University of Alberta

Changes Observed in Persons with Parkinson's Disease Pre- and Post-Voice Choral Singing Therapy

by

Sable Jade Chan

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

in

Speech-Language Pathology

Department of Speech Pathology and Audiology

©Sable Jade Chan

Spring 2012
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Dedication

This thesis is dedicated to the participants who had the strength and determination to complete the original study on which this thesis is based. Though you were unaware of it the time, the time and dedication you gave throughout the treatment would fuel my project and passion five years later. Thank-you for your participation as we continue to investigate treatment options for those still struggling with Parkinson's Disease.

Abstract

Voice Choral Singing Therapy has been gaining interest as an alternative form of voice therapy for persons with Parkinson's disease. Often Voice Choral Singing Therapy employs individual and/or group singing activities to improve phonation. The present thesis follows this trend. A professional singing teacher provided individual singing instruction three times per week and group singing instruction once per week for one month, resulting in six and one-half hours individual instruction and four hours group instruction for every one of five participants. Interview data, singing, and speech samples were collected and analyzed to observe outcomes from pre- to post therapy.

The interview data were analyzed making novel use of a global rating change scale (Kamper, Maher, & MacKay, 2009) to detect changes from pre- to post-therapy in swallowing, coughing, speech, and facial-expression domains. The findings were inconclusive, but a trend was detected for some change in the speech domain. The singing data were submitted to a descriptive analysis using several acoustic measurements. After therapy, participants took fewer breaths, increased mean intensity, maximum intensity, and range of intensity following therapy. In the speech data mean intensity, maximum intensity, and the range of intensity also increased. These three intensity measures are not independent variables; they are interrelated and changes in one would

be expected in all. Pre- and post-therapy sentences were paired in a discrimination listening task for 33 naïve listeners. Naïve listeners significantly chose the post-treatment speech samples as sounding better. [McNemar test > chi2= 0.0000].

Interview data yielded detection of only a trend toward post-treatment speech changes, however, singing and speech data showed positive acoustic and auditory-perceptual differences following Voice Choral Singing Therapy.

Acknowledgements

I would like to express my sincere gratitude to the following people who all played crucial roles in helping me carry out my research.

♪

My committee members: Professor Harold Wiens, Dr. Megan Hodge, Dr. Benjamin Tucker, and Dr. Melanie Michelle Campbell. Whether it was graciously providing me with the data for my research, posing of thought-provoking questions, or the countless hours of reading and motivational words---all these elements challenged me to become a better student and researcher.

♪

My financial benefactors: the Speech Pathology and Audiology department and the Faculty of Rehabilitation Medicine.

♪

My statistical gurus: Susan Armijo Olivio & Jorge Fuentes Contreras in the Rehabilitation Research Centre for all of their guidance and support.

♪

The Alberta Phonetics Laboratory research assistants: Kaitlin Mackie, Dallin Mendenhall, Shelley Lim, Ryan Podlubny and Alexandra Schellenberger for helping me run all my listeners and welcoming me into the lab.

♪

My role model: Merrill Tanner who inspired this vein of Parkinson's research for showing me how to unite my clinical, educational, and musical interests.

♪

To my parents, Stephen and Ann Chan, who have always unconditionally supported all of my educational and life passions.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

FRC-Functional Residual Capacity

GRC-Global Rating of Change

IC-Inspiratory Capacity

LSVT-Lee Silverman Voice Therapy

MID-Minimal Important Difference

PD-Parkinson's Disease

PPD-Persons with Parkinson's Disease

RET-Respiratory Exercise Therapy

SD-Standard Deviation

SIT-Speech Intelligibility Test

S-LP-Speech-Language Pathologist

TLC-Total Lung Capacity

VC-Vital Capacity

VCST-Voice Choral Singing Therapy

Literature Review

Introduction

Parkinson's Disease (PD) is a neurodegenerative adult disorder that presents with a loss of motor function. It is suspected that PD is the result of an interaction between genetic susceptibility and environmental toxic factors (Duffy, 2005). PD is the second most common neurodegenerative disease, next to Alzheimers (Shulman & DeJager, 2009). The mean age of adult onset is approximately 50 years of age (Duffy, 2005). About one to two per 100,000 of the population is affected (Duffy, 2005). Korell and Tanner (2005) found that the incidence of PD increases once individuals are over the age of 50. There are approximately nine years of survival from symptom onset (Duffy, 2005). It is twice as frequent in men as in women (Przedborski, 2007).

Bartels and Leenders (2009) note that main symptoms of PD include resting tremor, unsteady gait, hypokinesia (decreased amplitude of movement), bradykinesia (slowness of movement), rigidity, akinesia (slow initiation of movement), postural instability, masked facies (frozen facial expression), festination (rapid, shuffling walking steps), and micrographia (small handwriting).

Visuo-motor coordination is an area of deficit in persons with Parkinson's Disease (PPD), reflected in a decreased ability to plan, generate, and execute motor functions. Visuo-motor coordination problems are difficulties coordinating vision with gross and fine motor movement, for example,

handwriting. This is due to the neurodegeneration in nondopaminergic areas in PPD. Inzelberg, Schechtman, and Hocherman (2008) argue that visuo-motor changes seen in PD are another effect of the multi-component role of the basal ganglia in executing motor, cognitive, and computational roles in normal functioning (Inzelberg et al., 2008).

Depression is the main non-motor symptom of PD and it occurs in approximately 40% of PD cases (Duffy, 2005). It is also important to note that some PD symptoms overlap with the symptoms of depression. These symptoms include fatigue, psychomotor slowness, sleep disturbance, and withdrawal from social activities due to dyskinesias, which are abnormal and involuntary movements (e.g., hand tremor) (Duffy, 2005), and insecurity regarding frozen facial expressions (McDonald, Richard, & DeLong, 2003).

In addition to the previously mentioned symptoms, Nakashima et al. (1997) state that the most common cause of death in PPD is pneumonia. Therefore, pneumonia should be treated more intensively to increase the survival rate of PPD. Troche, Rosenbek, and Sapienza (2008) found that PPD have a decreased ability to properly clear material from their airways. Furthermore, an ineffective voluntary cough contributes to penetration (when food or liquid contacts the vocal folds) and aspiration (when food or liquid moves past the level of the vocal folds). The swallowing and coughing deficits have an impact on their quality of life (Plowman-Prine et al., 2009).

Facial expression is also a challenge for PPD as they display a masked or expressionless, unblinking facial expression that is a characteristic trademark of PD (Duffy, 2005). This lack of facial expression hinders a PPD's ability to express non-verbal communication and becomes an "acquired pragmatic deficit". Justice (2010) defines pragmatics as the rules governing how language is used for social purposes. Thus the term "acquired pragmatic deficit", as used in this thesis, describes the way in which PPD are unable to utilize non-verbal communication means (e.g., facial expression) to appropriately engage in social interactions, both for talking and listening.

The most common voice disorder for a PPD is hypokinetic dysarthria; it includes characteristics of reduced vocal loudness, reduced pitch variability, short phonation time, imprecise articulation, variable speech rate, and a monotone or breathy or hoarse voice. Hypokinetic dysarthria often results in reduced intelligibility for listeners (Sapir, Ramig, & Fox, 2008).

It is seen that PD is a multi-factorial disorder, which requires the efforts from a wide spectrum of disciplines to manage disease symptoms: from neurology to diagnose the disorder to physiotherapy for the management of hypokinesia (reduced mobility). Although PD can introduce a wide range of potential symptoms, the aim of this thesis is to observe pre/post-therapy changes in areas of deficit traditionally identified as therapeutic areas within

the domain of Speech-Language Pathology. Swallowing and coughing are closely related areas that lie within the scope of practice for Speech Language Pathologists (S-LPs) to address as swallowing specialists. The area of facial expression should be a targeted area for S-LPs to address as facial expression conveys non-verbal meaning and there is an acquired pragmatic deficit in PPD. The parts of speech, articulation, voice and prosody, are all target areas for S-LPs since they all play a role in affecting intelligibility and overall expressive language.

Implicated Neural Pathways

PD is typically due to the degeneration of the dopaminergic neurons in the mid brain. The substantia nigra is one of five subcortical nuclei that comprise the basal ganglia. The top right hand portion of Figure 1 shows a coronal slice (any vertical plane that divides the body into a ventral and dorsal (belly and back) sections) of the cortex. The structure on the bottom right hand portion is a transverse slice (any horizontal plane that divides the body into superior and inferior (top and bottom) sections) and highlights the substantia nigra pars compacta and substantia nigra pars reticulata regions. The boxes with the arrows on the left hand side of Figure 1 represent the stimulation and inhibition of the basal ganglia motor loop required when sending signals for motor movement. The pars compacta and pars recticulata refer to the two main sections in the substantia nigra. The basal ganglia are composed of the corpus striatum (which includes the caudate nucleus and putamen), globus

pallidus, subthalamic nucleus, and substantia nigra. The thalamus has the ventral lateral (VL) nucleus and the ventral anterior (VA) nucleus. The corpus striatum refers to the caudate and putamen as a group. The corpus striatum serves as an input centre for the basal ganglia (Purves et al., 2001). The pars compacta is also an input centre for the basal ganglia, in addition to the caudate and putamen. The flow of the arrows represents the flow of messages in the circuit needed to carry out an action. There are no excitatory or inhibitory signals between the substantia nigra pars compacta, caudate, and putamen because they are all input centres to the basal ganglia and they share information amongst these connected regions. The basal ganglia initiates voluntary motor movements, facilitates the suppression of motor movements, and provides feedback concerning motor commands; "+" signals excitatory connections and "-" signals inhibitory connections (See Figure 1.) The input unit (substantia nigra pars compacta, caudate, and putamen) of the basal ganglia has the power to inhibit the signals sent to the output unit (globus pallidus, substantia nigra pars reticulata and subthalamtic nucleus). The pars reticulata and globus pallidus are the output centre from the basal ganglia to other cortex areas. Depending on the signals coming from the input centre, the globus pallidus and sustantia nigra pars reticulata can send excitatory or inhibitory signals to the subthalamic nucleus. Therefore, it is seen that the motor circuit in the cortex is finely controlled through stimulation and inhibition feedback loops systems in the basal ganglia. A

disturbance at any one of these levels would result in a disruption of the motor circuit loop to plan and carry out motor actions.

Figure 1. Basal Ganglia Divisions.

Figure 1 has been removed because of copyright restrictions. The material contained inhibitory and excitatory connections of the basal ganglia motor circuit, basal ganglia divisions, as well as anatomical cross-sections of the brain. Source: Basal Ganglia Divisions from "Projections to the Basal Ganglia," by Purves, D., Augustine, G.J., Fitzpatrick, D., Katz, L.C., LaMantia, A-S, McNamara, J.O., & Williams, S.M, 2001, Neuroscience, 2. From http://www.ncbi.nlm.nih.gov/books/NBK10988/bin/ch18f1.jpg Retrieved 12 November 2011. Copyright 2001 by Sinauer Associates, Inc.

The death of the dopaminergic neurons in the substantia pars compacta results in the characteristic features of PD that include akinesia, bradykinesia, and tremor. Therefore, the degeneration of the dopaminergic neurons in the substantia nigra affects the direct and indirect motor circuits surrounding the basal ganglia. In the absence of dopamine, the cells fire rapidly within the striatum and this imbalance causes the pyramidal motor system with axons in the corticospinal (cortex to spine) and corticobulbar (cortex to brainstem) tracts and extrapyramidal pathways (neural network in cortex for the coordination of motor movement) to misfire signals for motor activities (Daube, Reagan, Sandok, & Westmoreland, 1986). Bartels and Leenders (2009) argue that neurodegeneration begins in the nondopaminergic area, such as in the enteric nervous system (a subdivision of the peripheral nervous system that is involved in autonomic function innervation such as reflexes), and rises through the spinal cord, brainstem and finally the nigral and subcortical brain regions, which may better account for the various other symptoms, such as cognitive and visuo-motor deficits, experienced by PPD. Although the specific neurological mechanisms highlight the potential etiology of PD, it is still unclear as to the extent of neurological changes that develop as a result of PD.

Swallowing and Coughing

Dysphagia (swallowing disorder) is a symptom that occurs during the course of PD and aspiration pneumonia is reported as the leading cause of death in

PPD (Nakashima et al., 1997). Aspiration pneumonia is the inflammation of the lungs that develops from the presence of foreign material (e.g., food or liquid). Figure 2 outlines the physical structures for swallowing. Any abnormalities in the swallowing process can lead to dysphagia, including oral preparation (containing food and liquid in the mouth with good lip closure), with delayed oral transition (tongue movement of the bolus (mass of food or liquid) to the back of the mouth) and bolus preparation (e.g., chewing), to a delayed triggering of the pharyngeal stage, with poor tongue base retraction and prolonged opening of the upper esophageal sphincter (also known at the cricopharyngeal sphincter). The pharyngeal stage details the point in which a sensory signal from the presence of the food bolus is sent to the subcortex and a signal is sent back to initiate the following swallow movements. The tongue base is retracted to meet the pharyngeal wall and that generates positive pressure to move the bolus through the pharynx (portion of the throat directly behind the mouth). The upper esophageal sphincter is a circular muscle that delineates the end of the pharynx and the beginning of the esophagus and controls the entry of the bolus from the pharynx to the esophagus.

Figure 2. Physical Structures for Swallowing

Figure 2 has been removed because of copyright restrictions. The material outlined the physical structures for swallowing. Source: From Crystal's Website from "Diagram of the Anatomy of Swallowing." From **http://www.our-sma-angels.com/crystal/web%20pages/medicalpractices/therapists/Speech/diagramswallow.htm** Retrieved November 12, 2011.

Figure 3. Hypopharyngeal Anatomy

Figure 3 has been removed because of copyright restrictions. The material contained a diagram of hypopharyngeal anatomy. Source: *MedScape Reference* from "Hypopharyngeal Cancer" by Quon, H. & Goldberg, D (2011). From **http://img.medscape.com/pi/emed/ckb/otolaryngology/834279-846565-1375268-1376463.jpg** Retrieved November 12, 2011. Copyright 1994-2011 by WebMD LLC.

Swallowing is composed of a neurologically complex set of interconnected signals and movements; thus, any deficits in the previously mentioned stages can lead to pooling of food residue in the valleculae and pyriform sinuses, which can lead to aspiration (Miller, Noble, Jones, & Burn, 2006). Figure 3 is a posterior view (looking from behind) of the larynx and it shows the location of the pyriform sinus. The pear-shaped groove of the pyriform sinus, in addition to the valleculae (not shown in Figure 3), is susceptible to the pooling of food residue if there is inadequate swallow strength to clear it. Robbins, Logemann, and Kirshner (1986) found that many PPD are "silent aspirators," so they are aspirating food and liquids, but there are no overt signs or symptoms that would indicate that it is occurring. As a result, they exhibit a decreased voluntary cough reflex. Ebihara et al. (2003) found that in the early stages of the disease there were impairments in the motor component of the cough and in later stages of the disease there were impaired motor and sensory components of coughing function. This most likely contributed to the development of aspiration pneumonia (Ebihara et al., 2003). The voluntary cough is an essential component in airway protection and clearance. Pitts, Bolser, Rosenbek, Troche, and Sapienza (2008) argue that PPD have a decreased ability to properly clear material from their airways and that an ineffective voluntary cough contributes to penetration and aspiration of food materials and liquids. The swallowing function in PPD also was found to have an impact on their quality of life (Plowman-Prine et al., 2009). Significant effects were found in the domain of

swallowing-specific quality of life and general health-related quality of life. PPD frequently exhibited depression. In fact, the areas of social function and mental health showed the greatest impact of swallowing disorders (Plowman-Prine et al., 2009). While PD is ultimately a degenerative disorder, it is important to note that the maintenance of swallowing function is important to a patient's realization of his/her full potential for functioning and independence (Sapir et al., 2008).

Facial Expression

Masked facies is defined as the masked or expressionless, unblinking facial expression that is a characteristic trademark of PD (Duffy, 2005). Smith, Smith, and Ellgring (1996) found that PPD showed less facial reactivity when viewing emotional video clips. As well, the results of Simons, Pasqualini, Reddy, and Wood (2004) provide further evidence that PPD showed reduced spontaneous facial expression and that they had more trouble with posed facial expressions. That is, PPD had difficulty generating facial expressions in response to prompts to elicit emotion prototypes (e.g., "Show an angry face.") and imitating non-emotional facial movements (e.g., smile and pucker). Bowers et al. (2006) argue that diminished facial expression in PPD is due to bradykinesia (slowness of motor initiation) of the facial muscles that normally require rapid and brief changes and found that their PPD had a decreased range of facial expression mobility. These results show that facial

emotion is a behaviour that PPD continually struggle to express, and that it is an acquired non-verbal communication pragmatic deficit. As previously mentioned, Justice (2010) defined pragmatics as the rules governing language for social purposes and the term "acquired pragmatic deficit", as used in this thesis, describes how PPD are unable to utilize non-verbal communication means (e.g., facial expression) to appropriately engage in social interactions.

Voice and Speech

Sapir et al. (2008) indicate that 80%-90% of PPD develop a voice and speech disorder throughout the course of the disease progression (Ho, Iansek, Marigliani, Bradshaw, & Gates, 1998). PPD often note that their voices are quiet or weak and their speech rate is too fast, resulting in blurred speech (Duffy, 2005). Trademark speech and voice characteristics include reduced vocal intensity, reduced frequency variability, and short phonation time (suggesting insufficient breath quantity to sustain speech phrases in everyday communication). The normal range of phonation time for adults aged 49 to 87 is 18.6±0.1 seconds (Dagli, Mahieu, & Festen, 1997). Other typical speech and voice characteristics include imprecise articulation, variable speech rate, and a monotone or breathy or hoarse voice. Collectively these characteristics are typical of hypokinetic dysarthria (Sapir et al., 2008). Hypokinetic dysarthria is the trademark dysarthria most commonly found in PPD (Duffy, 2005). Blanchet and Snyder (2009) state that the features of

hypokinetic dysarthria are related to the underlying etiology of PD. The reduced range of muscular motion could easily give rise to monopitch, monoloudness, reduced stress, variable rate, rushed speech segments, and imprecise articulation.

PPD with voices that exhibit reduced frequency variability and reduced absolute intensity experience difficulty conveying emotions during their speech (Wan, Rüber, Hohmann, & Schlaug, 2010). For example, a person who is unable to vary his loudness when he is telling a story is not able to utilize the dramatic effect of different loudness levels.

Dysfluencies can also occur due to muscular stiffness and a reduced ability to initiate movement, but PPD often attribute their dysfluencies to nervousness or anxiety (Duffy, 2005). Dysfluencies can come in part word (e.g., p-p-put) and whole word (e.g., that-that-that one) repetitions, sound prolongations (e.g., that's mmmine), as well as blocks (moments where there is additional laryngeal tension and hard glottal attack on vowels as in "th----at" for "that") Goberman, Blomgren, and Metzger's (2008) results showed that PPD had more frequent and severe dysfluencies than did age-matched or older control speakers and that the PPD dysfluencies resembled developmental dysfluencies. Depending on the severity of the dysfluencies, there may be an effect on intelligibility or even social perceptions regarding the competence and intelligence of the speaker.

Velopharyngeal port dysfunction has an effect on vocal resonance (Wan et al., 2010) in PPD, which also may affect intelligibility. Resonance refers to the vibration of air molecules within the oral and sinus cavities that shape the sound source from the vocal folds into the sound that is heard as voice. Velopharyngeal port dysfunction occurs when the movement of the velopharynx (the soft palate and surrounding tissues) is not coordinated adequately. The soft palate is a valve that controls the flow of air through the nose or the mouth. For high-pressure consonants, such as “p,” “s,” and “sh”, the soft palate needs to be raised so that air is directed out of the mouth and not the nose. Conversely, for nasal sounds like “m,” “n”, “ng”, the soft palate needs to be lowered so that air is directed out of the nose. If the soft palate is not able to move in a coordinated fashion, speech intelligibility decreases. For example, velo-pharyngeal inadequacy could result in weak pressure consonants due to nasal air emission during words or phrases requiring high-pressure sounds (e.g., “eat a peach”).

Streifler and Hofman (1984) state that some of these voice problems are significant enough to have an impact upon daily communication and the quality of life. Jaywant and Pell (2010) investigated listener impressions of PPD, and found that PPD were perceived as less interested, less involved, less friendly and less happy than healthy speaker controls. Their findings further emphasize the negative social perception that listeners can form based upon

the voice quality of PPD, which may limit those PPD's opportunities for social interaction and communication.

Despite the widespread prevalence of voice disorders in PPD, only 3-20% seek speech therapy treatment (Clarke, Zobkiw, & Gullaksen, 1995; Mutch, Strudwick, Roy, & Downie, 1986; Oxtoby, 1982; Yarrow, 1999). Currently, there are no studies that explain why PPD do not seek voice treatment. This author hypothesizes that it could be due to PPD's lack of voice awareness in addition to the degenerative nature of PD, which can spawn feelings of apathy due to the gradual worsening of symptoms. There could also be a lower priority given to voice therapy options for PPD compared to other symptoms like dysphagia. Other contributing factors to the low uptake of treatment services might be related to reports that pharmacological interventions and traditional speech therapy techniques have not been successful in treating voice abnormalities in PPD (Weiner & Singer, 1989). Thus, it is important to develop strategies to maintain and manage PD voice symptoms using voice rehabilitative strategies based on evidence that they have a positive effect.

Lee Silverman Voice Treatment

One method of voice treatment that has been supported by evidence to be highly effective is the Lee Silverman Voice Treatment (LSVT), which is an intensive voice program for PPD. LSVT was developed upon the principles of

motor learning, skills acquisition and muscle training. The theme of the program is a focus on increasing the production of vocal loudness and on developing self-monitoring of vocal loudness in PPD. The foundation of LSVT comprises five main ideas:

1. Focusing on increasing vocal loudness
2. Improving sensory perception, self-monitoring, and the self-regulation of vocal effort and loudness
3. Employing a high-effort treatment style
4. Using an intensive treatment style
5. Quantifying and measuring speech and voice outcomes

The LSVT approach also has accompanying daily tasks and a hierarchy of practice items that comprise the homework portion of the program. The daily tasks include the maximum phonation task (the maximum length a sustained "ah" can be held on one breath), pitch glides from modal voice (most frequent voice register used for singing and speech) to both the highest and the lowest pitches to establish voice frequency range, and the reading aloud of ten functional phrases chosen by the client and clinician to practice speech skills in a structured format. Homework and carryover assignments for generalization of skills are required. The focus during all of these activities is on multiple repetition and high effort. The structure of LSVT treatment sessions occurs mainly in private therapy contexts between a client and clinician and is not designed around a group therapy context.

Therapy occurs four times per week and these sessions are normally 50-60 minutes in length. The homework portion of the LSVT program can incorporate the use of other individuals during some carry-over activities, but the daily activities and hierarchy items were designed for independent practice. Hierarchies are structured from least to most complex tasks or from most to least amount of clinician support (e.g., providing cues→no cues). Thus a client can be given tasks that increase in complexity, as in producing speech at incrementally louder levels. Figure 4 illustrates a possible hierarchy for the construction of speech-practice tasks:

single words→phrases→sentences→paragraph→conversation

Figure 4. Speech Tasks Hierarchy Example

In effect, LSVT is a highly intensive vocal loudness program that drives the phonatory and respiratory system to their limits in addition to altering self-perceptions of vocal loudness (Sapir et al., 2008). LSVT has been shown to improve disordered voices, speech, swallowing, tongue mobility and facial expression (Sapir et al., 2008). In a study by Ramig, Countryman, O'Brien, Hoehn and Thompson (1996) it was shown that speech abnormalities in PPD decreased following LSVT when compared to a placebo group. Additionally, LSVT was found to produce perceptible changes in voice quality (Baumgartner, Sapir, & Ramig, 2001). They obtained the perceptual ratings of two expert listeners who were blinded to whether speakers were healthy

or had PD. The listeners were presented with pre- and post- treatment audio speech samples of PPD and individuals in a control group reading the "Rainbow Passage." (The Rainbow Passage is a standardized passage commonly used in Speech-Language Pathology because all the sounds in English are represented within the passage. Thus, by using the standardized passage, clinicians can ensure that all the English sounds are produced and available for later analysis.) It was determined that there was a statistically significant difference between the LSVT and the control group in perceived hoarseness and breathiness of the voice samples. The long-term efficacy data of Ramig et al. (1996) and Sapir et al. (2002) support the use of LSVT treatment because its effects have been seen to persist 12 and 24 months after treatment.

A pilot study by Sharkawi et al. (2002) showed that following LSVT, eight PPD had improved neuromuscular control of their upper aerodigestive tracts and better oral preparation and tongue-based retraction during the oral and pharyngeal phases of swallowing in addition to improved vocal intensity. Overall, they reported that there was a 51% decrease in the number of swallowing motility disorders. An example of a swallowing motility disorder is achalasia. Achalasia involves impaired esophageal peristalsis, the smooth muscle contractions that move food and liquid through the esophagus. Symptoms involve the regurgitation of undigested food and progressive dysphagia.

Indeed, there is an interrelatedness of physiology connecting swallowing, coughing and phonation. All require increased glottal resistance and sub-glottal pressure. Studies of the effects of LSVT that collected evidence in each of these domains emphasize the logical correlation among them with respect to increased ability to forcefully adduct the vocal folds and increase subglottal pressure.

While Fox et al. (2006) noted the distributed LSVT effects of improved articulation, facial expression, and swallowing, there have been no comprehensive studies detailing the extent of the LSVT effects on facial expression.

Although the LSVT protocol has been shown to be a popular treatment option for PPD, with high efficacy and long-term effects, alternate therapy options also appear in the literature. These alternative methods have been designed to realize the same voice benefits as LSVT but, at the same time, to emphasize engagement in social interaction and enhanced opportunities for the generalization of learned vocal skills.

Voice Choral Singing Therapy

It is recorded in the literature that over twenty years ago investigators began to consider the use of aspects of music to treat PPD. The title “Voice Choral

Singing Therapy" generally describes any form of voice therapy that incorporates elements of individual and group singing activities. In 1988, Crozier and Hamill implemented rhythm and repetition treatment with PPD in order to improve speech production and rate. In the same year, Selman showed that a patient with PD and dysarthria benefited from singing therapy. Muscle rigidity of the jaw, tongue, and respiratory mechanism were reduced. Pacchetti, Aglieri, Mancini, Martignoni, and Nappi (1998) and Pacchetti et al. (2000) investigated the effects of choral singing, voice exercise, rhythmic and free-body movements on quality of life and the completion of daily living tasks by PPD, but no measures of vocal quality were taken. Haneishi (2001) implemented a music therapy voice protocol for a group of four female PPD that consisted of conversation (three minutes on current events), warm-up (e.g., facial muscle massage), vocal warm-ups (e.g., singing from lowest to highest comfortable pitch on syllables "mah" and "pah"), singing exercises (e.g., two or three songs selected by the individual, varying in phrase length and complexity) and a focus on phonation and breathing. He found that following 12-14 sessions, PPD were shown to have a significant increase in vocal intensity and speech intelligibility (as rated by caregivers). Learning how to stagger and sustain breath through a sung musical phrase helped to develop respiratory capabilities (Haneishi, 2001). None of the studies reviewed in this paragraph collected acoustic measures of voice change.

Studies of VCST have also been conducted locally. These studies *have* collected voice acoustic data. In February-April 2003, Merrill Tanner-Semple, an Edmonton Speech-Language Pathologist, conducted a pre-pilot study researching the potential effects of singing and voice therapy (in an out-patient rehabilitation setting) on PPD. She had two treatment subjects (female PPD) and five control subjects (three male and two female PPD). The only results recorded from this study were by Nicks (2004). Tanner-Semple measured eight dependent variables: 1) speech intelligibility using the Speech Intelligibility Test (SIT) (Yorkston & Beukelman, 1981); 2) intensity range (dB) – the range between the loudest and softest possible level at which a person can speak (e.g., shouting to almost whispering while still maintaining voice); 3) habitual intensity (dB) - the average intensity level of a speaker during a regular speech task.; 4) frequency range (Hz) – the range between the lowest and highest frequencies a person can produce on a held vowel; 5) habitual frequency (Hz) - the average frequency of a person's speaking voice in a regular speech task.; 6) phonation duration (sec) - the amount of time a person is able to phonate a sustained "ah" vowel; 7) s/z ratios - the time a participant can hold an "s" divided by the time he can hold a "z" consonant. The norm is a ratio value around 1 and values above 1.40 may be indicative of a vocal pathology. It is thought that people having no problems with their vocal folds will sustain "s" and "z" for the same amount of time. However, if there is vocal pathology present, it is suspected that a " z" voiced consonant will not be held as long as "s" which is a voiceless

consonant (Eckel & Boone, 1981). The explanation is that a vocal pathology would affect the vibratory pattern in "z", a voiced consonant; and 8) frequency perturbation (% F_0 variability in Hz) – a measure of the cycle-to-cycle variability in frequency during a sustained "ah" using the Visipitch II software and hardware from Kay Elemetrics (Kay Elemetrics Corporation, 1996). In a healthy individual there should be minimal perturbation because there is no vocal pathology affecting the movement of the vocal folds. The singing/voice therapy did not produce statistically significant changes in any of the eight variables measured; however, there was an anecdotal report by one of the two participants that her mood improved as a result of therapy.

Following that pre-pilot study, in June-July 2003, Tanner, Wiens, and Campbell (2003) then conducted a study that once again examined the benefits of a short period of intensive singing lessons on the vocal quality of PPD, but this time with more participants. They met in a campus-based-house setting two to three times per week for a treatment period of four weeks. They kept six of the eight measures done by Tanner-Semple earlier that year, but removed s/z ratio and frequency perturbation and replaced them with a self-assessment questionnaire and qualitative data gathered from individual interviews. This time it was found that frequency range increased. Self-assessments by the PPD revealed that they felt that they were more easily understood following treatment. Qualitative comments showed that there was reduced stuttering, improved coughing and swallowing, and

enhanced facial expression; however, once again their findings did not show the measured intelligibility and loudness benefits demonstrated in the LSVT method. One year later, in the fall of 2004, Wiens conducted singing training for a group of five PPD. It built upon the work of Tanner et al. (2003). Wiens' work will be discussed in detail in the next section of this *Introduction*. (It should also be noted that a PhD study by Tanner, now underway, is investigating speaking voice quality, as measured by acoustic parameters, before and after VCST.)

Most recently, studies of VCST have been conducted in Italy. A pilot study implemented by Di Benedetto et al. (2009) provided a VCST program to PPD. In total, PPD received 20 hours of speech therapy (two one-hour sessions per week for ten weeks) and 26 hours of choral singing (one two-hour session per week for 13 weeks). Speech therapy activities included muscular relaxation exercises and respiratory, laryngeal, facial, and prosodic exercises (e.g., speaking while invoking particular emotions such as anger or happiness). During the choral sessions they used simplified chants to work on music rhythm and proprioceptive cueing (i.e., changing the tension in facial structures). Di Benedetto et al. (2009) found that there was a significant decrease of functional residual capacity (volume of air left in the lungs after a normal exhalation) and a significant *increase* in maximum inspiratory pressure and maximum expiratory pressure. The investigators hypothesized that training in using a better diaphragm position might have

resulted in the lowered functional residual capacity following phonation. They further hypothesized that VCST respiratory muscle training might have resulted in increased maximum inspiratory and expiratory pressure values. Perception by listeners of the PPD's prosody while reading a passage was measured using a 10-cm visual analog scale with no subdivisions, where the left end represented the worst prosody patterns and the right side represented the best patterns of prosody. It was discovered that prosody improved during the passage reading but not the monologue. The researchers believed that the monologue was less reproducible than the passage reading and was more influenced by external and internal factors such as interest and emotion, which would make it a source of less reliable measures.. Listeners also reported less perceptual detection of vocal fatigue. Di Benedetto et al. (2009) proposed that VCST was not only effective, but was an engaging and stimulating alternative to traditional voice therapy for PPD. Firstly, VCST was administered over a long period of time with great compliance because participants enjoyed working on their voices within a musical context. They also found that most of their participants were eager to continue with choral singing following the therapy, which shows lasting interest in the activity. The authors also emphasized that VCST could potentially be a lower-cost model because it gives more treatment to more patients; the treatment comprised 20 speech therapy sessions (1 hour per session) and 13 choral singing sessions (two hours per session) for a group of PPD compared to 16 individual LSVT sessions (50-60 minutes per

session). Di Benedetto et al. (2009) further suggested that there were quality of life improvements as a result of therapy. This quality of life hypothesis agrees with that of Hallam (2010), who studied the benefits of singing for children. Hallam noted that singing within a choir setting resulted in physical relaxation and release of tension; improved physical well-being; emotional release and reduction in stress levels, improved emotional well-being, positive mood and enjoyment; increased energy levels; improved cognitive processes of attention, concentration, memory and learning; and increased self-confidence and self-esteem.

This literature review revealed that while there has been increased interest in VCST as a rehabilitative voice technique for PPD, there are still limited numbers of studies that detail its outcomes and effectiveness, especially with respect to acoustic and perceptual measures. Although only small sample sizes have been used in the studies reviewed, there are positive indications that VCST can ameliorate voice abnormalities within PPD. Future testing will be required to determine dose effects and a randomized control trial will be needed to determine efficacy (Wan et al., 2010). Russell, Ciucci, Connor, and Schallert (2010) noted that VCST has been shown to result in an increase in respiratory volume and pressure along with perceptions of decreased vocal fatigue, but they criticized the fact that there have been no tests of generalization of these skills to vocal quality and no tests of long-term effects of this treatment. The general consensus of the literature is that VCST

appears to have an effect, but it is unclear as to the degree of the effect or whether differences before and after treatment are perceptible to listeners.

Details of the Project Providing the Data for the Current Study

In the fall of 2004, Professor Harold Wiens, a voice professor at the University of Alberta Faculty of Music, sought to investigate whether singing lessons alone (without speech therapy) were beneficial for PPD. He recruited five idiopathic (unknown cause) PPD (three males, two females); they had no concomitant health conditions and did not have moderate or severe hearing loss. They all gave their written formal consent to participate in the study. The Arts Ethics Review Board approved the study on September 25, 2004.

The video and audio recording for Wiens' (2004) study began on October 7 and 14, 2004. (Two sessions were scheduled to collect one set of pre-treatment measures per participant.) The first treatment session began on October 18, 2004 and the last ended on December 2, 2004. Treatment occurred four times per week for one month. There were three individual sessions per week for 0.5h and one group session for one hour per week. In total there were 13 individual voice sessions and four group choral treatment sessions. All sessions were held in Corbett Hall at the University of Alberta.

The group choral-treatment sessions were led by Wiens and a music-student assistant; sessions included physical (e.g., arm and shoulder stretches) and vocal warm-ups (e.g., vocalizing on scales) and group singing. Emphasis was placed on breathing and vocalization exercises using varying pitches (e.g., yawn-sigh technique where participants breathed in and then yawned using a vocalized and sustained "ah") and the singing of a standard "Silent Night" passage.

Tanner-Semple (2003) proposed that voice training could be an effective therapeutic tool for individuals with PD. Later in 2003, for the Tanner, Wiens and Campbell study, Wiens followed up on this idea by conducting individual singing treatment sessions using vocal training exercises that he developed throughout his 32 years as a singing teacher:

> Wiens used a holistic form of therapy that engaged participants at the physical, mental, emotional, and spiritual levels. Further, the procedures drew upon exercises from Alexander and Feldenkrais, designed to develop, align, and engage the entire body. Before that 2003 study, Wiens had not worked with PPD. He did not have a clear idea what effect, if any, these exercises would have upon any of the symptoms of the disease. Each session was built upon the previous, and the exercises were very experimental in nature, not following any particular protocol. Nevertheless, procedures began to form and there was an initial effort to codify these procedures, but no record remains. Further

> research will be necessary to refine and describe a protocol that another singing teacher or therapist could accurately replicate. Such an attempt will require a substantial commitment of time and energy. (Wiens, personal communication)

The Wiens (2004) singing treatment evolved from the thinking and experimentation of 2003. The voice studio became a setting for a vocal workout of sorts, involving much movement. The sessions focused on the following areas:

- Incorporating physical action and gesture corresponding to/supporting expression and emotion
- Increasing vocal intensity
- Increasing pitch range
- Increasing respiratory support (using the diaphragm to support vocalization instead of forcing and pushing sounds from the vocal folds)
- Increasing movement of the jaw and tongue; humming; trilling the lips
- Improving posture
- Improving pitch-matching ability

All voice sessions were audio- and video-recorded in a soundproof booth. A music-student assistant conducted physical and vocal warm-up exercises prior to the individual voice lessons given by Professor Wiens. It should also

be noted that all PPD participants were highly motivated individuals, which is a contributor to positive therapeutic change.

Quantitative and qualitative measures were collected in the Speech and Hearing Sciences Teaching laboratory at Corbett Hall before, during, and after the treatment sessions for a total of 10 weeks. Speech intelligibility stimuli were audio and video-recorded by the music-student assistant. The participants read ten SIT sentences following every voice treatment session. A semi-structured interview portion was also conducted; the following questions were asked after each individual voice treatment session:

a. How would you describe your ability to swallow? Have you noticed any changes in your ability to swallow?

b. How often do you need to cough? How would you describe your cough? Have you noticed any changes in the quality (strength) or frequency of your cough?

c. How would you describe your facial expression? Have you or have others noticed any changes in your facial expression?

d. How would you describe your speech or articulation? Have you or have others noticed any changes in your speech or articulation?

Due to the fact that the interviews were semi-structured, the previously mentioned questions guided the interview, but other topics during the interview were allowed. Furthermore, the interviewer frequently asked

follow-up, closed-ended questions after the initial open-ended questions described above. The qualitative interviews were audio- and video-recorded on VHS tape and subsequently transcribed, word-for-word, with no interpretation; however, participant comments were sorted into topic areas (e.g., swallowing) by the transcriber.

In summary, the changes introduced between Tanner, Wiens, and Campbell's study in 2003 and Wiens' 2004 study were a change in dose of the treatment as Tanner, Wiens, Campbell's treatment group met two to three times per week for four weeks for a total of 8 to 12 sessions whereas Wiens' treatment group met four times each week for a total of 17 sessions. In addition, four of the 17 sessions were group choral work. The earlier study did not feature group treatment. Furthermore, Wiens conducted a semi-structured interview following every individual voice treatment session (interviews did not follow the weekly choral session) for a total of 17 interviews; interviews were not a feature of the earlier 2003 study.

Thesis Statement

This thesis was inspired by the Edmonton-based studies reviewed in the previous section. It continued the work of Wiens (2004) by using unanalyzed data from his study to answer Wiens' original question of whether singing lessons are beneficial to PPD. This thesis brought together the data collected from various sections of the 2004 project. This thesis united three data sources: singing, speech and interviews. The singing and speech samples underwent descriptive acoustic analysis. Measures of number and duration of breath groups, sentence duration, average intensity, intensity range, and voice quality at the end of phrases were taken from recordings of participants singing *Silent Night* and reading *SIT* sentence samples. The speech samples were also used to obtain auditory-perceptual judgments. This study obtained data from the perceptual impressions of listeners of the vocal quality of PPD who had received VCST to determine if there were auditory-perceptual differences in PPD voices before and after treatment. The auditory-perceptual portion was a "better or worse" discrimination task for pre- and post VCST speech samples. Thirdly, an examination of semi-structured interview transcript statements was conducted by judges naïve to the purpose of the study and to whether the statements were obtained before or after therapy. Statements were drawn from four themes: swallowing, coughing, facial expression and speech/articulation.

Materials and Methods

Research Question

This study answered the following research question:

What are the changes observed in PPD from Pre- to Post-Voice Choral Singing Therapy? ("Therapy" being the Independent Variable) The following Dependent Variables served to answer the question:

- Nine dependent variables from "Silent Night" singing samples: the average number of breath groups used during the song; the average duration of breath groups; the mean intensity, maximum intensity and intensity range averaged over breath groups (interrelated variables because changes in one measure will be seen in another); and the measurements of word duration, mean intensity, average maximum intensity, and presence or absence of pitch breaks at two specific pre-pausal phrase endings (the ends of phrases before a breath) for the words "child" and "mild". Breath groups were defined for this study as the period from inspiration to the end of the sung or spoken word before the next inspiration.
- Eight dependent variables from the SIT speech sentences: the average duration of sentences; mean intensity, average maximum intensity and the average intensity range across all sentences, as well as the word duration, mean intensity,

average maximum intensity and presence or absence of pitch breaks in the last word of the sentence.

- Four dependent variables from the auditory-perceptual portion of the study: the naïve listeners' choice of which of a pre- or post- treatment SIT speech sample pair sounded better on first presentation; the naïve listeners' choice of which of a pre- or post- treatment SIT speech sample pair sounded better on second presentation; the self-rated confidence level of the listener regarding his choice on the first presentation; and the naïve listeners' choice of which of a pre- or post- treatment SIT speech sample pair sounded better on second presentation.

- One dependent variable gathered from the responses of unfamiliar readers rated the amount of change described in semi-structured interview data collected before, during and after treatment to reveal the PPD perspective on the personal benefits of VCST in the areas of swallowing, coughing, facial expression, and speech and articulation.

Design

This study employed one independent variable of treatment (two levels: pre and post). The multiple available dependent variables that resulted from the

Wiens 2004 study culminated in this eclectic thesis project making secondary use of three different sources of data from the Wiens study: singing samples from Silent Night, speech samples from SIT sentences, and interview data. A multiple-methods approach was used to analyze the three sources of data, including comprehensive pattern analysis. The multiple-methods-of-data analysis provided a means to examine the degree of consistency among the findings (Maxwell & Satake, 2006). The singing samples were descriptively analyzed using the *Silent Night* pre- and post-acoustic measures of the nine dependent variables previously listed. The SIT speech sentences were used for the descriptive analysis of pre- and post-acoustic measures of eight dependent variables, previously listed, as well as the auditory-perceptual rating portion of the study, which looked at four dependent variables. That perceptual-rating portion yielded two categories of nominal data; a)listeners chose which item of a pair of randomized samples sounded better and b) rated how confident they were about those choices ("high" or "low"). The interview data yielded ordinal ratings by naïve raters of their perceptions of the degree of change revealed in statements from interview transcripts. It was the aim of this thesis to unite the various data sources to provide a holistic view of the outcomes of treatment.

Participant Privacy (Ethics)

The primary researcher in this study received all de-identified data from the PPD subjects from Harold Wiens. All the digital audio files and VHS tapes were labeled using the de-identified initials provided by Wiens. All five PPD signed consent forms and the Arts Ethics Review Board approved that study on September 25, 2004.

Ethical permission was sought for the secondary use of the Wiens data and the recruitment of listeners for the auditory-perceptual portion of this study. Ethics approval for this portion of the study was granted on January 11, 2011 (study identification: Pro00016088). Listeners signed consent forms before testing began. They were informed that they had the right to withdraw from the study at any time, if they chose to do so, and that this choice had no effect on the grade or the course credit they would receive in their introductory linguistics class. The researcher removed all identifying information from the consent form and identified the participant by an arbitrary identification number (e.g., 11100, 11101, etc.). There were no known risks associated with participating in this study.

Part I. Singing Analysis

A. "Silent Night" Singing Samples

Two singing samples from each of the five participants were used for analysis: a pre-treatment session sample from October 6, 2004 and a post-

treatment session from December 3, 2004. These sessions were video recorded and transferred to DVD. Therefore, in total, five pre-treatment and five post-treatment samples, for a total of ten "Silent Night" singing samples, underwent acoustic analysis. A DVD Ripper (MP4Converter Software Studio, 2002-2007) was used to convert the DVD singing sessions into WAV audio format (sampling rate 44100 Hz) for analysis.

Improvement in subglottal pressure and forceful adduction of the larynx has been shown to improve vocal intensity as well as voice quality. For example, the LSVT study by Baumgartner et al. (2001) found that LSVT exercises that emphasized phonatory-respiratory effort resulted in improvements in both vocal intensity and voice quality. The study by Di Benedetto et al. (2009) showed that there was a significant difference in the areas of functional residual capacity, maximum inspiratory pressure, and maximum expiratory pressure as a result of VCST treatment. It is important to note, however, that measuring the duration of a maximum phonation task does not reveal where in the vital capacity or how much lung volume is being used. Functional tasks like conversational speaking are typically produced within an operating range of lung volume. In contrast, this study examined the number and duration of breath groups taken throughout "Silent Night"; mean intensity, maximum intensity and intensity range (interrelated variables) averaged across all the breath groups taken throughout "Silent Night"; and voice-quality measures specifically for the words "child" and "mild". (See Figure 5.)

Silent night, holy night
All is calm, all is bright
Round yon virgin mother and **child**
Holy infant so tender and **mild**
Sleep in heavenly peace
Sleep in heavenly peace

Figure 5. Silent Night Lyrics

The reason for using "child" and "mild" was because the words occur in the middle of the song, enabling participants' voices to be warmed up by that time. "Peace" was not chosen because of its rising pitch in the piece, which is often challenging, even for experienced singers. It was expected that there should be an improvement in vocal quality at "child" and "mild" with increased subglottal air pressure, increased laryngeal opposing force and greater mouth opening following treatment. That is, there should be a generally higher intensity and fewer pitch breaks at "child" and "mild." Voice quality was defined here as a reflection of four separate measures: word duration of "child" or "mild"; mean intensity in the words "child" or "mild"; maximum intensity reached at the words "child" and "mild", and the presence or absence of pitch breaks within "child" and "mild". Praat (Boersma & Weenink, 2010) takes intensity measurements at each time point in the word, so it was decided to take the maximum intensity peak in the word to compare values pre- and post- VCST treatment.

These data values were examined using a descriptive analysis method. The reason for looking at the acoustic values in a descriptive way was because of

the quality of the acoustic recordings. The sound recordings were converted from video samples, and while the videos were recorded in a soundproof booth, there was persistent keyboard music and non-standardized microphone-to-mouth distance while obtaining the recordings. Thus, it was decided that a descriptive analysis would be better suited to convey the preliminary results of any acoustic changes without making generalizations about the acoustic results.

Praat acoustic software (Boersma & Weenink, 2010) was used for all acoustic singing measurements. Adding text grid boundaries in Praat segmented the sound samples. All the segmentation was done by hand to maintain consistency in the measurements. A consistent protocol was followed where a vowel onset or consonant in the spectrogram (a visual graph representation of sound where time is on the x-axis and frequency is on the y-axis) would mark the start of a phrase boundary and the end of the phrase boundary would be where the spectrogram formants (dark bands in a spectrogram that correspond to vocal tract resonance) would end. To maintain consistency, the primary researcher segmented all phrase portions. Thus in the *Silent Night* sample, each of the phrases was segmented with text grid boundaries to allow for analysis by a script. Individual productions were marked up on a text grid using Praat. Subsequently, a script was used to automatically extract maximum and minimum intensity of the utterance. Utterances were extracted with a 30ms pad on either side; this was done to

prevent any effects of uncertainty, due to edge effects, in the algorithm used to track intensity. Furthermore, the time steps and window size in Praat was calculated based on the "minimum pitch" which was set uniquely to fit each participant. Measurements for breath-group duration, end-word duration, average intensity, maximum intensity, and so forth, were obtained in this way for each of the sound segments. Therefore, depending on the number of breath groups of each of the participants, there were multiple corresponding acoustic measurements for each of the phrases. These multiple acoustic values were averaged to obtain an overall measurement value for each participant.

The number of breath groups was measured for each participant. Based on the singing experience of the researcher, it was expected that there would be a range of 5-10 breaths taken throughout the *Silent Night* sample. The pre- and post- treatment average of breath group duration was calculated within and across PPD. Due to the fact that Haneishi (2001) found that active respiration exercise played a key role in participants' learning how to stagger and sustain breath through a musical phrase, we expected there to be decreased breath group numbers and increased breath group durations following VCST.

There are three physiological variables that alter the perceptible loudness levels of speakers: subglottal pressure for phonation, the degree of adduction

of the vocal folds and the degree of mouth opening. The values of mean intensity, maximum intensity and intensity range were averaged across all the *Silent Night* breath groups, within and across participants. It is important to note that the measures of mean intensity, maximum intensity, and intensity range are all interconnected since an increase in one would affect another area; however, for the purposes of this thesis, it was decided that an examination of all three areas would provide more sensitivity when measuring the parameter of intensity. An increase in all measures of intensity would be expected following VCST, a therapy that results in laryngeal muscle strengthening. As a result, there should be increased vocal fold resistance, resulting in a build-up of sub-glottal pressure, yielding an increase in vocal intensity. The duration of the closed phase of vocal fold movement and the medial compression of the vocal folds increase and offer greater resistance at the larynx to subglottal pressure. Increased sub-glottal and expiratory pressure are required to overcome these laryngeal muscle forces. It was reasoned that a larger intensity range post-treatment should be indicative of an increased ability to employ respiratory support, increased mouth opening or increased laryngeal airway resistance during singing.

Method of Analysis

The data were analyzed descriptively by obtaining the mean measures across all singing acoustic measurements and by comparing them in pre- and post-treatment tables. As well, the data were looked at descriptively using a

pattern analysis to aid the investigator to visualize clustering and common trends. Any increases following treatment were indicated with a "+," any decreases were indicated with a "-," and no changes were indicated with "0". A pattern analysis was done on each of the following areas for the *Silent Night* singing samples: a) breath group number; b) breath group duration; c) intensity average across each breath group; d) intensity range from minimum to maximum values within each breath group; and e) voice quality at "child" and "mild" (defined as word duration of "child" and "mild"; average and maximum intensity reached at those words; and presence or absence of pitch breaks at those pre-pausal words). The pattern analysis was chosen to synthesize any change patterns revealed from the singing samples.

Part II: Speech Analysis

A. Acoustic Analysis of SIT Sentence Speech Samples

The SIT sentences used in this study come from a standardized test battery of sentences. The SIT was designed to measure the speech intelligibility of dysarthric speakers. The test contains lists of sentences of different length, numbered accordingly. Samples of a portion of the SIT sentences are in Figure 6. These happen to be from the "10" series – sentences containing ten words.

10.1 Most overweight people need to learn to be more assertive.

10.2 Most weeds can now be put on the compost pile.

10.3 There are effective ways to conserve both energy and water.

10.4 They are some of the best vegetable protein foods known.

10.5 At certain times, I like being strong for someone else.

10.6 Spending time with the family is really my favorite activity.

10.7 I had all the usual tests, and everything was fine.

10.8 The park has separate areas for cars and recreational vehicles.

10.9 Agriculture did not necessarily tie men to sedentary village life.

10.10 The contract provided an immediate raise in the hourly rate.

Figure 6. SIT speech sentences

Wiens (2004) recorded ten sentences from each participant at each voice treatment session. All participants recorded the same sentences at each time period, but different lists were used at each session. Thus, over time, the sentences used had progressively moved from ten-word to eleven-word lists. Thus the pre- and post- treatment sentences were different from one another not only in terms of actual words, but number of words used. The pre-treatment sentences were ten-word sentences and the post-treatment sentences were eleven-word sentences (SIT sentences numbers 10.1-10.10 and 11.61-11.70 respectively). The twenty sentences used from the SIT, pre- and post- treatment, are listed in Appendix A.

It was decided that parallel measures would be taken between the *Silent Night* and SIT speech samples to allow for comparison between the singing

and speech samples before and after VCST. Measures were taken of: duration of every sentence; mean intensity across every sentence; intensity range across every sentence; and voice quality on the final word of each sentence (e.g., the word "assertive" in sentence 10.1). The one difference between speech and singing dependent variables was that number of breath groups per sentence was not used, because participants were not expected to take more than one breath per sentence on average. As previously stated in the singing portion of this study, it was expected that if respiratory support and glottal resistance increased following treatment, results would reveal decreased sentence duration due to increased fluency; increased intensity range across every sentence; and greater mean and maximum intensity values in every sentence. For the purposes of this study, voice quality is a collective term looking across several parameters: the sentence duration; mean intensity across the final word; maximum intensity reached at the final word; and presence or absence of perceived pitch breaks at the final word.

Method of Analysis

For this study, Wiens (2004) recorded ten pre- and ten post- treatment sentences (20 sentences for each participant), yielding a total of 100 sentences that were used for acoustic analysis. The speech acoustic analysis was performed in a similar fashion to the singing acoustic analysis in that Praat acoustic software (Boersma & Weenink, 2010) was employed for all

acoustic measurements. A digital analog converter (ADVC 110) was used to transform the VHS analog audio into digital WAV files (sampling rate 48000 Hz) using iMovie (Apple Inc.) and these audio WAV files were imported into Praat. Thus, each sound file contained all ten sentences from the recording session for that day. The individual SIT speech samples were segmented by adding text grid boundaries in Praat. As with the *Silent Night* sound samples, all the segmentation was done by hand to maintain consistency in the measurements. A consistent protocol was followed where a vowel onset or consonant in the spectrogram marked the start of the SIT sentence and the end of the sentence was designated as the point where spectrogram formants ended. To maintain consistency, the primary researcher segmented all SIT speech samples. A script was used to go through all the speech samples; acoustic values were obtained from the segmented sentences and breaks between sentences were not analyzed. As well, to increase acoustic analysis accuracy in Praat, the minimum and maximum frequency for each block of SIT speech sound samples, by speaker, was reset because each speaker differed from one another. Measurements of sentence duration, average intensity, maximum intensity, and so forth, were obtained in this way for each of the sentences. These multiple acoustic values for each of the ten sentences were averaged to obtain an overall measurement value for each participant. For the acoustic analysis portion, a 0.03 seconds sound buffer was added before analysis.

The speech acoustic data were averaged for pre- and post- treatment measures,shown in Table 5. As well, the values were added to a descriptive pattern analysis in Table 9 to aid the investigator to visualize clustering and common trends. As it was with the singing samples, any increases following treatment were indicated with a "+," any decreases were indicated with a "-," and no change was indicated with "0." All speech acoustic data were grouped in a comprehensive and contrastive pattern analysis table to compare the results with the ones obtained in the singing samples. The pattern analysis patterns of change across singing and speech samples. Just as with the singing samples, the speech acoustic data were analyzed descriptively because of the quality of the acoustic recordings. They were originally recorded on a video recorder using its internal microphone and they were not recorded in a soundproof booth. There was intermittent background noise during the recordings and a non-standardized video microphone-to-mouth distance was employed. The sound recordings were converted from VHS video samples to digital audio files. Thus it was again decided that a descriptive analysis would be better suited to document any acoustic changes given the reduced quality of the recordings.

B. Auditory-Perceptual Analysis of SIT Sentence Speech Samples

Baumgartner et al. (2001) found that hoarseness and breathiness were reduced following LSVT as judged by two SLP listeners from pre- and post-treatment recordings of the "Rainbow Passage," but there has yet to be any

auditory-perceptual testing done on voice samples of PPD following a VCST treatment. Although voice quality changes can be measured acoustically, the amount of auditory-perceptual change detectible to an unfamiliar listener has more functional validity for a PPD who has received treatment. The overall impression of voice quality was judged using a "better or worse" discrimination task on a pair of sentences, one recorded pre-treatment and the other recorded post-treatment. The order of pre- and post- treatment sentence sample presentation was randomized; listeners were blind to the fact that they were comparing sentences recorded at two different points in time. They simply selected the one that sounded best to them. The listeners then rated how confident they were about each choice. This acted as a means to deduce whether raters were guessing about which sentence sounded better or making a choice with a clear idea of a difference between the sentences.

SIT Speech Samples for Auditory-Perceptual Testing

The same SIT speech samples used in the acoustic analysis, which was described in the previous section, were used for the auditory-perceptual testing portion; however, the SIT sentences used in the auditory-perceptual testing portion did not have a 0.03 second standardized sound buffer (which was added by the script) because the segmented sentences were directly extracted from Praat and converted to digital WAV format to use in testing. Wiens (2004) recorded ten pre- and post- treatment sentences (20 sentences

for each participant); this yielded a total of 100 sentences that were used for the listening task.

The presentation of the samples were blocked by speaker, so that listeners heard paired sentences of only one PPD at a time, but the sentences by that one speaker were randomized (within speaker) with respect to both whether they a) contained ten or eleven words, and b) were recorded pre-treatment or post-treatment. The pre-treatment samples came from the first treatment session on October 18, 2004 and the post-treatment samples came from the final treatment session on November 26, 2004.

Participants and Training

A pool of 33 undergraduate students, enrolled in introductory linguistics classes at the University of Alberta, were recruited and performed the auditory-perceptual ratings of voice quality. The listeners obtained in this study needed to complete research credit as a component of their course grade. In order to investigate the power needed for this experiment, a generalized literature search of a variety of auditory-perceptual experiments yielded listener numbers from 10-25. Thus it was expected that 30 listeners would allow for auditory-perceptual changes to be revealed. These raters did not have any previous training in listening to and/or rating voice quality. The raters had normal hearing, as determined by self-report, were native English speakers, and had no knowledge of the speakers' health condition,

the pre- or post- treatment order assignment of the recordings, or the experimental hypothesis.

E-Prime 2.0 software was used to design the listening task (Schneider, Eschman, & Zuccolotto, 2002). To familiarize listeners with the testing procedure, the raters were given three practice trials on sentence sample pairs in a "Which is better, A or B?" discrimination task. An initial computer information screen was presented on which they were told to rate the voice quality of the speaker of the sentences, ignoring the content of the sentences. They were also informed that sequential blocks of sentences spoken by five different speakers would be presented to them. They were trained in the testing procedure to answer the question, "Which sample sounds better, A or B?" with a two-button choice option. They were also asked to choose a confidence rating of "high" or "low." The reason for obtaining the confidence rating was to see if listeners felt like they were guessing or if they genuinely felt that one of the samples sounded better.

To test for rater reliability, each pair of samples was replayed to the listeners. Thus in total, the listeners heard 200 sentences in 100 comparison tasks (5 PPD X 10 sentence pairs (10 pre-treatment sentence pairs and 10 post-treatment sentence pairs put together for ten decision tasks) X 2 (all sentence pairs replayed a second time)). The 200 decisions coupled to 200 confidence-rating decisions required less than a half hour of listening.

Rating procedure

All perceptual testing was done in the Alberta Phonetics Laboratory within the Department of Linguistics. The raters listened in a sound-treated booth. Lists were randomized for each individual speaker, but the blocks occurred in the same order. Thus the individual items were randomized (for pre- and post-treatment order), but the order of speakers was exactly the same. Each rater listened to the compiled master tracks and s/he was asked to answer, "Which sample sounds better, A or B?" Individuals were allowed to listen to each sample once (with no replay option) in order to rate which sample was better. All 33 listeners heard a different randomized order of items; there were 33 separate randomized sequences.

Once they chose a sample, listeners then rated how confident they were with their selection. They were asked the question, "What level of confidence do you feel about your previous sample choice?" and they were provided with two choices, "high" or "low."

Method of Analysis

Quantitative nominal data were obtained during this portion of the study. A one-way repeated measures non-parametric McNemar test was used to compare the number of pre- and post- treatment samples listeners rated as better and to see if there was a significant difference between the two related

groups (pre- and post- treatment). The results from this test were added to the descriptive pattern analysis.

A second one-way repeated measures non-parametric McNemar test was used to see if there was a significant difference in the number of post-treatment samples that were chosen depending on the first or the second presentation of the sound samples. This was done in order to determine whether the listeners were more or less likely to choose the post-treatment sample depending on the first or second presentation time.

Lastly, a third McNemar test was used again to determine whether there was a significant difference in confidence levels of the rater between the first and second presentation.

Part III: Interview Data Analysis

In this study, four topic areas were analyzed: swallowing, coughing, facial expression, and speech/articulation. The following questions were asked in Wiens (2004) study.

a. How would you describe your ability to swallow? Have you noticed any changes in your ability to swallow?

b. How often do you need to cough? How would you describe your cough? Have you noticed any changes in the quality (strength) or frequency of your cough?

c. How would you describe your facial expression? Have you or have other noticed any changes in your facial expression?

d. How would you describe your speech or articulation? Have you or have others noticed any changes in your speech or articulation?

Interview Tapes

Wiens (2004) conducted a total of 13 interviews, one from each of his sessions, but only two interview transcripts were selected for analysis. It was reasoned that if there were changes from session to session, they would be minimal. Thus, the mid- and post-treatment time points were picked for maximum contrast at equally occurring intervals throughout therapy. Interview statements were taken from the mid-point in treatment on November 1, 2004 and the last session on December 3, 2004. A total of 40 transcripts were rated and analyzed across all five participant transcripts (5 participants x 2 transcripts each x 4 topic areas=40 transcripts).

Qualitative data, elicited during interview, are supposed to be acquired using open-ended questions (Maxwell & Satake, 2006). Questions in the 2004 study sometimes began open-ended, but the follow-up questions frequently resulted in the interviewer guiding the participants' answers and asking close-ended, follow-up questions that became "leading questions."

Participants were not given the opportunity to deeply reflect upon their answers. Two examples have been provided below:

Example #1

Q: And you feel you can clear your throat a little easier? (Note: leading question)

A: Yes I can. Yeah, in the morning, right now there is something there that is not quite right. What it is I can feel it, but I just can't get rid of it, uh, but typically in the morning yes, I can clear the phlegm from my throat much easier than I could, uh, two weeks ago.

Q: And coughing is much stronger? (Note: leading question)

A: Yes. Yes. Uh, I guess the stronger coughing is what clears the phlegm more easy, easily I suppose, a combination of things there I guess but, uh, all in all I, I, I can notice in that area quite a change in my body habits so this has been good for me, this particular time that we've had here in the last few weeks.

Q; When you started you mentioned that when you coughed it was raspy and dry. (Note: leading question)

A: Uhuh.

Q; That has contributed to the hoarseness? Has that improved at all?

A: Yes. Yes. It has also improved considerably and right now when I cough it's only to bring up phlegm and not two or three coughs and it's over with. It doesn't need to be… there's no pressure behind it, there's enough humph there when you cough.

Q: Is your voice clearer than after that when you finish? (Note: leading question)

A: Oh yes. Yeah.

Example #2

Q: How would you describe your ability to swallow? (Note: open-ended question)

A: Sort of hard at times—phlegm in my throat is what it feels like—I have to swallow a lot too. I always have to drink water with my meals, that's not different but I find that I need water when I'm eating my food.

Q: Just to help it go down.

A: Yes, to help.

Q: Do you have to consciously think of the action of swallowing more than you used to? (Note: leading question)

A: I think so. My medication seems to want to stop right about there (points to middle of throat).

Q: So fluids are OK?

A: Well, even water—when I take a glass full of water to drink because I have to take medication 5x per day—just to take a pill with a glass of water, it takes about four attempts to empty it because I can't go from the top to bottom without…

Q: So you have to allow more time for the action of swallowing?

A: Yes

The resulting transcript was not rich enough to support the performance of a thematic analysis, so the qualitative data were analyzed making novel use of a global rating-of-change scale.

Method of Analysis

The change scale used in this study is an 11-point scale with end-point descriptors. Typically patients complete this kind of change scale to indicate their perceptions about change following treatment. Preston and Colman (2000) stated that global rating-of-change scales need to be 7 to 11 points to balance the need for patient preference, adequate discrimination ability, and test-retest reliability. Streiner and Norman (2003) also noted that, due to end-aversion bias, respondents rarely respond on the extreme ends of scales, thus, the rationale for an 11-point scale over a 7-point scale. Kamper et al. (2009) stated that the scale should have an "unchanged" midpoint and a positive and negative deterioration from that midpoint using a numerical scale. The scale used in this study incorporated these elements, but was adapted to include the descriptors "greatly worsened" and "greatly improved" at the extreme end points (See Appendix B).

The adapted scale was put to novel use. Wiens' (2004) patients did not complete the scale. Readers/judges used the scale to rate printed statements. Those statements were the interview segments described above. A research

assistant took the 40 sentence transcripts and randomized them for speaker, treatment time, and topic area. They were then assembled into a rating sheet. Readers/judges read the statements and used the scale to indicate their perceptions of whether or not the statements suggested that change had occurred and to what degree.

Ten first- or second-year Graduate Speech-Language Pathology students in the Masters of Science Speech-Language Pathology program at the University of Alberta rated the perceived level of change on the global change scale (see Appendix B) suggested by the 40 responses to the semi-structured interview questions. They started first by reading the instructions, reading all of the statements, and re-reading the final instruction sentence, "Please focus on whether or not these statements reflect change," before going on to rate the statements. There were two sections: training and experimental, and the readers/judges followed the same procedure for both sections. Initially, there was a training section of three statements to allow the readers/judges to familiarize themselves with the rating scale. They were allowed to ask the primary researcher any questions at that time. Afterward, readers/judges were allowed to take the experimental sections to rate on their own time while following the previously outlined procedure. Each finished his/her ratings in one session. The readers/judges did not have the same background knowledge of the study as the primary investigator, but they were informed that the interview data were obtained throughout voice therapy. After all

readers/judges had been through the data, the primary researcher assembled the responses and ordinal data obtained from the scale. The medians of the mid- and post- treatment values were taken for each participant because medians would have been less affected by outliers than if the mean values were taken. The medians from each of the four areas were graphed and descriptively analyzed to visualize trends.

Results

Part I: Clinical Significance and Minimal Important Difference

In this study, the evaluation of clinical significance relates to the pre- and post- treatment values of the five PPD in the areas of breath group number, breath group duration, sentence duration, end-word duration, mean intensity, maximum intensity and intensity range for the *Silent Night* singing samples and the SIT sentences. The effect size was calculated from the distribution of acoustic values from the five PPD participants and the minimal important difference (MID) was calculated using Cohen's MID criteria. Cohen (1988) described 0.2, 0.5, and 0.8 as small, moderate and large effect sizes respectively. It is noted that these Cohen coefficients are only a guideline and should not be used too rigidly when making statements regarding clinical significance (Callahan & Reio, 2006).

Cohen (1988) defines MID as the degree to which a phenomenon is present in a population; thus the larger the effect size, the greater the degree the phenomenon under study is present. Furthermore, the magnitude of the effect size has been interpreted as an index of clinical significance (Musselman, 2007). Therefore, the larger this effect size index, the larger the difference between groups and the larger the clinical significance of the results (Callahan & Reio, 2006). The standard error of the measurement was not used in this study because information regarding the reliability of the acoustic measurements taken in this study was not available.

Though the effect size is calculated on a normally distributed population, in this study it was used to serve as an indicator of clinical significance. If a clinically significant change occurred, it could serve as a platform for further research. This type of analysis was chosen for the results because the values measured in this study came from audio recordings with poor quality. Therefore, the effect size values should serve as a better indicator of clinically significant change.

The effect size is calculated by dividing the difference between group mean scores (e.g., pre-treatment scores and post-treatment scores) by the standard deviation at baseline, by the standard deviation of the control group, or by the pooled standard deviation of the two groups (Portney & Watkins, 2000). The following formula was used to calculate the effect size.

$$\overline{ES} = \frac{\overline{X}_{G1} - \overline{X}_{G2}}{s_{pooled}}$$

Figure 7. Effect Size Equation

Where:
ES= effect size
$\mathbf{X_{G1}}$=Mean group1
$\mathbf{X_{G2}}$=Mean group 2
S_{pooled}: Pooled standard deviation (SD)

To calculate the pooled standard deviation, the following formula was used:

$$s_{pooled} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

Figure 8. Pooled Standard Deviation Equation

Where the variables were defined as:

S_1=SD group 1
S_2 =SD group 2
$n_{1=}$ Sample size for group 1
$n_{2=}$ Sample size for group 2

Minimal Important Difference (MID)

The MID has been defined as "the smallest difference in score in the domain of interest that patients perceive as important, either beneficial or harmful, and which would lead the clinician to consider a change in the patient's management" (Guyatt, Osoba, Wu, Wyrwich, & Norman, 2002). Therefore, the MID allows clinicians to formulate treatment decisions. However, in this study, we used a 0.5 moderate effect size to calculate the value of the MID, which would be a possible indication of which values may imply clinical significance. This was done by multiplying the effect size of the difference obtained between groups considered as important (moderate effect size=0.5) by the pooled baseline standard deviation between the two groups ($S_{pooled\ baseline}$) (Lemieux, Beaton, Hogg-Johnson, Bordeleau, & Goodwin, 2007). The following formula was used:

$$MID= 0.5* S_{pooled\ baseline}$$

Figure 9. Minimal Important Difference Equation

For interpretation purposes, a mean difference between groups that is higher than the MID can be considered as clinically important (Musselman, 2007).

Part II: Singing Results

Question: What are the changes observed in PPD from Pre- to Post- Voice Choral Singing Therapy, as Revealed in Singing Data?

A. "Silent Night" Singing Samples

Table 1. Mean Values for Silent Night

Subject	# Breath Groups		Breath Group Duration (sec)		Mean Intensity (dB)		Maximum Intensity (dB)		Intensity Range (dB)	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
A	8	6	5.19	6.60	70.34	72.16	81.21	84.44	33.74	37.32
B	8	8	4.71	4.50	69.50	71.28	81.44	83.90	37.56	37.42
C	8	8	4.71	6.11	66.92	76.22	76.31	85.99	31.19	38.28
D	8	8	3.94	4.23	54.15	63.91	64.97	73.89	23.44	27.16
E	8	8	4.85	5.08	64.35	72.11	75.73	84.01	29.86	34.87
All*	8	7.2	4.68	5.31	65.05	71.13	75.93	82.45	31.16	35.01

*Mean of all subject values

It can be seen in Table 1 that after therapy one participant took fewer breaths to sing *Silent Night*. All but one had longer breath group durations and all had greater mean intensity across the whole song after VCST. They all achieved greater maximum intensity in the song after therapy. The range of intensity between soft and loud points in the song increased following therapy for four participants.

Table 2. Silent Night Effect Size

	MID*	Effect Size
Number of Breath Groups	0.38	0.93
Breath Group Duration	0.40	0.72
Mean Intensity	2.80	0.98
Maximum Intensity	2.92	1.01
Intensity Range	2.46	0.71

*MID= 0.5 x $S_{\text{pooled baseline}}$

Table 2 reports the effect size values and shows number of breath groups, breath group duration, and intensity values across all subjects. When interpreting the MID results, the values are compared to Cohen's (1988) guidelines, which describe 0.2, 0.5, and 0.8 as small, moderate and large effect sizes respectively. In accordance with these guidelines, the MID values which have a large effect are mean intensity, maximum intensity, and intensity range while number of breath groups and breath group duration have a smaller effect size. The actual effect sizes are reported in the third column; they should be compared to the value of 1 which indicates a large effect size. Maximum intensity shows the largest effect size, which means that it showed the greatest amount of likely change. The areas of breath group duration and intensity range show moderate effect size values of 0.72 and 0.71 respectively.

Table 3. Silent Night "Child" and "Mild" Mean Acoustic Values

Subject		Word Duration (sec)		Mean Intensity (dB)		Maximum Intensity (dB)		Pitch Breaks	
		Pre	Post	Pre	Post	Pre	Post	Pre	Post
A	Child	1.27	1.14	64.01	67.61	73.29	75.01	absent	absent
	Mild	1.16	1.04	64.33	68.81	69.31	73.65	absent	absent
B	Child	1.35	1.11	61.65	68.35	70.90	75.87	absent	absent
	Mild	0.89	1.16	60.02	66.22	67.14	72.73	absent	absent
C	Child	1.25	1.02	63.02	72.22	69.59	76.19	absent	absent
	Mild	1.04	1.09	62.48	72.08	69.09	77.36	absent	absent
D	Child	1.10	1.06	53.74	58.69	60.55	62.70	absent	absent
	Mild	---	1.03	---	59.88	---	66.15	absent	absent
E	Child	0.80	1.10	64.06	62.18	71.13	72.17	absent	absent
	Mild	0.95	1.22	62.18	71.20	67.41	79.74	absent	absent
All*		1.16	1.09	61.30	65.81	69.09	72.39	absent	absent
All**		1.01	1.19	62.19	67.64	68.24	73.93	absent	absent

*Mean of all subject values for “child”
**Mean of all subject values for “mild”

Table 3 shows the word duration, average intensity, maximum intensity, and presence or absence of pitch breaks at the ends of the phrase words “child” and “mild.” Measuring these parameters were expected to echo the measures taken for the entire *Silent Night* sample, but allow a more focused view of the two words, which would most likely show change following therapy. The word duration would show if PPD were able to hold the word longer after therapy and if the mean intensity and maximum intensity increased following therapy. Overall, the mean intensity and maximum intensity were greater following therapy on the words “child” and “mild.” The duration the word “child” was held decreased following therapy but increased on the word “mild” following therapy. There were no pitch breaks present in either

the “child” or “mild” pre-pausal phrase segments. Subject four had no pre-treatment value for “mild” as the word was not sung.

Table 4. Silent Night "Child" and "Mild” Effect Size

	MID*		Effect Size	
	Child	Mild	Child	Mild
Word Duration	0.08	0.05	0.40	0.85
Mean Intensity	2.44	1.84	0.83	1.28
Maximum Intensity	2.66	1.87	0.56	1.32

*MID= 0.5 x $S_{\text{pooled baseline}}$

The effect sizes were investigated for the words “child” and “mild” because it was hypothesized that these would be the words within the song that would most likely reveal changes following therapy. These words occur at a midpoint within the *Silent Night* song, and breath support would need to be engaged since they are at the end of the phrase. Table 4 states the MID and effect size values for the “child” and “mild” pre-pausal segments. Cohen’s (1988) guidelines were used again to describe small, moderate and large effect sizes (0.2, 0.5, and 0.8 respectively). It was seen that word duration was an area of small minimal important difference, but mean intensity and maximum intensity were areas of large minimal important difference since their effect sizes were over 0.8. The areas of largest effect size (numbers closest to 1) were mean intensity for “child” and word duration, mean intensity, and maximum intensity for “mild.” Therefore, it is seen that average intensity most likely changed for the “child” and “mild” words. It should also be noted that a data point was missing for one subject for “mild.”

It did not appear that word duration for "child" was indicative of clinical change since the effect size was small.

Part III: Speech Results

Question: What are the changes observed in PPD Pre- and Post- Voice Choral Singing Therapy, as Revealed in Speech Data?

A. Acoustic Results of SIT Sentences Speech Samples

Table 5. Mean Values for SIT Sentences

Subject	Sentence Duration (sec)		Mean Intensity (dB)		Maximum Intensity (dB)		Intensity Range (dB)	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
A	4.09	3.06	64.27	65.75	76.70	77.21	28.55	26.60
B	3.56	3.07	60.44	62.54	76.24	77.3	31.20	31.56
C	3.34	2.84	64.93	67.49	76.07	77.01	26.85	26.16
D	3.87	4.68	56.37	60.96	67.89	75.00	24.31	32.86
E	3.74	3.90	59.58	62.69	69.56	75.04	21.36	28.26
All*	3.72	3.51	61.11	63.89	73.29	76.31	26.45	29.09

*Mean of all subject values

The number of breath groups was not measured for the SIT sentences because each sentence could be said with one breath. PPD were able to produce sentences in a shorter amount of time following treatment; sentence duration decreased. The average intensity increased following therapy; participants were louder overall across the sentences. The maximum intensity also increased after therapy. The intensity range was greater, too, following therapy; PPD were able to access higher and lower loudness levels.

The reader is reminded that these three intensity measures are interrelated; it is expected that changes in one will realize changes in the others.

Table 6. SIT Sentences Effect Size

	MID*	Effect Size
Sentence Duration	0.29	0.33
Mean Intensity	1.56	0.8
Maximum Intensity	1.54	0.88
Intensity Range	1.71	0.70

*MID= 0.5 x $S_{\text{pooled baseline}}$

In the second column of Table 6 MID values are displayed. Using Cohen's (1988) MID coefficients of a small, moderate, and large effect size 0.2, 0.5, and 0.8, it is seen that mean intensity, maximum intensity, and intensity range values (all interrelated measures) were all over 0.8, indicating a large effect. The sentence duration MID indicated a small effect size. The actual effect sizes are displayed in the third column and should be compared to the value of 1 to indicate a large effect size. Sentence duration showed the smallest effect size, so there was likely no clinical change. The intensity range showed a moderate effect size, so it's possible this was an area of some change. The effect sizes for the interrelated intensity values in the SIT sentences were calculated from the mean values of all subjects. Mean intensity had the largest effect sizes, most indicative of clinical change. It is interesting to note that increased intensity would be expected to use breath supply more quickly, yet sentence duration appeared *not* to be greatly shortened.

Table 7. SIT Sentence Final Word Average of Voice Quality Acoustics

Subject	Word Duration (sec)		Mean Intensity (dB)		Maximum Intensity (dB)		Pitch Breaks	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
A	0.56	0.42	63.01	63.44	71.24	70.82	absent	absent
B	0.51	0.50	53.37	56.23	60.99	64.63	absent	absent
C	0.50	0.40	60.61	64.79	68.69	71.73	absent	absent
D	0.54	0.48	54.59	61.34	60.84	67.47	absent	absent
E	0.49	0.46	55.82	60.25	60.93	66.69	absent	absent
All*	0.52	0.45	57.48	61.21	64.54	68.27	absent	absent

*Mean of all subject values

Table 7 values showed that PPD held the final SIT word for less time than they did at the beginning of treatment. Following therapy, average intensity across the final word of the SIT sentence and peak intensity on the final word increased. There were no pitch breaks present in either the pre- or post-treatment SIT sentence final words.

Table 8. Voice Quality During Last SIT Sentence Word

	MID*	Effect Size
Word Duration	0.018	1.91
Mean Intensity	1.87	1.00
Maximum Intensity	2.06	0.90

*MID= 0.5 x $S_{\text{pooled baseline}}$

Table 8 values show that the MID values of mean intensity and maximum intensity were large (i.e., over the value of 0.8, which indicates a large effect size according to Cohen (1988)); word duration had a small MID value. The actual effect sizes appear in column three; the reader should compare them to a value of 1 that indicates a large effect size. The SIT sentence final words had large effect sizes for all three dependent variables, the last two of which

are interrelated. The values suggest the possibility of real clinical change in these parameters.

B. Auditory-Perceptual Results of SIT Sentence Speech Samples

The values reported in the following section are based on the observed and expected percentages of the responses given by the 33 listeners for the 5 PPD subjects (5 PPD subjects x 10 pre/post sentences x 2 presentations x 33 listeners=3300 responses).

The first area that was investigated in the auditory-perceptual portion of the experiments was the percentage of listeners that chose the post-treatment samples as sounding better. The percentage of listeners choosing the pre-treatment speech samples as sounding better was 32.6% and the percentage choosing the post-treatment speech samples was 67.4%. The McNemar test yielded a chi square value of 0.0000, which is statistically significant, and the difference between proportions was 0.35 or 35%. Therefore, the number of listeners who preferred the post-treatment samples was significant.

The second area that was calculated was a reliability level of the listeners' choices; that is, did the listeners consistently choose the same sentence (recorded either pre- or post-treatment) the first time and the second time they audited each pair of sentences? For example, the previous paragraph stated that more listeners preferred sentences produced post-treatment, as compared to those who preferred sentences produced pre-treatment. The

reliability calculation answered the question, “Did the listeners who preferred post-treatment sentences on the first presentation make the same choice when the pair-choice tasks were repeated?” This yielded a McNemar test value of 0.9227 and the percentage difference between the proportions was 0.18% [95%CI 0.23; 0.26%]. Although more listeners chose the post-treatment samples on the second presentation than on the first presentation, these results were not significant and there was a very small difference of 0.18% between the number of post-treatment samples chosen on the first and second presentation. The result suggests that listeners were choosing the same sentences both times they listened; they were not guessing haphazardly.

The third question that was asked in this area was the level of confidence that listeners reported upon the first and second presentation of the samples. The McNemar test value was 0.022 and the difference between proportions of those who indicated that they had either low or high confidence in their choices at first audit versus second audit was 33% [95% CI 4-61%]. Therefore, the difference in confidence level of the raters between the first and second presentation was significant; listeners had higher confidence when they audited sentences for the second time.

Part IV: Semi-Structured Interview Results

Question: What are the changes observed in PPD Pre- and Post- Voice Choral Singing Therapy, as Revealed in the Interview Data?

The following figures in this section detail the mean of median values for each of the four interview areas from 10 different raters who used the GRC scale designed for this study. The mid- and post- mean of the median values were represented visually with the following figures.

In order to describe the amount of change raters perceived, given the interview statements, the cut-offs used in Musselman (2007) was adapted and applied to the scale used in this study. The global rating scale used by Musselman describes ratings of ±1-3 as a small change (MID), ±4-5 as a moderate change, and ±6-7 as a large change. For the purposes of this study, ±1 will indicate a small change, ±2-3 will indicate a moderate change, and ±4-5 will indicate a large change. These qualitative descriptors were given in order to allow for comparison between the values.

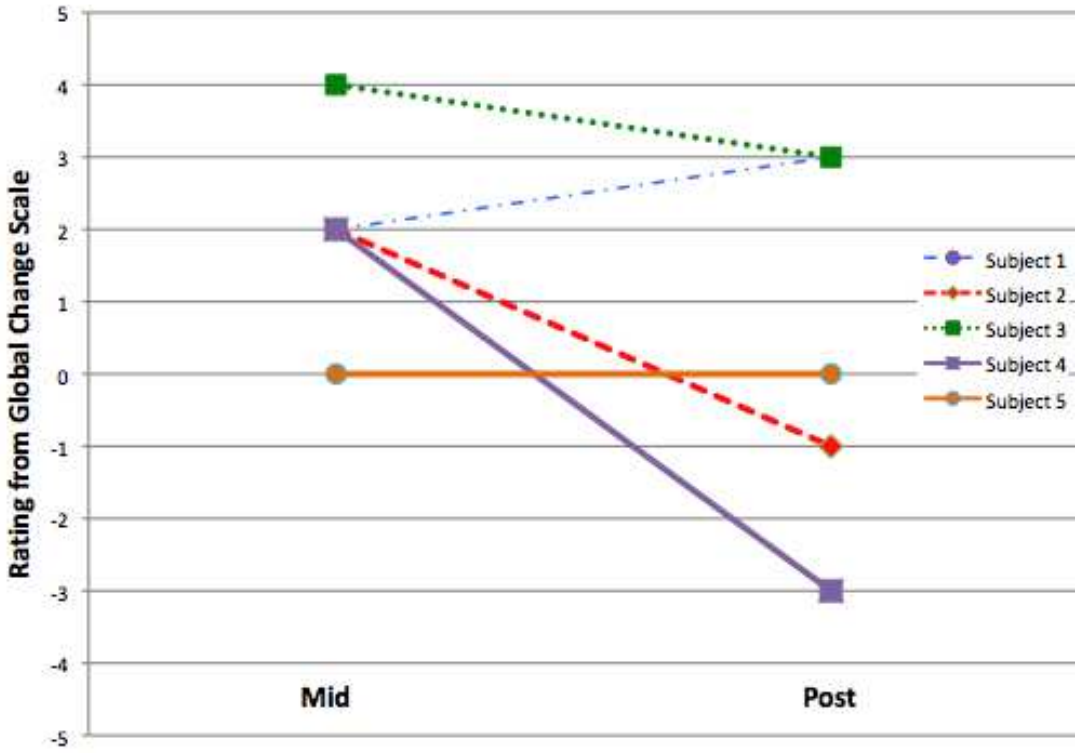


Figure 10. Swallowing Change Ratings Mid- and Post- Treatment

In the swallowing portion for the interview statements it was noted that only subject one was judged to reflect improvement in this area.

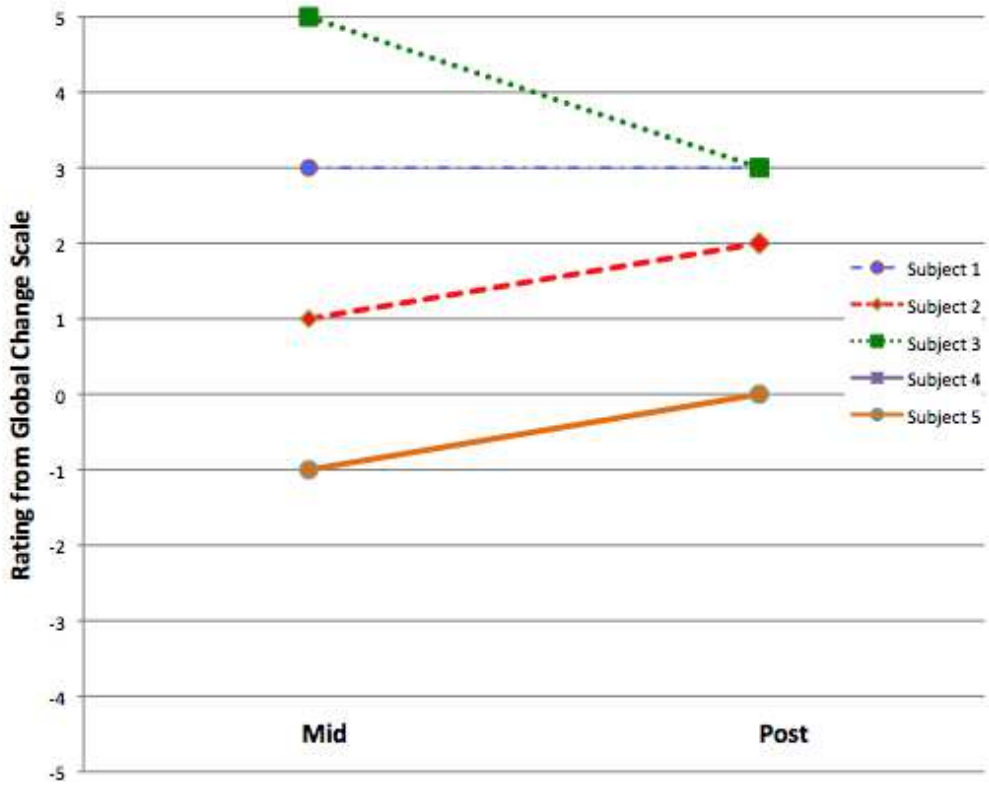


Figure 11. Coughing Change Ratings Mid- and Post- Treatment

In only two subjects (two and five) there was a small improvement seen in the domain of coughing. Subject three showed a decrease in the area of perceived change for coughing. Subject four's data are missing because there were no post-treatment statements available.

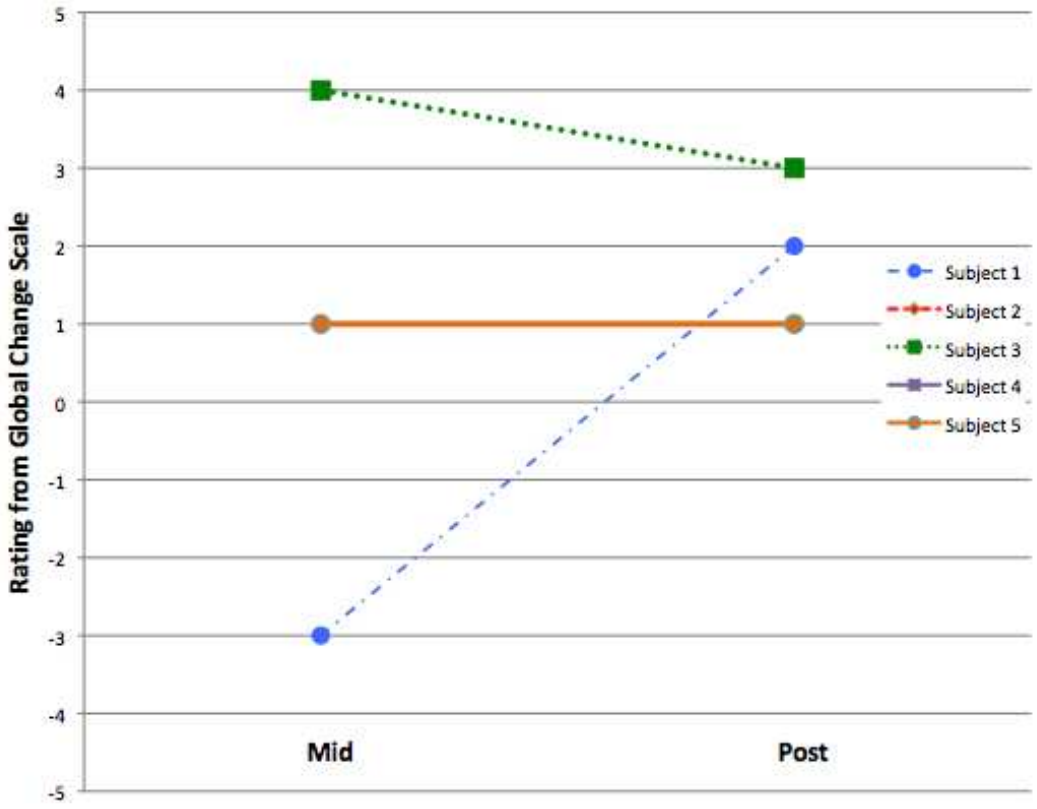


Figure 12. Facial Expression Change Ratings Mid- and Post- Treatment

In the domain of facial expression, only one subject was judged to show a large improvement in the area of facial expression. Data from subject two and four were missing because no post-treatment statements were available.

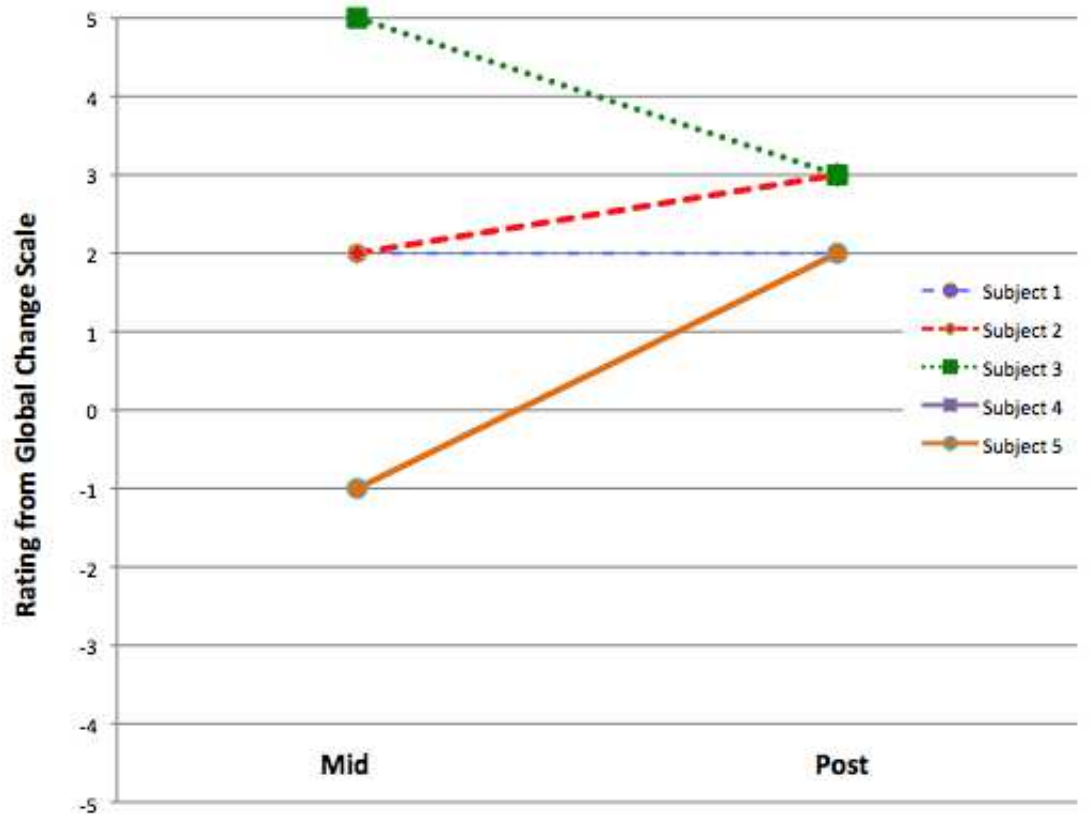


Figure 13. Speech Change Rating Mid- and Post- Treatment

Only two subjects (two and five) showed a change improvement as judged by raters. Subject three was judged to have a negative change in improvement as deemed by raters. There is no data value for subject four because the post-treatment response was not collected.

It was seen that amongst the four domains that were measured, only the swallowing, coughing, and speech domains showed that three participants were rated as having a small positive change. There were two subjects who were perceived as having a positive change in the area of speech. There was one subject who was judged to have a large positive effect in the area of facial

expression. All the other responses were either unchanged or had a small to large negative change.

Part V: Pattern Analysis Results

The trends seen in the comprehensive pattern analysis of Table 9, overall, show an increase in the areas of acoustic measures for singing and speech. There were more decreases for sentence duration in speech, but this trend was not seen in the singing samples. The interview area shows no clear trends and is more of a mix of positive, negative and no-change trends. Pitch breaks in the end words showed no change in either singing or speech samples. Overall, it was seen that for the areas of singing acoustics, speech acoustics, and listeners' auditory-perceptual impressions of sentences, there was a positive trend indicating desired changes in, and listener preference for, the post-VCST treatment samples.

Table 9. Pattern Analysis

Subject	Acoustics Singing									Acoustics Speech								A-P*	Interview			
	#BG	BGD (s)	MI (dB)	MXI (dB)	IR (dB)	Voice Quality				SD (s)	MI (dB)	MXI (dB)	IR (dB)	Voice Quality					S	C	F	S&A
						WD (s)	MI (dB)	MXI (dB)	+/- PB					WD (s)	MI (dB)	MXI (dB)	+/- PB					
A	+	+	+	+	+	-	+	+	0	-	+	+	-	-	+	-	0		+	0	+	0
B	0	-	+	+	-	+	+	+	0	-	+	+	+	-	+	+	0		-	+	/	+
C	+	+	+	+	+	-	+	+	0	-	+	+	+	-	+	+	0	+	-	-	-	-
D	0	+	+	+	+	-	+	+	0	+	+	+	+	-	+	+	0		-	/	/	/
E	0	+	+	+	+	+	+	+	0	+	+	+	+	-	+	+	0		0	0	0	+

"+"=Post better than pre
"-"=Post worse than pre
"0"=Post not better than pre
"/"=Missing value
*Auditory-Perceptual portion calculated using a grouped McNemar Test for all subjects

Legend

S=Swallowing
C=Coughing
F=Facial Expression
S&A=Speech & Articulation
BG=Breath Group
BGD=Breath Group Duration
SD=Sentence Duration
WD=Word Duration
MI=Mean Intensity
MXI=Maximum Intensity
IR=Intensity Range
PB=Pitch Breaks

Discussion

Part I: Singing Discussion

A. "Silent Night" Singing Samples

The comparison of the pre- and post- treatment acoustic values between the *Silent Night* singing samples show there was a decrease in the number of breath groups and increased breath-group duration. PPD needed to take fewer breaths and were able to hold phrases for a longer period of time, which support the findings of Haneishi (2001). Haneishi found that participants were able to sustain breath through musical phrases; although Haneishi did not measure them, sustaining musical phrases would logically have resulted in longer breath-group durations and decreased breath-group number.

A post-treatment increase in average intensity, maximum intensity and intensity range were also found in this study, again consistent with Haneishi's (2001) findings. It is important to note that the measures of average intensity, maximum intensity, and intensity range are interconnected variables; you cannot have an increase in one without a cascading effect on the others. For example, if only one measure were to be taken in the future, maximum intensity would suffice since an increase in maximum intensity would increase the average intensity as well as the intensity range. It was expected that VCST would strengthen the laryngeal muscles; the increased vocal-fold adduction required for phonation would also increase the closed phase and medial compression of the vocal folds. Since there was increased adduction of the vocal folds, it is logical that there would be an increase in sub-

glottal pressure and expiratory pressure that would be required for phonation. Furthermore, the increase in vocal intensity seen following VCST could also be attributed to increased mouth opening.

Di Benedetto et al. (2009) showed that Maximum Expiratory Pressure increased following VCST. The article does not clarify precisely what instructions were given to the patients, but it is possible that they were instructed to "take big breaths" and to "expel all the air". If that were the case, the pressure measure would not reflect the expiratory pressure required to produce either speech or song, so it is difficult to make any comparison between the Di Benedetto et al. findings and our results in this study.

In this study the intensity measurements that were obtained pre-treatment were within acceptable norms for typical speakers. Although the voice intensity of PPD increased following therapy, they were still speaking at an acceptable intensity level before treatment was administered. Nonetheless, the increased maximum intensity showed that PPD were able to engage production of subglottal pressure, more forceful adduction of the vocal folds, greater mouth opening, or some combination of these factors to reach higher levels of their intensity capacity.

Increases in intensity were not as large for *Silent Night* as they were for speech, but this might be explained by the fact that singing is more often a maximal performance task. In other words, there might have been a "ceiling effect"; pushing

for maximum sound in song at pre-treatment might have left less room for increase at post-treatment.

Effect sizes of acoustic parameters were obtained in this study. They do not reflect the distribution of those that might be seen in a larger group of participants; the effect sizes serve more as a means to reveal the areas in which change most likely occurred. The changes appeared meaningful for number of breath groups, mean intensity, and maximum intensity; emphasis should be placed in future research on further investigation of those parameters that show potential for change as a result of treatment.

The reason for looking specifically at "child" and "mild" was because those words occur at the end of musical phrases that are at a mid-point in the piece (allowing voices to be warmed up by that time) and in the modal voice register. As such, those words were chosen because it was hypothesized that any changes in vocal quality would occur once voices were warmed-up in the mid-point of the piece. As well, the notes would have been in a comfortable mid-range as compared to a later end-phrase word on a rising pitch ("peace") and may have required the use of a higher vocal register, like falsetto.

Results showed that there was an overall increase of 3 to 6 dB in average intensity and maximum intensity for the pre-pausal words "child" and "mild". Furthermore, average intensity and maximum intensity showed the greatest effect sizes. Such increases are consistent with an interpretation of increased respiratory support or

vocal fold adduction during a singing task. These results are especially meaningful because these values were taken at the end of the phrase and likely at the end of a breath group. Taken together with the finding of an overall increase in average intensity and maximum intensity across the entire *Silent Night* song, it may be fairly concluded that the participants exhibited greater respiratory capacity for singing tasks post-VCST treatment or acheived a higher sub-glottal pressure in order to overcome the increased vocal fold resistance as a result of VCST laryngeal strengthening.

Conversely, the average word duration of the musical-phrase-ending words "child" and "mild" ***decreased*** from 1.16 seconds to 1.09 seconds. It was expected that an observation of an increase in word duration might have been made post-VCST treatment; however, a difference of less than a tenth of second might be due to standard error or an artistic choice that the PPD subject exercised in this study. It appeared that word duration and maximum intensity had only moderate effect sizes. Given the intensity increases, there may have been a slight trade-off between intensity and word duration; as such it is reasonable to emphasize that those intensity increases were made at negligible word duration differences. It was also noted that the post-treatment word duration values were clustered closer together. This decrease in variability may be due to the fact that participants were better at giving the note duration its full worth or they were able to better adhere to the tempo Wiens set during the song, potentially indicating increased muscular control.

Finally, there were no pitch breaks during these words; that is, there was no perceptual evidence that breath was waning at the "child"/"mild" pre-pausal endings. The finding of no pitch breaks was made both pre- and post-VCST treatment, so there was no evidence of change in this parameter.

Implications for future research

The dimensions by which to acoustically measure the differences in "voice quality" at pre-pausal words is still unclear. Perhaps the voice quality at the ends of ***all*** phrases must be viewed to gain a comprehensive indirect assessment of breath support and vocal fold adduction force at pre-pausal words. The challenge is to determine which acoustic parameters correlate with what listeners perceive as improved voice quality. Measures of shimmer and jitter may be useful in future, too.

Areas of Improvement

In future studies of pre- and post VCST changes, several areas of methodological improvement are recommended to improve internal validity and reliability. Firstly, the acoustic samples from *Silent Night* in this study were obtained from VHS video/audio files that were then digitized as audio WAV files for analysis. It is recommended that high quality acoustic recordings be obtained directly via a digital video/audio or audio recorder pre- and post- treatment to most accurately obtain and measure acoustic data. Controlling a standard mouth-to-microphone distance to calibrate intensity measurements is also recommended.

Secondly, in this study, the VCST instructor played an electric keyboard in the background and the chord sounds were intermittently present in the acoustic data, which would have affected the accuracy of the intensity measurements. Future studies should ensure that participants have the ability to hear the musical accompaniment, but that only the speech signal be recorded. Thus, it is recommended that participants wear headphones when their singing samples are recorded in order to reduce the interference of the musical accompaniment.

Thirdly, in this study the microphone-to-mouth distance for the *Silent Night* samples probably varied slightly, which would have affected the intensity readings. While the video camera, on which the *Silent Night* singing was recorded, was always placed in precisely the same position in the sound booth and participants were instructed to stand in the same spot each session, there would have been slight variations in body orientation as they were singing and consequent effects on intensity readings. In order to remedy this in future studies, all acoustic recordings should be done with a standardized microphone-to-mouth distance, best realized with use of a head-mounted microphone. The configuration will need to be carefully thought out if participants wear headphones to hear accompaniment. It is also recommended that a high-quality microphone and a high-quality recorder be used.

There was also possible human error in this study in taking the acoustic measurements, even though the analyst tried to maintain consistency in terms of where to mark word boundaries (e.g., starting at the beginning of formants of

vowel- onset words). This would have affected all the acoustic results if there were a margin of human error that was introduced during marking word and phrase boundaries. In future research, this could be remedied by having multiple analysts conducting acoustic measurements and then testing inter-rater reliability.

Although methodological improvements could have added to the internal validity and reliability of this study, it succeeded in documenting acoustic changes that were observed post-treatment and thus serves as a useful guide for future research. It appears important to continue to focus on functional differences following VCST in subjects' singing. Di Benedetto et al. (2009) found respiratory differences following VCST but did not offer information about how much lung volume is typically used in functional daily activities. In future, measures of sub-glottal pressure or vocal fold closure or airflow might allow more detailed observation of post-VCST changes. It appears justified to continue evaluating acoustic areas, because they appear to indirectly indicate an increase in sub-glottal pressure and muscular strength in the larynx in order to produce changes in breath-group duration, mean intensity, maximum intensity, and intensity range in singing activities pre- and post-VCST. Measuring perceptible vocal differences by listeners also appears promising.

Part II: Speech Discussion

A. Acoustics Discussion of SIT Sentence Speech Samples

The acoustic measures made from SIT sentences showed an increase from pre- to post-treatment in the interrelated intensity dependent variables: average intensity,

maximum intensity, and intensity range. Overall, intensity levels of speech increased. In addition, the use of a larger range of intensity is consistent with wider variation of stress patterns used for animation, interest and emphasis in speech delivery.

There was a decrease in sentence duration following treatment. This could be attributed to the fact that speakers had more fluent articulation post-treatment and were able to fit more syllables per second within the sentence duration. Syllables produced per second were not counted in this study, but could be, to check this hypothesis. It is also possible that they had faster onsets and offsets, which may indicate a decrease in bradykinesia. This, too, could be measured. Individuals may take in more breath at the beginning of a planned longer sentence in order to prepare for the extended duration, creating greater sub-glottal pressure and forceful vocal-fold adduction to manage the breath stream. Possible increased speed of articulation might also be associated with the production of longer sentences, in order to deliver them on one breath.

Additionally, effect sizes were calculated for measurements from the SIT sentences; again the interrelated dependent variables mean intensity, maximum intensity and intensity range were largest. This outcome was consistent with findings for singing. It appears that these three measures should be areas that are included in future research. Sentence duration did not show a significant decrease, but given that most participants produced the ten- and eleven-word SIT sentences on one breath, the

samples used may not have taxed their respiratory systems sufficiently to reveal change pre- and post -treatment. It is recommended that speech samples of greater length, like paragraphs, are included in future research in order to provide a maximum performance task.

Further investigation of the end words in the SIT sentences revealed an overall decrease in word duration across all participants and an increase in mean and maximum intensity. The increase in the mean and maximum intensity is consistent with the increase in values obtained for the entire SIT sentence, so it appears that at the final word of the sentence, where one might expect deteriorating voice quality due to diminishing air supply, intensity increased from the pre- to post-treatment time points. Furthermore, it was seen that across all participants there was a decrease in the duration in the sentence final word, which matches the trend seen across the entire SIT sentence. As previously discussed, this may be due to the individual variations in pronunciations of the sentence and may not be directly related to VCST changes. There could also be increased fluency and better motoric control in the speakers so that the sentence took a shorter amount of time to express post-treatment.

The effect sizes were calculated for acoustic measures of the end words in the SIT sentences. It was seen that each domain of sentence duration, mean intensity, and maximum intensity had a large effect size, so it suggests that they would be worthwhile areas to continue investigating. The outcomes are also consistent with

the large effect sizes for mean and maximum intensity seen across the entire SIT sentences.

Suggestions for Future Research

More valid comparisons could have been made in this study if the acoustic measures had been taken from the same SIT sentences pre- and post- treatment. The pre-treatment sentences were ten words long and the post-treatment sentences were eleven words long. (It is worth noting here that despite the use of longer sentences post-treatment, sentence durations were not lengthened.) In future research, it would be recommended to use the same pre- and post- sentences. Furthermore, it would be interesting to have standard sentences of increasing length to visualize any potential changes in fluency by looking at the number of syllables expressed during the sentence duration. In this way, the sentence length and content would be controlled. In addition, including some very long sentences for pre- and post-measures would create a condition in which respiratory systems are taxed to a greater degree, leaving more room for change to be detected pre- and post.

As with the singing samples, the SIT-sentence samples were also derived from VHS video/audio tapes that were subsequently digitized as WAV files. In future it would be advisable to make direct digital recordings from which acoustic measures could be drawn.

Finally, as with the singing samples, the microphone-to-mouth distance was not completely controlled in this study. Although the camera was placed at the same

distance away from the speaker at each session, the participants often looked down at the page to read the SIT sentences and this change in orientation might have affected the pick-up of the acoustic signal and, consequently, the intensity readings. In future studies, it would be advisable to ensure a standardized microphone-to-mouth distance and to make sure the participant is speaking directly into the microphone. Again, a head-mounted microphone would remedy these distance and direction recording concerns, but as stated earlier, careful thought would need to be given to how to keep microphone distance standard and have participants wear headphones to listen to accompaniment.

B. Auditory-Perceptual Evidence from SIT-Sentence Speech Samples

Russell et al. (2010) noted that VCST up until this point had no functional measures of voice quality. This portion of the study yielded new findings; the auditory-perceptual change in voice quality of speech samples from pre- to post-VCST has not been investigated in previous research. The PPD-produced, voice-sample preference of unfamiliar listeners was determined by playing randomized pre- and post-sentence samples. The proportion of listeners choosing the post-treatment sample as sounding better was significantly higher than the proportion of listeners choosing the pre-treatment speech samples as sounding better. These results showed that on a gross perceptual level, native English speakers, who had no previous training in rating voice quality, chose the post-treatment samples as sounding better. Although the results do not show why listeners thought those samples sounded better, during the debriefing portion of the study, listeners were asked to list the characteristics

that determined which samples they thought sounded better. Listeners stated that samples that had greater clarity, loudness, pronunciation, flow, expressivity, inflection, naturalness, confidence, energy, pitch range or higher pitch were the samples they were listening for in order to make a choice. The debriefing information showed that a combination of criteria, like those listed above, shaped listener perception of the voice quality of participants' voices. In a public lecture given by Doyle (2011), he stated that acceptability, naturalness, and pleasantness are all highly correlated descriptive measures. As well, he noted that his research shows that naïve listeners are effective at giving global assessments of voice quality and may, in fact, offer less biased judgments than do clinicians, who have a vested interest in hearing positive change as a result of the therapy they offer.

Doyle (2011) also reported that it is difficult for naïve listeners to distinguish between intelligibility and voice quality; a speaker could be perfectly intelligible but exhibit very poor voice quality that can have a serious impact on the success of communicative interactions. Thus, poor voice quality in speakers could result in their experiencing a communication effectiveness disorder. A communication effectiveness disorder means that speakers could have disrupted communicative interactions and decreased social participation due to listener perceptions in response to their poor voice quality. That evidence underlines the importance for future research to continue to use the perceptions of naïve listeners as a metric for detecting voice change from pre- to post-VCST.

The confidence ratings of the listeners showed that there was a statistical difference between the first and second presentation of the samples. That is, listeners expressed that they made their choices with greater confidence on the second presentation of speech samples (perhaps mere training builds confidence); however, listeners were consistent in their sample choices, no matter if it was the first or the second time they were hearing an item. Therefore, in future studies it may not be necessary to double the presentation of test items; re-testing 20% would likely be adequate to check listener reliability. As well, it may not be essential to obtain a confidence level in the future. Though raters felt more confident with the second presentation of items, it appears that this did not affect their reliability in choosing items that they considered to exhibit better voice quality.

Suggestions for Future Research

The data from this perceptual portion of the study were nominal level data, with either an A or B voice sample choice. To increase the capacity to perform higher-level statistics in the future, a visual analog scale could be used for raters to rate how much they preferred each of the voice samples when they were presented in pairs. Due to the fact that a visual analog scale would be on an open-ended 100 mm line, a listener could mark where they felt their judgments fell and this could be measured to obtain ratio-level data.

Listeners were not completely randomized in this study; they were students in a linguistics course who could choose which studies to participate in given the study

posting. This preferential study choice and this educated population could have introduced listener bias in the study. In the future, using a more randomized pool of listeners with varying educational backgrounds may decrease this area of potential bias and result in a listening group more representative of the general population.

Although the listening portion of the study took under 30 minutes, there were four listeners who reported fatigue and boredom during the listening task near the end of the session. This could have resulted in guessed responses, which would have affected choice accuracy throughout the experiment and may have been remedied by not repeating samples twice. It is advised in future research that repetition of samples be kept to twenty percent. As well, since we attempted to calculate intra-rater reliability, it was later recommended by our statistics consultant that the same listener should have repeated the listening task but with a week in between tests to act as a buffer. Due to logistical reasons for this study, this was not possible, but this would be an area to consider in the future.

Part III: Evidence from Semi-Structured Interviews

Of the four interview topic areas that were investigated, only participants' comments about the area of coughing were perceived by judges as showing some amount of change in two of the subjects. Participants' comments from the three other interview areas (swallowing, speech/articulation and facial expression) were perceived by judges as exhibiting some change in only one participant.

Areas of Improvement

The challenge in this portion of the study was designing a methodology to analyze a set of interview data that did not lend itself to its initial aim of qualitative analysis. Therefore, in this study, we attempted to design an analysis method that would evaluate the changes and find a way to rate their degree of change and graph the changes that resulted from mid- to post- treatment times. As well, in the initial study, the interview statements were not asked in a way in which change could be revealed, so oftentimes the statements were not so much statements of change as statements regarding current health status. For example: "My swallowing has worked very well." This statement could be rated as +2, which is slightly improved, but it could also be rated as 0 to reflect that it is unchanged. Thus, these statements of health status could be rated as either positive or negative depending on the rater. It is due to this subjectivity that makes it difficult to interpret some of the unexpected trends seen throughout the data.

As well, the raters were all graduate students in Speech-Language Pathology, so while they were unfamiliar with the subjects in the study, they have had experience in evaluating change as a result of their experience in the program. Therefore, they could have been rating with some pre-existing bias as they read through the interview statements.

To have more accurately assessed this area, a qualitative approach should have been used in the initial design of this study, utilizing more open-ended questions to

produce a rich transcript that would reveal therapeutic VCST themes. Instead, the semi-structured interview questions in the original study began as open-ended but follow-up questions quickly began to be shaped by the interviewer as close-ended. Thus it is important, in future qualitative research in this area, to get participants to comprehensively discuss the changes they have experienced as a result of VCST and not to guide them towards particular answers. In order to improve this area a specially trained qualitative interviewer should conduct all interviews. Furthermore, in order to quantify a perceived change, it would also be advisable to conduct only one pre- and one post- interview session rather than asking the same questions each week in order to view a change; however, if the researcher wished to ask the same questions, they could ask the participants to fill out the same rating scale each session but only conduct comprehensive interviews at the beginning and end of treatment.

The GRC scale was employed to quantify the amount of change PPD experienced throughout the experiment. GRC scales are popularly used in clinical research and allow for researchers to quantify a patient's improvement or deterioration over a period of time in response to treatment (Kamper et al., 2009). The scale is traditionally used as a tool to quantify a patient's self-perceived level of change; however, due to the fact that a GRC was not used during the initial stages of the VCST when the data were obtained, it was used as a tool in this thesis for secondary data analysis. In using an adapted GRC scale of change for this thesis, the ten raters documented their thoughts after rating the PPD subjects' interview statements.

Overall, the comments detailed that the scale felt extremely subjective and that it was hard to rate without knowing the question that was asked or having a reference point for the statement. While there was a training session and items for the raters to learn how to use the scale, it appeared that more training was required; however, the scale was also an adapted scale and was piloted for a particular use within this study. The interview statements were chosen from standardized time-points and often the statements did not lend themselves to revealing change experienced as a result of therapy. There was confusion as to whether the 0 midpoint of the scale was considered normal or if +5 signaled the return to a normal state of functioning.

The area of facial expression was also a difficult one for the PPD participants to answer, as there was very little self-awareness on the subject. In order to specifically test this area, a caregiver questionnaire may have been preferable. This would be similar to the method utilized by Haneishi (2001) to report on the speech and articulation skills of PPD. Haneishi recognized that participants may have had less self-awareness in regards to their own speech and articulation, (or, as in this case, facial expression) that it would be advisable to recruit the opinion of caregivers who spend the most time with participants. As well, videotapings of participants' emotional responses could have been recorded and viewed by unfamiliar viewers to rate facial expression. Furthermore, using video recordings of facial responses could be measured using the Facial Action Coding System (FACS) (Ekman & Friesen, 1978). The FACS allows researchers to systematically categorize facial expressions. With the expression of any emotion, there are different muscle contraction and

relaxation patterns and these respective movements correspond to action units, which can be measured. These methods may have yielded more insightful observations of facial expression change than direct self-reflection.

It is due to the combination of the challenges listed above that we experienced difficulty in interpreting some trends that resulted from this portion of the study. In addition to the fact that multiple data points were missing for the areas of coughing and facial expression, which did not allow for mid- and post- treatment comparison, there were challenges for the raters to apply the rating scale to statements that did not reflect change. This would have been remedied by the initial methodological design of the study in which the rating-of-change scale would have been given directly to the PPD participant after each treatment session, or at the end of treatment, and this would eliminate the need to use the subjective ratings of unfamiliar Speech-Language Pathology judges. As well, a true qualitative approach, such as grounded theory, should have been used at the start and end of treatment with the interview being conducted by a skilled interviewer in order for change themes to be revealed. Employing the improvements highlighted above would greatly increase ability to produce valid and reliable results and comprehensively analyze the resulting change themes that participants experienced during the course of VCST.

Pattern Analysis

Overall, by combining the data from these three separate sources, the data showed that there were positive changes in the domain of singing and speech acoustics and auditory-perceptual differences that listeners were able to hear as a result of VCST in PPD. It is unclear whether the interview areas of Swallowing, Coughing, Facial Expression and Speech/Articulation show an effect, but it would be a worthwhile future investigation area. It may be that the five PPD in this study did not cite much change, but that may not be indicative of all PPD. Furthermore, since there were increases in maximum intensity, it suggests that there was an increase in laryngeal muscle strength as a result of VCST and that these physiological changes would also have implications for coughing and swallowing functions. If there is increased vocal fold resistance to sub-glottal pressure and strength of vocal fold adduction, then these forces would transfer to coughing ability since the same physiological mechanisms would need to be employed in order to clear secretions. The acoustic results indicate that there are laryngeal and respiratory physiological changes as a result of VCST, so it was logical to see the transfer of these skills to other functions. It is also unclear the level of change that is seen given the smaller acoustic measurement differences seen at end-of-phrase words in singing and speech; however, the large effect sizes in the acoustics areas, in general, indicate that it is still worthwhile to continue investigation. It may be deemed, in the future, that taking acoustic measures across the entire phrase is indicative enough and that fine examination of the voice quality at the ends of phrases may not be revealing, unless acoustic parameters other than breath-group/sentence duration, mean intensity, and maximum intensity are chosen to indicate change. Ultimately, the area that

holds the most promise is the significant voice quality preference by naïve listeners of PPD speakers' post-VCST speech productions.

Final Summary

The speech and voice difficulties exhibited by PPD can be significant enough to impact functional communication abilities as well as quality of life (Jaywant & Pell, 2010). Hypokinetic dysarthria can develop and result in decreased vocal loudness that can decrease speech intelligibility (Sapir et al., 2008). Furthermore, decreased frequency and intensity range can also reduce a PPD's ability to convey emotion during speech (Wan et al., 2010) and to communicate in social situations.

The results obtained from this study showed that there are multiple areas in which PPD might benefit from a VCST approach. A return to the Pattern Analysis Table 9 will enable the reader to see that there were positive changes in the domains of singing acoustics and speech acoustics and auditory-perceptual judgments. These results highlight the assumed increases in glottal resistance and sub-glottal pressure following VCST, since they are correlated with perceptible loudness. While interview data could not be interpreted, there was no indication that qualitative measures would not be a worthwhile future investigation area. It may be fruitful to promote self-reflection of therapy benefits by the participants. As well, all PPD participants were highly motivated individuals, which is a contributor to positive therapeutic change.

There are therapeutic effects of singing within group contexts. Bailey and Davidson (2003) showed the holistic health benefits of singing in society. A quasi-experimental study by Cohen et al. (2006) also showed the possible benefit of singing on the physical and mental health of the elderly. These findings are meaningful given that PPD are perceived as being less interested, less involved, less friendly and happy than speaker controls (Jaywant & Pell, 2010). VCST really aims to increase the social interaction and communication opportunities within a functional group therapy context. The breath-support strengthening and voice exercises that are used in treatment could easily be integrated into the pedagogy of local choirs, combining vocal benefits with social benefits. Mood, in relation to VCST, is a promising area of future research; the reader is reminded that Nicks (2004) reported that out of the eleven acoustic variables measured by Tanner-Semple in her VCST treatment study in 2003, there was no significant change but one – a participant indicated an improvement in his mood.

Di Benedetto et al. (2009) argue that VCST could be a more cost-effective treatment. Potential cost-effectiveness (an area for future research) as well as the accessibility of VCST services makes it an appealing voice therapy option for PPD. Both are promising areas for research.

Neuroimaging is also expanding the potential for research to explore the relation between music and brain structures. For example, cerebral blood flow could be analyzed in participants who have undergone VCST. Narayana et al. (2010) found

increased bloodflow in the right hemisphere in the speech motor region following LSVT treatment. Wan and Schlaug (2010) found that instrumental musicians had strengthened connections between the auditory and motoric regions of their brains through the arcuate fasciculus as a result of repetitive practice and continuous multisensory feedback that occurs while playing a musical instrument. The evidence suggests that there is plasticity for acquiring sensori-motor and cognitive functions as a result of making music. Music might offer an interactive treatment form for maintaining normal aging or addressing developmental and neurological disorders. Functional imaging studies might be a way to investigate the further neurological changes that result from therapy.

Wan et al. (2010) stated that there still need to be randomized control trials to determine efficacy and dose effects of VCST, further tests of generalization of VCST to voice quality, and long-term treatment effects. The results from this study showed a promising improvement trend in terms of acoustic measurements of mean intensity, maximum intensity, and the auditory-perceptual responses of unfamiliar listeners to speech of those who have undergone VCST; however, while there are still many facets of research to investigate before VCST can become a standardized voice-treatment protocol for PPD it remains a promising area of voice therapy research.

References

Apple Inc. (2007-2008). iMovie'08 (version 8.1.4) [Computer software]. Available from www.apple.com.

Bailey, B.A., & Davidson, J.W. (2003) Perceived holistic health effects of three levels of music participation. In R. Kopiez, A.C. Lehmann, I. Wohther, & C. Wolf (Eds.), *Proceedings of the 5th Triennial ESCOM Conference*, 8-13 September 2003, Hanover University of Music and Drama, Germany.

Bartels, A. L., & Leenders, K. L. (2009). Parkinson's disease: The syndrome, the pathogenesis and pathophysiology. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior,* 45, 915-921.

Baumgartner, C. A., Sapir, S., & Ramig, T. O. (2001). Voice quality changes following phonatory-respiratory effort treatment (LSVT) versus respiratory effort treatment for individuals with Parkinson disease. *Journal of Voice: Official Journal of the Voice Foundation*, 15(1), 105-114.

Blanchet, P. G., & Snyder, G. J. (2009). Speech rate deficits in individuals with Parkinson's Disease: A review of the literature. *Journal of Medical Speech-Language Pathology*, 17(1), 1-7.

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer [Computer program]. Version 5.2.03, [retrieved 1 November 2010 from http://www.praat.org/]

Bowers, D., Miller, K., Bosch, W., Gokcay, D., Pedraza, O., Springer, U., & Okun, M. (2006). Faces of emotion in Parkinsons disease: Micro-expressivity and bradykinesia during voluntary facial expressions. *Journal of the International Neuropsychological Society: JINS,* 12, 765-773.

Callahan J.L., & Reio, Jr, T.G. (2006). Making subjective judgments in quantitative studies: The importance of using effect sizes and confidence intervals. *Human Resource Development Quarterly*. 17, 159-73.

Clarke, C. E., Zobkiw, R. M., & Gullaksen, E. (1995). Quality of life and care in Parkinson's disease. British Journal of Clinical Practice, 49: 288-293.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, New jersey: Academic Press Inc.

Cohen, G.D., Perlstein, S., Chapline, J., Kelly, J., Firth, K.M. & Simmens, S. (2006) The impact of professionaly conducted cultural programs on the physical health, mental health and social functioning of older adults-2-year results. *Journal of Aging, Humanities and the Arts*, 1: 5-22.

Crozier, E., & Hamill, R. (1988). The benefits of combining speech and music therapy. *Speech Therapy Practitioner.* Nov., 9-10.

Crystal's Website. Diagram of the Anatomy of Swallowing. Retrieved November 12, 2011 from Google.ca. http://www.our-sma-angels.com/crystal/web%20pages/medicalpractices/therapists/Speech/diagramswallow.htm.

Dagli, A.S., Mahieu, H.F., & Festen, J.M. (1997). Quantitative analysis of voice quality in early glottic laryngeal carcinomas treated with radiotherapy. European Archives of Otorhinolaryngology, 254, 78 -80.

Daube, J. R., Reagan, T. J., Sandok, B. A., & Westmoreland, B. F. (1986) *Medical Neurosciences*. Boston, MA: Little, Brown and Company.

Di Benedetto, P., Cavazzon, M., Mondolo, F., Rugiu, G., Peratoner, A., & Biasutti, E. (2009). Voice and choral singing treatment: A new approach for speech and voice disorders in Parkinson's disease. *European Journal of Physical and Rehabilitation Medicine, 45*(1), 13-19.

Doyle, P. C. "Voice quality and speech intelligibility measurement in laryngectomy." University of Alberta, Public Lecture. Edmonton, AB. 16 Sept. 2011.

Duffy, J. R. (2005). *Motor Speech Disorders*. St. Louis, MO: Elsevier Mosby.

Ebihara, S., Saito, H., Kanada, A., Nakajoh, M., Takahaski, H., Arai, H., & Sasaki, H. (2003). Impaired efficacy of cough in patients with Parkinson disease. *CHEST*, 124, 1009-1015.

Eckel, F. C., & Boone, D. R. (1981). The S/Z ratio as an indicator of laryngeal pathology. *The Journal of Speech and Hearing Disorders*, 46(2), 147-149.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system.* Palo alto: Consulting Psychologist Press.

Fox, C. M., Ramig, L. O., Ciucci, M. R., Sapir, S., McFarland, D. H., & Farley, B. G. (2006). The science and practice of LSVT/LOUD: neural plasticity-principled approach to treating individuals with Parkinson disease and other neurological disorders. *Seminars in Speech and Language*, 27(4): 283-299.

Goberman, A. M., Blomgren, M., & Metzger, E. (2008). Characteristics of speech disfluency in Parkinson disease. *Journal of Neurolinguistics*, 23, 470-478.

Grass Valley. (2004-2007) High-Quality Analog/Digital Video Conversion Canopus ADVC 110.

Guyatt, G. H., Osoba, D., Wu, A. W,. Wyrwich, K. W., Norman, G. R., & Clinical Significance Consensus Meeting Group. (2002). Methods to explain the clinical significance of health status measures. *Mayo Clinic Proceedings*, 77, 371-383.

Hallam, S. (2010). The power of music: Its impact on the intellectual, social, and personal development of children and young people. *International Journal of Music Education*, 28(3), 269-289.

Haneishi, E. (2001). Effects of a music therapy voice protocol on speech intelligibility, vocal acoustics measure, and mood of individuals with Parkinson's disease. *Journal of Music Therapy*, 38, 272-290.

Ho, A.K., Iansek, R., Marigliani, C., Bradshaw, J.L., & Gates, S. (1998). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology*, 11(3), 131-137.

Inzelberg, R., Schechtman, E., & Hocherman, S. (2008). Visuo-motor coordination deficits and motor impairments in Parkinson's disease. *PLoS ONE*, 3(11): e3663. doi:10.1371/journal.pone.0003663

Jaywant, A., & Pell, M. D. (2010). Listener impressions of speakers with Parkinson's disease. *Journal of the International Neuropsychological Society,* 16, 49-57.

Justice, L. M. (2010). *Communication Sciences and Disorders A Contemporary Perspective*. (2nd ed.). Boston, MA: Allyn & Bacon.

Kamper, S. J., Maher, C. G. & MacKay, G. (2009). Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design. *The Journal of Manual & Manipulation Therapy*, 7(3), 163-170.

Kay Elemetrics Corporation (1996). Visi-Pitch II Model 3300 [Computer software]. Lincoln Park, NJ.

Korell, M. & Tanner, C.M. (2005). Epidemiology of Parkinson's Disease: An overview. In M. Ebadi & R.F. Pfeiffer, (Eds.), *Parkinson's Disease*, CRC Press, New York, pp. 39–50.

MP4Converter Software Studio. DVD Ripper (Version 4.0.35.0719) [Computer Software]. Retrieved November 19, 2010. Available from http://www.mp4converter.net/.

Lemieux, J., Beaton, D. E., Hogg-Johnson, S., Bordeleau, L. J., & Goodwin, P. J. (2007). Three methods for minimally important difference: No relationship was found with the net proportion of patients improving. *J Clin Epidemiol*, 60, 448-455.

Maxwell, D.L. & Satake, E. (2006). *Research and Statistical Methods in Communication Sciences and Disorders.* New York: Thomson.

McDonald, W. M., Richard, I. H., & DeLong, M. R. (2003). Prevalence, etiology, and treatment of depression in Parkinson's disease. *Biological Psychiatry*, 54(3), 363-375.

MedScape Reference, by Harry Quon, David Goldberg. 2011, Retrieved from http://img.medscape.com/pi/emed/ckb/otolaryngology/834279-846565-1375268-1376463.jpg. Retrieved November 12, 2011. Copyright 1994-2011 by WebMD LLC.

Miller, N., Noble, E., Jones, D., & Burn, D. (2006). Hard to swallow: Dysphagia in Parkinson's disease. *Age and Ageing, 35*(6), 614-618.

Musselman, K. E. (2007). Clinical significance testing in rehabilitation research: What, why, and how? *Physical Therapy Reviews*, 12, 287-296.

Mutch, W. J., Strudwick, A., Roy, S. K., & Downie, A. W. (1986), Parkinson's disease: Disability, review, and management. *British Medical Journal*, 293, 675-7.

Nakashima, K., Maeda, M., Tabata, M., Adachi, Y., Kusumi, M., & Ohshiro, H. (1997). Prognosis of Parkinson's disease in Japan. Tottori University Parkinson's Disease Epidemiology (TUPDE) Study Group. *European Neurology*, 38 (suppl), 60-63.

Narayana, S., Fox, P. T., Zhang, W., Franklin, C., Robin, D. A., Vogel, D. & Ramig, L. O. (2010), Neural correlates of efficacy of voice therapy in Parkinson's disease identified by performance–correlation analysis. *Human Brain Mapping*, 31, 222–236.

Nicks, S. (2004). Outcomes of a vocal & singing therapy program for PD. (Unpublished SPA 900 Project). University of Alberta, Edmonton, AB, Canada.

Oxtoby, M. (1982). *Parkinson's disease patients and their social needs.* Parkinson's Disease Society.

Pacchetti, C., Aglieri, R., Mancini, F., Martignoni, E., & Nappi, G. (1998). Active music therapy in Parkinson's Disease: Methods. *Functional Neurology*, 13, 57-67.

Pacchetti, C., Mancini, F., Aglieri, R., Fundaro, C., Martignoni, E. & Nappi, G. (2000). Active music therapy in Parkinson's disease: An integrative method for motor and emotional rehabilitation. *Journal of Psychosomatic Medicine*, 62(3), 386-93.

Pitts, T., Bolser, D., Rosenbek, J., Troche, M., & Sapienza, C. (2008). Voluntary cough production and swallow dysfunction in Parkinson's disease. *Dysphagia,* 23, 297-301.

Plowman-Prine, E. K., Sapienza, C. M., Okun, M. S., Pollock, S. L., Jacobson, C., Wu, S. S., & Rosenbek, J. C. (2009). The relationship between quality of life and swallowing in Parkinson's disease. *Movement Disorders : Official Journal of the Movement Disorder Society,* 24, 1352-1358.

Portney, L. G., & Watkins, M. P. (2000). *Foundations of Clinical Research: Applications to Practice* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psych*, 104, 1-15.

Przedborski, S. (2007). Etiology and pathogenesis of Parkinson's disease. In J. Jankovic & E. Tolosa (Eds.), *Parkinson's Disease and Movement Disorders* (pp. 77-92). Philadelphia, PA: Lippincott Williams and Wilkins.

Purves, D., Augustine, G.J., Fitzpatrick, D., Katz, L.C., LaMantia, A-S, McNamara, J.O., & Williams, S.M. (2001). Projections to the basal ganglia. *Neuroscience* (2) [Retrieved 26 September 2010, from Google.ca. http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=neurosci&part=A1223]

Quon, H. & Goldberg, D (2011). Hypopharyngeal cancer. Retrieved November 12, 2011 from MedScape. http://img.medscape.com/pi/emed/ckb/otolaryngology/834279-846565-1375268-1376463.jpg.

Ramig, L. O., Countryman, S., O'Brien, C., Hoehn, M., & Thompson, L. (1996). Intensive speech treatment for patients with Parkinson disease: Short and long-term comparison of two techniques. *Neurology*, 47, 1496–1504.

Robbins, J.A., Logemann, J.A., & Kirshner, H.S. (1986). Swallowing and speech production in Parkinson's disease. *Annals of Neurology,* 19, 283-287.

Russell, J. A., Ciucci, M. R., Connor, N. P., & Schallert, T. (2010). Targeted exercise therapy for voice and swallow in persons with Parkinson's disease. *Brain Research,* 1341, 3-11.

Sapir S, Ramig LO, Hoyt P, Countryman S, O'Brien C, & Hoehn, M. (2002): Speech loudness and quality 12 months after intensive voice treatment (LSVT LOUD) for Parkinson's disease: A comparison with an alternative speech treatment. *Folia Phoniatrica Logopedia,* 54, 296–303.

Sapir, S., Ramig, L. O., & Fox, C.M. (2008). Voice, speech, and swallowing disorders. In S. A. Factor & W. J. Weiner (Eds.), *Parkinson's Disease* (pp. 77-97). New York, NY: Demos Medical Publishing.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime User's Guide (Version 2.0) [Computer software]. Pittsburgh: Psychology Software Tools, Inc.

Selman, J. (1988). Music therapy with Parkinson's. *British Journal of Music Therapy,* part 2, 5-9.

Sharkawi, A. El., Ramig, L., Logemann, J.A., Pauloski, B. R., Rademaker, A. W., Smith, C.H., Pawlas, A., Baum, S., & Werner, C. (2002). Swallowing and voice effects of Lee Silverman Voice Treatment (LSVT): A pilot study. *Journal of Neurology Neurosurgery Psychiatry*, 72, 31-36.

Shulman, J. M., & De Jager, P. L. (2009). Evidence for a common pathway linking neurodegenerative diseases. *Nature Genetics*, 41(12), 1261-1262.

Simons, G., Pasqualini, M. C. S., Reddy, V., & Wood, J. (2004). Emotional and nonemotional facial expressions in people with Parkinson's disease. *Journal of the International Neuropsychological Society : JINS,* 10(4), 521-535.

Smith, M.C., Smith M K, & Ellgring, E. H. (1996). Spontaneous and posed facial expression in Parkinson's disease. *Journal of the International Neuropsychological Society, 2*(5), 383-391.

Streifler, M., & Hofman, S. (1984). Disorders of verbal expression in Parkinsonism. In R. Hassler & J. Christ (Eds.), *Advances in Neurology* (pp. 385-393). New York: Raven Press.

Streiner, D. L., & Norman, G. R. (2003). Health Measurement Scales: A Practical Guide to Their Development and use (3rd ed.). New York: Oxford University Press.

Tanner, M.,Wiens, H., & Campbell, M.M. (2003). The benefits of singing on the vocal quality of individuals with Parkinson's Disease. (Unpublished study). Glenrose Rehabilitation Hospital and University of Alberta, Edmonton, AB, Canada.

Tanner-Semple, M. (2005) Singing therapy for people with Parkinson's Disease: A pilot study. Poster at the Canadan Speech-Language Pathologists and Audiologists Annual Conference, Regina, May, 2005.

Troche, M., Rosenbek, J. C., & Sapienza, C. (2008) Effects of bolus consistency on timing safety of swallow in patient's with Parkinson's disease. *Dysphagia*, 23(1), 26-32.

Wan, C. Y., Rüber, T., Hohmann, A., & Schlaug, G. (2010). The therapeutic effects of singing in neurological disorders. *Music Perception*, 2010 27(4), 287-295.

Wan, C. Y., & Schlaug, G. (2010). Music making as a tool for promoting brain plasticity across the life span. *The Neuroscientist*, 16(5), 566-577.

Weiner, W., & Singer, C. (1989). Parkinson's disease and nonpharmacological treatment programs. *Journal of American Geriatric Society*, 37, 359-363.

Wiens, H. (2004). Are singing lessons beneficial for Parkinson's clients with Parkinson's disease? (Unpublished study). University of Alberta, Edmonton, AB, Canada.

Yarrow, S. (1999). Members' 1998 survey of the Parkinson's Disease Society of the United Kingdom. In R. Percival & P. Hobson (Eds.), *Parkinson's disease: Studies in psychological and social care* (pp.79-92). Leicester: BPS Books (British Psychological Society).

Yorkston, K. M, & Beukelman, D. (1981). *Assessment of Intelligibility of Dysarthric Speech*. Austin, TX: Pro-Ed.

Appendices

Appendix A: Sentence Intelligibility Test (SIT) Sentences

10.1 Most overweight people need to learn to be more assertive.

10.2 Most weeds can now be put on the compost pile.

10.3 There are effective ways to conserve both energy and water.

10.4 They are some of the best vegetable protein foods known.

10.5 At certain times, I like being strong for someone else.

10.6 Spending time with the family is really my favorite activity.

10.7 I had all the usual tests, and everything was fine.

10.8 The park has separate areas for cars and recreational vehicles.

10.9 Agriculture did not necessarily tie men to sedentary village life.

10.10 The contract provided an immediate raise in the hourly rate.

Post-treatment

11.61 You need something very sharp to get through all of that.

11.62 There was a sort of exuberance in our relationship that autumn.

11.63 He is fiercely independent, headstrong, and a hard man to contain

11.64 There are plenty of advantages to having an apartment to myself.

11.65 You know perfectly well, I would have been bored to death.

11.66 Like others, I'm doing my very best with what I've got.

11.67 He rewarded the ape with peanuts and oranges for its work.

11.68 Wordlessly, I clutched hundreds of little bodies and held them tight.

11.69 For your work, you will earn a great deal of money.

11.70 Then, perhaps, we would be able to combine safe with reason.

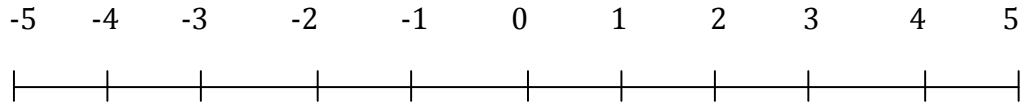
Appendix B: A Guide to Coding the Transcripts

1. Read the instructions.
2. Read **all** of the interview statements from beginning to end.
3. Re-read the last sentence: "Please focus on whether or not these statements reflect change".
4. Rate **all** responses on the scales provided.
5. Finish all ratings in one session.
6. Complete the test section and ask the researcher any questions.
7. Rate the experimental section on your own time and the return forms by **March 17, 2011**.

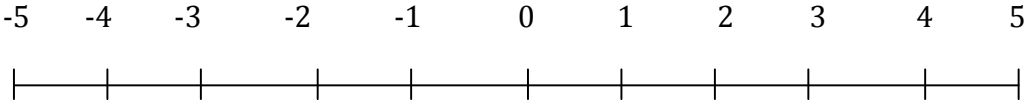
Note: Please do not discuss the content of the interview statements or discuss your ratings with the other raters to protect the privacy of the participants and their condition.

Test Section: Coding Scale for Interview Transcripts

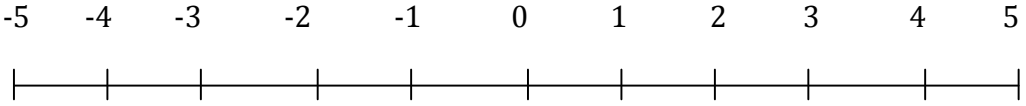
The following statements were collected from recordings of participants throughout the course of voice therapy. Please rate you perception of the following answers that individuals provided for each of the areas outlined below. *Please focus on whether or not these statements reflect change*.



Greatly worsened — Unchanged — Greatly improved



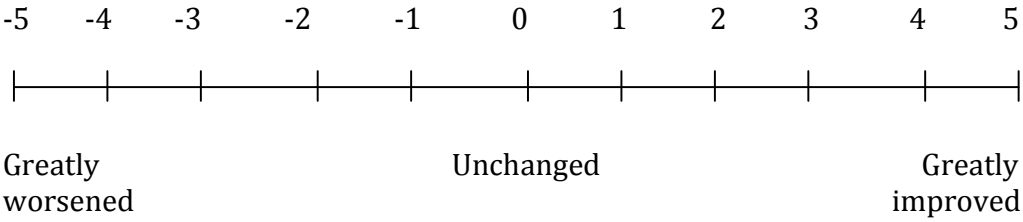
Greatly worsened — Unchanged — Greatly improved



Greatly worsened — Unchanged — Greatly improved

Experimental Section: Coding Scale for Interview Transcripts

The following statements were collected from recordings of participants throughout the course of voice therapy. Please rate you perception of the following answers that individuals provided for each of the areas outlined below. *Please focus on whether or not these statements reflect change*.



*Scale repeated for the number of interview statements