)	8.71 (6.02)	7.52 (2.74)	7.87 (2.94)	0.393

SD=standard deviation; IQ=interquartile range.

Acoustic Measures and Auditory-Perceptual Judgments of Voice Quality

Voice quality measures, including acoustic measures and CAPE-V overall severity rating, are reported in Table 14.

Results from the Shapiro-Wilk test revealed a normal distribution for three of the five acoustic measures: SPL during running speech, CPPS during /a/, and CPPS during reading. Mean SPL for the total sample was 71.7 dB (SD=4.5); mean CPPS during /a/ was 21.28 dB (SD=3.32); and mean CPPS during reading was 16.29 dB (SD=1.46). NHR and APQ during /a/ were not normally distributed

across the sample. The median for NHR /a/ was 0.16, and the interquartile range was 0.05. The median for APQ /a/ was 4.06%, and the interquartile range was 2.53%.

Results from the Shapiro-Wilk test revealed a normal distribution for the CAPE-V parameter of overall severity. The mean for the total sample was 45.57 (SD=23.80) on a visual analog scale (VAS) of 100 mm.

Results from the Mann-Whitney U test indicated that the values of the acoustic measures were similar in males and females, since no significant differences were found. However, the parameter of overall severity almost reached significance (p=0.051), with the females having a higher mean than the males. Table 14 reports the acoustic values for the total sample as well as for males and females separately, and the p-values from the Mann-Whitney U test.

Table 14. Descriptive Statistics for Voice Quality Measures

Acoustic Parameter	ALL		Males		Females		p-value
	Mean (SD)	Median (IQ)	Mean (SD)	Median (IQ)	Mean (SD)	Median (IQ)	
SPL speech (dB)	71.7 (4.5)	71.20 (6.0)	72.13 (2.70)	71.2 (3.9)	71.23 (6.03)	71.3 (8.7)	0.605
CPPS /a/ (dB)	21.28 (3.32)	21.78 (3.73)	21.30 (4.23)	21.78 (6.78)	21.27 (2.14)	21.81 (2.56)	0.756
CPPS reading (dB)	16.29 (1.46)	16.15 (2.25)	16.04 (1.14)	16.15 (1.0)	16.56 (1.78)	16.46 (3.09)	0.705
NHR /a/	0.19 (0.11)	0.16 (0.05)	0.20 (0.15)	0.14 (0.06)	0.19 (0.05)	0.18 (0.05)	0.152
APQ /a/ (%)	5.39 (4.47)	4.06 (2.53)	6.20 (5.99)	3.53 (3.48)	4.51 (1.69)	4.08 (1.77)	0.809
Overall Severity (100 mm VAS)	45.57 (23.80)	43.00 (42.00)	36.36 (26.15)	33.00 (39.00)	55.70 (16.77)	59.00 (25.00)	0.051

SD=standard deviation; IQ=interquartile range; SPL=sound pressure level; CPPS=smoothed cepstral peak prominence; NHR=noise-to-harmonic ratio; APQ=amplitude perturbation quotient; VAS=visual analog scale.

Aerodynamic Measures

Results for aerodynamic measures are presented in Table 15.

Aerodynamic measures were not normally distributed across the sample. The median for airflow during voicing was 0.17 L (IQ range=0.16). The median for subglottal pressure (mean peak air pressure) was 6.47 cmH_20 (IQ range=2.46). The median for aerodynamic resistance was 35.36 $cmH_20/(l/s)$ (IQ range=27.02). No significant difference between males and females were found for the aerodynamic parameters.

Table 15. Descriptive Statistics for Aerodynamic Measures

Aerodynamic Parameter	All		Males		Females		p-value
	Mean (SD)	Median (IQ)	Mean (SD)	Median (IQ)	Mean (SD)	Median (IQ)	
Mean airflow during voicinG (L/s)	0.21 (0.15)	0.17 (0.16)	0.25 (0.18)	0.23 (0.23)	0.15 (0.08)	0.14 (0.05)	0.201
Subglottal pressure (cmH_20)	7.15 (3.57)	6.47 (2.46)	7.20 (4.21)	5.87 (2.71)	7.07 (2.56)	7.00 (2.06)	0.791
Aerodynamic resistance ($cmH_20/(l/s)$)	43.93 (30.98)	35.36 (27.02)	36.67 (21.83)	30.17 (25.96)	55.34 (40.96)	47.04 (52.47)	0.246

*Mann-Whitney U test significant at an alpha level of 0.05
SD=standard deviation; IQ=interquartile range.

Self-Assessments Measures

Table 16 displays the results for the three self-assessment questionnaires administered. Total scores for Voice Handicap Index-10 (VHI-10), Glottal Function Index (GFI), and Communicative Participation Item Bank (CPIB) were not different across males and females, as indicated by the results of the Mann-Whitney U test. The scores of the three questionnaires were normally distributed across the total sample of participants. The mean score for VHI-10 was 19 (SD=8.14). The mean score for GFI was 10.90 (SD=3.71). The mean score for CPIB was 17.65 (SD=7.88).

Table 16. Descriptive Statistics for Self-Assessment Measures

Self-assessment questionnaire	All		Males		Females		p-value
	Mean (SD)	Median (IQ)	Mean (SD)	Median (IQ)	Mean (SD)	Median (IQ)	
VHI-10	19 (8.14)	20.00 (10.00)	17.27 (7.30)	17.00 (9.00)	20.90 (8.96)	25.00 (9.00)	0.173
GFI	10.90 (3.71)	12.00 (6.00)	10.91 (3.56)	13.00 (7.00)	10.90 (4.07)	11.5 (5.00)	0.809
CPIB	17.65 (7.66)	17.50 (9.00)	16.45 (8.15)	15.00 (10.00)	19.11 (7.20)	19.00 (9.00)	0.331

SD=standard deviation; IQ=interquartile range; VHI-10=Voice Handicap Index-10; GFI=Glottal Function Index; CPIB=Communicative Participation Item Bank.

Respiratory Factors

Two distinct factors emerged from the factor analysis conducted on all eight respiratory variables. Factor 1 included: MIP, MEP, FVC, and FEV1. Factor 2 included FEV1%, FEV1/FVC, and FEV1/FVC%. FVC% was the only respiratory parameter to not make it into one of the factors. Factor 1 was named “raw respiratory” because it included the raw measures for respiratory muscle strength, FVC, and FEV1. Factor 2 was named “standard respiratory” because it included percent predicted values, as well as the FEV1/FVC ratio. The factor loading scores for each respiratory variable are displayed in Table 17 (only variables with a loading score above the 0.71 cut-off value were included).

Table 17. Respiratory Factors

Respiratory Parameter	Factor 1 (Raw respiratory)	Factor 2 (Standard respiratory)
MIP	0.89	
MEP	0.88	
FVC	0.93	
FVC%	-	-
FEV1	0.87	
FEV1%		0.80
FEV1/FVC		0.88
FEV1/FVC%		0.93

Voice Factors

Firstly, a factor analysis was conducted on all voice measures (except for the laryngeal parameters) (Table 18). Factor 1, representing the construct of perceived handicap, included scores from the three self-assessment questionnaires: VHI-10, GFI, and CPIB. Factor 2 included NHR and APQ during a sustained /a/ and was therefore named the "perturbation" factor. Overall severity and CPPS for sustained /a/ were not included in this factor (or any factor), because the loading scores were below the cut-off value of 0.71. Factor 3 included mean airflow during voicing and mean peak air pressure (subglottal pressure) and was named the "aerodynamic" factor. Factor 4 included SPL during running speech and CPPS during reading and was named "speech". Lastly, factor 5 included only the variable of aerodynamic resistance and was therefore named "resistance". The factor loading values for each variable are presented in Table 18.

A second factor analyses was conducted on the laryngeal parameters from the VALI form and the bowing index measure. Three factors emerged from this analysis. The factor "pliability" included measures of amplitude of vibration, mucosal wave amplitude, and regularity of vibration. The factor "integrity" included phase symmetry and bowing index (multiplied by -1 to obtain a negative value, because a lower bowing index is indicative of a greater vocal fold integrity). Therefore, a higher score on the integrity factor would be indicative of a greater phase symmetry and less bowing of the vocal folds. Lastly, the factor "hyperfunction" included both types of supraglottic activity: anteroposterior and mediolateral. Factor loading scores are presented in Table 19.

Table 18. Voice Factors

Voice Parameter	Factor 1 (handicap)	Factor 2 (perturbation)	Factor 3 (aerodynamics)	Factor 4 (speech)	Factor 5 (resistance)
VHI-10	0.82				
GFI	0.73				
CPIB	-0.89				
CPPS /a/	-	-	-	-	-
NHR /a/		0.97			
APQ /a/		0.98			
Airflow			0.80		
Subglottal pressure			0.93		
SPL speech				0.91	
CPPS reading				0.85	
Resistance					0.86
Overall severity	-	-	-	-	-

Table 19. Voice Factors (Laryngeal Parameters)

Laryngeal Parameter	Factor 1 (pliability)	Factor 2 (integrity)	Factor 3 (hyperfunction)
Amplitude right	0.88		
Amplitude left	0.80		
Mucosal wave right	0.86		
Mucosal wave left	0.83		
Supraglottic activity (anteroposterior)			0.91
Supraglottic activity (mediolateral)			0.80
Phase symmetry		0.88	
Regularity	0.82		
Bowing Index		-0.83	

Cluster Analysis

Three distinct clusters emerged from the cluster analysis including the two respiratory factors, the eight voice factors, as well as age and gender. The overall R-squared value for the cluster analysis was 0.63 (F=13.34; cubic clustering criterion=-0.69). Table 20 presents mean values on each factor for the three clusters. Since the factors were weighted measures of standardized values, these means do not have any specific units (with the exception of age, which remained on the original scale).

Mean age was the highest in cluster 1 (78.9 years old, SD=5.28) and the lowest in cluster 3 (60.40 years old, SD=3.91). Cluster 2 had a mean age of 69.17 years old (SD=4.26). Cluster 1 was comprised of five males and five females, cluster 2 was comprised of one male and five females, and cluster 3 was constituted of males exclusively (five males). In total, cluster 1 had 10 participants, cluster 2 had six, and cluster 3 had five. Table 21 presents the age and gender repartition in each cluster.

Cluster 1 and 3 (the oldest and youngest clusters) had a similar mean for the standard respiratory factor, while the raw respiratory value of cluster 1 was the highest of all three clusters. Cluster 2 had the lowest means for both respiratory factors. Cluster 2 also had the highest mean value for handicap, resistance, and hyperfunction, and a considerably lower value for the pliability factor. As for the aerodynamic factor, it was the highest in cluster 3. Cluster 3 also had the highest value for integrity of the vocal folds, and the lowest value for hyperfunction. Cluster 1 had the lowest value for integrity, and a slightly higher value for perturbation.

Table 20. Factor Means by Cluster

Cluster	Raw respiratory	Standard respiratory	handicap	Pertur-bation	Aero	Speech	Resistance	Hyper-function	Integrity	Pliability	Age
1	3.98	5.42	4.47	**5.45**	**4.44**	5.24	4.88	4.69	**4.20**	6.07	78.9
2	2.89	3.55	**6.14**	4.42	4.80	4.99	**5.63**	**6.02**	5.30	**0.84**	69.17
3	9.57	5.90	4.40	4.80	**6.31**	4.54	4.57	4.38	**6.26**	6.64	60.4

Table 21. Age and Gender Repartition in Clusters

	Males	Females	Total	Age (SD)
Cluster 1	5	5	10	78.9 (5.28)
Cluster 2	1	5	6	69.17 (4.26)
Cluster 3	5	0	5	60.4 (3.91)
Total	11	10	21	71.71 (8.94)

Table 22. Correlations Between Age and Respiratory and Voice Factors

	Perturbation	Aero	Speech	Resistance	Handicap	Integrity	Pliability	Hyper function	Raw Respiratory	Standard Respiratory
age (r)	0.24	-0.49	0.08	0.15	-0.09	-0.56	-0.10	-0.03	-0.43	0.02
p-value	0.304	**0.039****	0.730	0.544	0.700	**0.020****	0.710	0.898	**0.050****	0.918
N	21	18	21	18	20	17	16	21	21	21

**Significant at an alpha level of 0.05

Table 23. Correlations Between Factors

Factor	Perturbation	Aero	Speech	Resistance	Handicap	Integrity	Pliability	Hyper function	Raw Respiratory	Standard Respiratory
Perturbation (r) **p-value** **N**	1.00000 21									
Aero	-0.18885 0.4529 18	1.00000 18								
Speech	-0.44675 **0.0423**** 21	0.08153 0.7478 18	1.00000 21							
Resistance	-0.02993 0.9062 18	-0.44066 **0.0672*** 018	0.00722 0.9773 18	1.00000 18						
Handicap	-0.09624 0.6865 20	-0.01225 0.9628 17	0.05714 0.8109 20	0.70098 **0.0017**** 17	1.00000 20					
Integrity	-0.43873 **0.0781*** 17	0.66071 **0.0073**** 15	0.12990 0.6192 17	-0.12143 0.6664 15	-0.20000 0.4577 16	1.00000 17				
Pliability	0.09131 0.7366 16	0.06381 0.8284 14	0.23122 0.3889 16	-0.62266 **0.0174**** 14	-0.28086 0.3106 15	-0.03038 0.9144 15	1.00000 16			
Hyperfunction	-0.01373 0.9529 21	-0.13736 0.5868 18	0.22092 0.3359 21	0.26120 0.2951 18	0.40365 **0.0776*** 20	-0.14207 0.5865 17	0.25836 0.3340 16	1.00000 21		
Raw respiratory	-0.16234 0.4820 21	0.12693 0.6157 18	-0.29351 0.1966 21	-0.40970 **0.0913*** 18	-0.27970 0.2323 20	0.25000 0.3332 17	0.19293 0.4741 16	-0.15294 0.5081 21	1.00000 21	
Standard respiratory	0.13766 0.5518 21	-0.14345 0.5701 18	0.28312 0.2136 21	-0.36636 0.1348 18	-0.53985 **0.0140**** 20	-0.07843 0.7648 17	0.66274 **0.0051**** 16	0.24772 0.2790 21	0.24805 0.2783 21	1.00000 21

Correlations Between Respiratory Factors, Voice Factors, and Age

Spearman correlations were computed to better understand the relationships between the respiratory and voice factors, as well as between age and all the factors. Correlation coefficients and corresponding p-values are presented in Tables 22 and 23.

Significant correlations were found between the standard respiratory factor and two voice factors: handicap (r=-0.54, p=0.014) and pliability (r=0.66 p=0.005). The raw respiratory factor was negatively correlated with the resistance factor and the relationship was marginally significant (r=-0.41, p=0.091). Significant or marginally significant correlations were also found between the voice factors: perturbation and speech (r=-0.45, p=0.042); perturbation and integrity (r=-0.44, p=0.078); aerodynamic and resistance (r=-0.44, p=0.067); aerodynamic and integrity (r=0.66, p=0.007); resistance and handicap (r=0.70, p=0.002); resistance and pliability (r=-0.62, p=0.017); and handicap and hyperfunction (r=0.40, p=0.078).

Moreover, age was found to be significantly correlated with three factors: aerodynamic (r=-0.49, p=0.039), integrity (r=-0.56, p=0.020), and raw respiratory (r=-0.43, p=0.050). The direction of the correlations showed that a younger age was correlated with a higher aerodynamic factor (airflow and subglottal pressure), a higher integrity factor (greater phase symmetry and less bowing of the vocal folds), and a higher raw respiratory factor (higher MIP, MEP, FVC, and FEV1). These correlations were expected.

Linear Regressions Models

Linear regressions were computed based on the results from the correlations, between the respiratory variables and the voice factors that showed a moderate to strong correlation and a significant or marginally significant relationship ($p \leq 0.1$) with the respiratory factors.

Handicap Factor

Among all the respiratory variables, FVC% and FEV1% were the strongest predictors for the handicap factor (p=0.006 and p=0.007, respectively). For an increase of one standard deviation in FVC% or FEV1%, an associated decrease of 0.60 and 0.58 standard deviations in handicap factor score was observed, respectively. Age, sex, and height were tested in the model and were not found to be significant confounders (Figure 13, table 24).

Resistance Factor

The strongest respiratory predictor for the resistance factor were FEV1 and FVC (Figure 14, Table 25). Although they did not quite reach significance (p=0.106 and p=0.109, respectively), scatter plots show an inverse relationship with resistance for both FEV1 and FVC. Because the resistance factor was comprised of only one variable, aerodynamic resistance, scatter plots were drawn using the raw values for this variable instead of the standardized values from the factor. This allowed for easier interpretation of the results. An increase of one standard deviation in FEV1 (0.84 L) or in FVC (1.07 L) was associated with a decrease in aerodynamic resistance of 0.39 standard deviation, which corresponds to approximately 12 $cmH_20/l/s$, in this sample. Age, sex, and height were tested in the model and were not found to be significant confounders. The direction of the relationship was the same for percent predicted values, but with smaller beta-weight values and higher p-values.

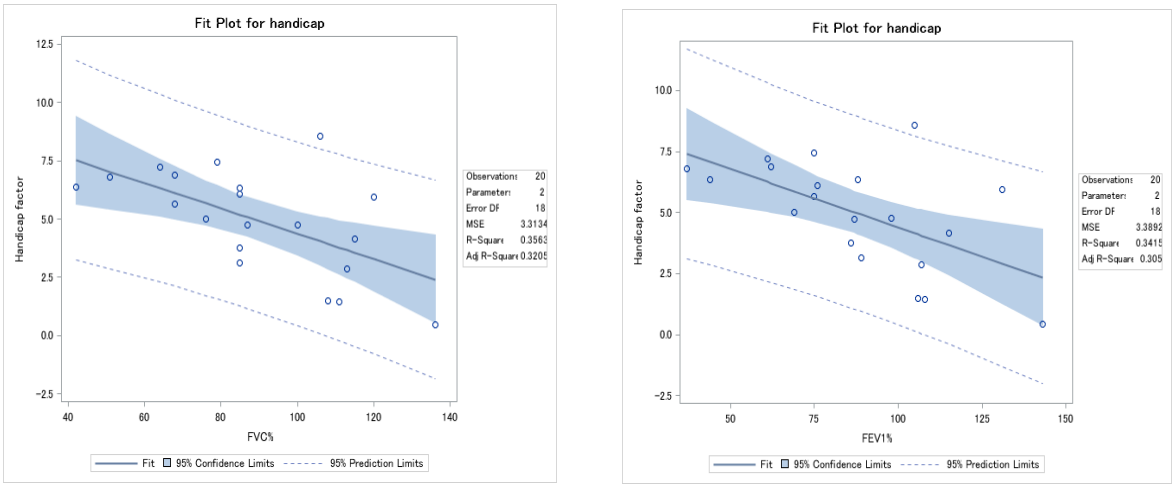


Figure 13. Scatter plots and linear regression fitted-lines for handicap factor (y-axis) and FVC% (x-axis) (left) and FEV1% (x-axis) (right).

Table 24. Linear Regression Outcomes for the Handicap Factor. (Above: results for FVC% as the respiratory predictor; below: results for FEV1% as the respiratory predictor).

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	9.80	1.59	0	6.17	<0.001	9.96 (0.006)	0.36	0.32
FVC%	-0.05	0.02	-0.60	-3.16	**0.006****			

**Significant at an alpha level of 0.05

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	9.16	1.44	0	6.37	<0.001	9.34 (0.007)	0.34	0.31
FEV1%	-0.05	0.02	-0.58	-3.06	**0.007****			

**Significant at an alpha level of 0.05

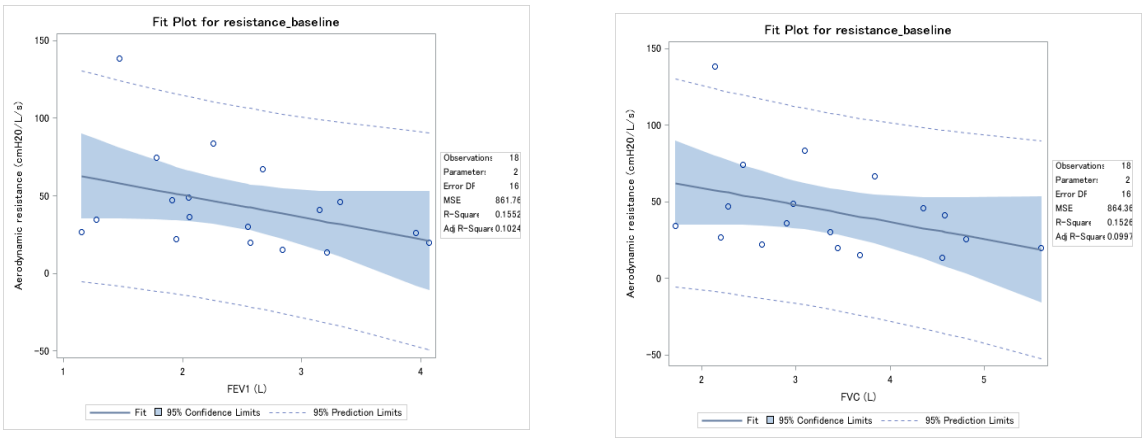


Figure 14. Scatter plots and linear regression fitted-lines for aerodynamic resistance (y-axis) and FEV1 (x-axis) (left) and FVC (x-axis) (right).

Table 25. Linear Regression Outcomes for Aerodynamic Resistance. (Above: results for FEV1 as the respiratory predictor; below: results for FVC as the respiratory predictor).

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	6.07	0.66	0	9.24	<0.001	2.94 (0.106)	0.16	0.10
FEV1	-0.43	0.25	-0.39	-1.71	0.106			

**Significant at an alpha level of 0.05

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	6.14	0.70	0	8.71	<0.001	2.88 (0.109)	0.15	0.10
FVC	-0.34	0.20	-0.39	-1.70	0.109			

**Significant at an alpha level of 0.05

A regression model was built with other voice factors as predictors to better understand what voice parameters impact resistance. The best explanatory model included the pliability factor (encompassing amplitude of vibration and of mucosal wave, as well as regularity of vibration) and the hyperfunction factor (encompassing anteroposterior and mediolateral supraglottic activity). The pliability factor had a beta-weight β of -0.65 (p=0.012), indicating that an increase in one standard deviation on this factor was associated with a decrease in 0.65 standard deviation in resistance. The hyperfunction factor had a beta-weight of 0.38 (p=0.102), indicating that an increase of one standard deviation on this factor was associated with an increase of 0.38 standard deviation in resistance (Table 26).

Figure 15 shows a scatter plot of the relationship between raw values of aerodynamic resistance and the hyperfunction factor. The plot shows that the four data points corresponding to the highest aerodynamic resistance values were all associated with an elevated degree of hyperfunction. These four points also seemed to play an important role in driving the relationship between resistance and handicap (Figure 16, Table 27). On the other hand, the presence of hyperfunction was not necessarily associated with an elevated aerodynamic resistance (Figure 15).

Table 26. Linear Regression Outcomes for Aerodynamic Resistance with Hyperfunction and Pliability Factors as Predictors

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	4.59	0.50	0	9.13	<0.001	5.52 (0.022)	0.50	0.41
Hyperfunction	0.18	0.10	0.38	1.78	0.102			
Pliability	-0.11	0.03	-0.65	-3.01	**0.012****			

**Significant at an alpha level of 0.05

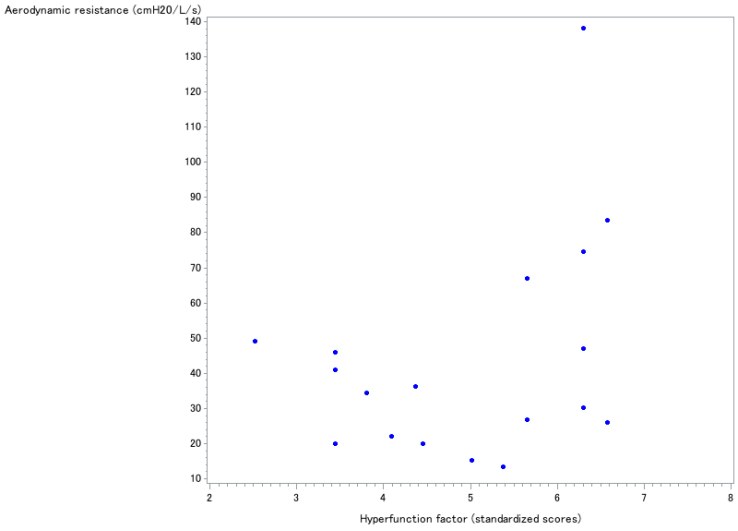


Figure 15. Scatter plot of aerodynamic resistance (y-axis) by handicap (x-axis)

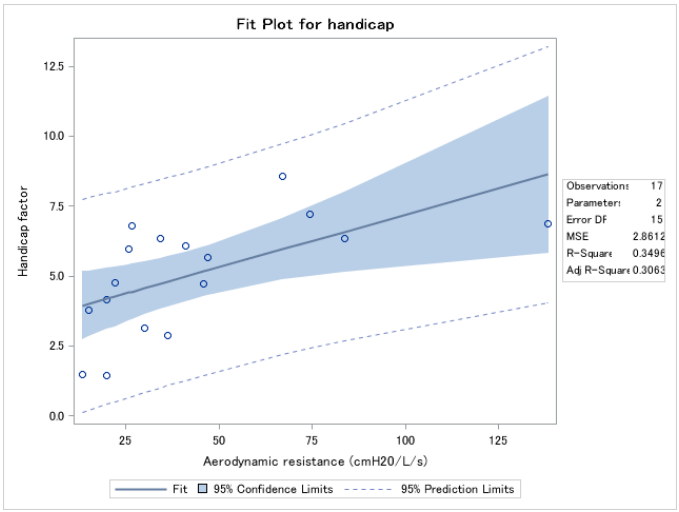


Figure 16. Scatter plot and linear regression fitted-line for handicap factor (y-axis) and aerodynamic resistance (x-axis)

Table 27. Linear Regression Outcomes for Handicap Factor with Aerodynamic Resistance as a Predictor

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	3.43	0.71	0	4.83	<0.001	8.06 (0.012)	0.35	0.31
Aerodynamic resistance	0.04	0.01	0.59	2.84	**0.012****			

**Significant at an alpha level of 0.05

Pliability Factor

Although the pliability factor was strongly correlated with the standard respiratory factor, the strongest individual predictor was the raw respiratory parameter MEP (p=0.092) (Figure 17, Table 28). An increase of one standard deviation in MEP (36.48 cmH_20) was associated with an increase of 0.44 standard deviations in the pliability factor. Age, sex, and height were not found to be significant confounders and were therefore not included in the model because they did not improve its explanatory ability.

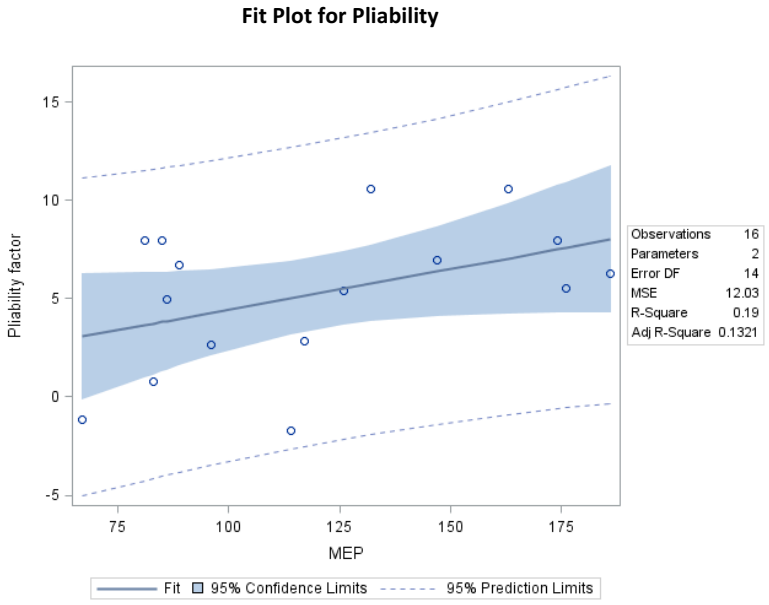


Figure 17. Scatter plot and linear regression fitted-line for pliability factor (y-axis) and MEP (x-axis)

Table 28. Linear Regression Outcomes for Pliability Factor with MEP as a Respiratory Predictor

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	0.27	2.89	0	0.09	0.93	3.28 (0.092)	0.19	0.13
MEP	0.04	0.02	0.44	1.81	**0.092***			

**Significant at an alpha level of 0.05

Results: Aim 3a

Results from Aim 2 revealed that respiratory function has an impact on various aspects of phonation and on voice-related handicap. The objective of aim 3a was to assess whether adding respiratory exercises to improve respiratory function would lead to improved voice outcomes. Results are presented below.

Participant Characteristics and Baseline Measures

Ten participants completed the intervention (five females, five males), aged 58 to 82 years old (mean=71.6, SD=7.68). Table 29 displays the characteristics of the participants (age, reflux symptom index scores, gender, smoking status, and professional voice use) for the total sample, as well as for each of the three intervention groups (IMST, EMST, and VFE). No significant difference between the groups were found for mean age and RSI scores. Chi-square values could not be computed for categorical variables because of the small sample size.

Table 29. Comparison of Groups Demographics at Baseline

	Total sample (N=10)	**IMST (N=4)**	**EMST (n=3)**	**VFE (n=3)**	**p-value**
Age **Mean (SD)**	72.33 (7.64)	67.50 (8.74)	72.33 (7.64)	76.33 (4.93)	0.354
RSI **Mean (SD)**	16.3 (6.91)	19.75 (3.59)	14.67 (7.09)	13.44 (10.21)	0.477
Gender **Females** **Males**	 5 5	 3 1	 1 2	 1 2	-
Smoking status **Never** **Quit**	 5 5	 1 3	 2 1	 2 1	-
Professional voice use **Yes** **No**	 5 5	 3 1	 0 3	 1 2	-

Table 30 presents the baseline means and standard deviations for respiratory, acoustic, auditory-perceptual, aerodynamic, and self-assessment measures. The values are provided for the total sample of participants who received the intervention, as well as for each intervention group separately. Values were compared across groups with a one-way ANOVA and the results confirmed that no significant differences were present between the groups. Results from the Shapiro-Wilk test indicated normal distributions for all variables except for FEV1/FVC (p=0.006), FEV1/FVC% (p=0.009), NHR /a/ (p=0.01), and APQ /a/ (p=0.03).

Table 30. Comparison of Respiratory and Voice Outcomes at Baseline Between Groups (one-way ANOVA)

Outcome Measure	Total sample Mean (SD)	IMST Mean (SD)	EMST Mean (SD)	VFE Mean (SD)	p-value
MIP (cmH_20) **Females** **Males**	79.70 (22.33) 70 (18.28) 89.40 (23.50)	88.50 (26.15)	75.33 (27.43)	72.33 (14.15)	0.644
MEP (cmH_20) **Females** **Males**	118.5 (38.59) 86.00 (13.06) 151.00 (23.22)	109.75 (46.55)	128.67 (44.06)	120.00 (34.60)	0.847
FVC (L) **Females** **Males**	3.01 (0.76) 2.53 (0.31) 3.48 (0.80)	2.97 (0.98)	2.78 (0.79)	3.28 (0.60)	0.770
FVC (%) **Females** **Males**	83.5 (17.16) 84.20 (16.35) 82.80 (19.85)	83.00 (15.85)	70.67 (17.62)	97.00 (10.82)	0.174
FEV1 (L) **Females** **Males**	2.15 (0.67) 1.80 (0.21) 2.51 (0.81)	2.17 (0.79)	1.88 (0.87)	2.39 (0.39)	0.704
FEV1 (%) **Females** **Males**	80.20 (20.26) 79.60 (16.24) 80.80 (25.67)	79.50 (14.82)	64.00 (24.88)	97.33 (8.02)	0.122
FEV1/FVC **Females** **Males**	0.71 (0.07) 0.71 (0.03) 0.70 (0.11)	0.73 (0.03)	0.65 (0.13)	0.73 (0.03)	0.355
FEV1/FVC (%) **Females** **Males**	94.6 (8.81) 94.40 (2.70) 94.80 (12.93)	95.75 (3.10)	87.67 (14.30)	100.00 (3.46)	0.291[a]
SPL speech(dB)	72.00 (5.22)	71.33 (8.12)	70.00 (1.56)	74.90 (1.25)	0.544
CPPS /a/	21.72 (2.72)	21.92 (1.05)	20.11 (1.26)	23.05 (4.84)	0.277[a]
CPPS reading(dB)	16.52 (1.52)	16.79 (1.28)	15.36 (1.93)	17.32 (1.05)	0.286
NHR /a/	0.15 (0.06)	0.15 (0.02)	0.16 (0.12)	0.14 (0.02)	0.812[a]
APQ /a/ (%)	3.90 (1.81)	3.09 (0.97)	4.77 (3.16)	4.12 (0.89)	0.423[a]
Overall severity **Females** **Males**	44.80 (22.73) 54.60 (19.60) 35.00 (23.21)	43.00 (28.58)	52.00 (30.41)	40.00 (7.55)	0.832[a]
Mean airflow (L)	0.18 (0.10)	0.12 (0.05)	0.26 (0.09)	0.16 (0.13)	0.199
Subglottal pressure (cmH_20)	6.69 (2.73)	8.66 (3.26)	6.47 (1.44)	4.87 (1.88)	0.261
Aerodynamic resistance ($cmH_20/l/s$)	41.35 (21.48)	56.52 (15.69)	21.01 (8.15)	39.73 (23.94)	0.202
VHI-10	21.00 (6.41)	24.00 (6.06)	17.00 (5.57)	21.00 (6.41)	0.408[b]
GFI	11.5 (3.44)	12.50 (3.11)	11.33 (3.79)	10.33 (4.51)	0.775[b]
CPIB	13.78 (6.28)	12.00 (6.08)	15.00 (4.00)	14.33 (9.87)	0.754[b]
Bowing Index	9.74 (2.90)	9.27 (1.92)	8.05 (0.53)	12.06 (4.37)	0.320[a]

[a]Result of the Welch test (when the assumption of homogeneity of variances was not met)
[b]Result from the Kruskal-Wallis test for ordinal data

The following sections describe the pre- to post-intervention changes on the respiratory, laryngeal, acoustic, auditory-perceptual, aerodynamic, and self-assessment measures. The various tasks and outcome measures that were recorded prior to and following the 4-week intervention period are summarized in Appendix I.

Results from the Interventions: Respiratory Measures

Prior to treatment, all participants except for one presented with a MIP and a MEP value above the lower limit of normal (LLN), as calculated with Enright's reference equations (Enright, Kronmal, Manolio, et al., 1994). One participant in the IMST group (subject 16) presented with a MEP value lower than the LLN.

Table 31 presents the pre- and post-treatment values for each participant for MIP and MEP. Two out of four participants in the IMST group improved their MIP, with a change greater than the SEM (as calculated in Aim 1) and with large effect sizes (SMD=0.87 and 0.94). One participant in the EMST group and one participant in the VFE group also improved their MIP with large and very large effect sizes (SMD=0.96 and 1.23, respectively). Two out of three participants in the EMST group improved their MEP with moderate to large effect sizes (SMD=0.66 and 1.11). In both cases, the change was greater than the minimum difference (MD) calculated for this outcome measure in Aim 1 (31.49 cmH_20 for females and 15.22 cmH_20 for males). One participant in the IMST group also improved their MEP with a change above MD, although the SMD was small (SMD=0.38). Two participants in the VFE group also improved their MEP considerably, with moderate to large effect sizes (SMD=0.62 and 0.98).

Table 33 presents the group effect sizes and the ANOVA results for respiratory outcomes. Interestingly, the VFE group increased both their MIP and MEP with moderate effect sizes

(d_w=0.60 and d_w=0.68, respectively). An increase in MIP, although slightly smaller, was also found in the IMST group, with an effect size of d_w=0.46. The smallest increase in MIP was observed in the EMST group (d_w=0.30). As for MEP, the largest increase was observed in the EMST group as expected (d_w=0.98), followed by the VFE group (d_w=0.68) and lastly the IMST group (d_w=0.29). A significant main time effect was found for MEP (p=0.012). A larger sample size may have allowed for detection of a time by group interaction effect.

Table 32 presents the pre- and post-treatment changes in raw and percent predicted pulmonary function values. No meaningful improvement in pulmonary function was noted for any of the participants following the intervention. In all three groups, the effect sizes for pulmonary function measures were small (below SMD=0.29). In addition, one participant in the IMST group (participant 16) presented with negative SMDs for all pulmonary function parameters (ranging from |SMD|=0.50 to 0.79), indicating a decrease in FVC, FEV1, and FEV1/FVC (raw and percent predicted values). One participant in the EMST group (participant 10) presented with large negative SMDs for the change in raw value and percent predicted value of FEV1/FVC (|SMD|=1.43 and 1.59, respectively). These two participants presented with particularly low pulmonary function values at baseline for FVC and FEV1, ranging between 37% and 64% of predicted values.

Group effect sizes and results from the two-way ANOVA confirmed the lack of change in pulmonary function measures in all three group. No main effect or interaction effect were found for any of the pulmonary variables. In addition, group effect sizes were small for all groups and all variables (Table 33).

Table 31. Change in Respiratory Muscle Strength (Single-Subject Results)

ID	Group	MIP (cmH_20)				MEP (cmH_20)			
		Pre	Post	Change	SMD_2	Pre	Post	Change	SMD_2
2	IMST	97	78	-19	-0.38	96	100	4	0.39
9	IMST	119	156	37**	0.87	176	221	45**	0.38
12	IMST	81	106	25*	0.94	100	102	2	-0.36
16	IMST	57	55	-2	0.11	67	61	-6	-0.17
10	EMST	60	57	-3	-0.04	126	180	54**	0.66
13	EMST	59	89	30*	0.96	86	143	57**	1.11
20	EMST	107	100	-7	-0.72	174	176	2	0.09
14	VFE	56	82	26*	1.23	81	109	28**	0.62
17	VFE	80	82	2	0.43	132	169	37*	0.98
18	VFE	81	93	12*	0.25	147	161	14	0.38

*Change greater than SEM

**Change greater than MD

MIP=maximum inspiratory pressure; MEP=maximum expiratory pressure; SMD_2=standard mean difference (two baseline and two follow-up data points were available)

Table 32. Change in Pulmonary Function (Single-Subject Results)

ID	Group	FVC Raw values (L) Percent predicted values (%)				FEV1 Raw values (L) Percent predicted values (%)				FEV1/FVC Raw values (L) Percent predicted values (%)			
		Pre	Post	Change	SMD	Pre	Post	Change	SMD	Pre	Post	Change	SMD
2	IMST	2.97 102	3.03 105	0.06 3	0.08 0.17	2.05 95	2.12 98	0.07 3	0.10 0.15	0.69 93	0.70 95	0.01 2	0.14 0.23
9	IMST	4.35 87	4.08 82	-0.27 -5	-0.36 -0.29	3.32 87	3.09 81	-0.23 -6	-0.34 -0.30	0.76 100	0.76 100	0 0	0 0
12	IMST	2.12 79	2.17 82	0.05 3	0.07 0.17	1.54 75	1.53 76	-0.01	-0.01 0.05	0.73 94	0.71 92	-0.02 -2	-0.29 -0.23
16	IMST	2.44 64	2.06 54	-0.38 -10	-0.5 -0.58	1.78 61	1.39 48	-0.39 -13	-0.58 -0.64	0.73 96	0.68 89	-0.05 -7	-0.71 -0.79
10	EMST	2.20 51	2.38 56	0.18 5	0.24 0.29	1.15 37	1.00 32	-0.15 -5	-0.22 -0.25	0.52 72	0.42 48	-0.10 -24	-1.43 -1.59
13	EMST	2.46 76	2.37 75	-0.09 -1	-0.12 -0.06	1.66 69	1.63 68	-0.03 -1	-0.04 -0.05	0.67 91	0.69 93	0.02 2	0.29 0.23
20	EMST	3.68 85	3.51 81	-0.17 -4	-0.21 -0.20	2.84 86	2.73 82	-0.11 -4	-0.14 -0.16	0.77 100	0.78 101	0.01 1	0.09 0.08
14	VFE	2.64 100	2.67 102	0.03 2	0.04 0.12	1.95 98	1.93 99	-0.02 1	-0.03 0.05	0.74 98	0.72 96	-0.02 -2	-0.29 -0.23
17	VFE	3.36 85	3.30 84	-0.06 -1	-0.08 -0.06	2.55 89	2.48 87	-0.07 -2	-0.10 -0.10	0.76 104	0.75 103	-0.01 -1	-0.14 -0.11
18	VFE	3.83 106	3.75 104	-0.08 -2	-0.11 -0.12	2.67 105	2.64 104	-0.03 -1	-0.04 -0.05	0.70 98	0.70 99	0 1	0 0.11

FVC=forced vital capacity; FEV1=forced expiratory volume in one second; SMD=standard mean difference

Table 33. Two-Way ANOVA Results for Respiratory Measures (Group Results)

Outcome Measure	IMST			EMST			VFE			F (p-value) Time effect Time*group Group effect
	Pre (SD)	Post (SD)	Diff. ES (d)	Pre (SD)	Post (SD)	Diff. ES (d)	Pre (SD)	Post (SD)	Diff ES (d)	
MIP (cmH_20)	88.50 (26.15)	98.75 (43.49)	10.25 d_w=0.46 d_b=-0.14	75.33 (27.43)	82.00 (22.34)	6.67 d_w=0.30 d_b=-0.30	72.33 (14.15)	85.67 (6.35)	13.34 d_w=0.60	2.288 (0.174) 0.077 (0.927) 0.400 (0.684)
MEP (cmH_20)	109.75 (46.55)	121.00 (69.29)	11.25 d_w=0.29 d_b=-0.39	128.67 (44.06)	166.33 (20.31)	37.66 d_w=0.98 d_b=0.29	120.00 (34.60)	146.33 (32.58)	26.33 d_w=0.68	**11.512 (0.012)**** 1.142 (0.372) 0.447 (0.657)
FVC (L)	2.97 (0.98)	2.84 (0.94)	-0.13 d_w=-0.17 d_b=-0.22	2.78 (0.79)	2.75 (0.66)	-0.03 d_w=-0.04 d_b=-0.09	3.28 (0.60)	3.24 (0.54)	0.04 d_w=0.05	1.337 (0.285) 0.402 (0.684) 0.310 (0.743)
FVC (%)	83.00 (15.85)	80.75 (20.87)	-2.25 d_w=-0.13 d_b=-0.11	70.67 (17.62)	70.67 (13.05)	0.00 d_w=0 d_b=0.02	97.00 (10.82)	96.67 (11.02)	-0.33 d_w=-0.02	0.294 (0.605) 0.214 (0.813) 2.124 (0.190)
FEV1 (L)	2.17 (0.79)	2.03 (0.77)	-0.14 d_w=-0.21 d_b=-0.15	1.88 (0.87)	1.79 (0.88)	-0.09 d_w=-0.13 d_b=-0.07	2.39 (0.39)	2.35 (0.37)	-0.04 d_w=-0.06	4.161 (0.081) 0.427 (0.668) 0.416 (0.675)
FEV1 (%)	79.50 (14.82)	75.75 (20.76)	-3.75 d_w=-0.19 d_b=-0.15	64.00 (24.88)	60.67 (25.79)	-3.33 d_w=-0.16 d_b=-0.13	97.33 (8.02)	96.67 (8.74)	-0.66 d_w=-0.03	2.665 (0.147) 0.366 (0.706) 2.691 (0.136)
FEV1/FEV	0.73 (0.03)	0.71 (0.03)	-0.02 d_w=-0.29 d_b=-0.14	0.65 (0.13)	0.63 (0.19)	-0.02 d_w=-0.29 d_b=-0.14	0.73 (0.03)	0.72 (0.03)	-0.01 d_w=-0.14	1.598 (0.247) 0.086 (0.919) 0.944 (0.434)
FEV1/FVC (%)	95.75 (3.10)	94.00 (4.69)	-1.75 d_w=-0.20 d_b=-0.12	87.67 (14.30)	84.00 (22.87)	-3.67 d_w=-0.42 d_b=-0.34	100.00 (3.46)	99.33 (3.51)	-0.67 d_w=-0.08	1.345 (0.284) 0.232 (0.799) 1.397 (0.309)

**Significant at an alpha level of 0.05

ES=effect size; d_w=Cohen's d within group; d_b=Cohen's d between groups; SD=standard deviation; MIP=maximum inspiratory pressure; MEP=maximum expiratory pressure; FVC=forced vital capacity; FEV1=forced expiratory volume in one second

Results from the Interventions: Laryngeal Features

Table 34 displays the change in bowing index for each of the participants. All but two participants for whom pre- and post-intervention data was available presented with a decrease in bowing. For the participants who did improve, effect sizes ranged from |SMD|=0.26 to |SMD|=3.07. Group analyses revealed that the largest mean reductions for bowing index occurred in the VFE group ($|d_w|$=1.19) and in the EMST group ($|d_w|$=0.68) (Table 35). Importantly, participant 17 in the VFE group improved considerably more than the others and strongly influenced the resulting group effect size.

Table 34. Change in Bowing Index (Single-Subject Results)

ID	**Group**	**Bowing Index**		
		Pre	Post	Change SMD
2	IMST	9.60	-	
9	IMST	11.27	-	
12	IMST	9.57	11.40	1.83 0.63
16	IMST	6.66	6.67	0.01 0.00
10	EMST	8.63	7.15	-1.48 -0.51
13	EMST	7.92	5.53	-2.39 -0.82
20	EMST	7.60	5.54	-2.06 -0.71
14	VFE	10.38	9.64	-0.74 -0.26
17	VFE	17.03	8.12	-8.9 -3.07
18	VFE	8.78	8.04	-0.75 -0.26

SMD=standard mean difference

Regarding laryngeal features, all participants had concave vocal fold edges and a predominantly open phase, both at baseline and after treatment. At baseline, five participants had a spindle-shaped glottal gap, two had a posterior gap, one had an anterior gap, one had complete closure, and one had unavailable data. No participant had a complete closure at post-treatment. No participant had non-vibrating segments.

Table 36 presents the following VALI features: vibration amplitude, mucosal wave amplitude, supraglottic activity, phase symmetry, and regularity of vibration. All but two participants had a vibration amplitude within a normal range (between 40% and 60%) at baseline. Participants 14 and 17 presented with an increased vibration amplitude (80%) for at least one vocal fold pre-treatment. These participants were both in the VFE group and both reduced their amplitude of vibration to a more normal value following the intervention. They also both reduced their mucosal wave amplitude on at least one vocal fold. Participant 16 (IMST group) had a reduced mucosal wave amplitude at baseline which increased following treatment, while participant 20 (EMST group) had an enhanced mucosal wave amplitude which did not change following treatment. Subject 13 (EMST group) increased their mucosal wave on one vocal fold and decreased on the other fold. Other participants had a normal mucosal wave at baseline or unavailable data.

All participants presented with a certain amount of supraglottic activity (hyperfunction), either anteroposterior or mediolateral or both, at baseline. In the IMST group, two participants decreased their hyperfunction while one participant increased and one had unavailable post-treatment data. In the EMST group, one participant slightly increased, one slightly decreased, and one did not change. In the VFE group, two participants decreased and one increased. In summary, the IMST and the VFE group had similar results regarding change in supraglottic activity, while the EMST group had the least amount of change.

Regarding phase symmetry, one participant in each of the intervention groups improved following treatment. The greatest improvements (+30% and +20%) were found in the IMST and the VFE group, respectively. However, this could be explained by the lower phase symmetry values at baseline, which may have left more room for improvement. As for regularity of vibration, the two participants with the lowest baseline values (40% and 60%) improved considerably following the intervention (in the IMST and EMST group, respectively).

Table 35. Two-Way ANOVA Results for Bowing Index (Group Results)

Outcome Measure	**IMST**			**EMST**			**VFE**			**F (p-value) Time effect Time*group Group effect**
	Pre (SD)	Post (SD)	Diff. ES (d)	Pre (SD)	Post (SD)	Diff. ES (d)	Pre (SD)	Post (SD)	Diff ES (d)	
Bowing Index	8.11 (2.06)	9.03 (3.35)	0.92 d_w=0.32 d_b=1.51	8.05 (0.53)	6.07 (0.93)	-1.98 **d_w=-0.68** d_b=0.51	12.06 (4.37)	8.60 (0.90)	-3.46 **d_w=-1.19**	1.884 (0.228) 1.247 (0.364) 2.338 (0.192)

ES=effect size; d_w=Cohen's d within groups; d_b=Cohen's d between groups; SD=standard deviation

Table 36. Change in Laryngeal Features (Single-Subject Results)

ID	Group	Amplitude (%) Right Left			Mucosal wave (%) Right Left			Supgraglottic activity Anteroposterior Mediolateral			Symmetry (%)			Regularity (%)		
		Pre	Post	Change	Pre	Post	Change	Pre	Post	Change	Pre	Post		Pre	Post	
2	IMST	40 40	40 40	0	40 60	40 40	-20	0 1	2 1	+2	60	90	+30	90	90	0
9	IMST	60 60	- -	-	60 40	- -	-	1 1	- -	-	100	-	-	90	-	-
12	IMST	- -	60 60	-	- -	60 40	-	3 4	2 3	-2	-	90	-	-	90	-
16	IMST	40 40	20 40	-20	20 20	- 80	+40	2 4	0 3	-3	80	70	-10	40	90	+50
10	EMST	60 60	60 60	0	60 60	60 80	+20	2 3	1 3	-1	80	90	+10	60	100	+40
13	EMST	60 40	40 40	+20	80 40	60 80	+20	0 1	0 2	+1	70	50	-20	70	50	-20
20	EMST	60 60	60 60	0	80 80	80 80	0	2 2	2 2	0	90	90	0	90	90	0
14	VFE	60 80	40 40	-60	60 80	20 60	-40	1 2	0 1	-2	90	50	-40	90	90	0
17	VFE	80 80	40 60	-60	80 80	60 80	-20	2 4	1 2	-3	50	70	+20	90	70	-20
18	VFE	60 40	- -	-	80 80	- -	-	2 3	5 3	+3	30	-	-	90	-	-

Results from the Interventions: Acoustic and Auditory-Perceptual Judgement of Voice Quality Measures

Table 37 presents the pre- and post-treatment acoustic measures for each participant. All participants showed a clinically meaningful improvement in at least one of the acoustic measures, except for one participant in the IMST group, one participantin the EMST group, and one participant in the VFE group, who didn't show any meaningful improvement.

Two participants in the IMST group increased their SPL during running speech with a moderate or large effect size (SMD=0.75 and 1.00). Another participant in the IMST group decreased their SPL, which was elevated at pre-treatment (from 82.4 dB to 80.3 dB; |SMD|=0.69). No meaningful changes in SPL were found for participants in the other intervention groups.

All participants increased (improved) their CPPS during reading to some extent, except for one participant in the EMST group (participant 20). These results were also reflected in the ANOVA results, which revealed a significant time effect (p=0.023). Although no time by group interaction were found, effect sizes were the greatest for the IMST (d_w=0.82) and the EMST (d_w=0.93) groups. As for the VFE group, the pre- to post-treatment effect size was small (d_w=0.28) and the mean difference was below the SEM. When the RMST groups were compared to the VFE group, the effect sizes were both medium although slightly superior for the EMST group (d_b=0.54 for IMST and d_b=0.66 for EMST). Group effect sizes and ANOVA results are displayed in Table 38.

As for measures of voice quality during a sustained /a/, one participant in the IMST group improved meaningfully for CPPS and APQ, and another one improved for NHR and APQ. Effect

sizes were moderate to large (|SMD|=0.58 to 0.81). In the EMST group, only one participant had a meaningful improvement on the sustained vowel: participant 13 improved their APQ and NHR, with very large effect sizes (|SMD|=1.59 and 3.67). This participant had perturbation measures considerably above normal values prior to treatment, which may have left more room for improvement and could explain the large effect sizes following the intervention. In the VFE group, one participant improved CPPS and APQ with large and very large effect sizes ((|SMD|=1.58 and 1.01), and one participant improved only APQ, with a moderate effect size (|SMD|=0.69). The two-way ANOVA and the group effect sizes revealed no differences between the groups for CPPS /a/ and NHR /a/. Moreover, the only meaningful effect size ($|d_w|$=0.72) was found within the VFE group, which improved on the APQ measure from pre- to post-treatment.

Table 39 presents the pre- and post-treatment values for perceptual judgment of overall severity, as rated on the CAPE-V form. The four participants in the IMST group had a decrease (improvement) in overall severity of voice quality, to different extents. Two of these participants had a decrease (improvement) beyond the MD calculated for this outcome (21.79 for females and 15.92 for males), which also corresponded to very large effect sizes (|SDM|=2.07 and 1.68). All participants in the EMST group also experienced a decrease (improvement) in overall severity. This decrease was important (beyond the MD) in one participant, with a moderate effect size (|SMD|=0.66). A similar scenario was observed in the VFE group, with one participant improving beyond the MD and with a large effect size (|SMD|= 1.03). In summary, improvements in overall severity of voice quality occurred in all three intervention groups, with at least one participant per group improving in a clinically meaningful way. These results were confirmed by the results from the ANOVA, which revealed a significant time effect. Although no significant time by group interaction was found, larger pre-treatment to post-treatment effect sizes were found in the IMST

group ($|d_w|$=0.97), followed by the EMST group ($|d_w|$=0.60), and lastly the VFE group ($|d_w|$=0.29). When compared to the VFE group, the IMST group had the largest effect size ($|d_b|$=0.67 versus $|d_b|$=0.31 for the EMST group). Group effect sizes and ANOVA results are displayed in Table 38.

Table 37. Change in Acoustic Measures (Single-Subject Results)

ID	Group	SPL speech (dB)			CPPS /a/ (dB)			CPPS reading (dB)			NHR /a/			APQ /a/ (%)		
		Pre	Post	Change SMD_2	Pre	Post	Change SMD_2	Pre	Post	Change SMD_2	Pre	Post	Change SMD_2	Pre	Post	Change SMD_2
2	IMST	62.9	70.9	8** 1.00	21.98	23.19	1.21 0.18	15.71	16.57	0.86* 0.33	0.15	0.13	-0.02 0.12	3.75	3.95	0.20 0.21
9	IMST	70.6	67.9	-2.7* -0.45	23.29	21.48	-1.81* -1.00	15.82	16.38	0.57 0.33	0.15	0.15	0 0.47	2.04	3.58	1.54 1.72
12	IMST	69.4	73.6	4.2* 0.75	21.65	22.17	0.53 0.35	17.22	20.01	2.79** 1.32	0.16	0.12	-0.04 -0.66	4.06	2.43	-1.63 -0.81
16	IMST	82.4	80.3	-2.1 -0.69	20.77	22.17	1.4 0.58	18.41	19.15	0.74* 0.06	0.12	0.10	-0.02 -0.46	2.50	1.69	-0.81 -0.62
10	EMST	71	73.1	2.1 0.31	18.85	20.32	1.47 0.26	14.58	16.07	1.49* 0.98	0.11	0.09	-0.02 0.12	3.53	2.63	-0.90 -0.23
13	EMST	68.2	70.3	2.1 0.23	20.10	19.93	-0.18 0.70	13.93	16.95	3.019** 1.53	0.30	0.22	-0.09 -3.67	8.36	7.71	-0.65 -1.59
20	EMST	70.8	70.2	-0.6 -0.03	21.38	22.278	0.90 0.37	17.55	17.306	-0.25 -0.41	0.08	0.138	0.061 1.01	2.41	2.04	-0.37 -0.12
14	VFE	73.9	72.8	-1.1 -0.25	17.48	22.17	4.70** 1.58	17.64	18.29	0.65* 0.18	0.15	0.13	-0.02 -0.13	4.66	2.32	-2.34 -1.01
17	VFE	74.5	75.5	1 0.19	26.06	26.85	0.80 0.45	16.15	16.29	0.14 0.34	0.13	0.14	0.01 -0.01	4.60	3.01	-1.60 -0.69
18	VFE	76.3	78.1	1.8 0.32	25.63	23.58	-2.05* -0.85	18.17	18.65	0.49 0.22	0.12	0.12	0 0.10	3.09	3.09	0 -0.05

*Change greater than SEM

**Change greater than MD

(SEM and MD were not available for NHR and APQ)

SPL=sound pressure level; CPPS=smoothed cepstral peak prominence; NHR=noise-to-harmonic ratio; APQ=amplitude perturbation quotient; SMD_2=standard mean difference (two baseline and two follow-up data points were available

Table 38. Two-Way ANOVA Results for Vocal Quality Measures (Group Results)

Outcome Measure	**IMST**			**EMST**			**VFE**			**F (p-value) Time effect Time*group Group effect**
	Pre (SD)	Post (SD)	Diff. ES (d)	Pre (SD)	Post (SD)	Diff. ES (d)	Pre (SD)	Post (SD)	Diff ES (d)	
SPL speech (dB)	71.33 (8.12)	73.18 (5.29)	1.85 d_w=0.35 d_b=0.25	70.00 (1.56)	71.20 (1.65)	1.20 d_w=0.23 d_b=0.12	74.90 (1.25)	75.47 (2.65)	0.57 d_w=0.11	1.122 (0.325) 0.112 (0.896) 0.859 (0.464)
CPPS /a/ (dB)	21.92 (1.05)	22.25 (0.71)	0.33 d_w=0.12 d_b=-0.30	20.11 (1.26)	20.84 (1.26)	0.73 d_w=0.27 d_b=-0.15	23.05 (4.84)	24.20 (2.40)	1.15 d_w=0.42	1.204 (0.309) 0.129 (0.881) 1.930 (0.215)
CPPS reading (dB)	16.79 (1.28)	18.03 (1.83)	1.24 d_w=0.82 d_b=0.54	15.36 (1.93)	16.78 (0.64)	1.42 d_w=0.93 d_b=0.66	17.32 (1.05)	17.74 (1.27)	0.42 d_w=0.28	**8.330 (0.023)**** 0.691 (0.532) 1.187 (0.360)
NHR /a/	0.15 (0.02)	0.13 (0.02)	-0.02 d_w=-0.33 d_b=-0.17	0.16 (0.12)	0.15 (0.06)	-0.01 d_w=-0.17 d_b=0	0.14 (0.02)	0.13 (0.09)	-0.01 d_w=-0.17	1.041 (0.342) 0.075 (0.929) 0.157 (0.857)
APQ /a/ (%)	3.09 (0.97)	2.91 (1.04)	-0.18 d_w=-0.10 d_b=0.62	4.77 (3.16)	4.13 (3.12)	-0.64 d_w=-0.35 d_b=0.37	4.12 (0.89)	2.81 (0.42)	-1.31 d_w=-0.72	4.027 (0.085) 0.903 (0.448) 0.593 (0.578)
Overall Severity (VAS 100mm)	43.00 (28.58)	21.00 (10.61)	-22 d_w=-0.97 d_b=-0.67	52.00 (30.41)	38.33 (30.50)	-13.67 d_w=-0.60 d_b=-0.31	40.00 (7.55)	33.33 (21.39)	-6.67 d_w=-0.29	**5.578 (0.050)**** 0.584 (0.583) 0.332 (0.728)

**Significant at an alpha level of 0.05

SPL=sound pressure level; CPPS=smoothed cepstral peak prominence; NHR=noise-to-harmonic ratio; APQ=amplitude perturbation quotient; VAS=visual analog scale; ES=effect size, d_w=Cohen's d within groups; d_b=Cohen's d between groups; SD=standard deviation

Table 39. Change in Auditory-Perceptual Judgment of Overall Severity (Single-Subject Results)

ID	Group	Overall Severity (VAS 100 mm)		
		Pre	Post	Change SMD_2
2	IMST	25	21	-4 -0.31
9	IMST	14	8	-6* -0.24
12	IMST	76	21	-55** -2.07
16	IMST	57	34	-23** -1.68
10	EMST	72	69	-3 -0.37
13	EMST	67	38	-29** -0.66
20	EMST	17	8	-9* -0.39
14	VFE	48	52	4 0.26
17	VFE	33	10	-23** -1.03
18	VFE	39	38	-1 0.09

*Change greater than SEM
**Change greater than MD
SMD_2=standard mean difference (two baseline and two follow-up data points were available) VAS=visual analog scale

Results from the Interventions: Aerodynamic Measures

Individual results for aerodynamic measures are presented in Table 40. In the IMST group, two out of four participants increased their aerodynamic resistance with large to very large effect sizes (SMD=0.90 and 1.46). In both cases, the increase was associated with a decrease in airflow (|SMD|=0.50 and |SMD|=0.90, respectively) and not with an increase in subglottal pressure. A decrease in subglottal pressure was in fact found in both participants, although the effect size was only important in participant 9 (|SMD|=0.88). Although pre-treatment data was unavailable regarding subglottal pressure and aerodynamic resistance for participant 12, the post-treatment profile was the following: a low mean airflow rate, an adequate subglottal pressure, and a high aerodynamic resistance. One participant in the IMST group (participant 16) had an abnormally high subglottal pressure before the intervention, which decreased following treatment (|SDM|=1.33), along with a small increase in mean airflow (SMD=0.30) and a considerable decrease in aerodynamic resistance (|SDM|=1.40).

In the EMST group, no considerable changes were reported for participant 10, who presented with an elevated glottal airflow rate at baseline and decreased aerodynamic resistance. Following treatment, a small increase in subglottal pressure was noted (SMD=0.44), with only a negligible increase in aerodynamic resistance (SDM=0.16). Mean airflow rate was slightly increased following therapy (SMD=0.10). Although pre-therapy data was unavailable for participant 13 regarding subglottal pressure and aerodynamic resistance, post-therapy data suggests a similar profile as participant 10: a slight increase in mean airflow rate (SDM=0.20) and a low aerodynamic resistance at post-treatment. Participant 20 had a very high airflow at pre-treatment, which did not decrease meaningfully after the intervention (|SMD|=0.10). In the same participant, an

insignificant increase in subglottal pressure and in aerodynamic resistance occurred following treatment (SMD=0.13 and 0.08, respectively).

Two out of three participants in the VFE group (participants 14 and 17) increased their aerodynamic resistance, with large effect sizes (SMD=1.10 and 1.07, respectively). Participant 14 decreased their mean airflow rate (|SMD|=1.8) and maintained a somewhat stable subglottal pressure (which was already within normal values), while participant 17 maintained a constant airflow and increased their subglottal pressure (SMD=0.77), which was very low at baseline. The third participant in the VFE group (participant 18) increased both airflow and subglottal pressure (SMD=1.4 and 0.71, respectively). However, aerodynamic resistance decreased following the intervention (|SMD|=1.73).

Results of the ANOVA (table 41) revealed a significant group effect for mean airflow during voicing (p=0.039), and post hoc Tukey test indicated that the difference was between the IMST and EMST groups (p=0.034). In fact, the EMST group had the greatest airflow value prior to the intervention and increased slightly following the intervention, while the IMST group had the smallest airflow and decreased following intervention, further widening the disparity between the groups. A significant group effect was also found for aerodynamic resistance (p=0.003), and post hoc Tukey test revealed that the differences were between IMST and EMST (p=0.002) between EMST and VFE (p=0.042), and between IMST and VFE (p=0.030). The EMST group had the lowest aerodynamic resistance at baseline and only had a slight increase following the intervention (d_w=0.12). The VFE group had the second largest resistance at baseline and also experienced only a slight increase (d_w=0.15). The IMST group had the largest resistance mean at baseline and also experienced an increase (d_w=0.32). Lastly, a significant time by group interaction was found for

subglottic pressure (p=0.057). As shown in Figure 18, both EMST and VFE groups increased their subglottal pressure (d=0.28 and d=0.41, respectively), while the IMST group had a mean decrease in subglottal pressure (|d|=0.84).

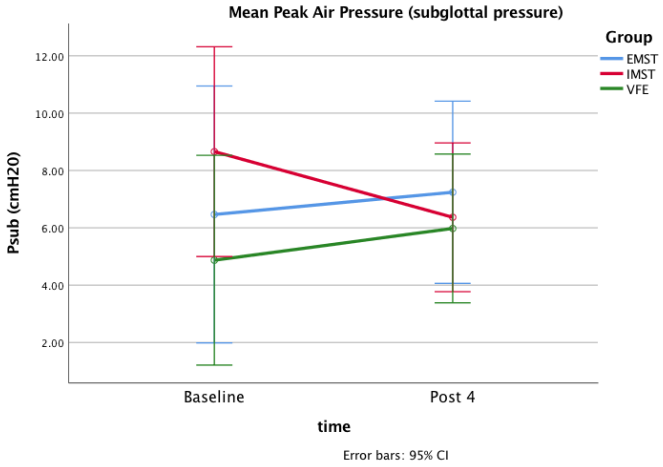


Figure 18. Change in Group Means for Subglottal Pressure

Table 40. Change in Aerodynamic Measures (Single Subject Results)

ID	Group	Mean airflow (L)				Subglottal pressure (cmH_20)				Aerodynamic resistance ($cmH_20/l/s$)			
		Pre	Post	Change	SMD	Pre	Post	Change	SMD	Pre	Post	Change	SMD
2	IMST	0.14	0.09	-0.05	-0.50	7	6.16	-0.84	-0.30	49.07	68.48	19.41	0.90
9	IMST	0.14	0.05	-0.09	-0.90	6.56	4.15	-2.41	-0.88	45.96	77.41	31.45	1.46
12	IMST	0.05	0.06	0.01	0.10	-	7.84	-		-	121.77	-	
16	IMST	0.16	0.19	0.03	0.30	12.41	8.79	-3.62	-1.33	74.52	44.41	-30.11	-1.40
10	EMST	0.27	0.28	0.01	0.10	7.48	8.68	1.2	0.44	26.77	30.15	3.38	0.16
13	EMST	0.17	0.19	0.02	0.20	-	5.37	-		-	27.61	-	
20	EMST	0.34	0.33	-0.01	-0.10	5.45	5.80	0.35	0.13	15.25	17.07	1.82	0.08
14	VFE	0.31	0.13	-0.18	-1.8	7	6.26	-0.74	-0.27	22.05	45.73	23.68	1.10
17	VFE	0.11	0.10	-0.01	-0.1	3.42	5.53	2.11	0.77	30.17	53.12	22.95	1.07
18	VFE	0.06	0.20	0.14	1.4	4.19	6.14	1.95	0.71	66.97	29.83	-37.14	-1.73

SMD=standard mean difference

Table 41. Two-Way ANOVA Results for Aerodynamic Measures (Group Results)

Outcome Measure	**IMST**			**EMST**			**VFE**			**F (p-value)** **Time main effect** **Time*group** **Group main effect**
	Pre (SD)	Post (SD)	Diff. ES (d)	Pre (SD)	Post (SD)	Diff. ES (d)	Pre (SD)	Post (SD)	Diff ES (d)	
Mean airflow (L)	0.12 (0.05)	0.10 (0.06)	-0.02 d_w=-0.20 d_b=0	0.26 (0.09)	0.27 (0.07)	0.01 d_w=0.10 d_b=0.30	0.16 (0.13)	0.14 (0.05)	-0.02 d_w=-0.20	0.154 (0.707) 0.102 (0.904) **5.375 (0.039)****
Subglottal pressure (cmH_20)	8.66 (3.26)	6.37 (2.33)	-2.29 d_w=-0.84 d_b=-1.25	6.47 (1.44)	7.24 (2.04)	0.77 d_w=0.28 d_b=-0.12	4.87 (1.88)	5.98 (0.39)	1.11 d_w=0.41	0.076 (0.794) **5.362 (0.057)*** 0.825 (0.490)
Aerodynamic resistance ($cmH_20/l/s$)	56.52 (15.69)	63.43 (17.07)	6.91 d_w=0.32 d_b=0.17	21.01 (8.15)	23.61 (9.25)	2.60 d_w=0.12 d_b=-0.03	39.73 (23.94)	42.89 (11.90)	3.16 d_w=0.15	0.151 (0.714) 0.016 (0.984) **23.244 (0.003)****

*Significant at an alpha level of 0.1

**Significant at an alpha level of 0.05

ES=effect size; d_w=Cohen's d within groups; d_b=Cohen's d between groups; SD=standard deviation

Results from the Interventions: Self-Assessment Measures

Individual results for the three self-assessment questionnaires are presented in Table 42.

All participants had a VHI-10 score above 11 prior to therapy, which is indicative of a meaningful perceived voice handicap (Arffa et al., 2012). **All participants in the IMST group improved (decreased) their total VHI-10 score with moderate to large effect sizes** ranging from |SMD|=0.62 to |SMD|=2.34. In the EMST group, only one participant decreased their VHI-10 score, and the effect size was small (|SMD|=0.31). One participant in the EMST group worsened (increased) their VHI-10 score by 10 points (SMD=1.56). In the VFE group, only one participant improved their VHI-10 score (|SDM|=1.40).

Three out of four participants in the IMST group improved (decreased) their GFI score with large effect sizes, ranging from |SMD|=1.16 to |SMD|=1.45. In the same group, one participant's score worsened (increased) by three points (SMD=0.87). For two participants in the EMST group, their GFI score improved but with small to moderate effect sizes, ranging from |SMD|=0.29 to |SMD|=0.58. In the VFE group, one participant had a meaningful improvement in GFI score (|SMD|=1.45).

In the IMST group, all participants for who CPIB scores were available improved (increased) their score with large to huge effect sizes, ranging from SMD=1.11 to SMD=3.18. In the EMST group, one participant improved considerably (SMD=1.11), while another participant remained the same and one got worse (SMD=-0.64). In the VFE group, two participants improved their CPIB scores, with moderate to large effect sizes (SMD=0.64 and 0.80), while the other participant decreased their CPIB score (SMD=-0.80).

The group effect sizes confirmed that larger improvements occurred in the IMST group. In fact, large to very large effect sizes were found in the IMST group on the three self-assessment measures, ranging from $|d_w|$=0.80 to 1.91. In opposition, only small effect sizes (or worsening of scores) were found in the EMST group. In the VFE group, small effect sizes were found for the VHI-10 and the CPIB questionnaires, and a medium effect size was found for the GFI questionnaire ($|d_w|$=0.58). The results from the ANOVA revealed a significant time by group interaction effect for VHI-10 (p=0.070). In fact, while the IMST group made a considerable improvement (reduced their mean VHI-10 score) ($|d_w|$=1.40), the VFE group made a small improvement ($|d_w|$=0.36) and the EMST group demonstrated a decline as evidenced by an increase in their mean perceived handicap at post-treatment ($|d_w|$=0.47). See Table 43 and Figures 19, 20, and 21 for the ANOVA results.

Correlations Between Improvements in Respiratory Muscle Strength and in Voice Outcomes

Tabl 44 displays the Spearman correlation coefficients between change in MIP and MEP and change in voice measures. Only one significant correlation was found between improvements in respiratory muscle strength and voice outcomes: a greater change in MIP was associated with a greater change in aerodynamic resistance (r=0.64, p=0.048).

Table 42. Change in Self-Assessment Measures (Single Subject Results)

ID	Group	VHI-10				GFI				CPIB			
		Pre	Post	Change	SMD	Pre	Post	Change	SMD	Pre	Post	Change	SMD
2	IMST	28	22	-6	-0.94	11	14	3	0.87	-	14	-	-
9	IMST	15	11	-4	-0.62	10	5	-5	-1.45	15	24	9	1.43
12	IMST	26	15	-11	-1.72	12	8	-4	-1.16	5	25	20	3.18
16	IMST	27	12	-15	-2.34	17	12	-5	-1.45	16	23	7	1.11
10	EMST	23	21	-2	-0.31	14	12	-2	-0.58	11	18	7	1.11
13	EMST	12	22	10	1.56	13	12	-1	-0.29	15	15	0	0
20	EMST	16	17	1	0.16	7	8	1	0.29	19	15	-4	-0.64
14	VFE	20	20	0	0	10	10	0	0	19	14	-5	-0.80
17	VFE	14	5	-9	-1.40	6	1	-5	-1.45	21	25	4	0.64
18	VFE	29	31	2	0.31	15	14	-1	-0.29	3	8	5	0.80

VHI-10=Voice Handicap Index-10; GFI=Glottal Function Index; CPIB=Communicative Participation Item Bank; SMD=standard mean difference

Table 43. Two-Way ANOVA Results for Self-Assessment Measures (Group Results)

Outcome Measure	IMST			EMST			VFE			F (p-value) Time main effect Time*group Group main effect
	Pre (SD)	Post (SD)	Diff. ES (d)	Pre (SD)	Post (SD)	Diff. ES (d)	Pre (SD)	Post (SD)	Diff ES (d)	
VHI-10	24.00 (6.06)	15.00 (4.97)	-9.00 d_w=-1.40 d_b=-1.04	17.00 (5.57)	20.00 (2.65)	3.00 d_w=0.47 d_b=0.83	21.00 (6.41)	18.67 (13.05)	-2.33 d_w=-0.36	2.403 (0.165) **3.993 (0.070)*** 0.034 (0.967)
GFI	12.50 (3.11)	9.75 (4.03)	-2.75 d_w=-0.80 d_b=-0.22	11.33 (3.79)	10.67 (2.31)	-0.66 d_w=-0.19 d_b=0.39	10.33 (4.51)	8.33 (6.66)	-2.00 d_w=-0.58	3.533 (0.102) 0.413 (0.677) 0.208 (0.817)
CPIB	12.00 (6.08)	24.00 (1.00)	12.00 d_w=1.91 d_b=1.70	15.00 (4.00)	16.00 (1.73)	1.00 d_w=0.16 d_b=-0.05	14.33 (9.87)	15.67 (8.62)	1.34 d_w=0.21	5.586 (0.056) 3.193 (0.114) 0.268 (0.774)

*Significant at an alpha level of 0.1

VHI-10=Voice Handicap Index-10; GFI=Glottal Function Index; CPIB=Communicative Participation Item Bank; ES=effect size; d_w=Cohen's d within groups; d_b=Cohen's d between groups; SD=standard deviation

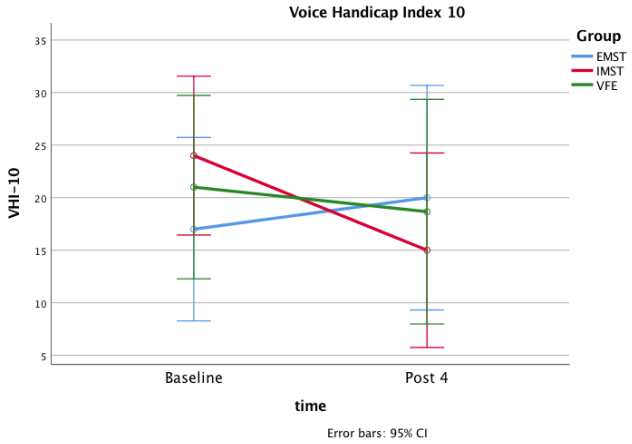


Figure 19. Change in Group Means for Voice Handicap Index-10

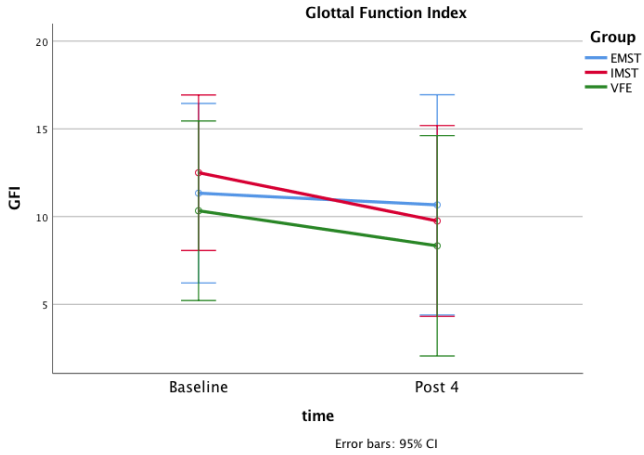


Figure 20. Change in Group Means for Glottal Function Index

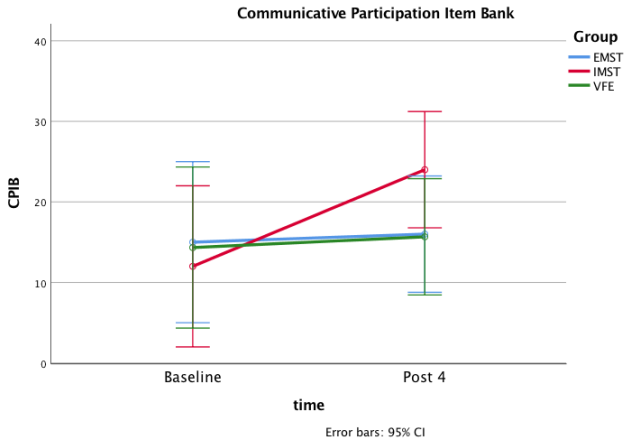


Figure 21. Change in Group Means for Communicative Participation Item Bank

Table 44. Correlations Between Improvements in Respiratory Muscle Strength and in Voice Outcomes

	SPL speech	CPPS /a/	CPPS reading	NHR /a/	APQ /a/	Overall severity	Mean airflow	Subglottal pressure	Aerodynamic resistance	VHI-10	GFI	CPBI	Bowing index
MIP (p-value)	-0.35 (0.327)	-0.47 (0.174)	0.19 (0.603)	-0.31 (0.385)	-0.10 (0.777)	-0.13 (0.713)	-0.09 (0.815)	0.21 (0.556)	**0.64 (0.048)****	0.22 (0.533)	-0.39 (0.260)	0.126 (0.748)	0.12 (0.779)
MEP (p-value)	0.07 (0.854)	-0.18 (0.626)	0.24 (0.510)	-0.19 (0.602)	0.05 (0.894)	0.16 (0.650)	-0.15 (0.686)	0.28 (0.434)	0.35 (0.318)	0.50 (0.137)	-0.09 (0.799)	-0.12 (0.763)	-0.66 (0.076)

**Significant at an alpha level of 0.05

Results: Aim 3b

The objective of aim 3b was to assess whether baseline respiratory function and severity of vocal fold atrophy impacted the response to the intervention in terms of change in voice-related handicap. Results are presented below.

Respiratory Predictors for the IMST Group

The IMST group demonstrated the largest improvement in VHI-10 score (a mean reduction of 9 points on the questionnaire). To assess whether participants with a lower respiratory function gained more from the IMST intervention, each respiratory variable was tested as a predictor for improvement in VHI-10 score following therapy. Table 45 displays the results.

In the IMST group, baseline MIP was found to have a significant relationship with change in VHI-10, with a beta-weight (β) of 0.98, an R-squared of 0.96, and a p-value of p=0.020 (Figure 22). FEV1% had a β of 0.92, an R-squared of 0.85, and a p-value marginally significant (p=0.076) (figure 23). FVC% was found to be a potentially good predictor for change in VHI-10 in the IMST group, with a β of 0.84 and an R-Squared of 0.71, although the p-value did not reach significance (p= 0.158) (Figure 24). MEP was also found to be a potentially meaningful predictor, with a β of 0.82 and an R-squared of 0.67 (p=0.184) (Figure 25). Raw values of FVC and FEV1 also did not reach significance (p=0.197 and p=0.239, respectively). The effect might have been too small to reach significance with this very small sample size (n=4). Nonetheless, the direction of the relationship was the same as for the other variables: a larger FVC and FEV1 were slightly associated with a smaller decrease in VHI-10 score. The FEV1/FVC ratio (raw and predicted percent), also did not reach statistical significance (p=0.930 for the raw value and p=0.675 for the percent predicted value) and presented with the smallest β values (Table 45).

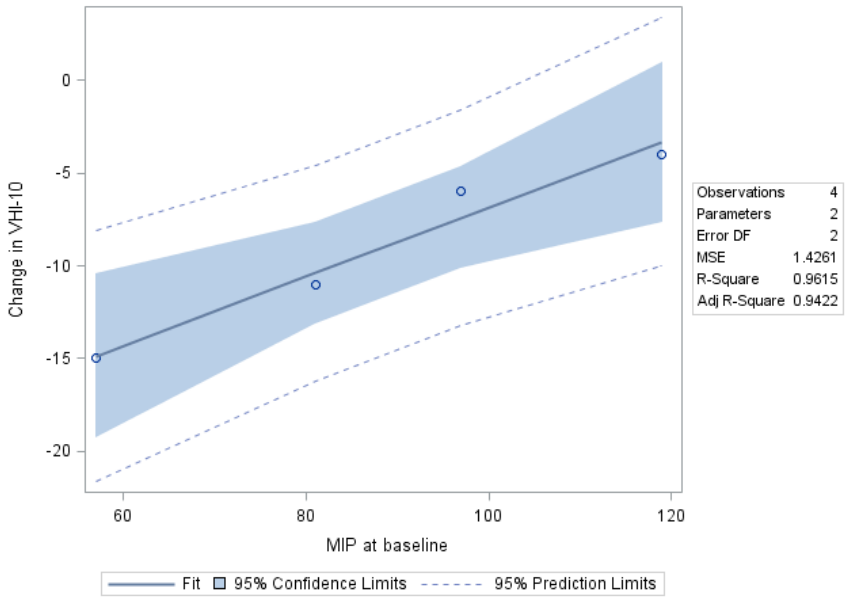


Figure 22. Scatter plot and linear regression fitted-line for change in VHI-10 (y-axis) and MIP at baseline (x-axis). A change towards the negative values is associated with a greater improvement.

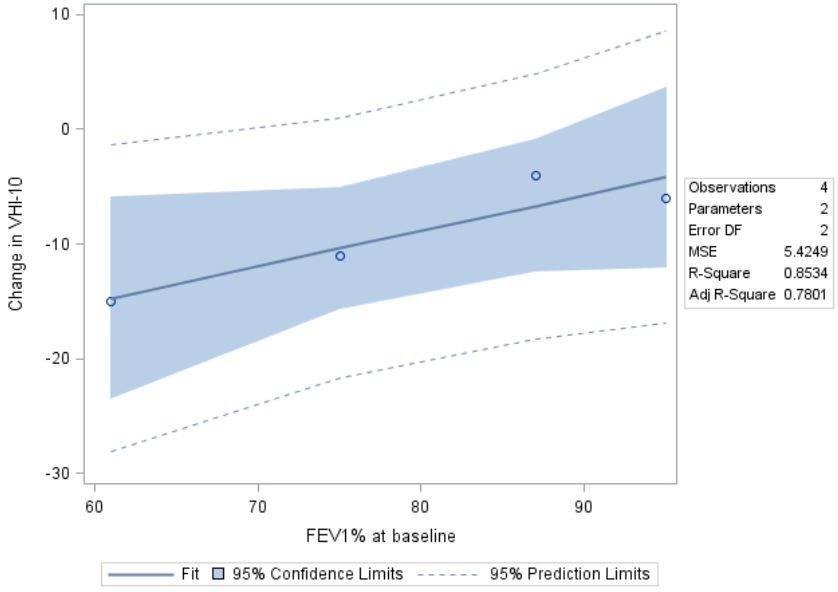


Figure 23. Scatter plot and linear regression fitted-line for change in VHI-10 (y-axis) and FEV1% at baseline (x-axis). A change towards the negative values is associated with a greater improvement.

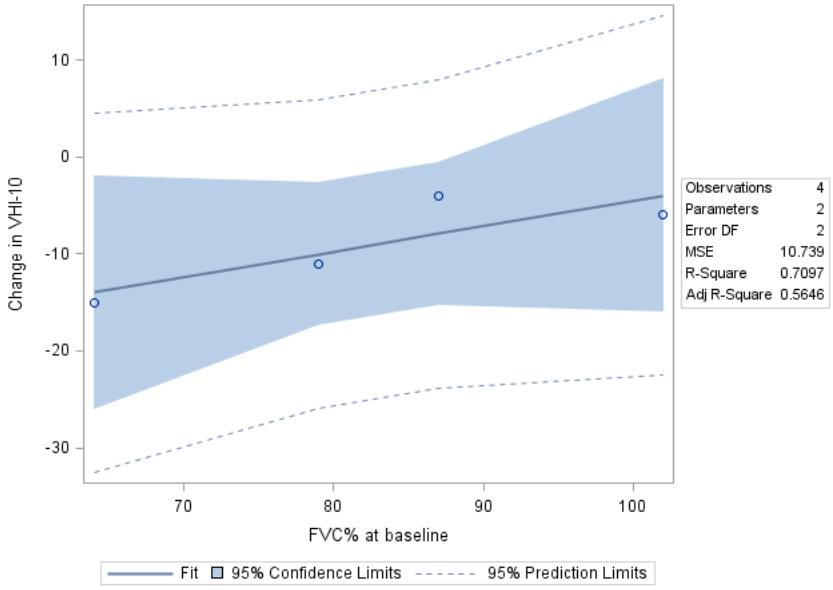


Figure 24. Scatter plot and linear regression fitted-line for change in VHI-10 (y-axis) and FVC% at baseline (x-axis). A change towards the negative values is associated with a greater improvement.

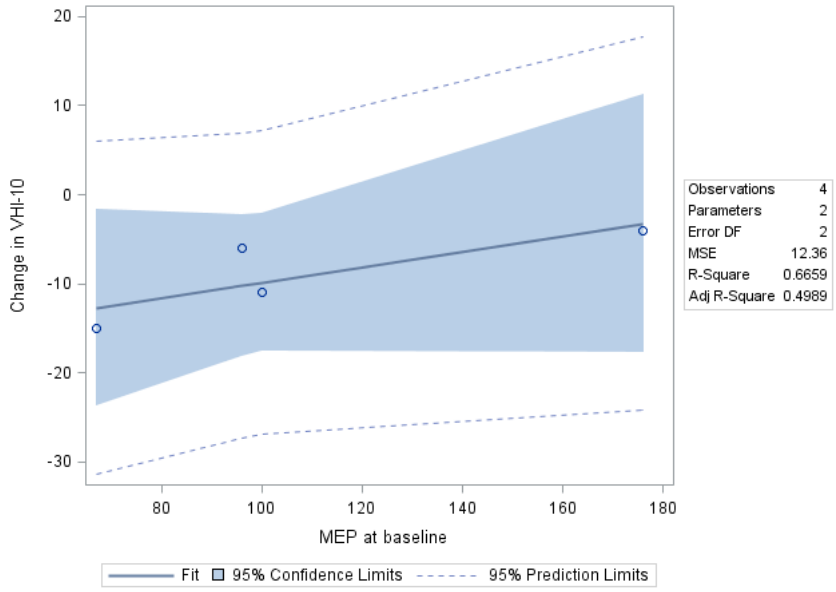


Figure 25. Scatter plot and linear regression fitted-line for change in VHI-10 (y-axis) and MEP at baseline (x-axis). A change towards the negative values is associated with a greater improvement.

Table 45. Linear Regression Outcomes for Change in VHI-10 in the IMST Group

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	-25.48	2.41	0	-10.58	0.009	49.89 (**0.020****)	0.96	0.94
MIP	0.19	0.03	0.98	7.06	**0.020****			
Intercept	-33.61	7.31		-4.60	0.044	11.64 (**0.076***)	0.85	0.78
FEV1%	0.31	0.09	0.92	3.41	**0.076***			
Intercept	-30.91	10.04	0	-3.08	0.091	4.89 (0.158)	0.71	0.56
FVC%	0.26	0.12	0.84	2.21	0.158			
Intercept	-18.56	5.10	0	-3.64	0.068	3.99 (0.184)	0.67	0.50
MEP	0.09	0.04	0.82	2.00	0.180			
Intercept	-21.03	6.57	0	-3.20	0.085	3.63 (0.197)	0.64	0.47
FVC	4.05	2.13	0.80	1.91	0.197			
Intercept	-19.36	6.55	0	-2.96	0.098	2.75 (0.239)	0.58	0.37
FEV1	4.77	2.87	0.76	1.66	0.239			
Intercept	-58.96	102.76	0	-0.57	0.624	0.24 (0.675)	0.11	-0.34
FEV1/FVC%	0.52	1.07	0.33	0.49	0.675			
Intercept	-17.82	88.78	0	-0.20	0.860	0.01 (0.930)	0.005	-0.49
FEV1/FVC	12.12	121.97	0.07	0.10	0.930			

*Significant at an alpha level of 0.1 **Significant at an alpha level of 0.05

B=parameter estimate; SE B= standard error of B β=standardized parameter estimate (beta-weight);

Respiratory Predictors for the Total Sample

To evaluate if respiratory measures were good predictors of treatment outcome regardless of the intervention received (IMST, EMST, or VFE), relationships were also assessed within the total sample. Because both IMST and EMST were strong predictors of the outcome (IMST predicted a greater decrease in VHI-10 score and EMST a lesser decrease) (Table 46), they were both tested as potential covariates with each respiratory variable. Age and gender were also tested in the models and controlled for if necessary.

Table 46. Linear Regression Outcomes: Effect of Intervention Group on Change in VHI-10

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	0.33	2.34	0	0.14	0.890	6.35 **(0.036**)**	0.44	0.37
IMST	-9.33	3.70	-0.67	-2.52	**0.036****			
Intercept	-6.14	2.30	0	-2.67	0.028	4.73 **(0.061***)	0.37	0.29
EMST	9.14	4.20	0.61	2.18	**0.061***			
Intercept	-3.86	2.89	0	-1.33	0.219	0.08 (0.780)	0.01	-0.11
VFE	1.52	5.28	0.10	0.29	0.780			

*Significant at an alpha level of 0.1
**Significant at an alpha level of 0.05

Two respiratory variables were found to be significant predictors for change in VHI-10, at an alpha level of 0.05: FVC% (p=0.013) and FEV1% (p=0.049). The R-squared for both models was 0.75. MEP was found to be marginally significant, at a 0.1 alpha level (p=0.076). Scatter plots for these three variables are presented below (Figures 26 to 28) and show the direction of the relationship: higher values of FVC%, FEV1%, and MEP at baseline were associated with less improvement in voice handicap index after the intervention. Lower values of FVC%, FEV1% and MEP were associated with greater improvement (more decrease in VHI-10 scores) following treatment.

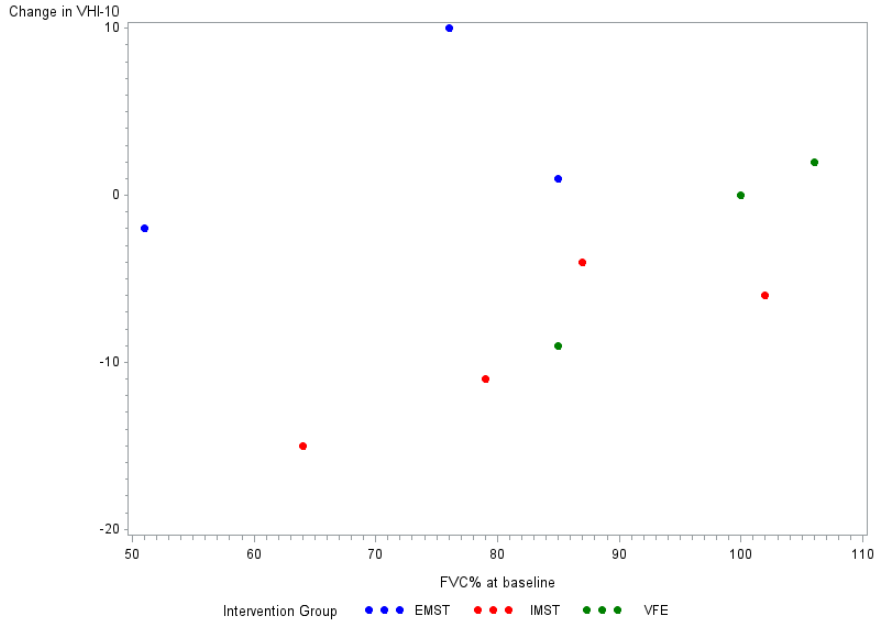


Figure 26. Scatter plot of the effect of FVC% at baseline (x-axis) on change in VHI-10 following treatment (y-axis), by intervention group. A change towards the negative values is associated with a greater improvement.

Table 47. Linear Regression Results for Change in VHI-10 with FVC% as the Respiratory Predictor

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	-33.24	8.38	0	-3.97	0.005	10.68 (**0.007****)	0.75	0.68
FVC%	0.30	0.09	0.72	3.29	**0.013****			
EMST	14.72	3.29	0.98	4.48	**0.003****			

**Significant at an alpha level of 0.05

Table 47 presents the linear regression results for FVC% as a predictor, when controlling for potential covariates (in this case the only significant covariate was EMST). A higher FVC% at baseline was associated with smaller improvement in VHI-10 following treatment. More specifically, an increase of 3.3% in baseline FVC% was associated with one point less in VHI-10 improvement following treatment. In terms of SDs, an increase of one SD in baseline FVC%

(17.16%) was associated with 0.72 SD less improvement in VHI-10 (corresponding to 5.22 points on the questionnaire).

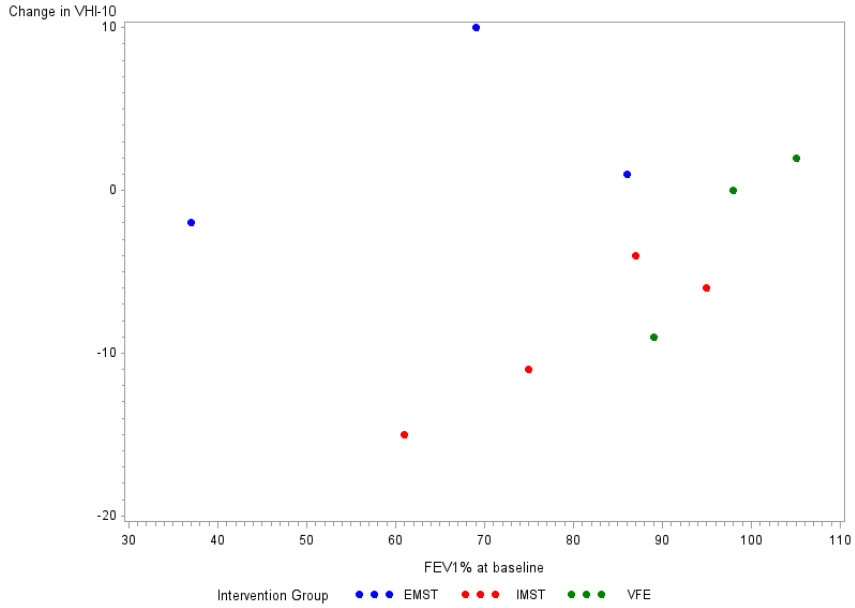


Figure 27. Scatter plot of the effect of FEV1% at baseline (x-axis) on change in VHI-10 following treatment (y-axis), by intervention group. A change towards the negative values is associated with a greater improvement.

Table 48. Linear Regression Results for Change in VHI-10 with FEV1% as the Respiratory Predictor

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	-42.01	14.67	0	-2.86	0.029	6.02 (**0.031****)	0.75	0.63
FEV1%	0.22	0.09	0.62	2.47	**0.049****			
EMST	14.00	3.72	0.93	3.76	**0.009****			
Age	0.23	0.20	0.25	1.18	0.281			

**Significant at an alpha level of 0.05

Table 48 presents the linear regression results for FEV1% as a predictor, when controlling for relevant covariates. A higher FEV1% at baseline was associated with smaller improvement in VHI-10 following treatment. More specifically, an increase of 4.55% in baseline FEV1% was associated with one point less in VHI-10 improvement following treatment. In terms of SDs, an increase of

one SD in baseline FEV1% (20.26%) was associated with 0.62 SD less improvement in VHI-10 (corresponding to 4.49 points on the questionnaire).

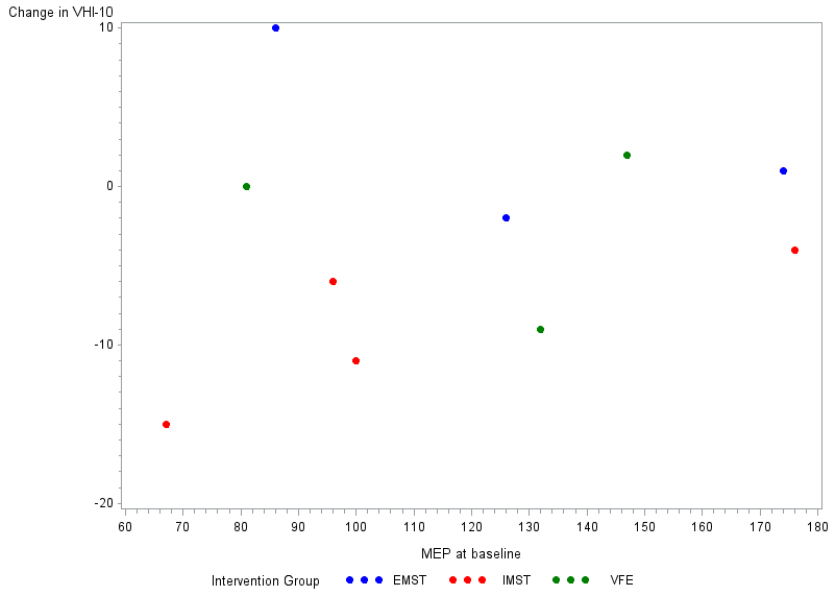


Figure 28. Scatter plot of the effect of MEP at baseline (x-axis) on change in VHI-10 following treatment (y-axis), by intervention group. A change towards the negative values is associated with a greater improvement.

Table 49. Linear Regression Results for Change in VHI-10 with MEP as the Respiratory Predictor

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	-14.82	8.29	0	-1.79	0.124	4.56 (**0.055***)	0.69	0.54
MEP	0.21	0.10	1.14	2.14	**0.076***			
IMST	-13.42	3.78	-0.96	-3.55	**0.012****			
Gender	-17.33	7.89	-1.26	-2.20	**0.070***			

*Significant at an alpha level of 0.1

Table 49 presents the linear regression results for MEP as a predictor, when controlling for relevant covariates. A higher MEP at baseline was associated with smaller improvement in VHI-10 following treatment. More specifically, an increase of 4.76 cmH_20 in baseline MEP was associated with a one point less in VHI-10 improvement following treatment. In terms of SDs, an increase of one SD in baseline MEP (38.59 cmH_20) was associated 1.14 SD less improvement in VHI-10 (corresponding to 8.25 points on the questionnaire).

Although the other respiratory variables did not reach statistical significance, the results indicated a relationship in the same direction. A greater sample size would be required to potentially reach statistical significance for these variables (Tables 50 to 54).

Table 50. Linear Regression Results for Change in VHI-10 with FVC as the Respiratory Predictor

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	-8.61	8.88	0	-0.97	0.370	2.71 (0.138)	0.58	0.36
FVC	4.53	3.57	0.48	1.27	0.252			
IMST	-12.05	4.34	-0.86	-2.78	0.032			
Gender	-7.16	5.66	-0.52	-1.26	0.253			

Table 51. Linear Regression Results for Change in VHI-10 with FEV1 as the Respiratory Predictor

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	-51.31	20.34	0	-2.52	0.045	3.92 (**0.0727***)	0.66	0.49
FEV1	4.85	2.84	0.45	1.71	0.138			
EMST	10.50	3.70	0.70	2.83	0.030			
Age	0.48	0.24	0.51	2.00	0.092			

Table 52. Linear Regression Results for Change in VHI-10 with MIP as the Respiratory Predictor

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	-8.32	7.35	0	-1.13	0.301	3.18 (0.106)	0.61	0.42
MIP	0.19	0.12	0.58	1.54	0.175			
IMST	-15.31	5.12	-1.09	-2.99	0.024			
Gender	-7.76	5.31	-0.56	-1.46	0.194			

Table 53. Linear Regression Results for Change in VHI-10 with FEV1/FVC as the Respiratory Predictor

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	-65.92	38.87	0	-1.70	0.141	2.70 (0.139)	0.57	0.36
FEV1/FVC	35.78	34.35	0.36	1.04	0.338			
EMST	11.39	4.72	0.76	2.41	0.052			
Age	0.47	0.28	0.50	1.66	0.147			

Table 54. Linear Regression Results for Change in VHI-10 with FEV1/FVC% as the Respiratory Predictor

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	-65.13	35.01	0	-1.86	0.112	2.88 (0.125)	0.59	0.38
FEV1/FVC%	0.31	0.26	0.37	1.16	0.289			
EMST	11.75	4.68	0.78	2.51	0.046			
Age	0.41	0.25	0.43	1.60	0.161			

Bowing Index as a Predictor

The influence of baseline bowing index on change in VHI-10 was also assessed with linear regression. The results did not reach statistical significance, indicating that the effect of bowing index on the response to the intervention was either not present or too small to be detected with the limited sample size (p=0.246) (Table 55). **In fact, meaningful improvements in VHI-10 scores (at least 5 points) were obtained for a large range of baseline bowing index (6.66 to 17.03**) (Figure 29).

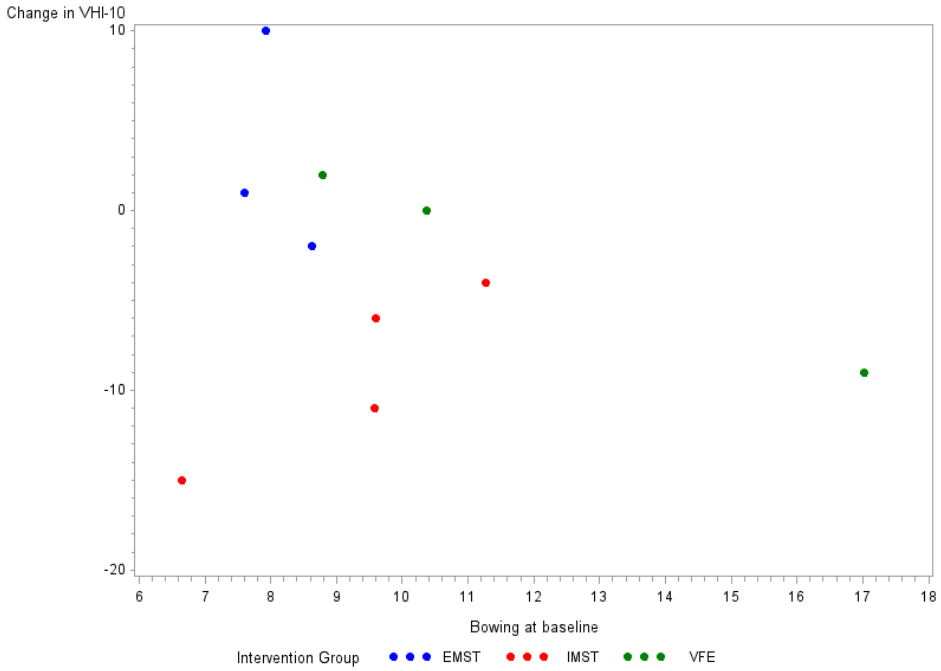


Figure 29. Scatter plot of the effect of bowing at baseline (x-axis) on change in VHI-10 following treatment (y-axis), by intervention group. A change towards the negative values is associated with a greater improvement.

Table 55. Linear Regression Results for Change in VHI-10 with Bowing Index as a Predictor

Variable	B	SE B	β	t	p	Model F (p-value)	R-Squared	Adjusted R-Squared
Intercept	8.52	6.85	0	1.24	0.254	4.21 (**0.063***)	0.55	0.42
Bowing Index	-0.81	0.64	-0.33	-1.26	0.246			
IMST	-9.97	3.61	-0.71	-2.76	0.028			

Controlling for Baseline Respiratory Measures

The effect of the intervention on change in VHI-10 score was assessed without controlling for baseline respiratory function, and was then re-assessed while controlling for the respiratory predictor that was the most significant influencer in the total sample: FVC%. Because a linear relationship was found between FVC% and the *difference* in VHI-10 score, a univariate ANCOVA was conducted with "change in VHI-10 score" as the dependent variable, "group" as the independent variable, and FVC% as the covariate.

Table 56 presents the results from a univariate ANOVA (without controlling for FVC%) and post hoc pairwise comparisons with Bonferroni correction. The group effect was marginally significant (p=0.070), and partial eta squared for the variable "group" was 0.533, meaning that the intervention group explained 53.3% of the variance in the outcome. Pairwise comparisons showed a marginally significant difference between the EMST and the IMST groups (p=0.080). No difference was found between the EMST and VFE groups, nor between the IMST and VFE groups.

Table 57 presents the results from the univariate ANCOVA when controlling for FVC%, as well as pairwise comparisons between estimated marginal means, when FVC% was held constant at 83.50%. In this model, the effect for the variable "group" became statistically significant (p=0.013), and the effect for FVC% at baseline was also significant at an alpha level of 0.05 (p=0.042). Partial eta squared for group increased when compared to the previous model (0.768 versus 0.533), meaning that controlling for baseline FVC% allowed for the detection of a greater effect of the intervention group on the dependent outcome. Pairwise comparisons showed a statistically significant difference between the EMST and the IMST groups (p=0.013). Moreover, a

marginally significant difference was found between the EMST and the VFE groups (p=0.089). No significant difference was found between the IMST and VFE groups.

Table 56. (a)ANOVA Results for the Effect of Intervention Group on Change in VHI-10 Without Controlling for Baseline Respiratory Function (b) Multiple Comparisons Between Intervention Groups

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	251.733[a]	2	125.867	3.993	**0.070***	0.533
Intercept	75.758	1	75.758	2.403	0.165	0.256
Group	251.733	2	125.867	3.993	**0.070***	0.533
Error	220.667	7	31.524			
Total	588.000	10				
Corrected Total	472.400	9				

a.

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
EMST	IMST	12.00	4.288	**0.080***	-1.41	25.41
	VFE	5.33	4.584	0.848	-9.00	19.67
IMST	EMST	-12.00	4.288	**0.080***	-25.41	1.41
	VFE	-6.67	4.288	0.492	-20.08	6.74
VFE	EMST	-5.33	4.584	0.848	-19.67	9.00
	IMST	6.67	4.288	0.492	-6.74	20.08

b. Adjustment for multiple comparisons= Bonferroni
*Significant at an alpha level of 0.1

Table 57. (a)ANOVA Results for the Effect of Intervention Group on Change in VHI-10 When Controlling for Baseline Respiratory Function (b) Multiple Comparisons Between Intervention Groups, adjustment for multiple comparisons= Bonferroni

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	367.745[a]	3	122.582	7.028	**0.022****	0.778
Intercept	143.153	1	143.153	8.207	**0.029****	0.578
FVC% baseline	116.012	1	116.012	6.651	**0.042****	0.526
Group	346.013	2	173.007	9.919	**0.013****	0.768
Error	104.655	6	17.443			
Total	588.000	10				
Corrected Total	472.400	9				

a.

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
EMST	IMST	15.312	3.439	**0.013****	4.008	26.616
	VFE	12.405	4.376	**0.089***	-1.980	26.790
IMST	EMST	-15.312	3.439	**0.013****	-26.616	-4.008
	VFE	-2.907	3.507	1.000	-14.437	8.623
VFE	EMST	-12.405	4.376	**0.089***	-26.790	1.980
	VFE	2.907	3.507	1.000	-8.623	14.437

b. Adjustment for multiple comparisons= Bonferroni
*Significant at an alpha level of 0.1
**Significant at an alpha level of 0.05

Chapter 5: Discussion

Discussion Aim 1

Which outcomes present with the strongest intra-subject reliability in a sample of patients with presbyphonia.

Out of the 29 variables assessed, only eight presented with sufficient intra-subject reliability for both females and males. These variables were: SPL during reading, CPPS /a/, CPPS during reading, NHR /a/, APQ /a/, MIP, MEP, and perceptual judgment of overall severity.

Respiratory Muscle Strength

A reliability study by Romer et al. (Romer & McConnell, 2004) revealed excellent agreement ratios for MIP and MEP. The authors explain these results by the fact that the participants in their experiment were thoroughly familiarized with the procedure a priori and that the measurements were all taken by the same investigator, therefore reducing secondary variance. Reducing bias is also likely to increase sensitivity to change of these measures, which have also proven to be strong. In fact, various studies have obtained statistically significant improvements in MIP and MEP after interventions targeting respiratory muscle strength (Souza et al., 2014). Our results also showed strong intra-subject reliability for MIP and MEP, both in males and females.

Measures of Intensity and Fundamental Frequency

Measures of SPL and F0 during reading as well as F0 during speech showed a significant trial effect, indicative of systematic error. This might have been caused by the fact that loudness and pitch were not controlled for during the assessment tasks. These measures are therefore not reliable enough to inform on pre- to post-treatment changes, despite presenting with excellent ICCs. These observations are consistent with Lee et al.'s results, which showed that participants for who frequency was not controlled produced significantly different F0 from one trial to another, 28 days apart (Linda Lee et al., 1999). The same study showed that other acoustic and aerodynamic measures' reliability were not affected by these changes in F0, when intensity remained constant. However, not controlling for intensity led to the most inconsistencies across trials for jitter (at low pitch), and for airflow rate and MPT (at high pitch) (Linda Lee et al., 1999). In fact, Huang et al. showed that measures of voice quality, including jitter, shimmer, and normalized noise energy, varied significantly with different intensity levels (D. Z. Huang, Minifie, Kasuya, & Lin, 1995).

In patients with dysphonia, the question of controlling for frequency and intensity is more complex than in vocally healthy participants, because they are part of the patients' clinical picture and may represent parameters that the therapy is aiming at modifying. This is especially true in patients with presbyphonia, who can present with reduced intensity and altered fundamental frequency. Controlling for these variables may improve the results of intra-subject variability but it may also obscure the positive effects of voice therapy on these variables and potentially on other voice measures. This hypothesis was also suggested by Awan and colleagues in their reliability study on aerodynamic measures, in which intensity and fundamental frequency were not controlled. The authors explained that, even though stronger intra-subject reliability values

may have been obtained, the validity of the measurements would have been reduced by masking the true phonatory behaviors (Awan et al., 2013; Higgins et al., 1994). For this reason, they suggested that the measurements be taken at comfortable loudness and pitch. A potential solution for systematic error could be to conduct multiple trials at baseline and at follow-up and to average those measures in order to obtain a more stable measurement while allowing participants to phonate at comfortable intensity and frequency.

Measures of Voice Quality

Acoustic measures have been criticized by various authors on the basis of a lack of reliability and sensitivity to change (Carding et al., 2009; Leong et al., 2013). Carding et al. revealed a poor reliability for jitter, shimmer, and NHR in dysphonic participants (ICCs=0.46, 0.40, and 0.33, respectively)(Carding et al., 2004). In the same study, ICCs were higher for normal voices (ICCs=0.73, 0.55, and 0.68, respectively). Leong et al. also found a good reliability for some perturbation measures in a sample of vocally healthy participants, when computed over 10 repeated trials (Leong et al., 2013). However, the measures were only reliable in male participants (ICC=0.91 for jitter; ICC=0.69 for relative average perturbation; ICC=0.71 for STD F0). Jitter and shimmer presented with ICCs lower than 0.60 in both males and females (Leong et al., 2013).

In our study, two measures of pitch perturbation (jitter and PPQ for sustained /i/) were also found to be reliable only in males. On the other hand, shimmer for sustained /a/ was reliable only in females. The only time-based vocal quality measures that showed acceptable reliability (ICC$\geq$0.60) for both genders were NHR and APQ for the vowel /a/.

The voice samples did not include any type III voice signals, which are too aperiodic to allow the acoustic software to compute perturbation measures. However, it is likely that type II voice signals were present. These voice signals can be analyzed by the computer but they present with considerable aperiodicity that can reduce the reliability of perturbation measures such as jitter and shimmer (Carding et al., 2004). This is a limitation inherent to the analysis of dysphonic voices with traditional acoustic voice measures.

As a response to critics regarding traditional acoustic measures, efforts to develop a more robust measure of voice quality have led to the development of cepstral peak prominence (CPP) measures. There is a growing body of evidence demonstrating that cepstral measures (frequency-based) are more appropriate than traditional acoustic measures (time-based) for analyzing connected speech, which makes them more ecologically valid (Maryn et al., 2009). Time-based measures rely on the periodic signal and thus their accuracy for severely dysphonic voices is not optimal (Gillespie et al., 2014). On the other hand, frequency-based measures inform on the dominance of F0 over superfluous noise in the signal and do not depend on the periodicity of the waveform (Gillespie et al., 2014). In a meta-analysis on acoustic measurement of voice quality, CPPS was the only acoustic measure to show acceptable concurrent validity with perceived overall severity in both sustained vowel and speech (Maryn et al., 2009). Lastly, with computer software for cepstral measure analyses becoming more accessible, ASHA advocates for the inclusion of cepstral measures in the acoustic analysis of voice quality (Awan et al., 2013). In our study, CPPS during sustained /a/ and during reading displayed a good reliability, with no significant difference between genders and no significant time effect.

Maximum Phonation Time

MPT demonstrated a poor reliability based on the ICC and the elevated values of CV and MD. In fact, a MD of 18 seconds would have to be obtained in order to ensure a true treatment effect, which is clinically unrealistic. Lee et al. found that MPT varied significantly across two trials (28 days apart) in the group in which intensity was not controlled (Linda Lee et al., 1999). Johnson et al. also found a significant trial effect when measuring MPT three times during a single session (Johnson & Goldfine, 2016). The second trial was in average two seconds longer than the first and third trials, and the authors hypothesized that variations were likely due to changes in laryngeal efficiency considering that phonatory volume (amount of air expelled during the task) remained constant across trials (Johnson & Goldfine, 2016). Despite this systematic variation, Johnson et al. did find a strong ICC for MPT (ICC=0.86), which is consistent with Awan et al.'s findings (ICC=0.87)(Awan et al., 2013). However, in these two studies, trials were recorded on the same day (Johnson & Goldfine, 2016) or at a one-week interval (Awan et al., 2013), and both experiments were conducted on vocally healthy participants. Patients with presbyphonia are likely to present with more variability, thus inflating the error term and reducing the ICC. This could explain why MPT had poor reliability in our study.

Reliability in Presbyphonia

There are a lot of discrepancies in the literature surrounding the intra-subject reliability of objective voice measures. This could be due to the differences in population samples and in instrumentation across studies. This highlights the importance of assessing the outcome measures' reliability specific to a study before conducting pre- to post-treatment analyses on those variables. More importantly, a majority of reliability studies have been conducted on young

and vocally healthy participants (Awan et al., 2013; Linda Lee et al., 1999; Leong et al., 2013) and therefore their results are not necessarily applicable to dysphonic populations. In fact, in can be expected that disordered voiced will present with a higher biological variability when compared to normal voices, thus leading to lower ICCs (Carding et al., 2004; Leong et al., 2013). In addition, as previously stated, time-based acoustic measurements' reliability decreases with more aperiodic signals, which is likely to be the case for many dysphonic voices (Leong et al., 2013).

In a paper on MPT, Johnson et al. suggested that documenting variability from one trial to another may provide relevant information on a patient's abilities, especially when it comes to elderly individuals (Johnson & Goldfine, 2016). In fact, older voices show more oscillation instability, as shown in short-term perturbations (increased jitter and shimmer), and long-term perturbations (variations in intonation contours when a utterance is repeated) (Baken, 2005). Baken hypothesized that an increased "biomechanical nonlinearity" of the vocal folds (the fact that a certain amount of stretch does not lead to a proportional amount of stress/tension) could be the main cause for the perceptual characteristics of the aging voice (Baken, 2005). The author demonstrated, with a mathematical model, that an increase in nonlinearity indeed led to increases in jitter and shimmer, as well as to changes in glottal closure and transglottal airflow (Baken, 2005). A decreased voice stability, and therefore increased variability, may therefore be an important characteristic of older voices and it should be taken into account when using voice outcomes to measure pre- post-therapy changes. An increased stability of the voice could be deemed a sign of improvement, and future studies could consider taking repeated post-therapy measurements (even on the same day) to assess the stability of the voice as an outcome measure.

Limitations for Aim 1

Many parameters can be controlled in order to obtain better intra-subject reliability outcomes. For example, previous studies have controlled for room temperature (Leong et al., 2013), time of the day (Leong et al., 2013), and intensity and frequency of the voice samples (Linda Lee et al., 1999). Although the need for a controlled assessment environment is crucial to be able to detect the true effect of a treatment, many biological factors are difficult to control and are part of the voice's inherent variability. The goal of therapy is to yield an improvement that will be noticeable despite the variability due to factors such as time of the day, hydration level, or fatigue/sleep time. In the context of this study, in which the goal was to determine which outcomes are the most sensitive to detect a significant change in voice in a specific sample of participants, controlling for less parameters served a purpose of ecological validity. The SEMs and MDs that were computed are therefore conservative in the sense that they take into account variability that can arise from these uncontrolled parameters.

When using reliability measures to estimate sample size for a study, it has been suggested that the number trials be representative of the duration of the study (Hopkins, 2000). In the present study, the interval of time between the first and the second assessment varied between five and 62 days and averaged 21 days. Therefore, it is possible that the expected error in measurement that could occur between baseline and follow-up was underestimated by measuring intra-subject reliability at an average of a three-week interval instead of four.

Discussion Aim 2

What is the respiratory function of patients with presbyphonia and how is it correlated to voice measures in this population?

The objectives of this aim were to 1) describe the voice and respiratory profile of a sample of treatment-seeking presbyphonic individuals, and 2) examine if relationships could be drawn between respiratory function and phonation in these participants.

Baseline Respiratory Function in Patients with Presbyphonia

Pulmonary function and respiratory muscle strength decrease as a function of aging, and a decreased respiratory function is therefore part of the clinical picture of presbyphonia. However, the extent to which it is related to poor voice outcomes remains unclear. The first question that arises is: do treatment-seeking presbyphonic patients have poorer respiratory function than the general elderly population?

One third of the participants in the studied sample had FVC and FEV1 values below 80% of the predicted values, which is often used as the definition of abnormal (Rabe et al., 2007). FVC and FEV_1 are known to decrease as a function of age (Janssens et al., 1999; Kim & Sapienza, 2005), and in some individuals the aging process may be more advanced and compounded by disease processes, leading to lower percent predicted values and potentially to a decreased FEV1/FVC ratio. Five of the participants in the sample presented with a FEV1/FVC ratio below the 0.70 threshold that defines obstructive pathology as suggested by the Global Initiative for Chronic Obstructive Lung Disease (GOLD)(Rabe et al., 2007). However, Enright et al. showed that this ratio

tends to decrease with age, and found lower limits of normal (LLN) between 0.56 and 0.64 in a cohort of elderly individuals (Enright, Kronmal, Manolio, et al., 1994). One participant had a FEV1/FVC ratio below 0.64 in our sample.

Regarding respiratory muscle strength, four participants had either a MIP or a MEP or both below the LLN as calculated with Enright et al.'s reference equations (Enright, Kronmal, Manolio, et al., 1994). When compared to individuals from the Cardiovascular Health Study (CHS) cohort, comprising MIP and MEP data for 5,201 and 756 functional individuals over 65 years old, respectively, our participants with presbyphonia presented with a further decline in MEP (Enright, Kronmal, Manolio, et al., 1994). In fact, 18% of the presbyphonic males (two participants) and 20% of the presbyphonic females (two participants) fell below the LLN, while only 11% of the males and 6.1% of the females fell below LLN in the CHS cohort. In addition, the mean MEP for former and never smokers in the CHS cohort was 141 cmH_20, which is considerably higher than the mean MEP from our sample of presbyphonic patients, 118.19 cmH_20. On the other hand, no atypical decline in MIP was observed: 9% of the males (one participant) and 0% of the females had a MIP below LLN, compared to 8.8% of the males and 6.9% of the females having a MIP below LLN in the CHS cohort. The mean MIP for former and never smokers in the CHS cohort was 66 cmH_20, compared to 82.67 cmH_20 in the presbyphonic sample (Enright, Kronmal, Manolio, et al., 1994).

In older individuals, inspiratory pressures are better preserved than expiratory pressures (Hoit & Hixon, 1987; Huber & Spruill, 2008). This can be explained by the loss of elastic recoil of the lungs, the decreased compliance of the ribcage, and by the atrophy process that has been shown to affect the expiratory muscles to a greater extent than the inspiratory muscles (Janssens et al.,

1999; Lalley, 2013). In addition, the small airways tend to close at a higher volume during expiratory effort, further reducing expiratory pressures (Lalley, 2013). However, this expected decrease in MEP is taken into account in the reference equations for respiratory muscle strength for elderly individuals. The fact that a higher percentage of participants in our sample had a MEP below LLN when compared to a general cohort of elderly participants potentially suggests a decrease in expiratory pressures greater than what would normally be expected based on normative values.

In summary, our sample of presbyphonic patient seemed to present values of FVC, FEV1, and MEP that are lower than what would be expected from normal aging. The hypothesis that a greater deficit in respiratory function may explain why some older people present with a "pathological" presbyphonia, while others are not bothered by their voice remains plausible and warrants further investigation (Crawley et al., 2018).

Baseline Voice Function in Patients with Presbyphonia

Laryngeal Features

Our sample of presbyphonic participants was representative of patients with presbyphonia with regards to laryngeal parameters that are considered as hallmarks of the disorder. Results for bowing index were similar to those obtained by Kaneko et al. in a sample of 16 participants with age-related vocal fold atrophy (Kaneko et al., 2015). This was expected since the mean age was similar and the bowing index is thought to be an indicator of age-related changes in the vocal folds. The presence of bowing is consistent with the fact that 95% of our sample had concave vocal fold edges. Moreover, all participants except for three had an incomplete glottal closure,

and the most common type of gap was a spindle gap, which is characteristic of presbyphonia (Pontes et al., 2006). In addition, all participants presented with a certain amount of supraglottic hyperfunction, either anteroposterior or mediolateral, or both. This is consistent with Yamauchi et al.'s finding that patients with vocal fold atrophy tend to present with greater supraglottic activity when compared to vocally healthy speakers (Yamauchi et al., 2015).

Regarding amplitude of vibration and mucosal wave amplitude, values below 40% or above 60% were considered to be abnormal based on clinical experience. Approximately 25% of the participants had an amplitude of vibration above 60% for at least one of the vocal folds. On the other hand, more than 50% of the sample had abnormal results regarding mucosal wave amplitude. Four participants had a reduced mucosal wave amplitude (below 40% on at least one vocal fold) and eight participants had an increased mucosal wave amplitude (above 60% on at least one vocal fold). As it will be discussed in the following sections, age-related changes in the vocal fold layers and in the TA muscle affect the pliability of the tissue can lead to increased amplitude of vibration and mucosal wave amplitude. On the other hand, compensatory mechanisms can limit the amplitude of vibration as well as the lateral movement of the mucosal wave. In addition, Yamauchi et al. found decreased amplitude of vibration in females with vocal fold atrophy and suggested that a poor respiratory function could lead to a decreased subglottal pressure and limit the excursion of the vocal folds (Yamauchi et al., 2015).

Consistent with the literature, phase symmetry was found to be decreased in our sample (Pontes et al., 2006; Yamauchi et al., 2015). This can be explained by the degenerative changes in the vocal folds' lamina propria and muscle affecting the tension, mass, and viscoelasticity of the vocal folds and leading to an asymmetric vibration (Yamauchi et al., 2015). This can also lead to an aperiodic vibration. The regularity of the vibration could further be affected by the unstable respiratory support due to age-related changes in the respiratory system.

Acoustic Measures

Normative values of voice quality in elderly speakers are sparse. Two studies have provided normative data for both APQ and NHR during a sustained /a/ specifically in older inviduals (Gorham-Rowan & Laures-Gore, 2006; Xue & Deliyski, 2001). The mean for NHR was 0.19 in Xue and Deliyski's study and 0.14 in Gorham-Rowan and Laures-Gore's study, when combining data for males and females. As for APQ, Xue and Deliyski obtained a mean of 4.04%, while Gorham-Rowan and Laures-Gore obtained a mean of 2.63%. These discrepancies could partly be related to the fact that the former study had an age mean higher by approximately five years when compared to the latter, which could explain the more severe perturbation values. Our participant sample had a mean age in between (72 years old), but the values for NHR and APQ were closer and even higher to those of the older speakers, indicating a more severe perturbation in the acoustic signal (Xue & Deliyski, 2001).

Although norms specifically for older speakers were not found for CPPS, some studies have provided information on values that can be expected from vocally healthy versus dysphonic adult speakers. Based on a reading sample, Sauder et al. obtained a mean value of 20.11 dB (SD=1.27 dB) in 70 non-dysphonic speakers and 17.49 dB (SD=1.52 dB) in 100 subjects with heterogeneous voice disorders (Phadke et al., 2018). Based on these results, our sample had a mean CPPS representative of dysphonic speakers for the reading task (mean CPPS=16.29 dB, SD=1.46 dB). Watts et al. obtained a mean CPPS of 20.07 dB (SD=3.33 dB) for a reading task and 22.86 dB (SD=4.07 dB) for a sustained vowel, when combining dysphonic and non-dysphonic subjects. Our results showed lower values, especially for the reading task, which is indicative of more dysphonic voices. On the other hand, our results are not consistent with those reported by Phadke et al.,

who obtained a mean CPPS of 13.8 dB (SD=2.1 dB) for a sustained vowel and 10.4 dB (SD=1.5 dB) for comfortable speech in 84 female teachers with no voice complaints (Phadke et al., 2018). Since a higher CPPS value is indicative of a better voice quality, it is not clear why we obtained higher values in our presbyphonic sample. Although we followed the procedures described by Phadke et al. obtain CPPS values, it is possible that other differences in methodologies could explain these discrepancies.

Normative values for SPL are challenging to obtain because many factors influence this measure, such as the distance at which the measure is taken and the level of background noise. Sundarrajan et al. found a mean of 77.7 dB during a reading and speaking task in older speakers. (Sundarrajan, Huber, & Sivasankar, 2017). The mean SPL obtained in our sample of presbyphonic participants was much lower (71.7 dB, SD=4.5 dB). Of note, the microphone in our experiment was placed at a slightly greater distance (14 cm) when compared to Sundarrajan et al.'s study (6 cm), which could partly account for the difference. Our results are closer to those obtained by Baker et al. who found a mean SPL of 71.44 dB in five older individuals, compared to 76.23 dB in younger participants, for a syllable repetition task (K. K. Baker et al., 2001). This could be explained by the fact that some participants in Baker et al.'s study also presented with vocal fold bowing and incomplete glottal closure. The authors, who also found reduced EMG amplitudes for the TA and LCA muscles in the older participants, explained that a reduced loudness in older speakers could be partly attributable to muscle atrophy and peripheral denervation, as well as to a reduced neural drive to the laryngeal motoneurons (K. K. Baker et al., 2001).

Auditory-Perceptual Measures

Heman-Ackhah et al. provided normative data for perceptual ratings of overall severity in a large sample of dysphonic and non-dysphonic voices (Heman-Ackah et al., 2003). Although the ratings were not based on the CAPE-V, the raters used a VAS of 100 mm to assess the parameter of dysphonia/normality, similarly to the CAPE-V methodology. The results indicated a 10^{th} percentile at 10 mm, which the authors considered as the cut-off value for a dysphonic voice (Heman-Ackah et al., 2003). This is also consistent with Angadi et al. who considered 10 mm on the CAPE-V to be the clinical threshold for normality (Angadi et al., 2018). Based on this cut-off, only one participant out of 21 (5% of the total sample) presented with a perceptually normal voice in our sample. Overall severity for the other 20 participants varied between 14 mm and 89 mm on the CAPE-V scale.

Aerodynamic Measures

Zraick et al. provided normative data by age group for all aerodynamic measures assessed with the Phonatory Aerodynamic System (PAS) (Zraick et al., 2012). Subglottal pressure in females with perceptually normal voices ranged between 3.79 and 18.39 cmH_20 (mean=7.78 cmH_20, SD=4.23 cmH_20). None of the females in our presbyphonic sample exceeded this range, and the mean was similar. Normative values for subglottal pressure in males were less variable and ranged between 2.02 and 9.46 cmH_20 (mean=6.31 cmH_20, SD=1.94 cmH_20) (Zraick et al., 2012). Two participants in our sample exceeded this range, and the mean for males was slightly higher. This could be explained by a higher phonation threshold pressure in the presence of either a large glottal gap or a significant amount of supraglottic activity, which are both hallmarks of presbyphonia. As for mean airflow during voicing, Zraick et al. reported a range between 0.02 and 0.33 L/s in females (mean=0.13 L/s, SD=0.07 L/s) and a range between 0.01 and 0.83 L/s in males (mean=0.15 L/s,

SD=0.18 L/s) (Zraick et al., 2012). In our presbyphonic sample, none of the females exceeded the range provided by Zraick et al., but the mean was slightly higher. Similarly, none of the presbyphonic males exceeded the range, but the mean and median were much higher, indicating that more participants had a mean airflow rate in the upper range, which is expected from patients with vocal fold atrophy, who often present with incomplete glottal closure. The normative data for aerodynamic resistance also shows a high inter-subject variability. Values for older females varied between 18 and 211.60 $cmH_20/L/s$ (mean=79.05 $cmH_20/L/s$, SD=52.05 $cmH_20/L/s$); and values for older males varied between 4.18 and 900.09 $cmH_20/L/s$ (mean=137.31 $cmH_20/L/s$, SD=221.50 $cmH_20/L/s$)(Zraick et al., 2012). None of the participants in the presbyphonic sample were outside of these ranges, but the means for females and for males were both lower than the normative values.

Even though the normative data described above were computed on individuals with perceptually normal voices, the resulting ranges were very large. This is indicative of high inter-subject variability for these measures, even in normal speakers. This emphasizes the importance of conducting a multidimensional assessment in order to better interpret aerodynamic parameters, their impact on other voice outcomes, and if the measures are indicative of abnormal physiology.

Self-Assessment Measures

Arrfa et al. calculated normative values for the VHI-10 and the results indicated that a score above 11 is indicative of a voice-related handicap (Arffa et al., 2012). In our sample, only three participants had a score below 11 at baseline, meaning that 86% of the participants were perceiving a meaningful voice-related handicap. A GFI score higher than 4 was determined as the cut-off value to differentiate normal and abnormal voices (Bach et al., 2005). 100% of the presbyphonic participants in our study had a GFI score greater than this threshold. This is expected

considering that this questionnaire is sensitive to deficits specifically experienced by patients with glottal insufficiency, as seen in presbyphonic patients. Lastly, the Communicative Participation Item Bank (CPIB) is a fairly new measure designed to assess the construct of communicative participation across various communication disorders and situations (Baylor et al., 2013). Although few studies are available for comparison, the form allows for a conversion to a standard T-score with a mean of 50 and a standard deviation of 10, based on the calibration sample (701 participants with four different types of communication disorders) (Baylor et al.). Base on the T-score, three presbyphonic participants in our sample presented with a score above one SD from the calibration mean and three participants had a score below one SD from the calibration mean. The rest of the participants had a score close to the mean, indicating that the sample was not too different from a larger sample of patients with different types of communication disorders (more specifically: multiple sclerosis, Parkinson's disease, amyotrophic lateral sclerosis and head and neck cancer) (Baylor et al.).

Clusters 1 and 3: The Impact of Age

The cluster analysis led to the formation of three clusters with distinct respiratory and voice characteristics. The formation of clusters 1 and 3 seem to have been greatly influenced by age: while cluster 1 was comprised of older individuals, with lower raw respiratory values, cluster 3 was comprised of younger individuals, with higher raw respiratory values. Moreover, despite very different raw respiratory means, clusters 1 and 3 had similar standard respiratory means, indicating similar respiratory health. Cluster 1 could therefore be referred to as the "healthy older cluster", and cluster 3 as the "healthy younger cluster", based on percent predicted values of respiratory function. Cluster 3 was comprised exclusively of males, which also explains the higher raw respiratory values.

The older cluster had the lowest mean for the integrity factor, while the younger cluster had the greatest mean. The integrity factor is a weighted measure including bowing index (multiplied by -1) and phase symmetry as rated on the VALI assessment form. Cluster 1 had the lowest aerodynamic factor (a weighted measure of subglottal pressure and glottal airflow) while cluster 3 had the highest one. Results from the correlations showed that the raw respiratory factor, the integrity factor, and the aerodynamic factor were significantly inversely correlated with age and therefore confirm the role of age in shaping these older and younger "healthy" clusters. Awan et al., who found correlations between vital capacity and some measures of vocal function, highlighted that these results do not necessarily imply a causal relationship, but may merely reflect a simultaneous decline in both the laryngeal and respiratory system with aging (Awan, 2006). This seems to be a plausible explanation regarding the results for the integrity and aerodynamic factors in clusters 1 and 3. Because age was normally distributed in the sample, with a wide range from 56 to 91 years old, its impact on these voice parameters was easily detectable in the cluster and correlation analyses.

The fact that age was correlated with the integrity factor was expected and reflects the effect of age-related changes in the larynx. In fact, a loss of fibers in the TA muscle has been reported in older vocal folds (Kersing & Jennekens, 2004; Malmgren et al., 1999), which is thought to be partly related to cell injury and death and a regeneration and re-innervation process that is insufficient to compensate for the loss of fibers as aging occurs (Malmgren, Lovice, & Kaufman, 2000). Age-related changes do not only affect the TA muscle, but also the composition of the layers of tissue surrounding it. For example, older vocal folds were found to contain a decreased number of cells forming the epithelium, resulting in a reduced thickness when compared to younger vocal folds

(Goncalves et al., 2016). These changes lead to atrophied vocal folds, which in turn is reflected in the bowing index. The fact that the parameter of phase symmetry varied closely with bowing index suggests that it represents another indicator of the severity of the laryngeal changes. Specifically, changes in the structure of the vocal folds affect its vibration with increased asymmetry as bowing increases.

Cluster 2: The Impact of an Impaired Respiratory Function

Cluster 2 had a mean age in between clusters 1 and 3 and a greater deficit in both respiratory factors as well as in some of the voice factors. This suggests that some voice factors were not influenced by age but may have been influenced by a lower respiratory function. Although cluster 2 was predominantly comprised of women, it is unlikely that gender drove the formation of this cluster, considering that women in the total sample had higher or similar standard respiratory values when compared to males, and that cluster 2 has the lowest standard respiratory factor. Therefore, it can be hypothesized that participants in cluster 2 present with specific respiratory and voice characteristics that tend to vary together. These relationships are elucidated in the following sections, and are schematized in Figures 30 and 31.

On Resistance and Hyperfunction

A linear relationship approaching significance was found between FVC and FEV1 and resistance, and therefore a lower respiratory function could explain why cluster 2 also had the highest mean for the resistance factor. An increase of one standard deviation in FEV1 (0.84 L) or in FVC (1.07 L) was associated with a decrease in aerodynamic resistance of 0.39 standard deviation, which corresponds to approximately 12 $cmH_20/l/s$. This is very close to the difference in resistance that was observed between a group of individuals with presbyphonia and a group of age and gender

matched subjects without presbyphonia in Angadi et al.'s study (Angadi et al., 2018), thus suggesting a clinically meaningful relationship.

Participants who present with a reduced amount of air available for phonation are potentially increasing their laryngeal airway resistance (R_{law}) for airflow conservation purposes (Zhang, 2016c). Increasing R_{law} and reducing glottal airflow prevents them from reaching their expiratory reserve volume too quickly, where natural expiratory pressures are low and muscular respiratory requirements are high (Zhang, 2016c). Participants with a lower respiratory function were therefore compensating for a decreased FVC by making laryngeal adjustments to remain at a higher lung volume and make use of natural expiratory pressures to decrease respiratory effort. Increasing R_{law} could also allow speakers to build up subglottal pressure with less expiratory effort. Solomon et al. found that R_{law} increased linearly during a maximum phonation task in some participants, as lung volume decreased (Solomon, Garlitz, & Milbrath, 2000). SPL was controlled during the experiment, requiring subjects to make adjustments to maintain the same loudness even at low lung volume. Increasing resistance allows speakers to maintain adequate subglottal pressure with less airflow and therefore with less expiratory effort (Zhang, 2015). However, as mentioned by Solomon et al., compromising glottal airflow is not characteristic of comfortable speech, during which speakers usually maintain a constant airflow (Iwata, Von Leden, & Williams, 1972; Solomon et al., 2000; Zhang, 2016c). This strategy may nonetheless be employed by patients who present with a decreased respiratory drive and who have difficulty building up sufficient subglottal pressure for speech. While increasing R_{law} may help build up subglottal pressure, an elevated R_{law} also generates the need for a greater subglottal pressure in order to overcome the increased phonation threshold pressure and initiate or maintain vocal fold vibration (Higgins & Saxman, 1991; Zhang, 2015). In this case, patients may also attempt to augment their

expiratory effort in addition to maintaining a low glottal airflow, which can lead to an imbalance between subglottal pressure and glottal airflow and further increase resistance and laryngeal effort.

Increasing resistance therefore seems to, at least partly, serve the purpose of compensating for a decreased respiratory function. The paradox in the case of patients with presbyphonia is that they also present with age-related changes affecting the vocal folds' molecular composition and their ability to fully close and generate adequate resistance to airflow is diminished (Hammond et al., 2000). Involvement of the supraglottic structures has been observed in subjects with presbyphonia as a compensation for this loss in resistance (Higgins & Saxman, 1991). This could explain why hyperfunction was not correlated with respiratory function, since it was used by many subjects as a compensation for laryngeal deficits and in an attempt to normalize resistance, regardless of their respiratory status. However, subjects who presented with the highest resistance values all presented with a clinicially significant amount of hyperfunction. Although a correlation between the hyperfunction and handicap factors was present, the relationship between resistance and handicap was much stronger. In other words, hyperfunction does not necessarily cause a meaningful increase in handicap unless it is associated with an excessive aerodynamic resistance and therefore in the presence of a significant imbalance between subglottal pressure and glottal airflow. Even so, a smaller ventricular gap has been found to be associated with a decrease in phonatory efficiency because of its effect on transglottal pressure (Alipour & Scherer, 2012; Kniesburges et al., 2017), and therefore more research is needed to fully understand the positive and negative effects of supraglottic activity in patients with presbyphonia.

On Pliability

The pliability factor was comprised of the following VALI parameters: amplitude of vibration (right and left vocal folds), amplitude of mucosal wave (right and left vocal folds), and regularity of vibration. It had a strong relationship with the resistance factor: a greater resistance was associated with a reduced pliability. This finding is not surprising, considering that a higher resistance, which was accompanied by hyperfunction, is likely to restrain vocal fold vibration and mucosal wave movement. In fact, a study by Alipour and Scherer on excised larynges showed that medialization and motion of the false vocal folds generates positive air pressure exerting a vertical force on the true vocal folds, potentially inhibiting lateral movement of the tissue (Alipour & Scherer, 2012). This is also prone to affect their periodicity, which could explain why the parameter of regularity of vibration was associated with vibration and mucosal wave amplitude in the factor analysis, despite it being an indicator of severity similarly to bowing index and phase symmetry. Some presbyphonic individuals may also attempt to compensate for an atrophied TA by relying on cricothyroid muscle (CT) activation (Zhang, 2019), which increases the anteroposterior stiffness of the true vocal folds and might also reduce their pliability. On the other hand, a severely decreased resistance is associated with flaccid vocal folds and an abnormally increased vibration and mucosal wave amplitudes. This is caused by age-related changes in the tissue of the vocal folds, affecting their viscoelastic properties (Hammond et al., 2000). However, the pliability factor was not correlated with age nor with the integrity factor despite being affected by the same age-related changes in vocal fold tissue and TA muscle. This may be because of the compensation strategies (CT activation and supraglottic hyperfunction), which can increase aerodynamic resistance and reduce tissue pliability despite important vocal fold atrophy.

Cluster 2, which included subjects with the lowest respiratory function, had a substantially smaller mean on the pliability factor when compared to the two other clusters. Moreover, MEP was found to be a good respiratory predictor for the pliability factor: a larger MEP was associated with a greater pliability. These findings could be caused by the fact that a greater expiratory pressure 1) reduces the need to increase resistance as a compensation for insufficient respiratory capacity, which would in turn restrain pliability, and 2) induces a greater lateral movement of the vocal folds (greater vibration and mucosal wave amplitude) if resistance is held constant (Zhang, 2015).

While a significantly reduced pliability is not desirable for phonation and can be a symptom of insufficient airflow or excessive laryngeal stiffness or tension, a markedly large vibration and mucosal wave amplitude is most likely to be associated with a lack of vocal fold resistance, a predominantly open phase, and a large glottal airflow. As is was demonstrated in a computational model, an increase in subglottal pressure without a sufficient increase in glottal resistance further pushes the vocal folds apart, increases mean glottal airflow rate and glottal area amplitude (Zhang, 2015). This is particularly relevant in the case of presbyphonic patients, for who increasing resistance of the true vocal folds is sometimes physiologically unachievable because of the degeneration of the TA muscle (Kersing & Jennekens, 2004; Malmgren et al., 1999), but also because other changes in the vocal fold tissue. For example, collagen networks, although quantitatively increased, are disorganized and not fulfilling their role as well as in younger vocal folds, therefore leading to a decreased tissue resistance (Hammond et al., 2000).

In summary, a decreased respiratory function leads to increased resistance as a compensation strategy, which in turn can limit vibration and mucosal wave amplitudes and hinder regularity of

vibration. On the other hand, an imbalance between the respiratory drive and the ability of the vocal folds to provide sufficient resistance against it may lead to abnormally large vibration and mucosal wave amplitudes, especially in vocal folds that are already more pliable because of age-related changes. A balance between true vocal fold resistance and respiratory effort needs to be achieved for an efficient phonation, which can represent a challenge in patients with presbyphonia who are likely to present with impairments in both the laryngeal and the respiratory system.

On Handicap

In addition to having lower means for the respiratory and pliability factors, as well as higher means for the resistance and hyperfunction factors, cluster 2 had the highest handicap value. The handicap factor was a weighted measure of three disease-specific self-assessment questionnaires: VHI-10, GFI, and CPIB. Thus, subjects with a lower respiratory function also had a higher perceived voice handicap. These results are in line with those of Hunter et al., who found that female teachers with a lower FVC also had a higher Voice Fatigue Index factor (Hunter, Maxfield, & Graetzer, 2019). However, only raw values were found to be predictors of vocal fatigue in Hunter et al.'s study, while standardized values were better predictors in our sample. This could be because of the different subject samples: Hunter et al.'s experiment was conducted on otherwise healthy teachers, who would be expected to present with a healthy respiratory function. Percent predicted values may therefore not have presented with sufficient variation within the sample to detect a relationship with vocal fatigue. Alternatively, variations in raw spirometry values, driven by factors such as height, age, and weight, allowed for the detection of a relationship between raw pulmonary function and vocal fatigue (Hunter et al., 2019). On the other hand, a significant proportion of the participants in our study with presbyphonia presented

with a decreased respiratory health, with one third of the participants having FVC% and FEV1% values below 80% of the predicted value, which revealed that a reduced respiratory health was associated with a greater vocal handicap, in this sample.

Standardized respiratory values were stronger predictors of perceived vocal handicap than raw respiratory values. Raw values, representative of the amount of air available and the strength of the respiratory muscles, may impact handicap through its direct effect on phonation physiology (on resistance, hyperfunction, and pliability). On the other hand, standardized respiratory values inform on the health of the respiratory system, which may impact perceived handicap through both a physiological pathway and a general impairment pathway. For example, individuals with poor respiratory health may already experience impairment from shortness of breath (SOB), which can be exacerbated when speaking. This is likely to be the case in patients with presbyphonia, who present with incomplete glottal closure and are susceptible to run out of air quickly when speaking. SOB caused by the decline in respiratory function is likely to aggravate the discomfort associated with vocal atrophy and could lead to more avoidance behaviors when it comes to communication situations, making the individual less functional in that they are less inclined to participate. Many questions on the voice-related self-assessment questionnaires pertain to participation, as it is a crucial aspect of health-related quality of life (McDougall, Wright, & Rosenbaum, 2010). To use the International Classification of Functioning, Disability and Health's (ICF) terminology (*International Classification of Functioning, Disability and Health*, 2001), because the impairment affects many body functions and structures (which in addition are inter-related), the resulting participation restrictions are more important .

Individuals who are generally less healthy may also live the burden of their voice disorder more negatively because of personality traits. A study by Kubzansky et al. using longitudinal data demonstrated that a "pessimistic explanatory style" (the way an individual reacts to events in their life) was associated with a faster decline in FEV1 and FVC over time (Kubzansky et al., 2002). Alternatively, an optimistic explanatory style was associated with better pulmonary health and a slower decline in FEV1 and FVC over time (Kubzansky et al., 2002). The relationships were strong and were even comparable to the effect of smoking on lung function. Optimism is thought to have direct physiological effects on health (McEwen & Stellar, 1993), but can also influence health via indirect psychological and social pathways such as through a good support networks as well as proactive and healthy habits (Kubzansky et al., 2002). Although not assessed in our study, these characteristics could have influenced the participants' perception of their voice disorder: more optimistic subjects may experience less handicap when compared to their more pessimistic counterparts, in addition to having a better respiratory function.

Figure 30 is a handicap-focused model and summarizes the relationships between respiratory and voice function and how they impact perceived voice handicap in patients with presbyphonia, based on the studied sample. Wider arrows indicate stronger relationships, based on the Spearman correlation coefficients.

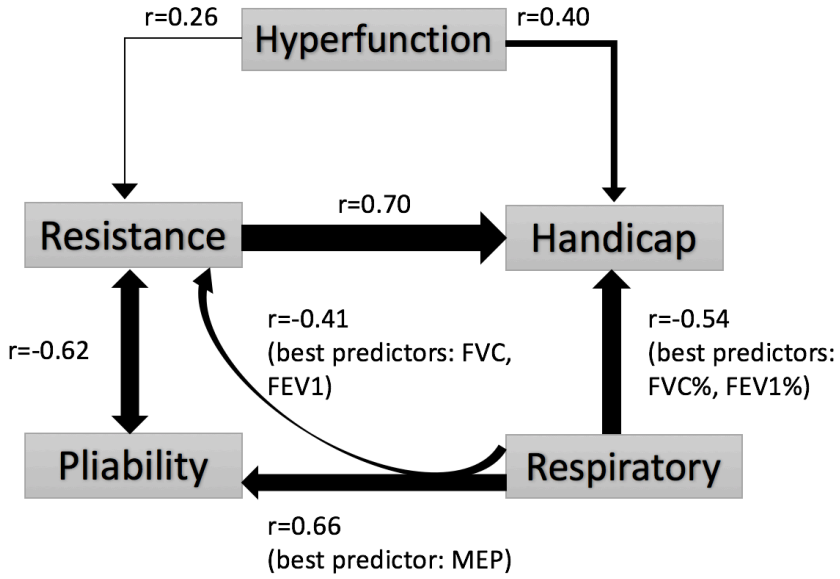


Figure 30. Voice handicap-focused model of the effect of respiratory function in presbyphonia. Wider arrows indicate stronger relationships, based on Spearman correlation coefficients (r).

Voice Quality

The two factors related to voice quality (perturbation and speech) were not included in the handicap model. In fact, no relationship was found between voice quality and perceived handicap in the studied sample, suggesting that the sound of the voice does not functionally affect patients with presbyphonia as much as the physical demands required to produce voice. This is despite measures of perturbation and noise being increased when compared to norms for young and middle-aged adults, and similar to other values reported for older individuals (Xue & Deliyski, 2001).

Respiratory function was not found to be a predictor for voice quality in our sample, which was only correlated with the integrity factor: a decreased integrity of the vocal folds was associated with more perturbation in a sustained vowel (smaller NHR and APQ values). This could explain why the older cluster (cluster 1), which had the lowest integrity value, also had the highest

perturbation value. Surprisingly, cluster 1 also had the highest mean speech factor. This was not expected, considering the inverse correlation between the speech and perturbation factors. The speech factor included SPL during conversation and CPPS during reading, and was therefore a measure of loudness and quality of speech during functional tasks. As it could be expected, a higher perturbation during the sustained vowel was correlated with a reduced voice quality and loudness during more complex tasks such as reading and speaking.

As shown in Figure 31, the aerodynamic factor (comprised of subglottal pressure and glottal airflow) was positively correlated with the integrity factor. However, it is very likely that this relationship was mainly driven by age, as both of these factors had an inverse correlation with age. The model also shows that age was correlated with respiratory function (raw values), which was expected. However, respiratory function did not impact voice factors in this model. The fact that respiratory function was not correlated with the aerodynamic factor is somewhat surprising, considering that subglottal pressure is in large part determined by the alveolar lung pressure (Zhang, 2016b). However, patients with a decreased respiratory function may still be able to achieve normal levels of subglottal pressure, but at the cost of great respiratory and/or laryngeal effort and potentially in an inconsistent manner. Another surprising result, in our small sample, is the fact that neither the respiratory factors or the aerodynamic factor were correlated with the speech factor, encompassing SPL during speech. The reason for this could also reside in the use of compensatory strategies, potentially allowing patients to achieve a conversational loudness without adequate airflow and respiratory drive, by increasing their laryngeal airway resistance. In fact, at low subglottal pressure, an increase in vocal fold resistance leads to a higher maximum flow declination rate (MFDR), which is a predictor of intensity (Zhang, 2016a, 2016b).

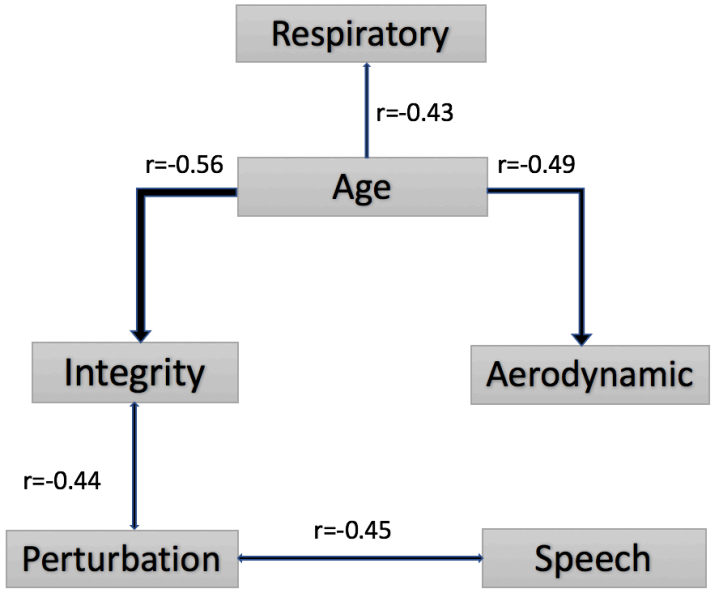


Figure 31. Voice quality-focused model of the effect of age in presbyphonia. Wider arrows indicate a stronger relationship, based on Spearman correlation coefficients (r).

Limitations for Aim 2

Factor and cluster analyses are usually conducted on large sample sizes, which was not the case in this study. Although these methodologies were helpful in extracting meaningful information from the multiple outcome measures, their results would be more robust with a much larger sample size. The small sample size could also explain why some of the relationships between respiratory and voice outcomes were not found to be significant despite having the same directionality as other significant results. Only the relationships with the largest effect sizes reached significance, which was to be expected in a pilot study. In order to detect smaller effect sizes, a future study with more participants should be conducted.

Another limitation from this study resides in the fact that different voice outcomes were recorded during different assessment tasks, and phonation physiology is likely to differ from one task to another. For example, laryngeal parameters were rated based on a sustained /i/ during

videostroboscopy, while aerodynamic variables were recorded during syllable repetition. Not only are these tasks different from each other, but they are also different from phonation during running speech. To solve this issue, a future study could attempt to measure all voice outcomes during a single task, although this would represent a significant methodological challenge.

Discussion Aim 3a

What is the effect of respiratory muscle strength training on respiratory and voice outcomes in patients with presbyphonia?

The rationale behind adding respiratory muscle training to voice therapy was that improving respiratory function would lead to improved voice outcomes. However, the mechanism behind this hypothesis remains unclear because of a lack of studies on respiratory exercises and voice outcomes. A recent literature review conducted by our research team concluded that, in order to better our understanding of the mechanism of action, future studies should 1) isolate respiratory training as the independent variable of the study, 2) report a comprehensive array of voice and respiratory outcomes, 3) assess the relationship between changes in respiratory and voice function, and 4) compare the effect of "voice therapy only" to "voice therapy + respiratory training" (Desjardins & Bonilha, 2019). The present study followed these recommendations and the results are discussed below.

Effect on Respiratory Measures

One of the limitations of the literature when assessing the effect of respiratory exercises on voice was that, in many studies, few or no respiratory outcomes were assessed. Therefore, even if change in phonation was observed, it was unclear whether it had been caused by a change in respiratory function (Desjardins & Bonilha, 2019). Just like phonation, respiration is multi-dimensional and can be assessed with various measures. In this study, measures of maximal effort were recorded, including maximal respiratory muscle strength (MIP and MEP) and pulmonary function tests (FVC, FEV1, and FEV1/FVC). Raw measures as well as LLN and percent predicted values were considered.

A mean increase in both MIP and MEP was observed in the three intervention groups. As expected, the EMST group had the largest increase in MEP, and this is consistent with current evidence supporting that EMST induces significant changes in MEP (Desjardins & Bonilha, 2019). The IMST group increased their MIP, but interestingly the effect size was slightly less than for the VFE group. Various studies have confirmed the efficiency of IMST for improving MIP (Beaumont, Forget, Couturaud, & Reychler, 2018; Lotters et al., 2002; Mueller, 2013; Ray, 2018; Silva et al., 2013). However, two subjects in the IMST group from our study did not have a meaningful improvement in MIP, which could potentially be explained by a ceiling effect. In fact, one of the subjects who did not improve was a recreational swimmer with a baseline MIP much higher than the predicted value for their age, gender, and weight. On the other hand, the other subject who did not improve had a baseline MIP very close to the LLN and had had a recent injury to the ribcage, which may have limited the gains from IMST.

Improvements in MIP and MEP in the VFE group were likely caused by training effects, since respiratory muscle strength measures were taken prior to and after each therapy session. Measurement error could also partly explain these results, although the increase was larger than the SEM in two out of three subjects for both MIP and MEP. It is also possible that, for some participants, VFE exerted a sufficient load on the respiratory muscles to induce a significant increase in strength. This could also explain the increase in MIP for some subjects in the EMST group, and the increase in MEP for some subjects in the IMST group, although these could also be the result of the RMST . In fact, previous studies have reported changes in MIP following EMST and changes in MEP following IMST and there is therefore some evidence for cross-training effects (Duruturk, Acar, & Dogrul, 2018; Gosselink, Kovacs, Ketelaer, Carton, & Decramer, 2000; Ray, 2018; Vorona et al., 2018).

No improvement in pulmonary function measures (raw values or predicted percent values) were observed in any of the intervention groups. This is consistent with many studies in the respiratory literature having found no improvement in spirometry measures after a RMST intervention (Cerny et al., 1997; Mota et al., 2007; Tsai et al., 2016; Weiner, Magadle, Beckerman, Weiner, & Berar-Yanay, 2003). Unless the decreased FVC was a result of respiratory muscle weakness, a change in pulmonary function would not necessarily be expected despite an increase in MIP and/or MEP. As for FEV1, another explanation could reside in the fact that, beyond a certain alveolar pressure at a given volume, maximal flow rate is mainly determined by the elastic recoil from the lungs and therefore an increase in MEP may not affect flow rate. This justification was suggested by Chiara et al. (Chiara, Martin, Davenport, & Bolser, 2006) who found limited improvements in FVC and FEV1 in subjects with multiple sclerosis, despite enhanced MEP following an EMST intervention.

Effect on Laryngeal Structures and Function

The majority of participants experienced a reduction in the amount of bowing of the vocal folds following the intervention, except for two participants, both in the IMST group (participant 12 and 16). Interestingly, these subjects also presented with the lowest compliance (76% and 77% of daily home practice completed). The largest effect sizes in terms of bowing reduction was found in the VFE group. Altogether, these results suggest that VFE does help counteract the atrophy of the TA muscle, and that more benefits regarding vocal fold atrophy can be gained from spending more time doing the voice exercises. However, as it will be discussed in the following sections, merely improving the integrity of the vocal folds with VFEs may not be sufficient to reduce the burden of presbyphonia and enhance communicative participation in all patients, and the addition of RMST (more specifically IMST), seems to provide benefits beyond those achieved with VFEs only.

Although no significant effects for bowing index were found based on the ANVOA results, a large effect size was noted in the VFE group, and a greater sample size may have revealed a significant time by group interaction effect. Nonetheless, this lack of significance is consistent with Kaneko et al.'s results, who reported no change in bowing index following a VFE intervention in 16 patients with vocal fold atrophy, despite improvement in various other voice measures (Kaneko et al., 2015). The authors hypothesized that vocal function might have been improved despite a change in bowing due to enhanced coordination between the respiratory and the laryngeal systems (Kaneko et al., 2015). As a matter of fact, improved coordination is one of the objectives of the VFEs, and as it will be explained later, may be further achieved by performing IMST prior to the voice exercises. This could also elucidate why the IMST group presented with the largest improvements on self-assessment outcomes despite having the smallest decrease in bowing index.

The VFEs also helped normalize vibration and mucosal wave amplitude in two participants in the VFE group, while at the same time reducing supraglottic activity. These results are consistent with the greater aerodynamic resistance obtained for these participants. Two participants in the IMST group also decreased their hyperfunction, as well as one participant in the EMST group. The decrease in bowing index and in supraglottic activity could help clarify why some participants in the IMST and in the VFE groups improved their perceived handicap despite an increase in aerodynamic resistance. In fact, it was shown in Aim 2 that a high resistance was strongly associated with a higher perceived handicap (Figure 30). The handicap-focused model also showed that hyperfunction was a predictor, to some extent, for both resistance and handicap. Results from this aim suggest that voice therapy helps reduce hyperfunction and rebalances resistance so that it is provided by an enhanced approximation of the true vocal folds and not

mainly through compensatory strategies. However, these benefits were not observed in the EMST group, in which no subjects were able to increase their resistance with a meaningful effect size, despite a reduction in bowing observed for all three participants. This could be due to the inability of the vocal folds to resist the increased subglottal pressures triggered by the EMST intervention, which could also be described as a lack of coordination between the respiratory and laryngeal systems.

Effect on Acoustic Measures and Auditory-Perceptual Judgements of Voice Quality

No meaningful change in SPL was observed in the EMST or the VFE groups. This could be explained by a lack of effect of the intervention on intensity, or in the case of the VFE group, by the fact that participants presented with a small deviation from normal at baseline. On the other hand, participants in the IMST group presented with larger deviations from normal and consequently there was greater room for improvement. Only participants in the IMST group achieved moderate to large effect sizes and raw changes above the SEM for this outcome measure. Participants 2 and 12 increased their SPL, while subject 16 decreased meaningfully. The changes occurred in opposite directions and this could explain why no significant result was found in the ANOVA when means were compared, despite the fact that these participants experienced a change towards a normalized SPL value.

SPL and subglottal pressure are closely related, since an increase in subglottal pressure is the main strategy used by healthy speakers to increase loudness of the voice (Zhang, 2016a). However, a computational study by Zhang showed that, at low subglottal pressures, laryngeal adjustments improving vocal fold resistance enhance vocal fold contact which in turn induces a higher MFDR and an increase in intensity (Zhang, 2016a, 2016b). This could explain why subject 2, in the IMST

group, who decreased their subglottal pressure but increased their aerodynamic resistance, was able to increase conversational SPL.

Although a reduced intensity is a common complaint of patients with presbyphonia, an inability to produce a soft vocal quality can also be an issue in this population and can occur because of two reasons. Firstly, a large glottal gap increases the subglottal pressure required to initiate and sustain phonation. Patients with severe glottal atrophy may therefore not be able to produce voice at low subglottal pressures and intensity. Alternatively, patients who compensate with excessive supraglottic activity are also increasing their phonation threshold pressure and reducing their ability to produce a soft voice (Zhang, 2015). A reduction in subglottal pressure and in SPL, if these were abnormally elevated at baseline as it was the case for participant 16 in the IMST group, can therefore be indicative of an improved balance between the respiratory and laryngeal systems.

Auditory-perceptual judgments of overall severity confirmed that subjects in the IMST group were potentially able to achieve a better phonato-respiratory coordination. In fact, single-subject and group effect sizes showed that the largest improvements occurred in the IMST group, followed by the EMST group, and lastly the VFE group. Regarding CPPS during reading, subjects in the VFE group did not improve, while one subject in the IMST group and two subjects in the EMST group made sizable improvements. Large group effect sizes and raw improvements above SEM were noted in the IMST and the EMST groups for this variable, which was not the case for the VFE group. One of the objectives of RMST was to provide subjects with more air available for speech, since sufficient airflow and subglottal pressure are necessary to sustain the vibration of the vocal folds. The hypothesis was that training the inspiratory muscles would allow to increase the inspiratory volume and to have a better control of the airflow, and that training the expiratory muscles would

allow to make a better use of the expiratory reserve volume (ERV). Subjects in both respiratory training groups increased their vocal quality during conversation and reading to a greater extent that subjects in the VFE, which could suggest that they had more air available to sustain a sufficient subglottal pressure and therefore a healthier vibration and greater voice quality during those tasks. In addition, the lack of improvement in perceived handicap in the EMST group suggest that participants in this group may have been using suboptimal strategies in terms of respiratory effort, as it will be discussed later.

Different results were obtained regarding the voice quality of a sustained vowel. The only meaningful difference between the groups was the larger pre- to post-treatment effect size found in the VFE group for the APQ measure. Participants in the VFE group spent more time practicing sustained sounds and therefore it is fitting that they were able to improve the vocal quality of the sustained /a/ slightly more than the other groups. Studies having assessed the effect of RMST on vocal quality are very sparse and therefore few comparisons can be made. Johansson et al. tested an EMST intervention on five subjects with multiple sclerosis and reported that three out of the five subjects experienced a decrease in coefficient of variation of F0 during a sustained vowel task, with effect sizes ranging from d=-0.7 to d=-0.12 (Johansson, 2012). Pereira et al. found no significant improvement in jitter, shimmer, GNE, and noise following an EMST intervention in teachers with and without voice disorders, and even found a significant increase in shimmer in the EMST group (Pereira, 2015). However, EMST was not combined with voice exercises in Pereira et al.'s experiment. As it was discussed previously, augmenting subglottal pressure without making any laryngeal adjustments results in an increase in glottal airflow and consequently in turbulence and noise (Zhang, 2015). This further highlights the importance of combining voice exercises with respiratory training.

Measures of vocal quality during speech and reading are more ecologically valid than those obtained from a sustained vowel task, and this is especially relevant for presbyphonic patients. In fact, Huber found that the effects of aging on the voice were more obvious when utterance length increased (Huber, 2008). For example, lung volume excursion was greater in older adults when compared to younger adults in longer utterances, but not in shorter ones. This implies that long utterances tax the respiratory system to a greater extent and older speakers have to provide more effort in order to compensate for changes in both the laryngeal and respiratory systems. Assessing vocal quality during conversation and reading tasks is therefore crucial in understanding the impact of the intervention on age-related deficits. This provides a potential explanation as to why greater improvements in overall severity and CPPS reading were found in the RMST groups, while these differences were not observed in the sustained vowel task.

Effect on Aerodynamic Measures

Aerodynamic measures are crucial in understanding the mechanistic effect of RMST interventions because they are the result of both respiratory and phonatory activities. However, they have to be interpreted carefully and while considering the whole clinical picture of the patient because an increase or a decrease in those measures do not necessarily imply that a meaningful improvement occurred.

Participants in the IMST group decreased their subglottal pressure following the intervention. This reduction was accompanied by a reduction in glottal airflow in two of the participants, who also considerably increased their aerodynamic resistance. Participant 16, in the IMST group, had a different response to the intervention: the decrease in subglottal pressure was accompanied by a slight increase in glottal airflow and a decrease in aerodynamic resistance. This could be

explained by the reduction in supraglottic activity, which was elevated at baseline and probably causing the abnormally elevated subglottal pressure observed at baseline by increasing the required phonation threshold pressure. Many patients with vocal fold atrophy engage supraglottic structures to reduce airflow and help build-up subglottal pressure, as discussed in Aim 2. Although this theoretically increases aerodynamic resistance, the resistance is provided mostly by compensatory mechanisms and not by a healthy vocal fold closure. In the case of participant 16, a longer intervention duration may have induced better results: once hyperfunction is reduced, the respiratory and the laryngeal system have to readjust their coordinated activity until an optimal balance between resistance, glottal airflow, and subglottal pressure is achieved.

In fact, the vocal folds modulate the flow of air that is directed towards them and respond to it by adjusting their resistance in order to maintain a glottal airflow that is adequate for phonation (Zhang, 2015). As subglottal pressure increases, a proportional increase in vocal fold approximation and/or stiffness needs to occur to maintain a small glottal airflow (Zhang, 2015). If subglottal pressure is unpredictable or too strong, the atrophied vocal folds may not be able to maintain their resistance against it and glottal airflow will likely be elevated. Optimal airflow conservation therefore relies on the balance between subglottal pressure and glottal resistance (Zhang, 2015). Training the inspiratory muscles is relevant in finding this balance because it promotes good airflow control: the inspiratory muscles have to be engaged during the expiratory phase of speech to control the natural expiratory pressures. The resulting subglottal pressure is more stable and can be more easily modulated by the vocal folds. Nonetheless, in the presence of substantial atrophy, a certain amount of supraglottic activity may be necessary to maintain a

reasonable airflow rate and to avoid excessive expiratory effort needed to produce a target subglottal pressure.

Participants in the EMST group slightly increased their subglottal pressure, but none of them reduced their airflow, nor had a meaningful increase in aerodynamic resistance, despite having baseline values below normal. It is possible that EMST helps increase subglottal pressure but also induces a slight increase in airflow because the vocal folds' valving function is impaired and cannot respond adequately to the intensified subglottal pressure. When subglottal pressure is increased against insufficient glottal resistance, two scenarios can occur: either the vocal folds are being further blown apart, resulting in a larger glottal airflow rate; or the larynx compensates with excessive hyperfunction. As it was mentioned above, a certain amount of supraglottic activity can be beneficial in reducing glottal airflow; however, excessive laryngeal hyperfunction can be a "defense mechanism" for erratic, and potentially elevated subglottal pressures (Rubin, Macdonald, & Blake, 2011) and can cause vocal fatigue and discomfort, in addition to hindering vocal fold vibration (Kniesburges et al., 2017).

This highlights the importance of voice exercises in this population: to strengthen the laryngeal muscles and help increase the ability of the true vocal folds to provide adequate resistance against subglottal pressure, while at the same time allowing to reduce supraglottic activity. There are two ways to augment vocal fold resistance: by enhancing vocal fold approximation, and by augmenting their antero-posterior stiffness (Zhang, 2015). Vocal fold approximation is regulated by the combined action of the lateral cricoarytenoid (LCA) and the inter-arytenoid (IA) muscles. In addition, the involvement of the thyroarytenoid (TA) muscle is necessary to achieve complete closure of the membranous vocal folds (Chhetri et al., 2012; Yin & Zhang, 2014). Enhancing

resistance by increasing vocal fold approximation is an effective strategy at low to medium subglottal pressures. At higher subglottal pressures, increasing stiffness becomes more efficient, and relies on the activation of the cricothyroid (CT) muscle (Zhang, 2015). Patients with presbyphonia tend to compensate for the atrophy of the TA muscle by relying mainly on CT muscle activation, which is not effective for providing adequate resistance and to control F0 in the lower register (Zhang, 2019). Building up the strength of the TA muscle and potentially reducing its atrophy is therefore important for enhancing vocal fold approximation and improving airflow modulation and F0 control. In fact, two of the participants who received only voice therapy increased their laryngeal resistance with large effect sizes. One of these subjects also increased their subglottal pressure and maintain a constant airflow, while the other participant maintained a somewhat constant subglottal pressure and reduced their airflow to a more normal value. Nonetheless, results from self-assessement measures suggest that adding IMST to voice exercises could lead to greater benefits for patients with presbyphonia, as it will be discussed in the next section.

Effect on Self-Assessment Measures

Despite the improvement observed in voice quality during conversation and reading, participants in the EMST group did not improve their self-assessment scores meaningfully. Importantly, an increase in VHI-10 scores, indicating an increased in perceived handicap, was observed for this group, with a medium effect size. On the other hand, the IMST group made meaningful improvement on all self-assessment questionnaires, with large to very large effect sizes. Although some improvement was also observed in the VFE group, improvements in the IMST group were much larger.

EMST promotes the use of the abdominal muscles to compress the air out of the lungs with greater force. As previously mentioned, the resulting increase in subglottal pressure is likely to be accompanied by an increase in glottal airflow if resistance of the vocal folds is not proportionally increased (Zhang, 2015). As a result, the speaker runs out of air quickly and has to further compress the ribcage to use the air from the expiratory reserve volume and maintain sufficient airflow and subglottal pressure for phonation. Preliminary findings from this study suggest that this may not be the preferable strategy to improve voice-related quality of life in this patient population. It could be hypothesized that increasing subglottal pressure and phonation duration by relying on the expiratory muscles comes at a great cost effort-wise, especially if the glottal gap remains significant. Subglottal pressure tends to decrease as lung volume decreases, as a result of reduced recoil pressure (Iwarsson, Thomasson, & Sundberg, 1998). The expiratory effort needed to maintain sufficient subglottal pressure below the resting expiratory level is even more important because the natural pressures become negative and need to be overcome by muscular effort (Huber, 2008). Despite training of the expiratory muscles, the required effort may be too great for older speaker in whom expiratory pressures are already decreased due to age-related changes in the respiratory system, more so than inspiratory pressures (Enright, Kronmal, Higgins, Schenker, & Haponik, 1994). Additionally, if the expiratory effort necessary to produce voice is not being maintained to successfully meet voice demands, more effort from the laryngeal muscles will be required in order to compensate.

It may be that a respiratory training capitalizing on the inspiratory strength, which is already better preserved in this population, is more effective in achieving an optimal balance between respiratory and laryngeal effort during phonation.

Activating the inspiratory muscles during the expiratory phase of speech allows the speaker to **maintain the natural recoil pressures at a higher level** for a longer period of time, which provides necessary airflow and subglottal pressure without requiring much effort from the expiratory or laryngeal muscles. It is also possible that participants were able to initiate phonation at a higher lung volume, to take advantage of greater recoil pressures. This strategy is already used by older speakers, who initiate speech at a higher lung volume when compared to younger adults, to compensate for reduced expiratory pressures (Huber, 2008; Huber & Spruill, 2008). As it was shown in Aim 2, participants with a lower respiratory function tend to increase their laryngeal airway resistance to slow down the decrease in lung volume and maintain higher natural expiratory pressure. However, this compensation was associated with laryngeal hyperfunction and with higher perceived handicap (Figure 30). Even though FVC was not increased following IMST, subjects may have learned to make a greater use of their inspiratory capacity, thus optimizing their phonatory volume. Having more air available to provide adequate airflow and subglottal pressure for phonation has been associated with decreased vocal fatigue (Hunter et al., 2019). In fact, more subjects in the IMST group improved on the questionnaire items related to vocal fatigue and discomfort, when compared to the EMST and the VFE groups: “I feel as though I have to strain to produce voice”; “Throat discomfort or pain after using your voice”; “Vocal fatigue (voice weakened as you talked)” (Appendix III).

In addition, the greater airflow and airflow control resulting from a stronger activation of the inspiratory muscles may have allowed the subjects to **extend utterance lengths between breaths**, without increasing laryngeal effort. In studies comparing older and young speakers, older adults were found to expend a greater percentage of the vital capacity per syllable and to produce less syllables per breath group when compared to younger adults (Huber, 2008; Huber & Spruill,

2008). A more effective inspiration prior to speaking, as well as a better use of the inspiratory muscles during phonation, may have reduced the number of pauses needed for breathing, which could in turn could have decreased the sensation of shortness of breath. Interestingly, no item from the administered self-assessment questionnaires addressed shortness of breath. The full version of the VHI does comprise the item "I run out of air when I talk" (Jacobson et al., 1997), however it is not present in the short version, which was used in the present study (Arffa et al., 2012). The CPIB questionnaire assesses different communication situations, some of which involve the production of more complex and thus longer utterances. More subjects in the IMST group improved on those items when compared to the other groups: "Having a long conversation with someone you know about a book, movie, show or sports event"; "Giving someone detailed information"; and "Trying to persuade a friend or family member to see a different point of view". A greater proportion of subjects in the IMST group also improved on the items "Communicating when you need to say something quickly" and "Getting your turn in a fast-moving conversation" (Appendix III). This could be related to an enhanced efficiency of the inspiratory muscles to take quick and yet efficient inspirations between utterances.

Correlations Between Changes in Respiratory and Voice Function

No correlations were found between improvements in MIP and MEP and improvements in voice outcomes, except for aerodynamic resistance which was positively correlated with change in MIP. Moreover, two subjects in the IMST group did not have a meaningful improvement in MIP, but did improve their voice and voice-related quality of life, more so than subjects in other groups. One reason for these findings may reside in the fact that spirometry and muscle strength measurements are maximal effort tasks, while phonation is not a maximal effort task. An improvement in maximal measures of respiration may not be required to witness improvements

in phonation mechanics. The participants were instructed to performed the respiratory training immediately prior to the voice exercises. We hypothesize that IMST may have prompted the engagement of the inspiratory muscles during the VFEs, and that an improvement in respiratory function rather than in maximal strength may be at the source of the changes in voice quality, aerodynamics, and voice-related quality of life. Similarly, even if FVC during spirometry did not improve, inspiratory volume prior to phonation may have been increased with IMST, thus providing more air during speech.

Limitations for Aim 3a

The hypothesis that lung volume initiation may have been increased could not be verified, because kinematic measures were not assessed in this study, which constitutes a limitation. Although integrity of the respiratory system, as measured by maximal effort tests, has been shown to impact voice outcomes (Aim 2), its functional aspects should also be considered and would help further elucidate the mechanisms of action through which IMST improves voice outcomes.

The very small sample size represents a significant limitation. The assumption of equality of variances and of normal distribution of the residuals were not always met, and therefore the results of the ANOVA may be less reliable. This is why interpretation of the results also relied on single-subject analyses as well, as well as on effect size computations. This is all the more important considering that comparisons of group means do not take into account the whole clinical picture of the participants. For example, when assessing aerodynamic measures and laryngeal parameters, both a decrease or an increase in the outcome may be desirable, based on the baseline status and the whole clinical picture of the patient.

Intra-subject variability is an important factor to consider, especially in patients with presbyphonia who already tend to present with an increased intra-individuals variability in voice measures. If the importance of the raw change in one outcome was not confirmed by a meaningful effect size, which was based on two baseline and two follow-up measurements (when available), it indicates a lack of stability in the improvement. A future study could take multiple baseline and follow-up measurement to increase the robustness of single-subject findings, and to assess the stability of the improvements following the intervention. Moreover, it is possible that a longer intervention duration could have yielded more stable results.

Discussion: Aim 3b

How do baseline measures of respiratory function influence the effects of RMST in patients with presbyphonia?

Results from Aim 2 confirmed that the state of the respiratory system (pulmonary function and respiratory muscle strength) does have an impact on phonation physiology and on perceived voice handicap. In fact, a lower respiratory function was associated with more functional handicap related to phonation. Additionally, results from Aim 3a revealed that adding respiratory exercises, more specifically IMST, to voice exercises led to greater improvements in voice outcomes, especially in self-assessment measures. However, it was still unclear if respiratory exercises should be recommended for all patients with presbyphonia, or only for patients who present with a decreased respiratory function.

Results from Aim 3b revealed that, although the state of the respiratory system at baseline did impact the response to the intervention, this was true for all intervention groups and not only for those including respiratory exercises. Moreover, all participants in the IMST group improved on self-assessment measures, suggesting that patients with and without reduced respiratory function may benefit from respiratory training. The results are discussed in detail in the following sections.

Impact of Baseline Respiratory Function

The respiratory variables that had the largest effect on the response to the intervention, as measured by the change in VHI-10 score, were: FVC%, FEV1%, and MEP. Additionally, MIP was found to be an important predictor within the IMST group. Although preliminary, these findings support the hypothesis that baseline respiratory function influences the response to voice

therapy. More specifically, a lower baseline respiratory function was associated with a greater improvement in VHI-10 following the intervention.

This is consistent with the results from Aim 2, which showed that a lower respiratory function tends to be associated with a higher perceived voice handicap, potentially because of a higher laryngeal effort to compensate for the decreased respiratory support, and/or a feeling of shortness of breath when speaking. However, the reason why lower respiratory values predicted improvements in voice handicap is not obvious, considering that 1) none of the interventions induced changes in pulmonary function, and 2) improvements in respiratory muscle strength were not correlated with improvements in perceived vocal handicap. It is possible that **participants with a greater respiratory deficit were less likely to make an optimal use of their respiratory support during phonation**, and that the intervention improved the functional use of their respiratory system without necessarily improving maximal values as measured by MIP, MEP, FVC, and FEV1. In other words, an impaired respiratory system could lead to a suboptimal respiratory support, which leaves more room for improvement from respiratory training. On the other hand, if a patient is already making an optimal use of their respiratory support, they may have less room for improvement and this could translate into less change in VHI-10 score.

Interestingly, respiratory variables were also found to be important predictors for change in VHI-10 in the total sample, when controlling for the intervention received. It is already acknowledged that certain voice exercises target both the respiratory and the laryngeal system, and especially the coordination between the two. VFEs are one of those, as they have been described as a series of specific exercises targeting strength, endurance, flexibility and stability of the laryngeal and respiratory systems as well as the balance between muscular effort and airflow (Stemple et al.,

1994; Tay et al., 2012). When performing VFEs, subjects are instructed to hold certain notes for increased time durations, while producing sounds that promote a healthy vocal fold vibration. In participants who presented with decreased respiratory function, these exercises may be sufficient to promote a more efficient use of respiratory support during phonation, leading to greater improvements when compared to patients presenting with better baseline respiratory function.

Importance of Controlling for Baseline Respiratory Status

The VFE group had the highest mean on two of the respiratory predictors (FVC% and FEV1%), while the EMST group had the lowest mean on these same variables. Even if not statistically significant, these differences could be clinically meaningful in a sense that they have the potential to affect the response to treatment. A univariate analysis of covariance tested the differences between the mean change in VHI-10 across the three groups, while controlling for baseline FVC%. The results were then compared to those obtained when baseline respiratory function was not taken into account. When adjusting the means to control for FVC%, the difference between the IMST and the EMST group was amplified (the IMST group improved significantly more than the EMST group) and the difference between the EMST and the VFE group was also amplified (the VFE group improved more than the EMST group, and the difference was marginally significant). The EMST group would have been expected to make the greatest improvements based on the low baseline FVC%, which was not the case. In fact, the EMST group had by far the least improvement in VHI-10 score following the intervention. On the other hand, the difference between the IMST and the VFE group was reduced, meaning that when adjusting for baseline FVC%, the two groups had a similar decrease in VHI-10.

Voice therapy studies rarely incorporate respiratory outcomes. Of the 23 studies that assessed the effect of a respiratory intervention on voice outcomes, only 12 of them measured at least one respiratory outcome (Desjardins & Bonilha, 2019). Of those, four studies included exclusively respiratory muscle strength measures (Chiara et al., 2007; Johansson, 2012; Ray, 2018; Roy et al., 2003), two studies included exclusively spirometry measures (Choi, 2016; Ramig et al., 1995), three studies included exclusively kinematic or other types of respiratory measures (Collyer, Kenny, & Archer, 2009, 2011; Schaeffer, 2017), and three studies included a combination of respiratory measures (Cerny et al., 1997; Darling-White & Huber, 2017; Mueller, 2013).

Results from Aim 3b demonstrated that baseline respiratory function impacts the response to voice therapy, with or without respiratory exercises. In the voice field, a lot of research has been directed towards finding the best outcomes to measure treatment effect on voice. It is well acknowledged that voice is multidimensional and that a variety of outcomes is crucial in obtaining a complete clinical picture. However, respiratory function has not been included in the recommended assessment protocol despite playing a major role in voice production. Aerodynamic measures such as mean glottal airflow, subglottal pressure, and aerodynamic resistance are the result of laryngeal and respiratory function (Vaca et al., 2017). While they provide essential information on the combined activity of these systems, they do not provide information on the status of the respiratory system and how it could impact the response to the intervention. Including respiratory function measures in voice studies could help reduce unexplained variance and could help researchers and clinicians gain a better understanding of the mechanisms of action of the studied interventions, especially in populations at risk of presenting with heterogeneous profiles of respiratory healthy such as patients with presbyphonia.

Is There a Need for Impairment-Specific Interventions?

Findings from this study do not support the premise that only patients with an impaired respiratory function are likely to benefit from an intervention loading the respiratory system. Even though the difference between the IMST and the VFE group was substantially reduced when controlling for baseline FVC%, there was an increase in standard deviations for VHI-10 in the VFE group (SD=6.41 at pre-treatment and SD=13.05 at post-treatment). This increase was not observed in the IMST group, for which the VHI-10 standard deviation decreased from 6.06 to 4.97 after therapy. This indicates that the inter-subject variability was reduced in the IMST group. In fact, every participant in the IMST group responded positively to the intervention, even though subjects with a lower respiratory function experienced the greatest improvement. Even participant 2, who presented with a FVC% of 100%, had a decrease of 6 points on the VHI-10, which has been considered a clinically meaningful improvement (J. Gartner-Schmidt & Rosen, 2011). On the other hand, the two participants in the VFE group who presented with a FVC% of 100% or more experienced no improvement at all in VHI-10 score. Therefore, the effect of VFEs may be limited in patients with a less severe respiratory baseline condition potentially, because of a ceiling effect. A more intense intervention such as IMST may "raise the ceiling" and further improve respiratory support, which might not be possible only with VFEs. Future studies are needed in order to gain a better understanding of the effects of VFEs and IMST in presbyphonic patients with various baseline respiratory profiles.

In summary, regardless of the baseline respiratory function, the IMST intervention yielded the best and least variable response. This is also in line with the results from Ziegler et al., who found that the intervention involving a greater load of the respiratory system (PhoRTE) led to greater self-reported improvements in the studied sample (Ziegler, 2014). Although interventions

involving only voice exercises may lead to meaningful self-reported improvements with a longer treatment duration (E. E. H. Berg, E.; Klein, A.; Johns, M. M., 3rd, 2008; Kaneko et al., 2015; Sauder et al., 2010), the results of our study suggest that using an intervention involving a greater load on the respiratory system leads to better self-reported improvements, for some patients, even with a short treatment duration (four weeks). In sum, results from this study and from the literature indicate that an intervention involving a greater load on the respiratory system may allow patients to achieve optimal results quicker, and potentially reduce the need for surgical interventions.

It is also possible that less severe patients seeking treatment for voice difficulties represent a subset of patients with an increased awareness of their voice problem and greater expectations. In this case, a surgical intervention may or may not yield the desired results for these patients. For example, participant 18, in the VFE group, presented with a well preserved respiratory function in comparison to the rest of the sample, and with a bowing index below the median, indicating a mild severity profile. This participant was an active recreational singer very aware of their voice difficulties and concerned about their ability to perform upcoming concerts. This participant was susceptible of presenting with high expectations regarding therapy outcomes. Following four weeks of VFEs, perceived handicap increased by two points on the VHI-10 and the participant decided to undergo a surgical intervention. An IMST intervention, combined with counseling aiming at adjusting the expectations of the participant, may have yielded more optimal results. In fact, Gartner-Schmidt and Rosen's results revealed that the rate of success for patients undergoing surgery after an unsuccessful voice therapy trial was only 17%. In the authors' discussion of this result, they indicated that unrealistic expectations from the patients may be a

major cause, and emphasized the importance of counseling prior to a surgical intervention to readjust expectations (J. R. Gartner-Schmidt, C., 2011).

Impact of Baseline Laryngeal Status

The impact of baseline laryngeal status on the response to behavioral voice therapy in patients with presbyphonia is not well known and yet is crucial in making clinical decisions. A surgical option is usually recommended for patients with severe atrophy (as shown by increased bowing of the vocal fold and prominence of the vocal fold processes), because of the assumption that these patients are less likely to benefit from behavioral voice therapy. The findings from this study do not support this assumption. In fact, baseline bowing index was not found to be a significant predictor for the change in VHI-10 following behavioral therapy. This is consistent with the results from Tanner et al., who studied the response to treatment in two monozygotic twins with severe age-related vocal fold bowing (Tanner et al., 2010). Limited functional improvements were observed following surgical management, as reported by the subjects. Moreover, while mid-membranous closure was improved, a significant posterior gap remained following the procedure. On the other hand, subsequent voice therapy led to an improved mid-membranous and posterior closure, as well as improvements on the VHI scores, especially for one of the twins. In addition, Lu et al. reported that patients with vocal fold bowing experienced limited benefits from thyroplasty type 1: while glottal gap was improved postoperatively, most of the other voice outcomes remained unchanged (Lu et al., 1998). In the light of their results, Tanner et al. suggested that behavioral voice therapy should be considered as the primary approach regardless of the severity of the laryngeal changes (Tanner et al., 2010). Our findings also support this recommendation, considering that a clinically meaningful decrease in VHI-10 score was noted for patients with a wide range of bowing index scores and that a greater baseline severity was not predictive of less functional improvement following behavioral voice therapy.

The reason why bowing index was not a good predictor for change in VHI-10 following treatment could reside in the fact that, although it is a measure of baseline laryngeal severity, it is not sufficiently informative of the patient's general clinical picture. Baseline laryngeal severity is a complex construct to measure, especially in patients with glottic insufficiency who use many compensation strategies to counteract the lack of glottic closure. While some of these strategies may help normalize the sound, they may also tax the laryngeal and respiratory systems and lead to increased effort and vocal fatigue (Zhang, 2019). As it was demonstrated in Aim 2, the integrity factor (comprised of bowing index and phase symmetry), was not included in the handicap-driven model because it didn't play a significant role on perceived handicap. Factors that were related to handicap (directly or indirectly) were: respiratory, pliability, resistance, and hyperfunction. These factors are related to phonation *physiology.* On the other hand, bowing index is related to vocal fold anatomy and structure, and although it does have an impact on physiology, many other factors come into play and should be considered when attempting to predict a patient's responsiveness to voice therapy.

Limitations for Aim 3b

Because of the small sample size, we were limited in the choice of statistical approaches for this aim. A future study with a larger sample size could allow researchers to dichotomize the outcome and to compute the odds ratio of improving on the VHI-10 score, based on the baseline respiratory function. Moreover, a larger sample size would allow for the identification of the best predictors within each of the intervention group separately.

Conclusions and Directions for Future Studies

Respiratory function was found to impact voice via two main pathways: through its physiological effect on voice and through its impact on general health and impairment. The raw respiratory strength and amount of air available for phonation were found to be predictors of physiological voice outcomes: a lower respiratory function was associated with an increased aerodynamic resistance accompanied by hyperfunction and reduced amplitude of vibration, mucosal wave amplitude and regularity of vibration. Standardized measures of respiratory function had a direct impact on perceived handicap, indicating that poor respiratory health may exacerbate the burden of the voice disorder. Respiratory function did not have an impact on voice quality, which was mostly influenced by the severity of vocal fold atrophy.

In line with those results, we also found that a lower baseline respiratory function was predictive of a greater improvement in self-reported outcomes following behavioral voice therapy. Those results do not imply that only patients with respiratory impairments would benefit from a respiratory intervention, considering that voice therapy including IMST led to the greatest improvements in VHI-10, even when adjusting for baseline respiratory function. The mechanism of improvement could be explained by a better subglottal pressure control, as well as more air available for phonation.

On the other hand, subjects who received EMST had the least improvement, potentially because of an imbalance between the resulting respiratory drive and the ability of the vocal folds to provide adequate resistance to modulate the airflow. Nonetheless, future studies could explore the effect of starting with IMST and VFE and then adding EMST to the training protocol once an optimal phonato-respiratory coordination has been established. Future studies should also assess the relationships between respiratory and voice measures in other voice disorders, such as muscle tension dysphonia, as well as the impact of IMST in these patients.

References

Abitbol, J., Abitbol, P., & Abitbol, B. (1999). Sex hormones and the female voice. *J Voice, 13*(3), 424-446.

Alipour, F., & Scherer, R. C. (2012). Ventricular pressures in phonating excised larynges. *J Acoust Soc Am, 132*(2), 1017-1026. doi:10.1121/1.4730880

Amenedo, E., & Diaz, F. (1998). Effects of aging on middle-latency auditory evoked potentials: a cross-sectional study. *Biol Psychiatry, 43*(3), 210-219. doi:10.1016/s0006-3223(97)00255-2

Angadi, V., McMullen, C., Andreatta, R., Dietrich, M., Uhl, T., & Stemple, J. (2018). Biobehavioral Measures of Presbylaryngeus. *Journal of Voice*, undefined-undefined. doi:10.1016/j.jvoice.2018.11.005

Arffa, R. E., Krishna, P., Gartner-Schmidt, J., & Rosen, C. A. (2012). Normative values for the Voice Handicap Index-10. *J Voice, 26*(4), 462-465. doi:10.1016/j.jvoice.2011.04.006

Awan, S. N. (2006). The aging female voice: acoustic and respiratory data. *Clin Linguist Phon, 20*(2-3), 171-180. doi:10.1080/02699200400026918

Awan, S. N., Novaleski, C. K., & Yingling, J. R. (2013). Test-retest reliability for aerodynamic measures of voice. *J Voice, 27*(6), 674-684. doi:10.1016/j.jvoice.2013.07.002

Bach, K. K., Belafsky, P. C., Wasylik, K., Postma, G. N., & Koufman, J. A. (2005). Validity and reliability of the glottal function index. *Arch Otolaryngol Head Neck Surg, 131*(11), 961-964. doi:10.1001/archotol.131.11.961

Baken, R. J. (2005). The aged voice: a new hypothesis. *J Voice, 19*(3), 317-325. doi:10.1016/j.jvoice.2004.07.005

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

Baker, S. E., Sapienza, C. M., & Collins, S. (2003). Inspiratory pressure threshold training in a case of congenital bilateral abductor vocal fold paralysis. *Int J Pediatr Otorhinolaryngol, 67*(4), 413-416.

Baker, S. E., Sapienza, C. M., Martin, D., Davenport, S., Hoffman-Ruddy, B., & Woodson, G. (2003). Inspiratory pressure threshold training for upper airway limitation: a case of bilateral abductor vocal fold paralysis. *J Voice, 17*(3), 384-394.

Barrichelo, V. M., & Behlau, M. (2007). Perceptual identification and acoustic measures of the resonant voice based on "Lessac's Y-Buzz"--a preliminary study with actors. *J Voice, 21*(1), 46-53. doi:10.1016/j.jvoice.2005.08.014

Baylor, C., Yorkston, K., Eadie, T., Kim, J., Chung, H., & Amtmann, D. (2013). The Communicative Participation Item Bank (CPIB): item bank calibration and development of a disorder-generic short form. *J Speech Lang Hear Res, 56*(4), 1190-1208. doi:10.1044/1092-4388(2012/12-0140)

Baylor, C., Yorkston K Fau - Eadie, T., Eadie T Fau - Kim, J., Kim J Fau - Chung, H., Chung H Fau - Amtmann, D., & Amtmann, D. The Communicative Participation Item Bank (CPIB): item bank calibration and development of a disorder-generic short form. (1558-9102 (Electronic)).

Beaumont, M. A.-O. h. o. o. X., Forget, P., Couturaud, F., & Reychler, G. A.-O. h. o. o. (2018). Effects of inspiratory muscle training in COPD patients: A systematic review and meta-analysis. *The Clinical Respiratory Journal.*(1752-699X (Electronic)).

Belafsky, P. C., Postma, G. N., & Koufman, J. A. (2001). The validity and reliability of the reflux finding score (RFS). *Laryngoscope, 111*(8), 1313-1317. doi:10.1097/00005537-200108000-00001

Belafsky, P. C., Postma, G. N., & Koufman, J. A. (2002). Validity and reliability of the reflux symptom index (RSI). *J Voice, 16*(2), 274-277.

Bellis, T. J., Nicol, T., & Kraus, N. (2000). Aging affects hemispheric asymmetry in the neural representation of speech sounds. *J Neurosci, 20*(2), 791-797.

Berg, E. E., Hapner, E., Klein, A., & Johns, M. M., 3rd. (2008). Voice therapy improves quality of life in age-related dysphonia: a case-control study. *J Voice, 22*(1), 70-74. doi:10.1016/j.jvoice.2006.09.002

Berg, E. E. H., E.; Klein, A.; Johns, M. M., 3rd. (2008). Voice therapy improves quality of life in age-related dysphonia: a case-control study. *Journal of Voice, 22*(1), 70-74 75p.

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

Bland, J. M., & Altman, D. G. (1996). Statistics notes: measurement error proportional to the mean. *Bmj, 313*(7049), 106.

Bless, D. M., & Welham, N. V. (2010). Characterization of vocal fold scar formation, prophylaxis, and treatment using animal models. *Curr Opin Otolaryngol Head Neck Surg, 18*(6), 481-486. doi:10.1097/MOO.0b013e3283407d87

Bloch, I., & Behrman, A. (2001). Quantitative analysis of videostroboscopic images in presbylarynges. *Laryngoscope, 111*(11 Pt 1), 2022-2027. doi:10.1097/00005537-200111000-00029

Bode, F. R., Dosman, J., Martin, R. R., Ghezzo, H., & Macklem, P. T. (1976). Age and sex differences in lung elasticity, and in closing capacity in nonsmokers. *J Appl Physiol, 41*(2), 129-135.

Bonilha H, D. M., Focht K & Martin-Harris B. (2017). *Parameters and Scales Used to Assess and Report Findings from Stroboscopy: A Systematic Review*. Paper presented at the The Voice Foundation Annual Symposium Care of the Professional Voice, Philadelphia, PA.

Bonilha, H. S., Focht, K. L., & Martin-Harris, B. (2015). Rater methodology for stroboscopy: a systematic review. *J Voice, 29*(1), 101-108. doi:10.1016/j.jvoice.2014.06.014

Boone, D., McFarlane SC, Von Berg SL & Zraick R. (2010). *The Voice and Voice Therapy 8th Edition*. USA: Pearson.

Boulet, M. J., & Oddens, B. J. (1996). Female voice changes around and after the menopause--an initial investigation. *Maturitas, 23*(1), 15-21.

Branski, R. C., Verdolini, K., Sandulache, V., Rosen, C. A., & Hebda, P. A. (2006). Vocal fold wound healing: a review for clinicians. *J Voice, 20*(3), 432-442. doi:10.1016/j.jvoice.2005.08.005

Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cereb Cortex, 18*(4), 837-845. doi:10.1093/cercor/bhm131

Buckmire, R. A., Bryson, P. C., & Patel, M. R. (2011). Type I gore-tex laryngoplasty for glottic incompetence in mobile vocal folds. *J Voice, 25*(3), 288-292. doi:10.1016/j.jvoice.2009.12.003

Cantillo-Banos, E., Jurado-Ramos, A., Gutierrez-Jodas, J., & Ariza-Vargas, L. (2013). Vocal fold insufficiency: medialization laryngoplasty vs calcium hydroxylapatite microspheres (Radiesse Voice(R)). *Acta Otolaryngol, 133*(3), 270-275. doi:10.3109/00016489.2012.728717

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. *Clin Otolaryngol Allied Sci, 29*(5), 538-544. doi:10.1111/j.1365-2273.2004.00846.x

Carding, P. N., Wilson, J. A., MacKenzie, K., & Deary, I. J. (2009). Measuring voice outcomes: state of the science review. *J Laryngol Otol, 123*(8), 823-829. doi:10.1017/s0022215109005398

Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. *Journal of applied psychology, 84*(6), 851.

Carroll, T. L., & Rosen, C. A. (2011). Long-term results of calcium hydroxylapatite for vocal fold augmentation. *Laryngoscope, 121*(2), 313-319. doi:10.1002/lary.21258

Casiano, R. R., Ruiz, P. J., & Goldstein, W. (1994). Histopathologic changes in the aging human cricoarytenoid joint. *Laryngoscope, 104*(5 Pt 1), 533-538.

Cerny, F. J., Panzarella, K. J., & Stathopoulos, E. (1997). Expiratory muscle conditioning in hypotonic children with low vocal intensity levels. *Journal of Medical Speech-Language Pathology, 5*(2), 141-152.

Chen, F. C., Ma, E. P., & Yiu, E. M. (2014). Facial bone vibration in resonant voice production. *J Voice, 28*(5), 596-602. doi:10.1016/j.jvoice.2013.12.014

Chen, S. H., Hsiao, T. Y., Hsiao, L. C., Chung, Y. M., & Chiang, S. C. (2007). Outcome of resonant voice therapy for female teachers with voice disorders: perceptual, physiological, acoustic, aerodynamic, and functional measurements. *J Voice, 21*(4), 415-425. doi:10.1016/j.jvoice.2006.02.001

Chhetri, D. K., Neubauer, J., & Berry, D. A. (2012). Neuromuscular control of fundamental frequency and glottal posture at phonation onset. *J Acoust Soc Am, 131*(2), 1401-1412. doi:10.1121/1.3672686

Chiara, T., Martin, A. D., Davenport, P. W., & Bolser, D. C. (2006). Expiratory muscle strength training in persons with multiple sclerosis having mild to moderate disability: effect on maximal expiratory pressure, pulmonary function, and maximal voluntary cough. *Arch Phys Med Rehabil, 87*(4), 468-473.

Chiara, T., Martin, D., & Sapienza, C. (2007). Expiratory muscle strength training: speech production outcomes in patients with multiple sclerosis. *Neurorehabil Neural Repair, 21*(3), 239-249. doi:10.1177/1545968306294737

Chinn, S. (1991). Statistics in respiratory medicine. 2. Repeatability and method comparison. *Thorax, 46*(6), 454-456.

Choi, J. Y. R., D. W.; Park, E. S. (2016). Change in Pulmonary Function after Incentive Spirometer Exercise in Children with Spastic Cerebral Palsy: A Randomized Controlled Study. *Yonsei Med J, 57*(3), 769-775. doi:10.3349/ymj.2016.57.3.769

Cohen, S. M. (2010). Self-reported impact of dysphonia in a primary care population: an epidemiological study. *Laryngoscope, 120*(10), 2022-2032. doi:10.1002/lary.21058

Cohen, S. M., Dinan, M. A., Kim, J., & Roy, N. (2015). Otolaryngology utilization of speech-language pathology services for voice disorders. *Laryngoscope*. doi:10.1002/lary.25574

Cohen, S. M., Kim, J., Roy, N., Asche, C., & Courey, M. (2012a). Direct health care costs of laryngeal diseases and disorders. *Laryngoscope, 122*(7), 1582-1588. doi:10.1002/lary.23189

Cohen, S. M., Kim, J., Roy, N., Asche, C., & Courey, M. (2012b). Prevalence and causes of dysphonia in a large treatment-seeking population. *Laryngoscope, 122*(2), 343-348. doi:10.1002/lary.22426

Cohen, S. M., & Turley, R. (2009). Coprevalence and impact of dysphonia and hearing loss in the elderly. *Laryngoscope, 119*(9), 1870-1873. doi:10.1002/lary.20590

Collyer, S., Kenny, D. T., & Archer, M. (2009). The effect of abdominal kinematic directives on respiratory behaviour in female classical singing. *Logoped Phoniatr Vocol, 34*(3), 100-110. doi:10.1080/14015430903008780

Collyer, S., Kenny, D. T., & Archer, M. (2011). Listener perception of the effect of abdominal kinematic directives on respiratory behavior in female classical singing. *J Voice, 25*(1), e15-24. doi:10.1016/j.jvoice.2009.10.006

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*: Psychology press.

Connor, N. P., Suzuki, T., Lee, K., Sewall, G. K., & Heisey, D. M. (2002). Neuromuscular junction changes in aged rat thyroarytenoid muscle. *Ann Otol Rhinol Laryngol, 111*(7 Pt 1), 579-586. doi:10.1177/000348940211100703

Constantinescu, G. T., Deborah; Russell, Trevor; Ward, Elizabeth; Wilson, Stephen; Wootton, Richard. (2011). Treating disordered speech and voice in Parkinson's disease online: a randomized controlled non-inferiority trial. *International Journal of Language & Communication Disorders, 46*(1), 1-16 16p. doi:10.3109/13682822.2010.484848

Crawley, B. K., Dehom, S., Thiel, C., Yang, J., Cragoe, A., Mousselli, I., . . . Murry, T. (2018). Assessment of Clinical and Social Characteristics That Distinguish

Presbylaryngis From Pathologic Presbyphonia in Elderly Individuals. *JAMA Otolaryngol Head Neck Surg, 144*(7), 566-571. doi:10.1001/jamaoto.2018.0409

D'Haeseleer, E., Depypere, H., Claeys, S., Van Borsel, J., & Van Lierde, K. (2009). The menopause and the female larynx, clinical aspects and therapeutic options: a literature review. *Maturitas, 64*(1), 27-32. doi:10.1016/j.maturitas.2009.06.009

Daniilidou, P. C., P.; Wilson, J.; Drinnan, M.; Deary, V. (2007). Cognitive behavioral therapy for functional dysphonia: a pilot study. *Ann Otol Rhinol Laryngol, 116*(10), 717-722.

Darling-White, M., & Huber, J. E. (2017). The Impact of Expiratory Muscle Strength Training on Speech Breathing in Individuals With Parkinson's Disease: A Preliminary Study. *Am J Speech Lang Pathol, 26*(4), 1159-1166. doi:10.1044/2017_AJSLP-16-0132

Davids, T., Klein, A. M., & Johns, M. M., 3rd. (2012). Current dysphonia trends in patients over the age of 65: is vocal atrophy becoming more prevalent? *Laryngoscope, 122*(2), 332-335. doi:10.1002/lary.22397

Davis, P. J., Zhang, S. P., Winkworth, A., & Bandler, R. (1996). Neural control of vocalization: respiratory and emotional influences. *J Voice, 10*(1), 23-38.

Deliyski, S. A. X., Dimitar. (2001). Effects of aging on selected acoustic voice parameters: Preliminary normative data and educational implications. *Educational Gerontology, 27*(2), 159-168.

Desjardins, M., & Bonilha, H. (2019). The Impact of Respiratory Exercises on Voice Outcomes: A Systematic Review of the Literature. *Journal of Voice*, undefined-undefined. doi:10.1016/j.jvoice.2019.01.011

Desjardins, M., Halstead, L., Cooke, M., & Bonilha, H. S. (2017). A Systematic Review of Voice Therapy: What "Effectiveness" Really Implies. *J Voice, 31*(3), 392.e313-392.e332. doi:10.1016/j.jvoice.2016.10.002

Detels, R., Coulson, A., Tashkin, D., & Rokaw, S. (1975). Reliability of plethysmography, the single breath oxygen test, and spirometry in population studies. *Bull Physiopathol Respir (Nancy), 11*(1), 9-30.

Ding, H., & Gray, S. D. (2001). Senescent expression of genes coding collagens, collagen-degrading metalloproteinases, and tissue inhibitors of metalloproteinases in rat vocal folds: comparison with skin and lungs. *J Gerontol A Biol Sci Med Sci, 56*(4), B145-152.

Dromey, C., Ramig, L. O., & Johnson, A. B. (1995). Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: a case study. *J Speech Hear Res, 38*(4), 751-764.

Durlak, J. A. (2009). How to Select, Calculate, and Interpret Effect Sizes. *Journal of Pediatric Psychology, 34*(9), 917-928. doi:10.1093/jpepsy/jsp004

Duruturk, N., Acar, M., & Dogrul, M. I. (2018). Effect of Inspiratory Muscle Training in the Management of Patients With Asthma: A RANDOMIZED CONTROLLED TRIAL. *J Cardiopulm Rehabil Prev, 38*(3), 198-203.

Enright, P. L., Kronmal, R. A., Higgins, M. W., Schenker, M. B., & Haponik, E. F. (1994). Prevalence and correlates of respiratory symptoms and disease in the elderly. Cardiovascular Health Study. *Chest, 106*(3), 827-834.

Enright, P. L., Kronmal, R. A., Manolio, T. A., Schenker, M. B., & Hyatt, R. E. (1994). Respiratory muscle strength in the elderly. Correlates and reference values. Cardiovascular Health Study Research Group. *Am J Respir Crit Care Med, 149*(2 Pt 1), 430-438. doi:10.1164/ajrccm.149.2.8306041

Euser, A. M., Dekker, F. W., & le Cessie, S. (2008). A practical approach to Bland-Altman plots and variation coefficients for log transformed variables. *J Clin Epidemiol, 61*(10), 978-982. doi:10.1016/j.jclinepi.2007.11.003

Evans, J. A., & Whitelaw, W. A. (2009). The assessment of maximal respiratory mouth pressures in adults. *Respir Care, 54*(10), 1348-1359.

Ferrand, C. T. (2002). Harmonics-to-noise ratio: an index of vocal aging. *J Voice, 16*(4), 480-487.

Friedrich, G., Dikkers, F. G., Arens, C., Remacle, M., Hess, M., Giovanni, A., . . . Gugatschka, M. (2013). Vocal fold scars: current concepts and future directions. Consensus report of the Phonosurgery Committee of the European Laryngological Society. *Eur Arch Otorhinolaryngol, 270*(9), 2491-2507. doi:10.1007/s00405-013-2498-9

Garrison, C. R. (2009). *Repeatability of aerodynamic measurements of voice.* Miami University.

Gartner-Schmidt, J., & Rosen, C. (2011). Treatment success for age-related vocal fold atrophy. *Laryngoscope, 121*(3), 585-589. doi:10.1002/lary.21122

Gartner-Schmidt, J. L., Roth, D. F., Zullo, T. G., & Rosen, C. A. (2013). Quantifying component parts of indirect and direct voice therapy related to different voice disorders. *J Voice, 27*(2), 210-216. doi:10.1016/j.jvoice.2012.11.007

Gartner-Schmidt, J. R., C. (2011). Treatment success for age-related vocal fold atrophy. *Laryngoscope, 121*(3), 585-589 585p. doi:10.1002/lary.21122

Gibson, G. J., Pride, N. B., O'Cain, C., & Quagliato, R. (1976). Sex and age differences in pulmonary mechanics in normal nonsmoking subjects. *J Appl Physiol, 41*(1), 20-25.

Gierut, J. A., Morrisette, M. L., & Dickinson, S. L. (2015). Effect Size for Single-Subject Design in Phonological Treatment. *J Speech Lang Hear Res, 58*(5), 1464-1481. doi:10.1044/2015_jslhr-s-14-0299

Gillespie, A. I., Dastolfo, C., Magid, N., & Gartner-Schmidt, J. (2014). Acoustic analysis of four common voice diagnoses: moving toward disorder-specific assessment. *J Voice, 28*(5), 582-588. doi:10.1016/j.jvoice.2014.02.002

Golub, J. S., Chen, P. H., Otto, K. J., Hapner, E., & Johns, M. M., 3rd. (2006). Prevalence of perceived dysphonia in a geriatric population. *J Am Geriatr Soc, 54*(11), 1736-1739. doi:10.1111/j.1532-5415.2006.00915.x

Goncalves, T. M., Dos Santos, D. C., Pessin, A. B., & Martins, R. H. (2016). Scanning Electron Microscopy of the Presbylarynx. *Otolaryngol Head Neck Surg, 154*(6), 1073-1078. doi:10.1177/0194599816637081

Gorham-Rowan, M. M., & Laures-Gore, J. (2006). Acoustic-perceptual correlates of voice quality in elderly men and women. *J Commun Disord, 39*(3), 171-184. doi:10.1016/j.jcomdis.2005.11.005

Gorman, S., Weinrich, B., Lee, L., & Stemple, J. C. (2008). Aerodynamic changes as a result of vocal function exercises in elderly men. *Laryngoscope, 118*(10), 1900-1903. doi:10.1097/MLG.0b013e31817f9822

Gosselink, R., Kovacs, L., Ketelaer, P., Carton, H., & Decramer, M. (2000). Respiratory muscle weakness and respiratory muscle training in severely disabled multiple sclerosis patients. *Arch Phys Med Rehabil, 81*(6), 747-751.

Gregory, N. D., Chandran, S., Lurie, D., & Sataloff, R. T. (2012). Voice disorders in the elderly. *J Voice, 26*(2), 254-258. doi:10.1016/j.jvoice.2010.10.024

Halpern, A. E. R., L. O.; Matos, C. E.; Petska-Cable, J. A.; Spielman, J. L.; Pogoda, J. M.; Gilley, P. M.; Sapir, S.; Bennett, J. K.; McFarland, D. H. (2012). Innovative technology for the assisted delivery of intensive voice treatment (LSVT(R)LOUD) for Parkinson disease. *Am J Speech Lang Pathol, 21*(4), 354-367. doi:10.1044/1058-0360(2012/11-0125)

Hammer, M. J., & Krueger, M. A. (2014). Voice-related modulation of mechanosensory detection thresholds in the human larynx. *Exp Brain Res, 232*(1), 13-20. doi:10.1007/s00221-013-3703-1

Hammond, T. H., Gray, S. D., & Butler, J. E. (2000). Age- and gender-related collagen distribution in human vocal folds. *Ann Otol Rhinol Laryngol, 109*(10 Pt 1), 913-920. doi:10.1177/000348940010901004

Hamnegard, C. H., Wragg, S., Kyroussis, D., Aquilina, R., Moxham, J., & Green, M. (1994). Portable measurement of maximum mouth pressures. *Eur Respir J, 7*(2), 398-401.

Hansen, J. K., & Thibeault, S. L. (2006). Current understanding and review of the literature: vocal fold scarring. *J Voice, 20*(1), 110-120. doi:10.1016/j.jvoice.2004.12.005

Hansen, J. T. (2014). *Netter's Anatomy Flash Cards 4th Edition*. Philadelphia, PA: Saunders Elsevier.

Heman-Ackah, Y. D., Heuer, R. J., Michael, D. D., Ostrowski, R., Horman, M., Baroody, M. M., . . . Sataloff, R. T. (2003). Cepstral peak prominence: a more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol, 112*(4), 324-333.

Herbst, C. T. (2017). A Review of Singing Voice Subsystem Interactions-Toward an Extended Physiological Model of "Support". *J Voice, 31*(2), 249.e213-249.e219. doi:10.1016/j.jvoice.2016.07.019

Higgins, M. B., Netsell, R., & Schulte, L. (1994). Aerodynamic and electroglottographic measures of normal voice production: intrasubject variability within and across sessions. *J Speech Hear Res, 37*(1), 38-45.

Higgins, M. B., & Saxman, J. H. (1991). A comparison of selected phonatory behaviors of healthy aged and young adults. *J Speech Hear Res, 34*(5), 1000-1010.

Hirano, M. (1974). Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatr (Basel), 26*(2), 89-94.

Hirano, M., Sato, K., & Nakashima, T. (2000). Fibroblasts in geriatric vocal fold mucosa. *Acta Otolaryngol, 120*(2), 336-340.

Hirano, S., Bless, D. M., del Rio, A. M., Connor, N. P., & Ford, C. N. (2004). Therapeutic potential of growth factors for aging voice. *Laryngoscope, 114*(12), 2161-2167. doi:10.1097/01.mlg.0000149450.37640.db

Hirano, S., Kishimoto, Y., Suehiro, A., Kanemaru, S., & Ito, J. (2008). Regeneration of aged vocal fold: first human case treated with fibroblast growth factor. *Laryngoscope, 118*(12), 2254-2259. doi:10.1097/MLG.0b013e3181845720

Hirano, S., Nagai, H., Tateya, I., Tateya, T., Ford, C. N., & Bless, D. M. (2005). Regeneration of aged vocal folds with basic fibroblast growth factor in a rat model: a preliminary report. *Ann Otol Rhinol Laryngol, 114*(4), 304-308.

Hirano, S., Tateya, I., Kishimoto, Y., Kanemaru, S. I., & Ito, J. (2012). Clinical trial of regeneration of aged vocal folds with growth factor therapy. *Laryngoscope, 122*(2), 327-331. doi:10.1002/lary.22393

Hiwatashi, N., Hirano, S., Mizuta, M., Kobayashi, T., Kawai, Y., Kanemaru, S. I., . . . Suzuki, S. (2017). The efficacy of a novel collagen-gelatin scaffold with basic fibroblast growth factor for the treatment of vocal fold scar. *J Tissue Eng Regen Med, 11*(5), 1598-1609. doi:10.1002/term.2060

Hixon, T. J. (1973). Kinematics of the chest wall during speech production: volume displacements of the rib cage, abdomen, and lung. *J Speech Hear Res, 16*(1), 78-115.

Hixon, T. J. (1987). *Respiratory Function in Speech and Song*. London: Taylor & Francis.

Hoit, J. D., & Hixon, T. J. (1987). Age and speech breathing. *J Speech Hear Res, 30*(3), 351-366.

Hoit, J. D., Plassman, B. L., Lansing, R. W., & Hixon, T. J. (1988). Abdominal muscle activity during speech production. *J Appl Physiol (1985), 65*(6), 2656-2664.

Honjo, I., & Isshiki, N. (1980). Laryngoscopic and voice characteristics of aged persons. *Arch Otolaryngol, 106*(3), 149-150.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Med, 30*(1), 1-15. doi:10.2165/00007256-200030010-00001

Huang, C. H., Yang, G. G., Wu, Y. T., & Lee, C. W. (2011). Comparison of inspiratory muscle strength training effects between older subjects with and without chronic obstructive pulmonary disease. *J Formos Med Assoc, 110*(8), 518-526. doi:10.1016/s0929-6646(11)60078-8

Huang, D. Z., Minifie, F. D., Kasuya, H., & Lin, S. X. (1995). Measures of vocal function during changes in vocal effort level. *J Voice, 9*(4), 429-438.

Huber, J. E. (2008). Effects of utterance length and vocal loudness on speech breathing in older adults. *Respir Physiol Neurobiol, 164*(3), 323-330. doi:10.1016/j.resp.2008.08.007

Huber, J. E., & Spruill, J., 3rd. (2008). Age-related changes to speech breathing with increased vocal loudness. *J Speech Lang Hear Res, 51*(3), 651-668. doi:10.1044/1092-4388(2008/047)

Hunter, E. J., Maxfield, L., & Graetzer, S. (2019). The Effect of Pulmonary Function on the Incidence of Vocal Fatigue Among Teachers. *J Voice*. doi:10.1016/j.jvoice.2018.12.011

International Classification of Functioning, Disability and Health. (2001). Geneva, Switzerland: World Health Organization.

Ishikawa, K., & Thibeault, S. (2010). Voice rest versus exercise: a review of the literature. *J Voice, 24*(4), 379-387. doi:10.1016/j.jvoice.2008.10.011

Iwarsson, J., Thomasson, M., & Sundberg, J. (1998). Effects of lung volume on the glottal voice source. *Journal of Voice, 12*(4), 424-433. doi:https://doi.org/10.1016/S0892-1997(98)80051-9

Iwata, S., Von Leden, H., & Williams, D. (1972). Air flow measurement during phonation. *J Commun Disord, 5*(1), 67-79.

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), 66-69.

Janssens, J. P., Pache, J. C., & Nicod, L. P. (1999). Physiological changes in respiratory function associated with ageing. *Eur Respir J, 13*(1), 197-205.

Johansson, K. M. K., Liselotte; Schalling, Ellika; Hartelius, Lena; Fredrikson, Sten. (2012). 'I Can Walk Briskly and Talk at the Same Time': Effects of Expiratory Muscle Strength Training on Respiration and Speech in Multiple Sclerosis. *Journal of Medical Speech-Language Pathology, 20*(4), 70-76.

Johns, M. M., 3rd, Arviso, L. C., & Ramadan, F. (2011). Challenges and opportunities in the management of the aging voice. *Otolaryngol Head Neck Surg, 145*(1), 1-6. doi:10.1177/0194599811404640

Johnson, A. M., Ciucci, M. R., & Connor, N. P. (2013). Vocal training mitigates age-related changes within the vocal mechanism in old rats. *J Gerontol A Biol Sci Med Sci, 68*(12), 1458-1468. doi:10.1093/gerona/glt044

Johnson, A. M., & Goldfine, A. (2016). Intrasubject Reliability of Maximum Phonation Time. *J Voice, 30*(6), 775.e771-775.e774. doi:10.1016/j.jvoice.2015.11.019

Jurgens, U. (2009). The neural control of vocalization in mammals: a review. *J Voice, 23*(1), 1-10. doi:10.1016/j.jvoice.2007.07.005

Jurgens, U., & von Cramon, D. (1982). On the role of the anterior cingulate cortex in phonation: a case report. *Brain Lang, 15*(2), 234-248.

Kaneko, M., Hirano, S., Tateya, I., Kishimoto, Y., Hiwatashi, N., Fujiu-Kurachi, M., & Ito, J. (2015). Multidimensional Analysis on the Effect of Vocal Function Exercises on Aged Vocal Fold Atrophy. *J Voice, 29*(5), 638-644. doi:10.1016/j.jvoice.2014.10.017

Kersing, W., & Jennekens, F. G. (2004). Age-related changes in human thyroarytenoid muscles: a histological and histochemical study. *Eur Arch Otorhinolaryngol, 261*(7), 386-392. doi:10.1007/s00405-003-0702-z

Kim, J., Davenport, P., & Sapienza, C. (2009). Effect of expiratory muscle strength training on elderly cough function. *Arch Gerontol Geriatr, 48*(3), 361-366. doi:10.1016/j.archger.2008.03.006

Kim, J., & Sapienza, C. M. (2005). Implications of expiratory muscle strength training for rehabilitation of the elderly: Tutorial. *J Rehabil Res Dev, 42*(2), 211-224.

Kniesburges, S., Birk, V., Lodermeyer, A., Schutzenberger, A., Bohr, C., & Becker, S. (2017). Effect of the ventricular folds in a synthetic larynx model. *J Biomech, 55*, 128-133. doi:10.1016/j.jbiomech.2017.02.021

Kobayashi, T., Mizuta, M., Hiwatashi, N., Kishimoto, Y., Nakamura, T., Kanemaru, S. I., & Hirano, S. (2017). Drug delivery system of basic fibroblast growth factor using gelatin hydrogel for restoration of acute vocal fold scar. *Auris Nasus Larynx, 44*(1), 86-92. doi:10.1016/j.anl.2016.04.005

Koufman, J. A., Amin, M. R., & Panetti, M. (2000). Prevalence of reflux in 113 consecutive patients with laryngeal and voice disorders. *Otolaryngol Head Neck Surg, 123*(4), 385-388. doi:10.1067/mhn.2000.109935

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. *J Voice, 25*(4), 395-405. doi:10.1016/j.jvoice.2010.02.001

Kromer, B., & Howard, D. (2013). *Labor Force Participation and Work Status of People 65 Years and Older*. United States Census Bureau.

Kubzansky, L. D., Wright, R. J., Cohen, S., Weiss, S., Rosner, B., & Sparrow, D. (2002). Breathing easy: a prospective study of optimism and pulmonary function in the normative aging study. *Ann Behav Med, 24*(4), 345-353. doi:10.1207/s15324796abm2404_11

Kwon, T. K., An, S. Y., Ahn, J. C., Kim, K. H., & Sung, M. W. (2010). Calcium hydroxylapatite injection laryngoplasty for the treatment of presbylaryngis: long-term results. *Laryngoscope, 120*(2), 326-329. doi:10.1002/lary.20749

Lagorio, L. A., Carnaby-Mann, G. D., & Crary, M. A. (2010). Treatment of vocal fold bowing using neuromuscular electrical stimulation. *Arch Otolaryngol Head Neck Surg, 136*(4), 398-403. doi:10.1001/archoto.2010.33

Lalley, P. M. (2013). The aging respiratory system--pulmonary structure, function and neural control. *Respir Physiol Neurobiol, 187*(3), 199-210. doi:10.1016/j.resp.2013.03.012

Larson, C. R. (1998). Cross-modality influences in speech motor control: the use of pitch shifting for the study of F0 control. *J Commun Disord, 31*(6), 489-502; quiz 502-483; 553.

Larson, C. R. (2017). *Production and Perception of Voice: An Overview.* Paper presented at the The Voice Foundation Annual Symposium, Philadelphia.

Larson, C. R., Altman, K. W., Liu, H., & Hain, T. C. (2008). Interactions between auditory and somatosensory feedback for voice F0 control. *Exp Brain Res, 187*(4), 613-621. doi:10.1007/s00221-008-1330-z

Lee, L., Stemple, J. C., & Kizer, M. (1999). Consistency of acoustic and aerodynamic measures of voice production over 28 days under various testing conditions. *J Voice, 13*(4), 477-483.

Lee, L., Stemple, J. C., & Kizer, M. (1999). Consistency of acoustic and aerodynamic measures of voice production over 28 days under various testing conditions. *Journal of Voice, 13*(4), 477-483.

Leeper, H. A., Jr., & Graves, D. K. (1984). Consistency of laryngeal airway resistance in adult women. *J Commun Disord, 17*(3), 153-163.

Leong, K., Hawkshaw, M. J., Dentchev, D., Gupta, R., Lurie, D., & Sataloff, R. T. (2013). Reliability of objective voice measures of normal speaking voices. *J Voice, 27*(2), 170-176. doi:10.1016/j.jvoice.2012.07.005

Levendoski, E. E., Leydon, C., & Thibeault, S. L. (2014). Vocal fold epithelial barrier in health and injury: a research review. *J Speech Lang Hear Res, 57*(5), 1679-1691. doi:10.1044/2014_jslhr-s-13-0283

Lexell, J., Taylor, C. C., & Sjostrom, M. (1988). What is the cause of the ageing atrophy? Total number, size and proportion of different fiber types studied in whole vastus lateralis muscle from 15- to 83-year-old men. *J Neurol Sci, 84*(2-3), 275-294.

Lindholm, P., Vilkman, E., Raudaskoski, T., Suvanto-Luukkonen, E., & Kauppila, A. (1997). The effect of postmenopause and postmenopausal HRT on measured voice values and vocal symptoms. *Maturitas, 28*(1), 47-53.

Linville, S. E. (1996). The sound of senescence. *J Voice, 10*(2), 190-200.

Linville, S. E., & Fisher, H. B. (1985). Acoustic characteristics of women's voices with advancing age. *J Gerontol, 40*(3), 324-330.

Liu, P., Chen, Z., Jones, J. A., Huang, D., & Liu, H. (2011). Auditory feedback control of vocal pitch during sustained vocalization: a cross-sectional study of adult aging. *PLoS One, 6*(7), e22791. doi:10.1371/journal.pone.0022791

Loth, D. W., Ittermann, T., Lahousse, L., Hofman, A., Leufkens, H. G., Brusselle, G. G., & Stricker, B. H. (2013). Normal spirometry values in healthy elderly: the Rotterdam Study. *Eur J Epidemiol, 28*(4), 329-334. doi:10.1007/s10654-013-9800-4

Lotters, F., van Tol, B., Kwakkel, G., & Gosselink, R. (2002). Effects of controlled inspiratory muscle training in patients with COPD: a meta-analysis. *Eur Respir J, 20*(3), 570-576.

Loucks, T. M., Poletto, C. J., Simonyan, K., Reynolds, C. L., & Ludlow, C. L. (2007). Human brain activation during phonation and exhalation: common volitional control for two upper airway functions. *Neuroimage, 36*(1), 131-143. doi:10.1016/j.neuroimage.2007.01.049

Lu, F. L., Casiano, R. R., Lundy, D. S., & Xue, J. W. (1998). Vocal evaluation of thyroplasty type I in the treatment of nonparalytic glottic incompetence. *Ann Otol Rhinol Laryngol, 107*(2), 113-119. doi:10.1177/000348949810700206

Lu, F. L., Presley, S., & Lammers, B. (2013). Efficacy of intensive phonatory-respiratory treatment (LSVT) for presbyphonia: two case reports. *J Voice, 27*(6), 786.e711-723. doi:10.1016/j.jvoice.2013.06.006

Lyon, M. J., & Malmgren, L. T. (2010). Age-related blood flow changes in the rat intrinsic laryngeal muscles. *Acta Otolaryngol, 130*(1), 145-149. doi:10.3109/00016480902968078

Lyon, M. J., Steer, L. M., & Malmgren, L. T. (2007). Stereological estimates indicate that aging does not alter the capillary length density in the human posterior cricoarytenoid muscle. *J Appl Physiol (1985), 103*(5), 1815-1823. doi:10.1152/japplphysiol.00030.2007

Mada, P. (n.d.). Mechanics of Respiration. *Functions of Cells and Human Body.* Retrieved from http://fblt.cz/en/skripta/vi-dychaci-soustava/2-mechanika-dychani/

Malmgren, L. T., Fisher, P. J., Bookman, L. M., & Uno, T. (1999). Age-related changes in muscle fiber types in the human thyroarytenoid muscle: an immunohistochemical and stereological study using confocal laser scanning microscopy. *Otolaryngol Head Neck Surg, 121*(4), 441-451. doi:10.1016/s0194-5998(99)70235-4

Malmgren, L. T., Lovice, D. B., & Kaufman, M. R. (2000). Age-related changes in muscle fiber regeneration in the human thyroarytenoid muscle. *Arch Otolaryngol Head Neck Surg, 126*(7), 851-856.

Mantyh, P. W. (1983). Connections of midbrain periaqueductal gray in the monkey. II. Descending efferent projections. *J Neurophysiol, 49*(3), 582-594.

Martins, R. H., Benito Pessin, A. B., Nassib, D. J., Branco, A., Rodrigues, S. A., & Matheus, S. M. (2015). Aging voice and the laryngeal muscle atrophy. *Laryngoscope, 125*(11), 2518-2521. doi:10.1002/lary.25398

Marx, R. G., Menezes, A., Horovitz, L., Jones, E. C., & Warren, R. F. (2003). A comparison of two time intervals for test-retest reliability of health status instruments. *J Clin Epidemiol, 56*(8), 730-735.

Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., & Corthals, P. (2009). Acoustic measurement of overall voice quality: a meta-analysis. *J Acoust Soc Am, 126*(5), 2619-2634. doi:10.1121/1.3224706

Maryn, Y., & Weenink, D. (2015). Objective dysphonia measures in the program Praat: smoothed cepstral peak prominence and acoustic voice quality index. *J Voice, 29*(1), 35-43. doi:10.1016/j.jvoice.2014.06.015

Mathers-Schmidt, B. A., & Brilla, L. R. (2005). Inspiratory muscle training in exercise-induced paradoxical vocal fold motion. *J Voice, 19*(4), 635-644. doi:10.1016/j.jvoice.2005.03.005

Mathieson, L., Greene, L.C. (2006). *Greene & Mathieson's The Voice & Its Disorders Sixth Edition*. London: Whurr.

Matsumoto, K., & Nakamura, T. (2001). Hepatocyte growth factor: renotropic role and potential therapeutics for renal diseases. *Kidney Int, 59*(6), 2023-2038. doi:10.1046/j.1523-1755.2001.00717.x

Mau, T., Jacobson, B. H., & Garrett, C. G. (2010). Factors associated with voice therapy outcomes in the treatment of presbyphonia. *Laryngoscope, 120*(6), 1181-1187. doi:10.1002/lary.20890

McConnell, A. K., & Romer, L. M. (2004). Respiratory muscle training in healthy humans: resolving the controversy. *Int J Sports Med, 25*(4), 284-293. doi:10.1055/s-2004-815827

McDougall, J., Wright, V., & Rosenbaum, P. (2010). The ICF model of functioning and disability: incorporating quality of life and human development. *Dev Neurorehabil, 13*(3), 204-211. doi:10.3109/17518421003620525

McEwen, B. S., & Stellar, E. (1993). Stress and the individual. Mechanisms leading to disease. *Arch Intern Med, 153*(18), 2093-2101.

McFarland, D., Tremblay, P. . (2008). *Anatomie et Physiologie de la Parole ORA 1530.* Paper presented at the Faculté de Médecine Université de Montréal, Montréal, Qc.

McMullen, C. A., & Andrade, F. H. (2009). Functional and morphological evidence of age-related denervation in rat laryngeal muscles. *J Gerontol A Biol Sci Med Sci, 64*(4), 435-442. doi:10.1093/gerona/gln074

Melcon, M. C., Hoit, J. D., & Hixon, T. J. (1989). Age and laryngeal airway resistance during vowel production. *J Speech Hear Disord, 54*(2), 282-286.

Melton, L. J., 3rd, Khosla, S., & Riggs, B. L. (2000). Epidemiology of sarcopenia. *Mayo Clin Proc, 75 Suppl*, S10-12; discussion S12-13.

Miller, M. R., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R., Coates, A., . . . Wanger, J. (2005). Standardisation of spirometry. *Eur Respir J, 26*(2), 319-338. doi:10.1183/09031936.05.00034805

Mills, D. E., Johnson, M. A., Barnett, Y. A., Smith, W. H., & Sharpe, G. R. (2015). The effects of inspiratory muscle training in older adults. *Med Sci Sports Exerc, 47*(4), 691-697. doi:10.1249/MSS.0000000000000474

Mortelliti, A. J., Malmgren, L. T., & Gacek, R. R. (1990). Ultrastructural changes with age in the human superior laryngeal nerve. *Arch Otolaryngol Head Neck Surg, 116*(9), 1062-1069.

Mota, S., Guell, R., Barreiro, E., Solanes, I., Ramirez-Sarmiento, A., Orozco-Levi, M., . . . Sanchis, J. (2007). Clinical outcomes of expiratory muscle training in severe COPD patients. *Respir Med, 101*(3), 516-524. doi:10.1016/j.rmed.2006.06.024

Mueller, G. H., M. T.; Perret, C. (2013). Comparison of respiratory muscle training methods in individuals with motor and sensory complete tetraplegia: a randomized controlled trial. *J Rehabil Med, 45*(3), 248-253. doi:10.2340/16501977-1097

Murty, G. E., Carding, P. N., & Kelly, P. J. (1991). Combined glottographic changes in the elderly. *Clin Otolaryngol Allied Sci, 16*(6), 532-534.

Nakai, T., Goto, N., Moriyama, H., Shiraishi, N., & Nonaka, N. (2000). The human recurrent laryngeal nerve during the aging process. *Okajimas Folia Anat Jpn, 76*(6), 363-367.

Nathans, L. L., Oswald, F. L., & Nimon, K. (2012). Interpreting multiple linear regression: A guidebook of variable importance. *Practical assessment, research & evaluation, 17*(9).

Nemr, K., Simoes-Zenari, M., Cordeiro, G. F., Tsuji, D., Ogawa, A. I., Ubrig, M. T., & Menezes, M. H. (2012). GRBAS and Cape-V scales: high reliability and consensus when applied at different times. *J Voice, 26*(6), 812.e817-822. doi:10.1016/j.jvoice.2012.03.005

Nemr, K., Souza, G. V., Simoes-Zenari, M., Tsuji, D. H., Hachiya, A., Cordeiro, G. F., . . . Dajer, M. E. (2014). Cognitive Vocal Program applied to individuals with signals presbylarynx: preliminary results. *Codas, 26*(6), 503-508. doi:10.1590/2317-1782/20142014108

Netter, F. H. (2014). *Atlas of Human Anatomy 6th Edition*. Philadelphia, PA: Saunders Elsevier.

Netterville, J. L., Stone, R. E., Luken, E. S., Civantos, F. J., & Ossoff, R. H. (1993). Silastic medialization and arytenoid adduction: the Vanderbilt experience. A review of 116 phonosurgical procedures. *Ann Otol Rhinol Laryngol, 102*(6), 413-424. doi:10.1177/000348949310200602

Nevill, A. M., & Atkinson, G. (1997). Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science. *Br J Sports Med, 31*(4), 314-318.

Oates, J., & Russel, A. (1997). Perceptual Voice Profile. *A Sound Judgement. CDROM*.

Oates, J. M. (2014). Treatment of dysphonia in older people: the role of the speech therapist. *Curr Opin Otolaryngol Head Neck Surg, 22*(6), 477-486. doi:10.1097/moo.0000000000000109

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. *J Voice, 28*(6), 733-741. doi:10.1016/j.jvoice.2014.02.004

Ogawa, M., Hosokawa, K., Yoshida, M., Yoshii, T., Shiromoto, O., & Inohara, H. (2013). Immediate effectiveness of humming on the supraglottic compression in subjects with muscle tension dysphonia. *Folia Phoniatr Logop, 65*(3), 123-128. doi:10.1159/000353539

Ohno, T., & Hirano, S. (2014). Treatment of aging vocal folds: novel approaches. *Curr Opin Otolaryngol Head Neck Surg, 22*(6), 472-476. doi:10.1097/moo.0000000000000096

Ohno, T., Yoo, M. J., Swanson, E. R., Hirano, S., Ossoff, R. H., & Rousseau, B. (2009a). Regeneration of aged rat vocal folds using hepatocyte growth factor therapy. *Laryngoscope, 119*(7), 1424-1430. doi:10.1002/lary.20497

Ohno, T., Yoo, M. J., Swanson, E. R., Hirano, S., Ossoff, R. H., & Rousseau, B. (2009b). Regenerative effects of basic fibroblast growth factor on extracellular matrix production in aged rat vocal folds. *Ann Otol Rhinol Laryngol, 118*(8), 559-564.

Omori, K., Slavit, D. H., Matos, C., Kojima, H., Kacker, A., & Blaugrund, S. M. (1997). Vocal fold atrophy: quantitative glottic measurement and vocal function. *Ann Otol Rhinol Laryngol, 106*(7 Pt 1), 544-551. doi:10.1177/000348949710600702

Ozdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage, 33*(2), 628-635. doi:10.1016/j.neuroimage.2006.07.013

Paulsen, F., Kimpel, M., Lockemann, U., & Tillmann, B. (2000). Effects of ageing on the insertion zones of the human vocal fold. *J Anat, 196 (Pt 1)*, 41-54.

Pellegrino, R., Viegi, G., Brusasco, V., Crapo, R. O., Burgos, F., Casaburi, R., . . . Wanger, J. (2005). Interpretative strategies for lung function tests. *Eur Respir J, 26*(5), 948-968. doi:10.1183/09031936.05.00035205

Penfield, W., & Boldrey, E. (1937). SOMATIC MOTOR AND SENSORY REPRESENTATION IN THE CEREBRAL CORTEX OF MAN AS STUDIED BY ELECTRICAL STIMULATION1. *Brain, 60*(4), 389-443. doi:10.1093/brain/60.4.389

Pereira, L. P. M., M. L.; Carvalho, F. M. (2015). Vocal warm-up and breathing training for teachers: randomized clinical trial. *Rev Saude Publica, 49*, 67. doi:10.1590/s0034-8910.2015049005716

Perie, S., St Guily, J. L., Callard, P., & Sebille, A. (1997). Innervation of adult human laryngeal muscle fibers. *J Neurol Sci, 149*(1), 81-86.

Pessin, A. B., Tavares, E. L., Gramuglia, A. C., de Carvalho, L. R., & Martins, R. H. (2017). Voice and ageing: clinical, endoscopic and acoustic investigation. *Clin Otolaryngol, 42*(2), 330-335. doi:10.1111/coa.12725

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.

Phadke, K. V., Laukkanen, A. M., Ilomaki, I., Kankare, E., Geneid, A., & Svec, J. G. (2018). Cepstral and Perceptual Investigations in Female Teachers With Functionally Healthy Voice. *J Voice*. doi:10.1016/j.jvoice.2018.09.010

Pitts, T., Bolser, D., Rosenbek, J., Troche, M., Okun, M. S., & Sapienza, C. (2009). Impact of expiratory muscle strength training on voluntary cough and swallow function in Parkinson disease. *Chest, 135*(5), 1301-1308. doi:10.1378/chest.08-1389

Poburka, B. J., & Bless, D. M. (1998). A multi-media, computer-based method for stroboscopy rating training. *J Voice, 12*(4), 513-526.

Poburka, B. J., Patel, R. R., & Bless, D. M. (2016). Voice-Vibratory Assessment With Laryngeal Imaging (VALI) Form: Reliability of Rating Stroboscopy and High-speed Videoendoscopy. *J Voice*. doi:10.1016/j.jvoice.2016.12.003

Pontes, P., Yamasaki, R., & Behlau, M. (2006). Morphological and functional aspects of the senile larynx. *Folia Phoniatr Logop, 58*(3), 151-158. doi:10.1159/000091729

Ptok, M., & Strack, D. (2008). Electrical stimulation-supported voice exercises are superior to voice exercise therapy alone in patients with unilateral recurrent laryngeal nerve paresis: results from a prospective, randomized clinical trial. *Muscle Nerve, 38*(2), 1005-1011. doi:10.1002/mus.21063

Rabe, K. F., Hurd, S., Anzueto, A., Barnes, P. J., Buist, S. A., Calverley, P., . . . Van Weel, C. (2007). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med, 176*(6), 532-555.

Ramig, L. O., Countryman, S., O'Brien, C., Hoehn, M., & Thompson, L. (1996). Intensive speech treatment for patients with Parkinson's disease: short-and long-term comparison of two techniques. *Neurology, 47*(6), 1496-1504.

Ramig, L. O., Countryman, S., Thompson, L. L., & Horii, Y. (1995). Comparison of two forms of intensive speech treatment for Parkinson disease. *J Speech Hear Res, 38*(6), 1232-1251.

Ramig, L. O., & Dromey, C. (1996). Aerodynamic mechanisms underlying treatment-related changes in vocal intensity in patients with Parkinson disease. *J Speech Hear Res, 39*(4), 798-807.

Ramig, L. O., Gray, S., Baker, K., Corbin-Lewis, K., Buder, E., Luschei, E., . . . Smith, M. (2001). The aging voice: a review, treatment data and familial and genetic perspectives. *Folia Phoniatr Logop, 53*(5), 252-265. doi:52680

Ramirez-Sarmiento, A., Orozco-Levi, M., Guell, R., Barreiro, E., Hernandez, N., Mota, S., . . . Gea, J. (2002). Inspiratory muscle training in patients with chronic obstructive pulmonary disease: structural adaptation and physiologic outcomes. *Am J Respir Crit Care Med, 166*(11), 1491-1497. doi:10.1164/rccm.200202-075OC

Ramsay, S. C., Adams, L., Murphy, K., Corfield, D. R., Grootoonk, S., Bailey, D. L., . . . Guz, A. (1993). Regional cerebral blood flow during volitional expiration in man: a comparison with volitional inspiration. *J Physiol, 461*, 85-101.

Ray, C. T., M. D.; McCoy, S. (2018). Effects of Respiratory Muscle Strength Training in Classically Trained Singers. *Journal of Voice, 32*(5), 644.e625-644.e634. doi:10.1016/j.jvoice.2017.08.005

Reychler, G., Delacroix, S., Dresse, D., Pieters, T., & Liistro, G. (2016). Randomized Controlled Trial of the Effect of Inspiratory Muscle Training and Incentive Spirometry on Respiratory Muscle Strength, Chest Wall Expansion, and Lung Function in Elderly Adults. *J Am Geriatr Soc, 64*(5), 1128-1130. doi:10.1111/jgs.14097

Reyes, A., Cruickshank, T., Nosaka, K., & Ziman, M. (2015). Respiratory muscle training on pulmonary and swallowing function in patients with Huntington's disease: a pilot randomised controlled trial. *Clin Rehabil, 29*(10), 961-973. doi:10.1177/0269215514564087

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport, 11*(9), 1997-2000.

Roberts, T., Morton, R., & Al-Ali, S. (2011). Microstructure of the vocal fold in elderly humans. *Clin Anat, 24*(5), 544-551. doi:10.1002/ca.21114

Rodeno, M. T., Sanchez-Fernandez, J. M., & Rivera-Pomar, J. M. (1993). Histochemical and morphometrical ageing changes in human vocal cord muscles. *Acta Otolaryngol, 113*(3), 445-449.

Romer, L. M., & McConnell, A. K. (2004). Inter-test reliability for non-invasive measures of respiratory muscle function in healthy humans. *Eur J Appl Physiol, 91*(2-3), 167-176. doi:10.1007/s00421-003-0984-2

Rosen, C. A., Gartner-Schmidt, J., Casiano, R., Anderson, T. D., Johnson, F., Remacle, M., . . . Zraick, R. I. (2009). Vocal fold augmentation with calcium hydroxylapatite: twelve-month report. *Laryngoscope, 119*(5), 1033-1041. doi:10.1002/lary.20126

Rosen, C. A., Lee, A. S., Osborne, J., Zullo, T., & Murry, T. (2004). Development and validation of the voice handicap index-10. *Laryngoscope, 114*(9), 1549-1556. doi:10.1097/00005537-200409000-00009

Rosner, B. (2005). *Fundamentals of Biostatistics*. Belmont, CA: Duxbury Press.

Roy, N., Kim, J., Courey, M., & Cohen, S. M. (2015). Voice disorders in the elderly: A national database study. *Laryngoscope*. doi:10.1002/lary.25511

Roy, N., Merrill, R. M., Gray, S. D., & Smith, E. M. (2005). Voice disorders in the general population: prevalence, risk factors, and occupational impact. *Laryngoscope, 115*(11), 1988-1995. doi:10.1097/01.mlg.0000179174.32345.41

Roy, N., Weinrich, B., Gray, S. D., Tanner, K., Stemple, J. C., & Sapienza, C. M. (2003). Three treatments for teachers with voice disorders: a randomized clinical trial. *J Speech Lang Hear Res, 46*(3), 670-688.

Rubens, A. B. (1975). Aphasia with infarction in the territory of the anterior cerebral artery. *Cortex, 11*(3), 239-250.

Ruotsalainen, J. H. S., J.; Lehto, L.; Jauhiainen, M.; Verbeek, J. H. (2007). Interventions for treating functional dysphonia in adults. *Cochrane Database Syst Rev*(3), Cd006373. doi:10.1002/14651858.CD006373.pub2

Sachs, M. C., Enright, P. L., Hinckley Stukovsky, K. D., Jiang, R., & Barr, R. G. (2009). Performance of Maximum Inspiratory Pressure Tests and Maximum Inspiratory Pressure Reference Equations for 4 Race/Ethnic Groups. *Respir Care, 54*(10), 1321-1328.

Sandnes, A., Andersen, T., Hilland, M., Ellingsen, T. A., Halvorsen, T., Heimdal, J. H., & Roksund, O. D. (2013). Laryngeal movements during inspiratory muscle training in healthy subjects. *J Voice, 27*(4), 448-453. doi:10.1016/j.jvoice.2013.02.010

Sapienza, C., Troche, M., Pitts, T., & Davenport, P. (2011). Respiratory strength training: concept and intervention outcomes. *Semin Speech Lang, 32*(1), 21-30. doi:10.1055/s-0031-1271972

Sapienza, C. M. (2008). Respiratory muscle strength training applications. *Curr Opin Otolaryngol Head Neck Surg, 16*(3), 216-220. doi:10.1097/MOO.0b013e3282fe96bd

Sapienza, C. M., Brown, J., Martin, D., & Davenport, P. (1999). Inspiratory pressure threshold training for glottal airway limitation in laryngeal papilloma. *J Voice, 13*(3), 382-388.

Sapienza, C. M., & Wheeler, K. (2006). Respiratory muscle strength training: functional outcomes versus plasticity. *Semin Speech Lang, 27*(4), 236-244. doi:10.1055/s-2006-955114

Sato, K., & Hirano, M. (1997). Age-related changes of elastic fibers in the superficial layer of the lamina propria of vocal folds. *Ann Otol Rhinol Laryngol, 106*(1), 44-48. doi:10.1177/000348949710600109

Sato, K., & Hirano, M. (1998). Age-related changes in the human laryngeal glands. *Ann Otol Rhinol Laryngol, 107*(6), 525-529. doi:10.1177/000348949810700612

Sato, K., Hirano, M., & Nakashima, T. (2002). Age-related changes of collagenous fibers in the human vocal fold mucosa. *Ann Otol Rhinol Laryngol, 111*(1), 15-20.

Sato, T., & Tauchi, H. (1982). Age changes in human vocal muscle. *Mech Ageing Dev, 18*(1), 67-74.

Sauder, C., Roy, N., Tanner, K., Houtz, D. R., & Smith, M. E. (2010). Vocal function exercises for presbylaryngis: a multidimensional assessment of treatment outcomes. *Ann Otol Rhinol Laryngol, 119*(7), 460-467.

Schaeffer, N. (2017). Pre- and Poststimulation Study on the Phonatory Aerodynamic System on Participants with Dysphonia. *Journal of Voice, 31*(2), 254.e251-254.e259.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., . . . Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods, 9*(7), 676-682. doi:10.1038/nmeth.2019

Schneider, B., van Trotsenburg, M., Hanke, G., Bigenzahn, W., & Huber, J. (2004). Voice impairment and menopause. *Menopause, 11*(2), 151-158.

Seino, Y., & Allen, J. E. (2014). Treatment of aging vocal folds: surgical approaches. *Curr Opin Otolaryngol Head Neck Surg, 22*(6), 466-471. doi:10.1097/moo.0000000000000099

Sellars, C., Carding, P. N., Deary, I. J., MacKenzie, K., & Wilson, J. A. (2006). Characterization of effective primary voice therapy for dysphonia. *The Journal of Laryngology & Otology, 116*(12), 1014-1018. doi:10.1258/002221502761698757

Shah, R. N., Deal, A. M., & Buckmire, R. A. (2013). Multidimensional voice outcomes after type I Gore-Tex thyroplasty in patients with nonparalytic glottic incompetence: a subgroup analysis. *Laryngoscope, 123*(7), 1742-1745. doi:10.1002/lary.23983

Shaheen Awan, J. B.-K., Mark Courey, Dimitar Deliyski, Tanya Edie, Robert Hillman, Rita Patel, Diane Paul, Jan Svec. (2005). Recommended Protocols for Instrumental Assessment of Voice - ASHA Committee on Instrumental Voice Assessment Protocols (IVAP).

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol Bull, 86*(2), 420-428.

Silva, I. S., Fregonezi, G. A., Dias, F. A., Ribeiro, C. T., Guerra, R. O., & Ferreira, G. M. (2013). Inspiratory muscle training for asthma. *Cochrane Database Syst Rev*(9), CD003792. doi:10.1002/14651858.CD003792.pub2

Silverman, E. P., Sapienza, C. M., Saleem, A., Carmichael, C., Davenport, P. W., Hoffman-Ruddy, B., & Okun, M. S. (2006). Tutorial on maximum inspiratory and expiratory mouth pressures in individuals with idiopathic Parkinson disease (IPD) and the preliminary results of an expiratory muscle strength training program. *NeuroRehabilitation, 21*(1), 71-79.

Smith, E., Verdolini, K., Gray, S., Nichols, S., Lemke, J., Barkmeier, J., . . . Hoffman, H. (1996). Effect of voice disorders on quality of life. *Journal of Medical Speech-Language Pathology, 4*(4), 223-244.

Solomon, N. P., Garlitz, S. J., & Milbrath, R. L. (2000). Respiratory and laryngeal contributions to maximum phonation duration. *J Voice, 14*(3), 331-340.

Souza, H., Rocha, T., Pessoa, M., Rattes, C., Brandao, D., Fregonezi, G., . . . Dornelas, A. (2014). Effects of inspiratory muscle training in elderly women on respiratory muscle strength, diaphragm thickness and mobility. *J Gerontol A Biol Sci Med Sci, 69*(12), 1545-1553. doi:10.1093/gerona/glu182

Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., & Toga, A. W. (2003). Mapping cortical change across the human life span. *Nat Neurosci, 6*(3), 309-315. doi:10.1038/nn1008

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. *Am J Speech Lang Pathol, 16*(2), 95-107. doi:10.1044/1058-0360(2007/014)

Stathopoulos, E. T., & Sapienza, C. (1993). Respiratory and laryngeal function of women and men during vocal intensity variation. *J Speech Hear Res, 36*(1), 64-75.

Stemple, J. C. (2005). A holistic approach to voice therapy. *Semin Speech Lang, 26*(2), 131-137. doi:10.1055/s-2005-871209

Stemple, J. C., Lee, L., D'Amico, B., & Pickup, B. (1994). Efficacy of vocal function exercises as a method of improving voice production. *J Voice, 8*(3), 271-278.

Stevens, S. (1975). *Psychophysics: introduction to its perceptual, neural, and social prospects*. New York: Wiley & Sons.

Sundarrajan, A., Huber, J. E., & Sivasankar, M. P. (2017). Respiratory and Laryngeal Changes With Vocal Loading in Younger and Older Individuals. *J Speech Lang Hear Res, 60*(9), 2551-2556. doi:10.1044/2017_jslhr-s-17-0106

Suzuki, T., Connor, N. P., Lee, K., Bless, D. M., Ford, C. N., & Inagi, K. (2002). Age-related alterations in myosin heavy chain isoforms in rat intrinsic laryngeal muscles. *Ann Otol Rhinol Laryngol, 111*(11), 962-967. doi:10.1177/000348940211101102

Takano, S., Kimura, M., Nito, T., Imagawa, H., Sakakibara, K., & Tayama, N. (2010). Clinical analysis of presbylarynx--vocal fold atrophy in elderly individuals. *Auris Nasus Larynx, 37*(4), 461-464. doi:10.1016/j.anl.2009.11.013

Tanner, K., Sauder, C., Thibeault, S. L., Dromey, C., & Smith, M. E. (2010). Vocal fold bowing in elderly male monozygotic twins: a case study. *J Voice, 24*(4), 470-476. doi:10.1016/j.jvoice.2008.10.010

Tay, E. Y., Phyland, D. J., & Oates, J. (2012). The effect of vocal function exercises on the voices of aging community choral singers. *J Voice, 26*(5), 672.e619-627. doi:10.1016/j.jvoice.2011.12.014

Thibeault, S. L. (2005). Advances in our understanding of the Reinke space. *Curr Opin Otolaryngol Head Neck Surg, 13*(3), 148-151.

Thomas, J. (Fall 2016). *Human Anatomy for Nurse Anesthetists AFN-527.* Paper presented at the Medical University of South Carolina, Charleston, SC.

Thomas, L. B., Harrison, A. L., & Stemple, J. C. (2008). Aging thyroarytenoid and limb skeletal muscle: lessons in contrast. *J Voice, 22*(4), 430-450. doi:10.1016/j.jvoice.2006.11.006

Thoms, G., & Jurgens, U. (1987). Common input of the cranial motor nuclei involved in phonation in squirrel monkey. *Exp Neurol, 95*(1), 85-99.

Tiago, R. S., Pontes, P. A., & Brasil Ode, O. (2008). Quantitative analysis of myelinic fibers in human laryngeal nerves according to age. *Braz J Otorhinolaryngol, 74*(1), 45-52.

Titze, I. R. (2001). Acoustic interpretation of resonant voice. *J Voice, 15*(4), 519-528. doi:10.1016/s0892-1997(01)00052-2

Titze, I. R. (2006). Voice training and therapy with a semi-occluded vocal tract: rationale and scientific underpinnings. *J Speech Lang Hear Res, 49*(2), 448-459. doi:10.1044/1092-4388(2006/035)

Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: theory. *J Acoust Soc Am, 123*(5), 2733-2749. doi:10.1121/1.2832337

Titze, I. R., & Worley, A. S. (2009). Modeling source-filter interaction in belting and high-pitched operatic male singing. *J Acoust Soc Am, 126*(3), 1530. doi:10.1121/1.3160296

Tolep, K., & Kelsen, S. G. (1993). Effect of aging on respiratory skeletal muscles. *Clin Chest Med, 14*(3), 363-378.

Tomita, H., Nakashima, T., Maeda, A., Umeno, H., & Sato, K. (2006). Age related changes in the distribution of laryngeal glands in the human adult larynx. *Auris Nasus Larynx, 33*(3), 289-294. doi:10.1016/j.anl.2006.01.001

Troche, M. S., Okun, M. S., Rosenbek, J. C., Musson, N., Fernandez, H. H., Rodriguez, R., . . . Sapienza, C. M. (2010). Aspiration and swallowing in Parkinson disease and rehabilitation with EMST: a randomized trial. *Neurology, 75*(21), 1912-1919. doi:10.1212/WNL.0b013e3181fef115

Tsai, Y. C., Huang, S., Che, W. C., Huang, Y. C., Liou, T. H., & Kuo, Y. C. (2016). The Effects of Expiratory Muscle Strength Training on Voice and Associated Factors in Medical Professionals With Voice Disorders. *J Voice, 30*(6), 759.e721-759.e727. doi:10.1016/j.jvoice.2015.09.012

Turkmen, S., Cansu, A., Turedi, S., Eryigit, U., Sahin, A., Gunduz, A., & Shavit, I. (2012). Age-dependent structural and radiological changes in the larynx. *Clin Radiol, 67*(11), e22-26. doi:10.1016/j.crad.2012.07.006

Turley, R., & Cohen, S. (2009). Impact of voice and swallowing problems in the elderly. *Otolaryngol Head Neck Surg, 140*(1), 33-36. doi:10.1016/j.otohns.2008.10.010

Vaca, M., Cobeta, I., Mora, E., & Reyes, P. (2017). Clinical Assessment of Glottal Insufficiency in Age-related Dysphonia. *J Voice, 31*(1), 128.e121-128.e125. doi:10.1016/j.jvoice.2015.12.010

Vaca, M., Mora, E., & Cobeta, I. (2015). The Aging Voice: Influence of Respiratory and Laryngeal Changes. *Otolaryngol Head Neck Surg, 153*(3), 409-413. doi:10.1177/0194599815592373

Verdolini, K., Druker, D. G., Palmer, P. M., & Samawi, H. (1998). Laryngeal adduction in resonant voice. *J Voice, 12*(3), 315-327.

Verdolini-Marston, K., Burke, M. K., Lessac, A., Glaze, L., & Caldwell, E. (1995). Preliminary study of two methods of treatment for laryngeal nodules. *J Voice, 9*(1), 74-85.

von Glass, W., & Pesch, H. J. (1983). [Ossification principle of the laryngeal skeleton of the human and mammals. Comparative anatomic studies]. *Acta Anat (Basel), 116*(2), 158-167.

Vorona, S., Sabatini, U., Al-Maqbali, S., Bertoni, M., Dres, M., Bissett, B., . . . Goligher, E. C. (2018). Inspiratory Muscle Rehabilitation in Critically Ill Adults: A Systematic Review and Meta-Analysis. *Ann Am Thorac Soc, 15*(6), 735-744.

Watts, C. R., Diviney, S. S., Hamilton, A., Toles, L., Childs, L., & Mau, T. (2015). The effect of stretch-and-flow voice therapy on measures of vocal function and handicap. *J Voice, 29*(2), 191-199. doi:10.1016/j.jvoice.2014.05.008

Watts, C. R. H., Amy; Toles, Laura; Childs, Lesley; Mau, Ted. (2015). A randomized controlled trial of stretch-and-flow voice therapy for muscle tension Dysphonia. *Laryngoscope, 125*(6), 1420-1425 1426p. doi:10.1002/lary.25155

Weiner, P., Magadle, R., Beckerman, M., Weiner, M., & Berar-Yanay, N. (2003). Specific expiratory muscle training in COPD. *Chest, 124*(2), 468-473.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res, 19*(1), 231-240. doi:10.1519/15184.1

Williams, P. L., Bannister, L. H., Berry, M. M., & al., e. (1995). *Gray's Anatomy*. London: Churchill Livingston.

Wingate, J. M., Brown, W. S., Shrivastav, R., Davenport, P., & Sapienza, C. M. (2007). Treatment outcomes for professional voice users. *Journal of Voice, 21*(4), 433-449.

Woo, P., Casper, J., Colton, R., & Brewer, D. (1992). Dysphonia in the aging: physiology versus disease. *Laryngoscope, 102*(2), 139-144. doi:10.1288/00005537-199202000-00007

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. *J Voice, 13*(4), 508-517.

Ximenes Filho, J. A., Tsuji, D. H., do Nascimento, P. H., & Sennes, L. U. (2003). Histologic changes in human vocal folds correlated with aging: a histomorphometric study. *Ann Otol Rhinol Laryngol, 112*(10), 894-898.

Xue, S. A., & Deliyski, D. (2001). Effects of aging on selected acoustic voice parameters: Preliminary normative data and educational implications. *Educational Gerontology, 27*(2), 159-168.

Xue, S. A., & Hao, G. J. (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study. *J Speech Lang Hear Res, 46*(3), 689-701.

Yamauchi, A., Yokonishi, H., Imagawa, H., Sakakibara, K., Nito, T., Tayama, N., & Yamasoba, T. (2015). Vocal Fold Vibration in Vocal Fold Atrophy: Quantitative Analysis With High-Speed Digital Imaging. *J Voice, 29*(6), 755-762. doi:10.1016/j.jvoice.2014.12.008

Yin, J., & Zhang, Z. (2014). Interaction between the thyroarytenoid and lateral cricoarytenoid muscles in the control of vocal fold adduction and eigenfrequencies. *J Biomech Eng, 136*(11). doi:10.1115/1.4028428

Yiu, E. M., & Ho, E. Y. (2002). Short-term effect of humming on vocal quality. *Asia Pacific Journal of Speech, Language and Hearing, 7*(3), 123-137.

Yiu, E. M., Lo, M. C., & Barrett, E. A. (2017). A systematic review of resonant voice therapy. *Int J Speech Lang Pathol, 19*(1), 17-29. doi:10.1080/17549507.2016.1226953

Zeitels, S. M., Mauri, M., & Dailey, S. H. (2003). Medialization laryngoplasty with Gore-Tex for voice restoration secondary to glottal incompetence: indications and observations. *Ann Otol Rhinol Laryngol, 112*(2), 180-184. doi:10.1177/000348940311200213

Zhang, Z. (2015). Regulation of glottal closure and airflow in a three-dimensional phonation model: implications for vocal intensity control. *J Acoust Soc Am, 137*(2), 898-910. doi:10.1121/1.4906272

Zhang, Z. (2016a). Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *J Acoust Soc Am, 139*(4), 1493. doi:10.1121/1.4944754

Zhang, Z. (2016b). Mechanics of human voice production and control. *J Acoust Soc Am, 140*(4), 2614. doi:10.1121/1.4964509

Zhang, Z. (2016c). Respiratory Laryngeal Coordination in Airflow Conservation and Reduction of Respiratory Effort of Phonation. *J Voice, 30*(6), 760.e767-760.e713. doi:10.1016/j.jvoice.2015.09.015

Zhang, Z. (2019). Compensation Strategies in Voice Production With Glottal Insufficiency. *J Voice, 33*(1), 96-102. doi:10.1016/j.jvoice.2017.10.002

Ziade, G., Semaan, S., Ghulmiyyah, J., Kasti, M., & Hamdan, A. L. (2016). Structural and Anatomic Laryngeal Measurements in Geriatric Population Using MRI. *J Voice*. doi:10.1016/j.jvoice.2016.06.008

Ziegler, A. V. A., Katherine; Johns, Michael; Klein, Adam; Hapner, Edie R. (2014). Preliminary data on two voice therapy interventions in the treatment of presbyphonia. *Laryngoscope, 124*(8), 1869-1876 1868p. doi:10.1002/lary.24548

Zraick, R. I., Kempster, G. B., Connor, N. P., Thibeault, S., Klaben, B. K., Bursac, Z., . . . Glaze, L. E. (2011). Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *Am J Speech Lang Pathol, 20*(1), 14-22. doi:10.1044/1058-0360(2010/09-0105)

Zraick, R. I., Smith-Olinde, L., & Shotts, L. L. (2012). Adult normative data for the KayPENTAX Phonatory Aerodynamic System Model 6600. *J Voice, 26*(2), 164-176. doi:10.1016/j.jvoice.2011.01.006

Appendices

Appendix I: Pre- and Post-Intervention Tasks and Outcome Measures

Assessment Category	Tasks	Outcome Measures
Respiratory Assessments	Spirometry testing Respiratory muscle pressure testing	FVC, FEV1, and FEV1/FVC (raw and percent predicted values MIP and MEP
Laryngeal Features	Visual examination of the larynx with videostrosbocopy	Bowing index Laryngeal parameters from the VALI form
Acoustic and Auditory-Perceptual Assessments	Sustained /a/ at comfortable pitch and loudness for approximately 5 seconds	NHR, APQ,CPPS, overall severity (CAPE-V)
	Read the 6 sentences from the CAPE-V	CPPS during reading, overall severity (CAPE-V)
	Natural speech for approximately one minute	Mean SPL during conversation, overall severity (CAPE-V)
Aerodynamic Assessment	Syllable repetition with the PAS	Subglottal pressure, mean airflow during voicing, and aerodynamic resistance
Self-Assessments	VHI-10, GFI, and CPIB	Scores from the VHI-10, GFI, and CPIB

Appendix II: Individual Results for Respiratory Measures

Participant	MIP	Predicted (LLN)	MEP	Predicted (LLN)	FVC%	FEV1%	FEV1/FVC	FEV1/FVC%
1	69	54.94 (22.94)	114	109.95 (57.95)	113	107	0.71	95
2 (interstitial disease)	97	49.70 (17.70)	96	96.24 (44.24)	102	95	0.69	93
3 (mild asthma)	121	65.35 (33.35)	121	137.22 (85.22)	68	61	0.68	89
4	133	97.72 (56.72)	163	210.45 (139.45)	120	131	0.83	109
5 (CREST syndrome)	144	108.08 (67.08)	117	231.80 (160.80)	85	76	0.69	90
6	45	73.85 (41.85)	89	159.18 (107.18)	68	75	0.84	109
7	113	80.81 (39.81)	160	169.75 (98.75)	108	106	0.71	97
8	49	43.23 (11.23)	85	79.06 (27.06)	136	143	0.76	106
9	119	103.18 (62.18)	176	221.4 (150.4)	87	87	0.76	100
10	60	83.67 (42.67)	126	175.75 (104.75)	51	37	0.52	72
11	71	49.42 (17.42)	90	95.50 (43.50)	115	115	0.75	101
12	81	62.96 (30.96)	100	131.08 (79.08)	79	75	0.73	94
13	59	52.49 (20.49)	86	103.46 (51.46)	76	69	0.67	91
14	56	57.19 (25.19)	81	115.84 (63.84)	100	98	0.74	98
16	57	79.17 (47.17)	67	172.88 (120.88)	64	61	0.73	96
17	80	86.92 (45.92)	132	181.95 (110.95)	85	89	0.76	104
18	81	73.07 (32.07)	147	151.3 (80.3)	106	105	0.70	98
19	21	82.64 (41.64)	89	173.25 (102.25)	42	44	0.74	103
20	107	94.25 (53.25)	174	201.2 (130.2)	85	86	0.77	100
21	114	95.47 (54.47)	186	203 (132)	111	108	0.73	97
22	59	73.66 (32.66)	83	152.95 (81.95)	85	88	0.73	103

Appendix III: Number of Subjects Who Improved on Each Item of the Self-Assessment Questionnaires Following Treatment

Voice Handicap Index-10 Item	IMST (n=4)	EMST (n=3)	VFE (n=3)
1. My voice makes it difficult for people to hear me	4		1
2. People have difficulty understanding me in a noisy room	3		1
3. My voice difficulties restrict personal and social life	3	1	
4. I feel left out of conversations because of my voice	3	1	
5. My voice problem causes me to lose income			
6. I feel as though I have to strain to produce voice	3		1
7. The clarity of my voice is unpredictable	3		1
8. My voice problem upsets me	2	1	1
9. My voice makes me feel handicapped	2	1	
10. People ask What's wrong with your voice?	3		1

Glottal Function Index Item	IMST (n=4)	EMST (n=3)	VFE (n=3)
1. Speaking took extra effort	2	1	3
2. Throat discomfort or pain after using your voice	2		
3. Vocal fatigue (voice weakened as you talked)	2	1	1
4. Voice cracks or sounds different	3	1	

Communicative Participant Item Bank Item Does your condition interfere with:	IMST (n=3)	EMST (n=3)	VFE (n=3)
1. Talking with people you know	3	1	2
2. Communicating when you need to say something quickly	2		1
3. Talking with people you do not know	2	1	2
4. Communicating when you are out in your community (e.g. errands; appointments)	2	1	2
5. Asking questions in a conversation	1	1	1
6. Communicating in a small group of people	3	1	
7. Having a long conversation with someone you know about a book, movie, show or sports event	3	1	2
8. Giving someone detailed information	2	1	1
9. Getting your turn in a fast-moving conversation	2	1	
10. Trying to persuade a friend or family member to see a different point of view	3	2	