2016

The Influence of Real-Time Visual Feedback Training on Vocal Control

Justeena N. Zaki-Azat
Wilfrid Laurier University, zaki6270@mylaurier.ca

The Influence of Real-Time Visual Feedback Training on Vocal Control

Justeena Nagy Zaki-Azat

Bachelor of Science Honours, York University, 2013

THESIS

Submitted to the Department of Psychology

in partial fulfillment of the requirements for

Master of Science in Cognitive Neuroscience

Wilfrid Laurier University

Supervisor

Dr. Jeffery A. Jones

Laurier Center of Cognitive Neuroscience

Department of Psychology, Wilfrid Laurier University, Waterloo, Canada

Examining Committee

Dr. Elizabeth Olds

Laurier Center of Cognitive Neuroscience

Department of Psychology, Wilfrid Laurier University, Waterloo, Canada

Dr. William Hockley

Laurier Center of Cognitive Neuroscience

Department of Psychology, Wilfrid Laurier University, Waterloo, Canada

Dr. Marta McCarthy

Choral Canada, Association of Canadian Choral Communities

School of Fine Arts and Music, University of Guelph, Guelph, Canada

Abstract

Trained singers have better vocal control when compared to singers without vocal training. The development of precise vocal control, like any motor skill, requires practice with some form of feedback, such as auditory feedback. In addition to auditory feedback, singing training programs use online visual feedback to improve performance accuracy. The purpose of this thesis is to investigate the recent body of literature concerning the cognitive processing of vocal control, and apply this knowledge practically to develop an effective real-time visual feedback training program that enhances vocal control. In the first of two studies, non-singers and singers were randomly assigned to one of two training conditions: one condition with visual feedback of vocal performance, and the other condition with no feedback. Changes in vocal control as a function of training condition were assessed by comparing measures of pitch accuracy, vocal variability, and responses to sudden frequency-altered perturbations in participants' pitch feedback, before and after training. In the second study, training sessions were doubled and tested with another group of non-singers, with results from this second study compared to the first study. Overall, there was no effect of real-time visual feedback training or length of training on measures of vocal control. These findings may contribute to a better understanding of vocal control, and assist in improving singing training programs.

Keywords: vocal control, sensorimotor control, frequency-altered feedback, singing training, visual feedback

Dedication

This thesis is dedicated in loving memory of my two grandparents who passed away in 2014, while I was completing my degree. The two angels who have always supported my endeavours and without a shadow of a doubt watched over me and strengthened me to come thus far:

Refka Soliman

&

Nagy Bibawy

Acknowledgments

Firstly, I would like to thank my supervisor, Dr. Jeffery Jones, for his guidance and mentorship throughout the course of my degree. After every one of our meetings, I left with more confidence, and a mission, two components key to the successful completion of my degree. I also thank the entire faculty on my defence committee for evaluating this thesis: Dr. Olds, Dr. Hockley, and Dr. McCarthy, as well as my class professors for teaching me not only knowledge but some good research skills. Thanks to my lab mates Asiya, Nikki, Zahra, Dani, and Lindsay whom I had the pleasure to work with. Obviously, a special mention goes to my fellow graduate students Omar and Anaya who shared this journey alongside me daily. I wish you all the best in your future endeavours and I hope to stay friends.

I would also like to acknowledge my friends, old and new who have kept in touch and supported me through this journey: my dearest friends of Waterloo The Fam and the Students and Graduates of St. Mary’s Coptic Orthodox Church. Furthermore, I’d like to thank the kids, the servants, and the fathers at St. Mary’s Coptic Orthodox Church in Kitchener for taking me in and supporting me over the course of these last two years; keeping me involved and engaged in the community and granting me the privilege to feel as though I have a gained a whole new family. Also, thanks to the servants and the fathers at St. Mary & St. Joseph Coptic Orthodox Church in Richmond Hill for encouraging me to finish and return home whenever I could.

I must also express my gratitude to my extended family, near and far, who video called me all the time to keep up with me and offered their moral support. And of course, where would I be without the most important people in my life? To my parents, Nagy and Ninette, my number one fans. Thank you for pushing me, believing in me, and supporting me, even though it killed you that I lived far away. To my almost twin brother Philopateer, words cannot express

how much I enjoyed and appreciated your visits to Waterloo with me. Even though we were far, I’m so glad that my time away got us to bond and become much closer siblings. I wish you all the joy, adventure and success in your life and hope that we will remain best friends.

Finally, but most importantly, I thank You, God, for this opportunity You granted to me. You have covered me, helped me, guarded me, accepted me, spared me, supported me, and brought me to this hour. Although there were many mountains to climb, You did not abandon me for a moment but rather strengthened, and held my hand along every step of the way. Let my will always follow Your will and be in accord with it in a perfect way. For You are my salvation, my physician, my strength, my comfort, and my happiness.

Table of Contents

List of Tables

List of Figures

List of Acronyms

AAF	Altered Auditory Feedback
CP	Critical Period
DIVA	Directions Into Velocities of Articulators
F_0	Fundamental Frequency
FAF	Frequency-Altered Feedback
KR	Knowledge of Results
RTVF	Real-Time Visual Feedback
PSR	Pitch Shift Reflex

General Introduction

From their first breath of air, most humans are able to communicate and express themselves using their personalised instrument: their voice. To produce a vocalisation, air from the lungs passes through the trachea and vibrates two vocal folds (also known as vocal cords) in the larynx. The sound produced by these vibrations resonates when passing through the throat, mouth and nose and transforms into unique sounds. All the varying sounds produced by the voice, including cries, laughter, conversation, and song, require varying degrees of control over the various mechanisms of the vocal system (Welch, 2005).

Pitch, an integral feature of the sound of the voice, is understood as the sum of the rate of the different vibrations of the vocal cords. As the frequency, the rate of these vibrations, increases, pitch of the voice is perceived as higher in frequency. The rate of these vibrations depends on the size (length, width and thickness) of the vocal cords, the tension of the muscles controlling them, and the airstream passing through them. Thus, controlling vocal pitch is no small feat, and yet it is only one of the many things the vocal control system is designed to accomplish (Guenther, 2006). As an individual normally develops, their vocal control system develops as well and their vocal pitch fine-tunes and becomes more precise (Guenther, 2006).

Good singing is characterised by heightened vocal pitch control, attained after undergoing some sort of training. Singing training most often follows the expert-apprentice model, where the vocal teacher gives the student instructional feedback to improve their voice during in-person sessions (Welch, 1985a). As with other professions, this aspect of the instructor's role could possibly be replaced by the more objective feedback from a computer. With technological advancements and high-performing computers made accessible, many intricate visual training computer programs have been developed and promoted to enhance vocal pitch control (Hoppe, Sadakara, & Desain, 2006; Welch, 1985a). At the inception of these training programs, the knowledge about the vocal control system was not as detailed, and was not supported by as much evidence as there is present today. Thus, the purpose of this thesis is to connect the recent body of literature concerning the cognitive processing of vocal control, and apply this knowledge practically to develop an effective training program that increases vocal control of pitch in singing.

The experiments examined in this thesis aimed to measure how real-time visual feedback training impacts the vocal control system using several measures. The training program designed for these experiments contained a unique combination of visual indicators as feedback, unlike the studies before it, and tested vocal control of pitch in singing using a combination of measures. Experiment 1 tested for changes in vocal control in non-singers compared to singers, immediately after a training session. A longer training session was tested with another group of non-singers in Experiment 2. However, there were no observed improvements in vocal control regardless of the additional training implemented.

A review of the literature relevant to these experiments is presented in the form of a three-chapter introduction. First, Chapter 1 discusses the training of singing skill by exploring vocal pedagogy and then focuses on the use of real-time visual feedback training programs tested in previous literature. In order to understand the vocal control system, Chapter 2 explores the cognitive processes underlying the vocal control model that dominates the current literature. Finally, Chapter 3 discusses the theories and measures that have been used to test vocal control specifically exerted during the act of singing. After reviewing previous studies, a new real-time visual feedback training program was designed to be more effective at improving vocal control and tested in the two studies presented in this thesis.

Chapter 1: Singing Training

As with any form of education, there are many schools of singing training and they date back centuries. The most prominent and distinct vocal schools for classical singing in the West include the Italian, French, German and English (Miller, 1970), which differ in exercises, techniques, and priorities. However, in all of these schools, singing training follows the master-apprentice model, with weekly lessons supported by private practice and performance. This teaching model requires an expert instructor who usually can demonstrate the skill and also give feedback to guide their students in the acquisition of the skill (Callaghan, 2000). Welch (1985a) formulated a model to portray the traditional interaction between a teacher and students (Figure 1A). Typically, vocal instructors are performers themselves and teach the student using scaffolding methods based on their own experiences and perceptual abilities (Kennel, 2013). They provide feedback on the student's voice, also referred to as Knowledge of Results (KR; Welch, 1985a). KR is external feedback that needs to be meaningful, in order to guide the singer's error-labeling schema. Thus, within one lesson, the singer is required to consolidate internal feedback from within the body with the external feedback provided by their instructor in order to improve vocalisations (Welch, Howard, Himonides, & Brereton, 2005).

Researchers have looked into contemporary training techniques drawn from the classical schools and tested in children's music classrooms. Kramer (1985) found that a speech-to-song approach, which was centered on a comfortable "personal note," was effective in increasing pitch-matching ability in middle school children. Maintaining singers' confidence by rewarding gradual improvement has also been found to make a difference in their performance (Dennis, 1975). Furthermore, vocal instruction that reinforces visual and kinaesthetic representations of pitch has also led to better pitch-matching abilities in children (Apfelstadt, 1984). These

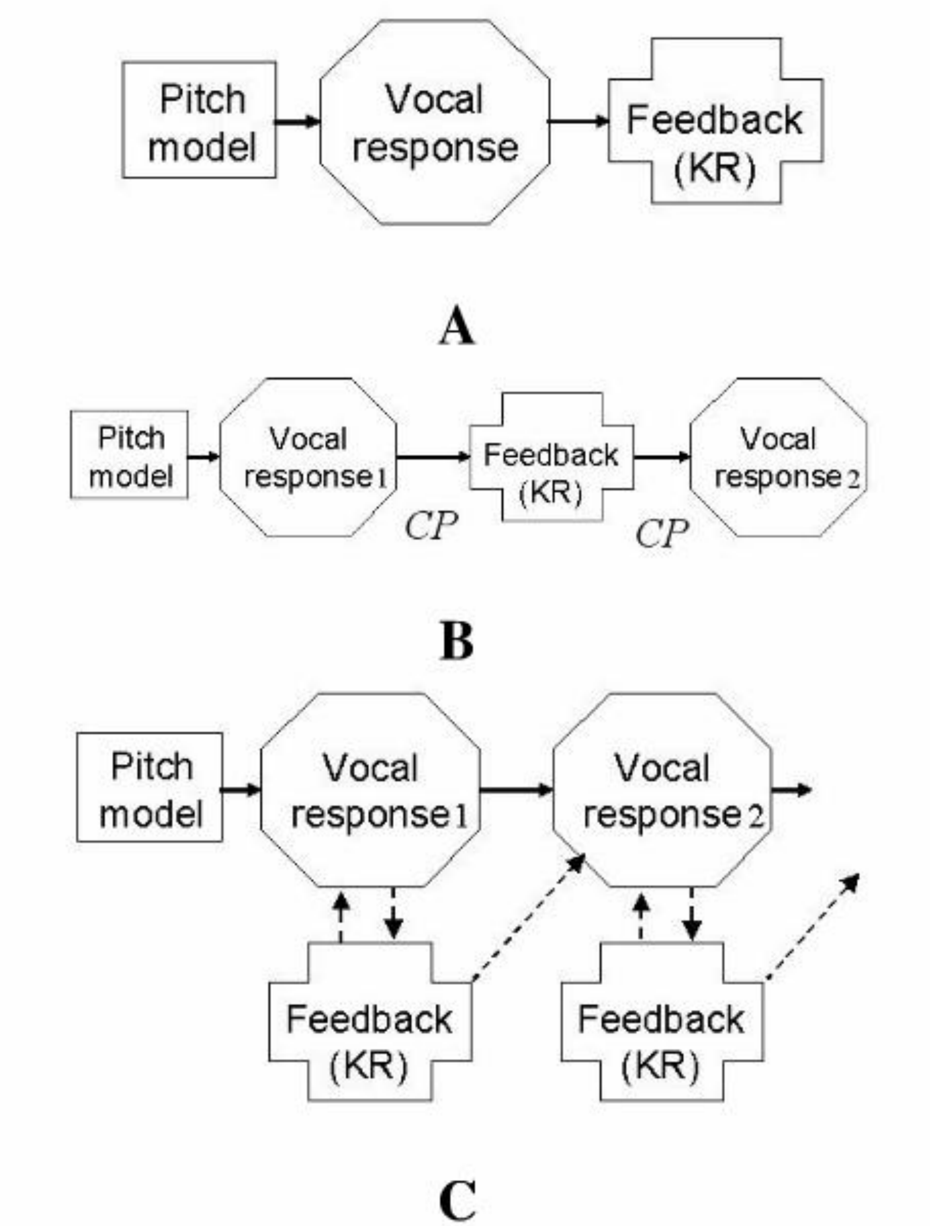

Figure 1. An illustration of the learning process for pitch in singing training based on Welch (1985) taken from Howard and his colleagues (2003). Time is from left to right in these diagrams. (A) A model of the traditional interaction between a student and their instructor; (B) the on-going learning process during a singing lesson; (C) the way in which real-time visual feedback can impact the learning process. KR = knowledge of results from an external source; CP = critical period for learning to occur.

techniques have led to an increased singing ability, however Welch (1985a) suggests that the KR provided by instructors in the traditional model is inefficient and lengthens the learning process for students. He critiques the conventional singing training model on two points: the quality of the feedback given by teachers and its timing.

Limited by the boundaries of language, many teachers attempt to describe the perceptual and production aspects of the voice to students. This is no simple task, as teachers frequently rely on auditory imagery and metaphoric language to translate their perception of the student's performance; which, in turn, the student must translate the verbal feedback into perceptual feedback (Welch, 1985b). Teachers' comments can be ambiguous and at worst frustrating for the student as it may be difficult to disassociate the identification of the instrument and themselves as performers (Callaghan, Thorpe, & van Doorn, 2004). Furthermore, the time delay between the KR provided by the teacher, defined as the critical learning period, is quite significant in comparison to perceptual and kinaesthetic feedback designed in the vocal control system of the student (Welch, 1985b). Thus, after the vocalisation, the student is required to accurately recall a detailed memory of their performance, interpret this feedback given to them and modify their motor plans for the next response as shown in Figure 1B. In an attempt to resolve these two weaknesses of the traditional singing training model, Welch (1985b) proposed that real-time visual feedback (RTVF) could impact the learning process.

Real-Time Visual Feedback

Modifying the traditional singing training model, Welch (1985b) suggested that the use of RTVF is advantageous in removing the time lag between a student's vocal response and their teacher's feedback (Figure 1C). This not only enables motor modifications to be made immediately, but it allows for further analysis of any effect caused by those modifications. The other advantage that RTVF addresses the ambiguity of feedback traditionally provided by instructors, as students are able to receive objective, quantitative information from a visual display (Welch, 1985b). Furthermore, feedback and motor skill learning literature indicates that it is more valuable for the learning process when one focuses attention externally to the consequences of one's movements rather than focusing internally on the movements (Wulf & Prinz, 2001). RTVF directs the singer's attention to the visual display of their auditory output rather than attention to the specific movements of the vocal tract (Hoppe et al., 2006). However, the concept of real-time visual feedback in singing training was not a novel one; it was previously tested by Seashore and Jenner (1910).

In an attempt to explore the use of an aid to shorten vocal training periods and to increase the effectiveness of the ear, participants were tested over the course of twelve days for forty-five minutes using a voice tonoscope (an instrument which converted sound vibrations into visual representation of pitch on a scale). Seashore and Jenner (1910) found improvement in vocal pitch-matching while participants sang using the aid and that transferred to after the aid was removed and they sang without it. Since that first experiment, many technological developments have allowed for better experimental designs to explore singing training techniques using RTVF. For instance, Howard and Welch (1989) compared children's pitch-matching ability using an oscilloscope screen called SINGAD, which plotted F_0 of the vocalisations. They found that although visual feedback facilitated accurate pitch production compared to no visual feedback, there was an additive benefit to the accuracy of vocalisations when KR was provided as a target pitch on the display. In the age of computers, not only did hardware improve, but also programming advancements resulted in endless options for RTVF interfaces: four of which were reviewed (Hoppe et al., 2006).

In this review, researchers examined the following programs: SINGAD (Howard & Welch, 1989), ALBERT (Rossiter & Howard, 1996), SING & SEE (Callaghan et al., 2004), and VOXed: WinSINGAD (Welch, Himonides, & Howard, 2004). In general, they all commonly include plots of F_0 against time, although these RTVF programs have improved and are now multifaceted with customisable functions for users (Hoppe et al., 2006). The program with the widest range of visual display features was WinSINGAD (the successor of SINGAD) with up to eight different parameters, including a side view webcam to examine posture (Welch et al., 2004; Hoppe et al., 2006). Although the use of some of these programs without supervision has resulted in improvements in pitch, results indicate that the most improvement occurs when teachers are included to guide the learning process (Welch, Howard, & Rush, 1989; Callaghan et al., 2004). Therefore, the information provided in RTVF itself may not be as useful if students do not properly understand it or know how to use it.

Wilson and her colleagues (2005) wanted to investigate whether the amount of information provided in RTVF, and the experience of the user, had an effect on the ability to sing in tune. Fifty-six participants with different singing skills were assigned to one of three conditions; one condition had a keyboard display with binary (right or wrong) feedback, another condition had a pitch display with detailed information, and finally the control condition, which was just a keyboard display with no feedback. Participants were tested before the training session, while using the RTVF, as well as after using it, and pitch error (the difference between what was sung and the target note) was calculated. Wilson et al. (2005) found that when comparing pre- and post-test performance, those with either RTVF displays (binary or detailed KR) improved more compared to the control group representing the effect of practice. Furthermore, the beginner singers seemed to benefit more from the detailed pitch graph display than from the binary keyboard display, and the opposite was the case for more experienced singers (Wilson, Lee, Callaghan, & Thorpe, 2005). However, characterising the differences between singers and non-singers in this study is not possible because singers were given more difficult pitch-matching exercises than non-singers during the test phases. Changing the level of difficulty of tasks between groups does not allow for objective comparisons.

Regardless of the improvements found between the pre-test and the post-test, Wilson and her colleagues (2005) found an overall difference between accuracy measures taken during training in the RTVF conditions compared to the no-feedback condition. The accuracy performance over the course of the training phase, however was different between singers and non-singers. Due to singers' already acquired vocal internal reference, they were the least inaccurate during training in the control condition. Singers who trained any RTVF display actually resulted in more vocal inaccuracy than without. This was the opposite for non-singers: they were the most inaccurate during the training in the control condition. Non-singers who trained with any RTVF display actually resulted in more vocal accuracy than training without (Wilson et al., 2005). All of the current studies examining RTVF in singing training have only been concerned with pitch-matching accuracy abilities, however that is only one of the many tasks the voice can do. To understand the underlying ingredients that have made RTVF training effective, the mechanics and cognitive processes underlying vocal control must be discussed.

Chapter 2: Vocal Control System

Vocalisations are produced by controlled actions of the respiratory system, the larynx, and all the structures of the vocal tract. These systems are complex with each component having its own configuration and function in speech. Contractions of over 50 tiny muscles are responsible for the movements of these structures, which result in the production of desired sounds with high precision and accuracy (Perkell, 2012). Researchers have developed theories to explain vocal control with evidence from cognitive, and neurophysiological data. Currently, the work of Guenther, Ghosh, and Tourville (2006) is very prominent in the literature to date. They developed the Directions Into Velocities Of Articulators (DIVA) Model, mapping out the neuronal network of the speech control system.

Vocal Control Model: DIVA

The DIVA model is a neural and computational model that maps out speech acquisition. Figure 2 is a schematic representation of the DIVA model. The boxes in the diagram each represent large structured neural networks with specific anatomical correlates based on previous neuroimaging and electrophysiological work (Guenther, 1994, 1995, 2006; Guenther et al., 2006). The model essentially posits that when a speaker wants to produce a specific sound, there are two mechanisms at work together: the feedforward loop and the feedback loop.

The feedforward system is driven by representations, which are detailed instructions stored as the speech sound map in the premotor cortex. This speech sound map contains previously acquired information about the relationships between the motor commands, the environment, and sensory feedback for the specific desired vocal production. When the feedforward loop is initiated, the brain selects and implements a speech sound map given the information available about the current condition of the voice. These instructions are sent to the articulator velocity and position maps which direct the articulators for the vocal production (Guenther et al., 2006). Thus, the initial vocalisation, which takes place between 0-100 ms, is attributed to open-loop control which does not rely on sensorimotor feedback (Burnett, Freedland, Larson, & Hain, 1998; Burnett, McCurdy, & Bright, 2008; Hain et al., 2000; Larson, Altman, Liu, & Hain, 2008; Patel et al., 2013).

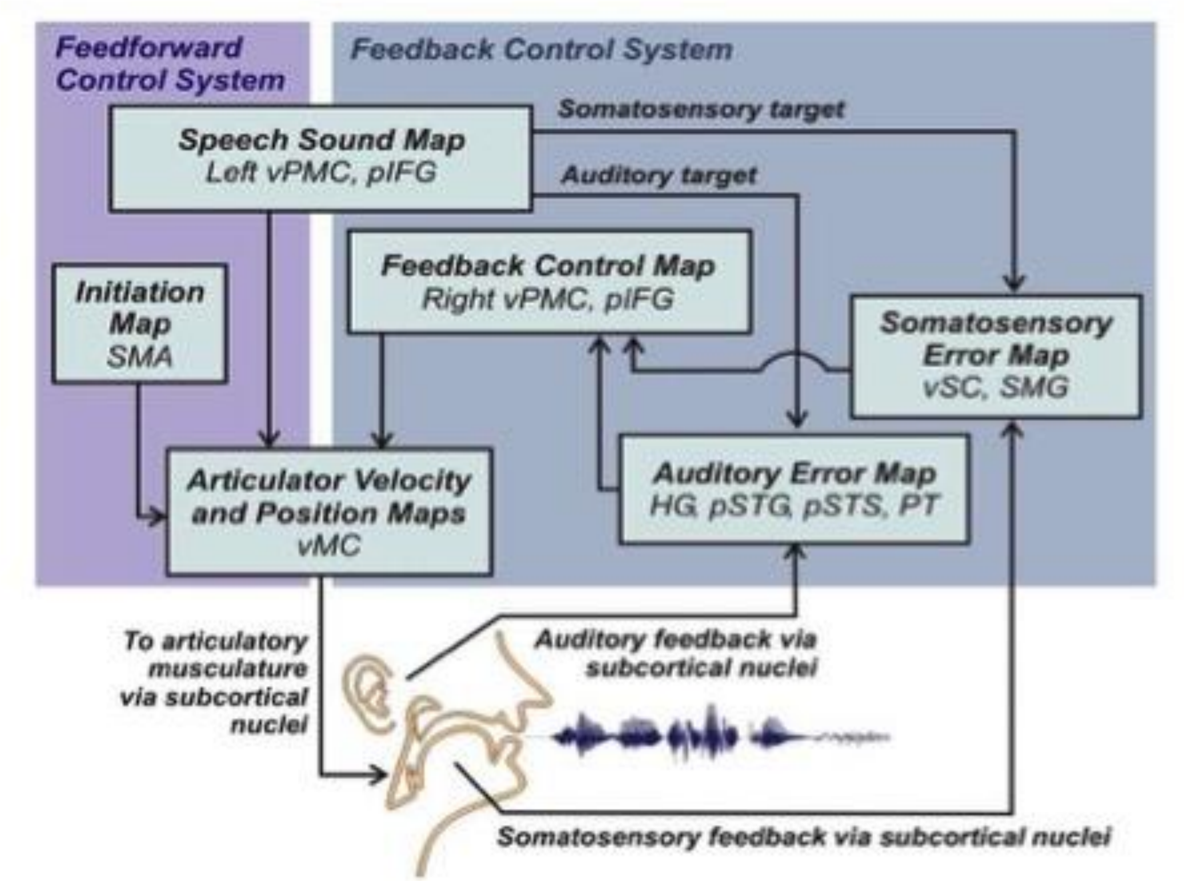

Figure 2. A schematic of the DIVA model taken from Guenther et al. (2006).

The feedback system relies on somatosensory and auditory feedback to detect production errors and correct them. During vocal production, the online auditory and somatosensory conditions, available through sensory feedback, are compared to learned targets previously initialized by the speech sound map. When the current and target sensory conditions match, no error signal arises. Otherwise, when the current and target sensory conditions do not match, an error signal arises in the error maps. These error signals guide the articulator velocity and position maps to appropriately correct the articulators (Guenther et al., 2006). The resulting compensatory responses, usually occurring between 150-250 ms after error detection, are the outcome of closed-loop control (Burnett et al., 1998; Hain et al., 2000; Hawco & Jones, 2009; Larson et al., 2008; Patel et al., 2013).

During the acquisition of speech, the feedback system provides a significant contribution to the vocalisation production. Over the course of development, the person experiences and learns different possible combinations of the different components of the sound of speech. Thus with practice, the speech sound map refines itself by monitoring the corrective motor commands sent from the feedback system and storing them for future use. By strengthening the speech sound map, the feedforward control becomes more dependable over time. Furthermore, with more consistent accurate productions, significant error signals in the feedback system become less frequent and thus, the role of feedback becomes less critical to the speech production process (Guenther et al., 2006). This change in weighting of the feedforward and feedback systems demonstrates the brain's plastic properties that are exploited by training methods.

Testing Vocal Control

One aspect of vocal control includes the ability to correct any errors of vocal production perceived through auditory feedback. Thus, when feedback does not match the desired production, the vocal control system adjusts in order to correct production. In order to test the relationship between auditory feedback and vocal-motor control, numerous speech studies have simulated vocal errors using altered auditory feedback (AAF) in an experimental setting (Elman, 1981; Burnett et al., 1998, 2008; Hain et al., 2000; Jones, & Munhall, 2002; Pfordresher & Mantell, 2014). AAF experiments typically involve participants vocalising into a microphone while they simultaneously receive AAF through headphones. Different experiments alter different components of speech such as timing (e.g. Howell & Sackin, 2002; Pfordresher & Palmer, 2002), loudness (e.g. Bauer, Mittal, Larson, & Hain, 2006; Heinks-Maldonaldo & Houde, 2005), formant frequency (e.g. Houde & Jordan, 1998; Purcell & Munhall, 2006; Tourville, Reilly, & Guenther, 2008) and fundamental frequency (F_0; e.g. Burnett et al., 1998, 2008; Elman, 1981; Jones & Munhall 2000, 2002, 2005; Larson et al., 2008; Scheerer & Jones 2012, 2014). Each of these AAF manipulations in the laboratory setting helps reveal how the vocal control system adjusts vocal production in different compensatory responses specific to the different manipulation applied to the auditory feedback. The AAF paradigm utilised in this thesis concerned the fundamental frequency of the voice.

Frequency-altered feedback. The human voice produces very complex sounds: vibrations of different frequencies at once. When these different frequencies are summated, they are perceived as ‘pitch’ by the brain (Titze & Martin, 1998). Of these frequencies, the vibration with the lowest frequency in the sound is known as the fundamental frequency (F_0). The frequency-altered feedback (FAF) paradigm has been found to elicit a reflex-like compensation; also known as the pitch-shift reflex (PSR; Burnett et al., 1998). As participants hear the F_0 of their voice shifted in one direction (up or down), they perceive this shift as a production error, and the corrective commands are sent to change the production. This results in an unintentionally produced shift of their F_0 in the opposite direction (down or up) in an attempt to compensate for the perceived error in their feedback. Once the shift is removed, participants are able to hear their unaltered feedback, and their pitch changes once more returning back to the pitch they were producing prior to the manipulation in feedback. Many different researchers have examined the PSR with two basic variations in paradigm: a FAF-perturbation paradigm (Burnett et al., 1998; Bauer, & Larson, 2003) and an FAF-adaptation paradigm (Jones & Munhall, 2000; 2002; 2005).

The perturbation paradigm is a short shift in the auditory feedback of the participant's F_0 over the course of one vocalisation. Thus, the participant begins to vocalise while hearing their unaltered feedback, and some time after the voice onset, they receive FAF for a short period of time (e.g. 200 ms). Then their auditory feedback returns to normal all before they complete their vocalisation. Due to its short length, a few perturbations can be presented in one vocalisation and still evoke the PSR reliably (Burnett et al., 1998). Accordingly, the FAF-perturbation paradigm allows researchers to examine the role of the feedback loop in vocal control by measuring the magnitude and latency of the reflexive response, and the timing of its occurrence.

In a repeated measures study conducted by Liu and Larson (2007), participants' vocal compensation responses were tested across different magnitudes of shifts for two different notes. Before each vocalisation, a high or a low target piano tone was presented and participants were instructed to match the note. Vocalisations were perturbed five times randomly, upward, downward or they were entirely unaltered. The perturbations were 200 ms in length and varied in magnitudes of 0, ±10, ±20, ±30, ±40, and ±50 cents (where 100 cents = 1 semitone). Responses to perturbations increased in magnitude as the shift magnitude increased.

Furthermore, the latency of the compensation response was found to decrease with the magnitude of the shift (Liu & Larson, 2007). These and other findings attest to the sensitivity of the feedback loop, as increasing compensation responses are elicited faster with increasing deviations of altered feedback (Burnett et al., 1998; Larson et al., 2008; Scheerer, Behich, Liu, & Jones, 2013).

The FAF-adaptation paradigm is used to investigate how feedback contributes to sensorimotor learning seen as the modification of the representations initiated by the feedforward control. This paradigm usually consists of three phases with multiple vocalisations in each one: the baseline phase, the shifted phase and the test phase. During the baseline phase, participants are asked to vocalise a few times while receiving unaltered auditory feedback of their voice. During the shifted phase, auditory feedback of the participant's F_0 is altered from the onset of the vocalisation to the end of it (deemed a full-utterance shift). Finally, during the test phase, participants received unaltered feedback once again as it was during the baseline phase. The difference between the F_0 produced during the baseline phase compared to that during the test phase represents any after-effects of prolonged exposure to altered feedback; a result of adaptation (Hawco, & Jones, 2010; Jones, & Munhall, 2000, 2002, 2005; Keough, Hawco, & Jones 2013; Keough & Jones, 2009). With respect to the DIVA model, adaptation is interpreted as an attempt by the vocal control system to reduce the consistent error signals triggered by incorrect feedback. In order to subsequently produce the correct vocalisation, a recalibration of the representation initiated by the feedforward loop is necessary (Guenther & Vladusich, 2012).

An FAF-adaptation study conducted by Hawco and Jones (2010) tested for multiple instances of adaptation within a single experimental session. The experiment used five different target notes in two blocks each, one shifted in frequency in the upward direction and another shifted in the downward direction for a total of 10 blocks. At the beginning of each trial, participants were presented with a target pitch recorded by trained singers then asked to produce two-second vocalisations. While vocalising, participants received unaltered feedback for the baseline and test trials, but for the middle shifted trials received shifted feedback by 100 cents. The analysis of the results indicated differences in F_0 production between the end of the baseline trials and the first few test trials after altered feedback was removed. This pointed to the conclusion that the sensorimotor mapping of the target F_0 had only required approximately 20 trials of altered feedback to modify previously learned representations of the notes. Although the FAF-adaptation paradigm may have evoked quick learning, it was not sustainable. By the end of the test trials, the same sensorimotor map of the target F_0 returned to the baseline pre-adaptation state (Hawco & Jones, 2010). It is possible that with more time, and practice, modified sensorimotor mapping can be learned and stored more permanently.

Other measures of vocal control. Other than the PSR and its measure of compensation magnitude and latency, two other measures of vocal control have been discussed in the literature: pitch accuracy and vocal variability. In a study conducted by Scheerer and Jones (2012), participants were asked to match 3 different target notes while being exposed to FAF-perturbations. Researchers were interested in the relationship between compensation, vocal variability, and accuracy at matching the notes. They measured accuracy as the deviation from the target note in cents, and vocal variability as the standard deviation of the F_0 produced. The results indicated that there was no correlation between compensation magnitude to FAF and pitch accuracy for producing the target notes. However, there was a positive correlation between vocal variability and compensation magnitude. The researchers suggested that this correlation supports current vocal control models, such that participants with more variable vocal productions depend on their feedback system a lot more than those who are not as variable. Those who have low variability in their productions vocalize more consistently and thus depend more on the feedforward control, which is related to lower compensation responses to FAF (Scheerer & Jones, 2012). Although there is no cumulative measure of vocal control to date, corrective compensation responses to FAF, in magnitude and latency, as well as vocal variability and pitch accuracy have all been used separately as indicators of vocal control.

Speech versus Song

Speech and song have been present in every society, irrespective of generation or location (Tsang, Friendly, & Trainor, 2011). From an evolutionary perspective, it is still unclear whether humans developed speech or song first (Titze & Martin, 1998). While one of the features distinguishing humans from animals is their development of language as a means of communication, singing is common between them. Moreover, birds and whales have been found to compose and improvise song as well as humans (Wallin & Merker, 2001).

From a developmental perspective, speech and song naturally emerge concurrently as they are two vocal behaviours with shared characteristics (but also differ in other characteristics; Welch, 2005). Their parallel emergence is considered possible due to the most obvious similarity between the two processes: they share common physical mechanical effectors, such as the throat, the larynx and the vocal cords (Sundberg, 2001). Acoustically, they show close patterns of pitch, stress, and rhythm. However, when analysing the acoustic differences between speech and song, spectrographic patterns show much more complexity when words are sung compared to when they are spoken (Sundberg, 2007). At the neurophysiological level, there is much debate about the overlap and different networks used in speech and song processing (Merrill, 2013) as well as production (Christiner & Reiter, 2013). Having stated this, behavioural and neural studies do show that musical training, whether it is vocal (Siupsinskiene & Lycke, 2011) or non-vocal (Stegemöller, Skoe, Nicol, Warrier, & Kraus, 2008), is directly advantageous for speech processing; supporting the notion of shared neural networks (Hutchins & Moreno, 2013, Özdemir, Norton, & Schlaug, 2006).

The role of fundamental frequency. In speech, F_0 has lexical and syntactic functions, though it is also important for the expression of affect and interpretation of other non-verbal cues (Elman, 1981). The function of F_0 is different in tonal languages such as Mandarin, compared to non-tonal languages such as English. In English, pitch within a syllable is not crucial to comprehension, and thus it is not necessary to tightly control F_0 when speaking (Natke, Donath, & Kaleveram, 2003). In contrast, to differentiate between words and grammatical categories, tonal languages require the speaker to aim for a relative target pitch allocated to a meaning (Jones & Munhall, 2002). Previous studies show evidence that when compared to non-tonal language control, tonal language speakers perceive musical pitches more accurately (Giuliano, Pfordresher, Stanley, Narayana, & Wicha, 2011) and also produce pitch more accurately when singing (Pfordresher & Brown, 2009).

In parallel to tonal language, accurate F_0 production is preferred in singing. Accurate singing is characterised by matching specific external pitches corresponding to musical notes. Therefore, deviations between the external reference F_0 and the personal voice F_0 need to be recognised and compensated for (Natke et al., 2003). A study by Natke and his colleagues (2003) investigated the differences between F_0 in speech and song in 24 non-tonal language speakers. Participants were asked to vocalise a nonsense word with a target rhythmic rate in the speaking condition, and a target piano pitch in the singing condition. While vocalising, participants received FAF and compensated between the two conditions differently. Results indicated that participants did not fully compensate for the 100-cent shift, but rather only compensated by an average of 47 cents in the speaking condition and 66 cents while singing. Therefore, researchers concluded that tighter control of F_0 is required when singing and that the accuracy of production influences amount of vocal compensation to perceived error (Natke et al., 2003).

Chapter 2: Vocal Control when Singing

Singing is usually considered a talent for the select few. Many people believe that without formal training or musical education, the inability to carry a tune is widespread (Pfordresher, Brown, Meier, Belyk, & Liotti, 2010). However, singing is natural for humans, as a universal form of vocal expression of affect, regardless of culture (Wallin & Merker, 2001). When singing is done with others, it is associated with a highly pleasurable experience, it promotes group cohesion, and it is therapeutic and used in many rehabilitation programs (Tsang et al., 2011). Singing emerges naturally through development and is important in viewing oneself as a musical being (Demorest & Pfordresher, 2015; Welch, 2005). Proficiency is usually determined by pitch accuracy, and contrary to popular belief, singing proficiency is not an attribute of a selected few but rather, singing proficiency is normally distributed in the general population (Dalla Bella, Giguère, & Peretz, 2007).

At the coarsest level of categorisation, individuals are divided into a dichotomy of singers and non-singers based on their vocal control. Over the years, researchers have divided each category further (Watts, Murphy, & Barns-Borroughs, 2003). Singers have been labeled as trained singers, talented singers, untrained talented singers, and accurate singers (Watts, Moore, & McCaghren, 2005, Watts et al., 2003). Non-singers have been sometimes specified as untrained non-talented singers, uncertain singers, imprecise singers, poor-pitched singers, monotones, and inaccurate singers (Pfordresher & Brown, 2007; Pfordresher & Mantell, 2014; Watts et al., 2005). These categories have been used in studies that try to identify the crucial variables contributing to accurate and inaccurate singing.

Vocal Control in Singing

Unlike the established DIVA model supported by plenty of evidence, there are few well-developed models for vocal control specific to singing (Granot, Israel-Kolatt, Gilboam & Kolatt, 2013; Hutchins & Moreno, 2013; Pfordresher & Mantell, 2014). The scholars who study singing organised a symposium to combine the evidence in the current literature. Their most recent efforts resulted in a mechanics of singing accuracy model Figure 3 (Pfordresher et al., 2015). This model outlines three functional representations related to the event of a vocal production. The first is a perceptual representation where the pitch, timbre and other quantitative information about the feedback from the external input are processed. Second, there is the categorical representation where the qualitative information about the feedback is processed and finally, third is a motor representation, which involves the articulator controls associated with the sound perceived. These representations are similar to those used in the DIVA model. Through either of the two simultaneous loops shown in Figure 3, the perceptual representation, can be translated or converted into another representation, such as the motor representation (also proposed in Linked Dual Representations theory, Hutchins & Moreno, 2013). All three representations are coupled together and become stored into memory as a vocal production event (Pfordresher et al., 2015).

The lower half of Figure 3 is referred to as the sensorimotor loop and is investigated in vocal imitation tasks. When imitating pitch, the initial target pitch is heard and a low-level perceptual representation is formed. That representation is translated into a motor plan, which is then executed and provides auditory feedback processed at the perceptual level (Pfordresher et al., 2015). Simultaneous to the vocal-motor translation, a categorical representation of pitch is

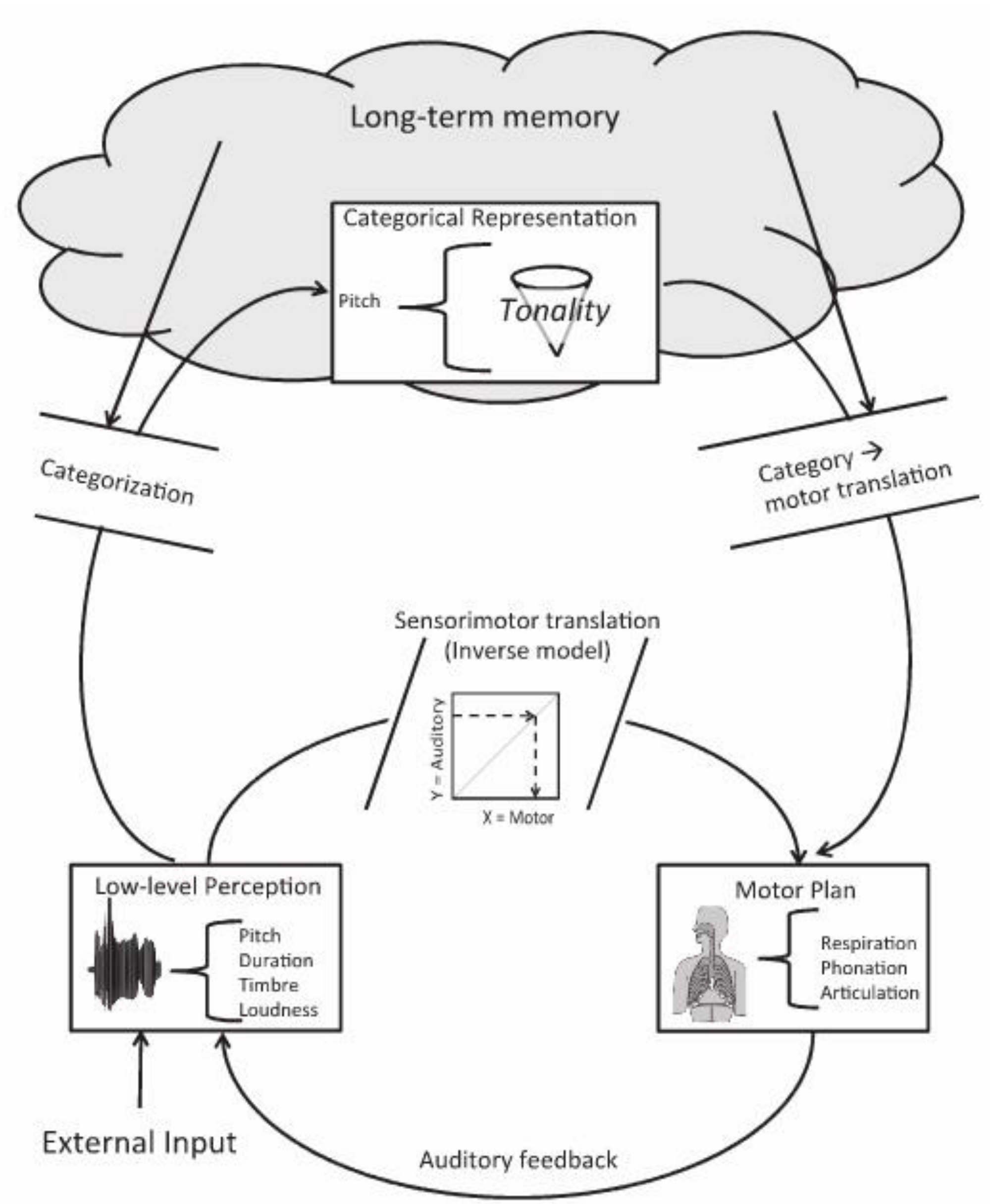

Figure 3. The functional architecture underlying the mechanics of singing accuracy proposed by Pfordresher and colleagues (2015).

formed and then translated once more into motor commands. With practice and time, categorical representations are learned and stored in long-term memory (Krumhansl, 1979). It has been hypothesised that memory helps guide the translations to and from the categorical representation (Pfordresher et al., 2015). Therefore, this symbolic loop allows for the direct categorical representation of a note to be retrieved from memory, translated into motor commands, and produced vocally. This singing accuracy model, in fact, results from numerous experiments examining inaccurate vocal production as a result of poor perception, and poor translation abilities (Ayotte, Peretz, & Hyde, 2002; Berkowska & Dalla Bella, 2009; Dalla Bella, Berkowska, & Sowiński, 2011; Dalla Bella et al., 2007, 2009; Hutchins & Moreno, 2013; Hutchins & Peretz, 2012; Peretz & Colheart, 2003; Pfordresher & Brown, 2007; Pfordresher et al., 2010; Pfordresher & Mantell, 2014; Welch, 1979).

Perception. Amusia is thought to be a neurogenetic disorder resulting in impaired ability to consciously detect and produce differences in pitches (Hutchins, Zarate, Zatorre, & Peretz, 2010; Peretz et al., 2008). People with amusia are not able to consciously discriminate pitch, even though studies show pitch-discrimination at the electrophysiological and neurophysiological level (Peretz et al., 2008; Zendel, Lagrois, Robitaille, & Peretz, 2015). Not surprisingly, they have a poor singing ability and have difficulty matching a pitch using their voice (Hutchins et al., 2010). In the amusia literature, however, there seems to be a few amusics found to have unimpaired vocal pitch-matching abilities despite their perceptual deficiencies (Dalla Bella et al., 2009; Hutchins et al., 2010). Thus, researchers cannot definitively conclude that their perceptual deficits cause impaired singing ability (Ayotte et al., 2002; Dalla Bella, et al., 2009).

Although perceptual deficits may be linked to most amusics' poor pitch-matching abilities, it is clear that accurately pitched singing depends on more than just the proper perception of pitch. One of the innovative experiments conducted by Hutchins and Peretz (2012) involved testing the perceptual ability of musicians and non-musicians using a visual representation of pitch in the form of an adjustable slider on the screen. Participants were presented with a target instrumental tone and were asked to move a pitch slider to match the pitch of the tone previously heard. In a later task, participants were asked to match the pitch of the tones they heard vocally. As long as the slider was in motion, or the participant was vocalising, the target tone was removed in order to prevent pitch matching through hearing the dissonance between the target and produced pitches. Participants were more accurate at matching the pitch with a slider (at ceiling) than with their voice (Hutchins & Peretz, 2012). Other studies support the conclusion that inaccurate singing cannot be solely attributed to the inability to perceive and discriminate between pitches (Bradshaw & McHenry, 2005; Dalla Bella et al., 2007; Lévêque, Giovanni, & Schön, 2012; Pfordresher & Brown, 2007).

Translation. Poor-pitch singing, described by inaccurate production, has been suggested to result from deficits in the connections between the different internal representations described in the mechanics of singing accuracy model (Pfordresher et al., 2015). Poor-pitch singers may acquire both accurate low-level perceptual representations and accurate motor plans, yet they may have faulty internalized rules that link them. As poor-pitched singers produce consistent inaccuracies while vocalising, researchers suggest that there is a possible deficit in the translations of representations which occur in the sensorimotor loop (Hutchins & Peretz, 2012; Pfordresher & Brown, 2007). In two experiments conducted by Pfordresher and Brown (2007), participants were asked to vocally imitate several target single notes, intervals, and melodies.

Also, participants completed these tasks under three conditions of feedback: normal feedback (the participant only hearing their voice), augmented feedback (the participant heard an accompanying voice), and finally, masked feedback (the participant only heard noise). The researchers found that poor-pitched singers consistently vocalised inaccurately regardless of the pitch of the notes (whether they were high or low pitch), and regardless of interval between the pitches (whether the differences in pitch between the notes was large or small). Furthermore, with respect to feedback, poor-pitched participants performed worse than others when accompanied by a reference voice. Similarly, in the second experiment, which provided target notes within their vocal range, poor-pitched singers still produced inaccurate vocalisations. Vocal accuracy improved in the interval trials and even more so in the melody trials, indicating that the ability to imitate one-tone pitches specifically relies on single, absolute pitch representations (Pfordresher & Brown, 2007).

Another study conducted by Pfordresher and colleagues (2010) sought to test the effectiveness of the sensorimotor loop with unfamiliar sequences compared to the familiar ones stored in long-term memory. Similar to Pfordresher and Brown (2007), participants were asked to imitate several single notes, intervals and unfamiliar melodies. The researchers measured accuracy by taking the average difference between target pitches and the actual produced pitch. They also measured precision by using the standard error of the produced pitch irrespective of the target pitch (similar to the vocal variability measure used in Scheerer & Jones, 2012), representing consistency in the production. Researchers found that accuracy and precision were correlated, and further analysis suggested that accuracy predicted precision in unfamiliar sequences. Interestingly, the relationship between accuracy and precision was weaker for familiar sequences. Therefore, the authors concluded that, even though both measures represent aspects of sensorimotor translation, they are partially independent of one another. Inaccuracy represents the consistent incorrect link between perception and action, while imprecision is related to noise, or the variability in that link (Pfordresher et al., 2010). Both measures demonstrating a lack of vocal control.

These and other studies have concluded that poor singers are not able to properly convert the different representations of pitch, whether they are perceptual, motor, or categorical representations (Pfordresher & Beasley, 2014; Pfordresher & Brown, 2007; Pfordresher et al., 2010; Pfordresher et al., 2015). This can also be understood as poor singers lack the ability to accurately predict the outcomes of their vocalisations. With imprecise predictions, they have incorrect comparisons to their actual production and thus, incorrectly change their productions (Pfordresher & Brown, 2007; Pfordresher & Beasley, 2014).

The singing voice experiences developmental changes as vocal control improves with age and can also be enhanced through specific vocal training. With age, physical development progresses and affects the articulators of the vocal tract; the vocal control system adapts accordingly. For instance, boys experience a change in their vocal range during puberty (Harries, Griffin, Walker, & Hawkins, 1996). Also, during the acquisition of speech, the fine motor control of the articulators is learned and motor commands become more detailed and accurate to produce the intricate sounds of any language (Guenther et al., 2006). In a study examining the developmental trajectory of vocal control, Scheerer and her colleagues (2013) collected a sample of 100 English speaking participants and divided them into five different age groups. Statistical analysis provided evidence that vocal variability differed between the children (4-6) and adults (18-30). As expected, adults, with more experienced vocal control, displayed much less variability. The researchers hypothesized that this improvement with age reflects the diminishing dependency on the feedback and the increasing importance of the internal representations stored in memory for vocal production (Scheerer, Liu, & Jones, 2013).

These conclusions have been drawn from research conducted with non-singers, but research investigating trained singers has also contributed to a clearer picture of the level of vocal control required while singing. Singing training, like any other skill training, has been used to further enhance voice quality and production above norms (Hoppe, Sadakata, & Desain, 2006; Saitou & Goto, 2009; Siupsinskiene & Lycke, 2010; Smith, 1963; Stegemöller et al., 2008). Thus, it is singers' heightened perceptual sensitivity and integration of sensorimotor feedback, which has demonstrated their enhanced ability to translate between different internal representations of vocal production in order to accurately vocalise when singing.

Singers

Whether it is singing solo, in a choir, a cappella or with instrumental support, a key skill of singers is their ability to accurately and quickly control their F_0 with an accuracy of less than 1 Hz (Sundberg, 1987; Mürbe, Friedemann, Hofmann, & Sundberg, 2002). Grell and her colleagues (2009) conducted a study comparing highly and moderately skilled choral singers' responses to a change in a pitch reference. These researchers found that the more experience the singers had, the more their resistance in their responses; eliciting more delayed responses (227 ms) than the quicker responses of less experienced singers (206 ms). In an attempt to slow the less experienced singers' responses, their vocal cords were anesthetised. This inhibited the kinaesthetic feedback usually available during vocal production and it did, in fact, slow down their corrections to pitch error. These results are indicative of the differential ability, even among singers, in the processing speed required to detect and correct for perceived vocal production errors (Grell, Sundberg, Ternström, Ptok, & Altenmüller, 2009).

An FAF-adaptation study by Keough and Jones (2009) investigated the sensitivity of singers' established sensorimotor representations, by testing singers' and non-singers' ability to integrate feedback and adapt their vocalisations. Participants were instructed to match a musical target over the course of 210 trials divided into three blocks of 70 trials each. In one of the three blocks, the target note remained the same across all 70 trials. In the other two blocks, participants' target note remained the same during the shifted trials only while the baseline and test trials had a changed target note one whole tone (200 cents) above and below the target note. These blocks were used to test whether adaptation to the shifted trials transferred to other unaltered notes around the altered target note. The first 10 and last 10 trials of each block represented the baseline phase and test phase, respectively. Over the course of the 50 trials in between, participants' vocal frequencies were increasingly altered at increments of 2 cents all the way up to 100 cents. Participants performed the procedure twice, once with the feedback of their voice shifted upwards and once shifted downward on separate days. Researchers calculated the mean F_0 of the first 1500 ms of every trial to represent the compensatory response to the FAF. The results showed the heightened sensitivity of singers, who began to compensate after shifts of 6 cents as compared to non-singers who began to compensate after approximately 22 cents. Researchers also calculated the median of the first 50 ms of every vocalisation to measure the accuracy of the pitch at which participants initialised their productions. Results showed that singers, compared to non-singers, gradually and more accurately adjusted to the FAF manipulations by initialising their vocalisations at the F_0 they were producing in the preceding trial. Furthermore, when testing differences between baseline trials and test trials, researchers found aftereffects in singers as they incorporated the altered feedback into their internal representations. This effect generalised to other notes that were not actually altered during the experiment, meaning the representations of other pitches also changed relative to the newly adapted pitch. The results of this study display singers' proficient ability to translate perceptual information into internal representations and accurately adjust motor plans accordingly (Keough & Jones, 2009).

On the contrary, Jones and Keough (2008) also showed that singers are much more reluctant to incorporate feedback and translate it accurately into motor plans. Singers and non-singers were compared in a different FAF-adaptation paradigm where their feedback was shifted by 100 cents for 30 trials in between 10 baseline trials and 20 test trials. While at the baseline phase, there were no differences between singers and non-singers, singers and non-singers differed in the shifted phase. When provided with FAF of their whole utterances, singers did not compensate entirely for the shift by 100 cents, but rather, compensated significantly less than the non-singers. The authors theorised that this effect was attributed to singers' higher dependency on their feedforward loop control. However, regardless of their reluctance, it became evident that a full recalibration of the sound map occurred, and singers were unable to return their pitch back to the baseline once FAF was removed in the test phase. This finding shows further evidence that singers depend on their feedforward control using their stored internal representations as a more reliable source than their feedback (Jones & Keough, 2008).

When considering this evidence in the context of the vocal control models, the deficits involved in poor-pitched singing are not necessarily in the sound, auditory, or somatosensory maps themselves, but rather in connections between them. In order to improve singing performance, and train the singing skill, the connections between these representations must be established through practice and learning. Although unnatural to normal vocal production, visual representations of the voice (RTVF) can be learned and used in a closed feedback loop to strengthen feedforward commands and lead to better vocal control through training.

Chapter 4: Current Studies

Vocal control system theories suggest that better vocal productions are a result of a reliable feedforward control loop with minimal contribution from the feedback control loop (Guenther et al., 2006). Conversely, poor vocal productions are suggested to be a result of comparably higher reliance on the feedback control loop (Scheerer & Jones, 2012). The singing literature suggests that singers have a better ability than non-singers to accurately translate perceptual representations of pitch into categorical representations stored in memory, as well as into accurate motor representations for more precise production (Pfordresher et al., 2015). As most non-singers start out producing less accurate vocalisations, RTVF training programs have been found to specifically improve pitch-matching accuracy (Wilson et al., 2005). However, the ameliorating effects of RTVF training programs have not yet been analysed using vocal control measures, other than accuracy, such as compensation to perceived vocal errors in speed and magnitude, as well as vocal variability.

In light of the literature reviewed, these questions remain: do vocal control measures, including (a) magnitude of compensation to error, (b) latency of compensation to error, (c) pitch-matching accuracy, and (d) vocal variability improve as a result of RTVF training? And how do some of the measures change during the RTVF training session? The two training studies presented in this thesis were conducted in order to answer these questions. Both studies consisted of a pre-test phase where all four measures of vocal control were initially measured, a training phase where participants were randomly assigned to either the feedback training condition or the control condition, followed by a post-test phase which was identical to the pre-test phase. The first study compared the measures between singers and non-singers, while the second study was an attempt to see if additional training contributed to any vocal control changes in non-singers.

Singers' compensatory responses to FAF have been found to be smaller in magnitude (Jones & Keough, 2008) and corrections to pitch errors occur later (Grell et al., 2009) compared to non-singers' responses due to their stronger reliance on their feedforward control. Furthermore, singers have been found to match pitches more accurately (Watts et al., 2003) and with less variability than non-singers (Pfordresher et al., 2010). Therefore, as sensorimotor representations of pitch are quite plastic and subject to learning (Hawco & Jones, 2009), RTVF training should help increase vocal control across all four measures from the pre-test to the post-test. It was hypothesised that only training non-singers to become more singer-like using RTVF would cause their compensation magnitude to decrease, their compensation latency to increase, their accuracy to increase, and their vocal variability to decrease. Additionally, as non-singers benefitted from a similar RTVF program (Wilson et al., 2005), it was expected that RTVF training would have more of an effect on non-singers compared to singers at post-test. Moreover, that effect was expected to increase further when non-singers were exposed to a longer training period.

Improvements in pitch accuracy have been found between pre-test and post-test, even though during the training phase RTVF has been found to impair performance (Wilson et al., 2005). Furthermore, singers were less impaired than non-singers during the training phase (Wilson et al., 2005). To support the results of these previously conducted studies, the first study in this thesis tested the impact visual feedback had in the progression of vocal accuracy during the training phases of each condition between singers and non-singers. It was expected that the singers would perform better than non-singers across the entire training phase due to their already improved vocal control. As found in Wilson et al. (2005), we also expected that the knowledge of results given by the RTVF in the feedback condition would result in better accuracy in non-singers during training compared to those in the no-feedback condition. Due to the longer RTVF training phase, improved vocal control was expected among non-singers for the second study in this thesis.

Chapter 5: Experiment 1

Singing, like any other skill, can improve with training. Therefore, it is important to develop good training programs that are effective at fulfilling their purpose, and advance vocal control of non-singers to the vocal control of singers. The goal of this first study was to determine if, for one pitch, non-singers' vocal control improves as a result of training using a newly developed RTVF training program compared to singer controls as well as non-singers with no-feedback. Changes in four different measures of vocal control were examined among singers and non-singers who were randomly assigned to either the feedback condition, with a novel RTVF training program, or the no-feedback condition where participants practiced without any feedback.

After participants completed the training, improvements in vocal control were predicted to be greater in non-singers compared to singers' improvements. As found in previous studies, improvements were expected to be expressed as lower (Jones & Keough, 2008) and slower (Grell, et al., 2009) compensation responses to detected vocal errors during post-test in the feedback condition compared to the no-feedback condition. Also, it was expected that at post-test, vocal accuracy would increase (or deviations from the target note decrease; Wilson et al., 2005), and vocal variability would decrease (Pfordresher et al., 2010) after training in the feedback condition compared to the no-feedback condition. However, due to singers' already heightened vocal control, it was not expected that they would significantly improve as a result of RTVF training, relative to non-singers (as seen in Wilson et al., 2005).

During the training phase, previous training studies (Welch et al., 1989; Wilson et. al, 2005) found a benefit in vocal accuracy while using RTVF compared to no visual feedback. When compared to the pre-test, however, pitch accuracy was worse during the training phase (Wilson et al., 2005). These studies measured average training accuracy rather than examining the progression (or regression) of accuracy over the course of the training phase. The current experiment attempted to investigate the effects of RTVF over the course of the training phase and to see whether a different pattern emerged for singers compared to non-singers. Again due to the already improved vocal control of the singers, RTVF training was not expected to make a significant difference for them during the training phase.

Method

Participants. Fifty-six participants between the ages of 18 and 26 years (*M* = 19.98; *SD* = 1.67) were recruited to participate in the study. All participants reported they did not speak a tonal language and were right handed. Forty were considered non-singers (17 males and 23 females) as they reported no formal vocal training. The remaining sixteen participants were recruited as singers (1 male and 15 females) because they reported receiving some years of formal vocal training (*M* = 7.78; *SD* = 3.28). Prior to participating in the study, all participants gave written informed consent and upon completion of the study, all participants received either course credit or financial compensation for their involvement. The procedures of this study complied with the ethical standards of Wilfrid Laurier University Research Ethics Committee.

Apparatus. The participant recording sessions took place in a double-walled sound attenuated booth (Industrial Acoustic Company, Model 1601-01). Participants were given a headset with noise-cancelling headphones attached to a boom microphone (Sennheiser HMD 280-13) that was maintained at a fixed distance of approximately 3 cm from their mouth. The experiment was programmed and controlled by Max/MSP 5 (Cycling '74, San Francisco, CA) and presented on a 17-inch computer monitor.

During the experiment, vocalisations were sent to a mixer (Mackie Oynx 1220, Loud Technologies, Woodinville, WA), followed by a digital signal processor (DSP; VoiceOne, T.C. Hellicon, Victoria, BC), which shifted the pitch of the participant's voice. This pitch-shifted vocalisation was then presented back to the participant as auditory feedback in real-time. The target tone was triggered by the command to the DSP, along with the unaltered voice signal, and both were digitally recorded (TAS- CAM HD-P2, Montebello, CA) at a sampling rate of 44.1 kHz for later analysis. The Max/MSP program was designed to calculate the instantaneous F_0 of the voice using the analyser object (Center for New Music and Audio Technologies at the University of California, Berkeley, CA) and display a graphical representation of that frequency on the screen to participants during the training phase in the feedback condition (see Figure 4).

Procedure. Prior to commencing, participants were asked to complete a language and handedness questionnaire as well as a music experience questionnaire (adapted from Cuddy, Balkwill, Peretz, & Holden, 2005) as seen in Appendix A. Before beginning the experiment, participants were instructed to select their target note by vocalising the vowel sound /a/ at a comfortable pitch in order not to strain their voice. The researcher used the VoiceOne to determine the most consistent pitch produced by the participant's voice over the course of a few trials prior to beginning the experiment. With the pitch of the participant reported visually by

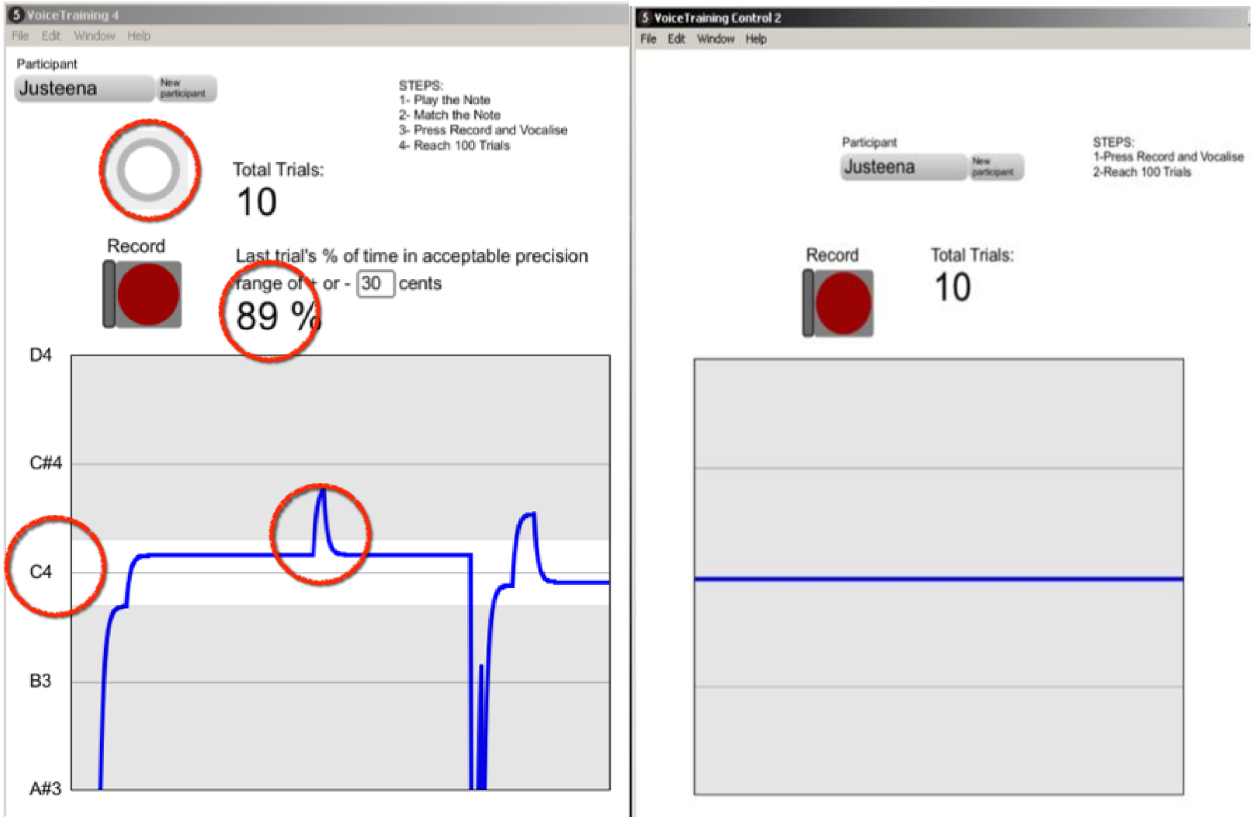

Figure 4. A screenshot of one trial in the training phase under: the feedback condition (left) and the no-feedback condition (right). Key features in the training program lacking from the control program are circled: target note play button, target note graph label, acceptable target range on graph depicted in white, real-time F_0 line plot, and evaluation of accuracy in percent.

VoiceOne, the researcher picked the most consistently displayed pitch and asked the participant to vocalise a few more times to confirm that pitch. This selected pitch was then entered into the program as the target note. The experiment consisted of three phases in the following order: a pre FAF test, a training phase, and a post FAF test. Participants were debriefed after the experimental procedure before leaving.

Test phases. Participants were shown a small box in the centre of the screen that alternated in colour from red to green. When the box was red, participants were instructed not to vocalise but rather to listen to their chosen target note presented for 5000 ms. (The target note was a MIDI recording of the piano available through Max/MSP). Following the presentation of the target note, the participants were instructed to begin vocalising when the box turned green. They were encouraged to try their best to match the target note in pitch by vocalising the vowel sound /a/ for the total duration of the green square, also 5000 ms. Participants were instructed to vocalise at a loud, but comfortable, volume. Vocalisations were played back to the participants in real time via headphones.

The FAF tests contained 4 blocks of 25 trials each, for a total of 100 trials per test and lasted approximately 20 min. After every block, participants were given a break to allow for a drink of water if needed. Out of the 100 trials, 20 were pseudo-randomly unaltered while the remaining 80 trials had FAF-perturbations. During the shifted trials, the pitch of the participant’s voice was perturbed downward 100 cents (1 semitone) three times for 200 ms each. The first shift occurred at a random time between 500 ms and 1000 ms after utterance onset. The second and third shift occurred at a random time between 700 ms and 900 ms after the previous shift just as was done by Scheerer and her colleagues (2013a) to avoid predictability effects. Figure 5 depicts the FAF paradigm used in this study.

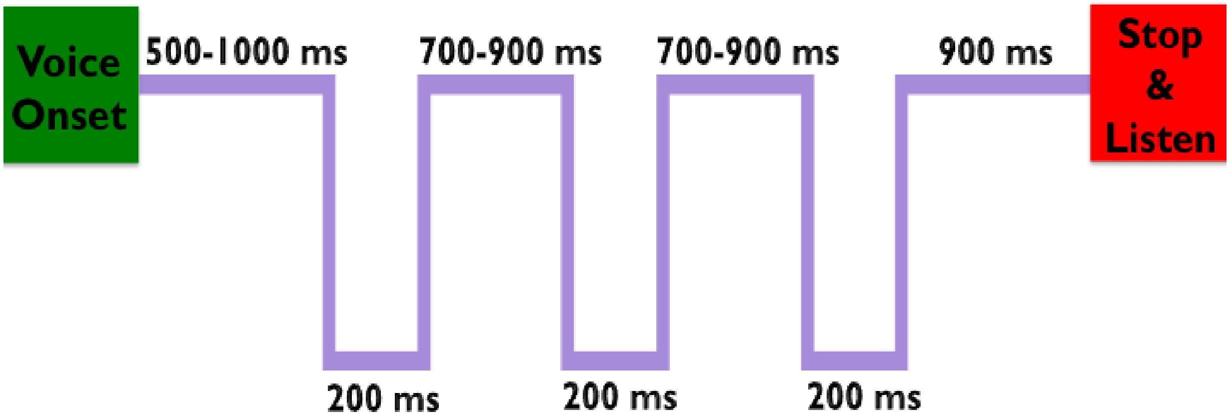

Figure 5. A visual depiction of one shifted trial in the FAF tests. The line represents auditory feedback presented to the participants. Out of the 100 trials in each test, 20 trials were not shifted and occurred pseudo-randomly. The other 80 trials had three downward shifts each. The shifts were unpredictable with varying delays in between (700 ms-900 ms). Every shift lasted for 200 ms with a magnitude of 100 cents (1 semitone). The green square indicated to the participant that they should vocalise, while the red square indicated that they should stop vocalising and listen to the target tone.

Training phase. The participants were randomly assigned to one of two conditions for the training phase: the feedback condition with RTVF or the no-feedback condition (see Figure 4). In the feedback condition, participants were instructed to click a button on the screen to hear their target note at least once before every trial. Thereafter, participants would click a red button on the screen to begin recording and they would begin vocalising for 5000 ms. While vocalising, participants were able to hear their unaltered vocal feedback online through the headphones. In addition, participants viewed their F_0 being plotted in blue on a graph with little to no perceivable delay. The graph was grey with the exception of a white target frequency range of ±30 cents around their target note. This lenient target range was arbitrarily chosen in order to avoid discouraging inaccurate non-singers from the task. Once the recording button turned off, the blue plot stopped graphing at the same time, and the participant was shown a percentage evaluation of their vocal accuracy for the duration of the recording. This percentage was immediately calculated using the time that the F_0 produced remained within the acceptable target ±30 cent range, divided by the total recording time (5000 ms).

In the no-feedback condition, participants were presented with a similar looking program; however, they lacked the important visual feedback. Participants in this condition were not reminded of the target note at all throughout the entire training phase; so they were required to produce their target note from memory. Without a target note, there was no target frequency range displayed in white on the grey plot on the screen. Furthermore, the blue line plotted was simply a straight line for the duration of the vocalisation to roughly match the visual load in the feedback condition. No evaluation was presented to participants at the end of the trials in order to remove any indication of how accurate or consistent their vocalisations were.

For both conditions, participants were instructed to complete 100 trials at their own pace with a trial counter displayed on the screen. Figure 4 shows screenshot examples of one trial from the training phase in each condition. Circles were drawn on the feedback screenshot to highlight the key visual indicators used to provide KR to the users. These visual indicators included the target note play button, the target note graph label, the acceptable target range on the graph depicted in white, the non-static real-time F_0 line plot, and the percent evaluation of time spent accurately producing the vocalisation. These indicators are clearly absent from the no-feedback condition to serve as a control condition. (For a full set of instructions used for each participant refer to Appendix B).

Design. This experiment was a mixed design with one within-subjects factor (Test Phase) and two between-subjects factors (Condition and Experience). Every factor had two levels: Test Phase (pre-test and post-test), Condition (feedback and no-feedback), and Experience (singer and non-singer). Four measurements were taken during the two test phases of the experiment: compensation magnitude, compensation latency, accuracy, and vocal variability.

Analysis. The digital recordings of the vocalisations during the pre and post FAF tests were segmented into separate utterances and F_0 values calculated for each utterance using the SWIPE algorithm (Camacho & Harris, 2008). F_0 values were normalized to their baseline vocalisations by converting Hz values to cents using the following formula:

$$cents = 100\left(12\ \log_2 \frac{F}{B}\right)$$

In the formula, F is the F_0 value in Hz and B is the mean frequency of the 100 ms prior to the shift onset also in Hz. Cents values were calculated for 200 ms before the pitch shift, and 500 ms after the shift onset. Graphical inspection of the vocalisations was done prior to averaging the signals in order to remove trials where F_0 was not properly traced digitally, or if there were any vocal interruptions such as a cough (Larson et al., 2008).

The mean of the F_0 trace for the 100 ms of unaltered voice before the pitch shift represents the baseline F_0. The standard deviation of this baseline mean F_0 represented a measure of vocal variability in each participant's vocalisation (Scheerer & Jones, 2012). For the 500 ms after the shift onset in the shifted trials, the average maximum pitch deviation from the corresponding pre-shift baseline F_0 represents the magnitude of the participant's compensation response (Scheerer et al., 2013a). The delay of this maximum pitch deviation represents a measure of compensation response latency (Patel et al., 2013). Only the shifted trials were used for these three measures. Finally, the median magnitude of deviation from the target note for the first 100 ms of every vocalisation represented a measure of accuracy (Keough et al., 2009).

In order to establish the natural occurrence of the PSR at the pre-test, the average peak F_0 difference between the 80 shifted trials and the 20 non-shifted trials were compared between singers and non-singers in a two-way ANOVA. Three-way ANOVAs were then performed for each of the four measures in order to determine if responses differed significantly before and after the training phase. Furthermore, Pearson correlations between the four measures were conducted to detect any relationships between them (as seen in Scheerer & Jones, 2012). Finally, a three-way ANOVA was conduct for the accuracy measure during the training phase (divided into quartiles) in order to gain insight into the progression of accuracy performance among singers and non-singers in each condition.

Results

Pre-test PSR. A two-way ANOVA was conducted in order to examine the presence of the PSR in singers and non-singers at pre-test prior to any conditions of training. There was a significant difference found between peak F_0 of the non-shifted trials compared to the shifted trials such that $F(1, 54) = 66.498$, $p < .001$, $\eta^2 = .552$. The average peak F_0 of the vocalisations during the shifted trials was significantly higher than the F_0 of the vocalisations during the non-shifted trials. There was no difference between singers and non-singers and no interaction between experience and shift ($p > .05$). Graphical representations of the average F_0 of all shifted trials compared to all non-shifted trials is shown in Figure 6A for non-singers and Figure 6B for singers.

Test phases.

Compensation and latency. A three-way mixed ANOVA considering the effects of singing experience, condition and test phase on peak compensation magnitude detected a main effect of test phase $F(1, 52) = 19.918$, $p < .001$, $\eta 2 = .277$. This means that overall compensation diminished from the time of the pre-test to the time of the post-test for both singers and non-singers. Furthermore, an interaction between condition and test phase approached significance $F(1, 52) = 4.019$, $p = .050$, $\eta^2 = .072$ such that regardless of experience, the feedback conditions diminished the compensation responses more than in the no-feedback conditions (Figure 7A). All other main effects and interactions failed to reach significance ($p > .05$).

For the measure of latency, a three-way mixed ANOVA detected a main effect of test phase $F(1, 52) = 9.187$, $p = .004$, $\eta^2 = .150$. Thus regardless of condition, compensation responses occurred sooner in the post-test phase than the pre-test phase. Also, a main effect of experience was found $F(1, 52) = 8.938$, $p = .004$, $\eta^2 = .147$ indicating that regardless of test phase or condition, compensation responses happened sooner in singers compared to non-singers (Figure 7B). All other interactions and main effects failed to reach significance ($p > .05$).

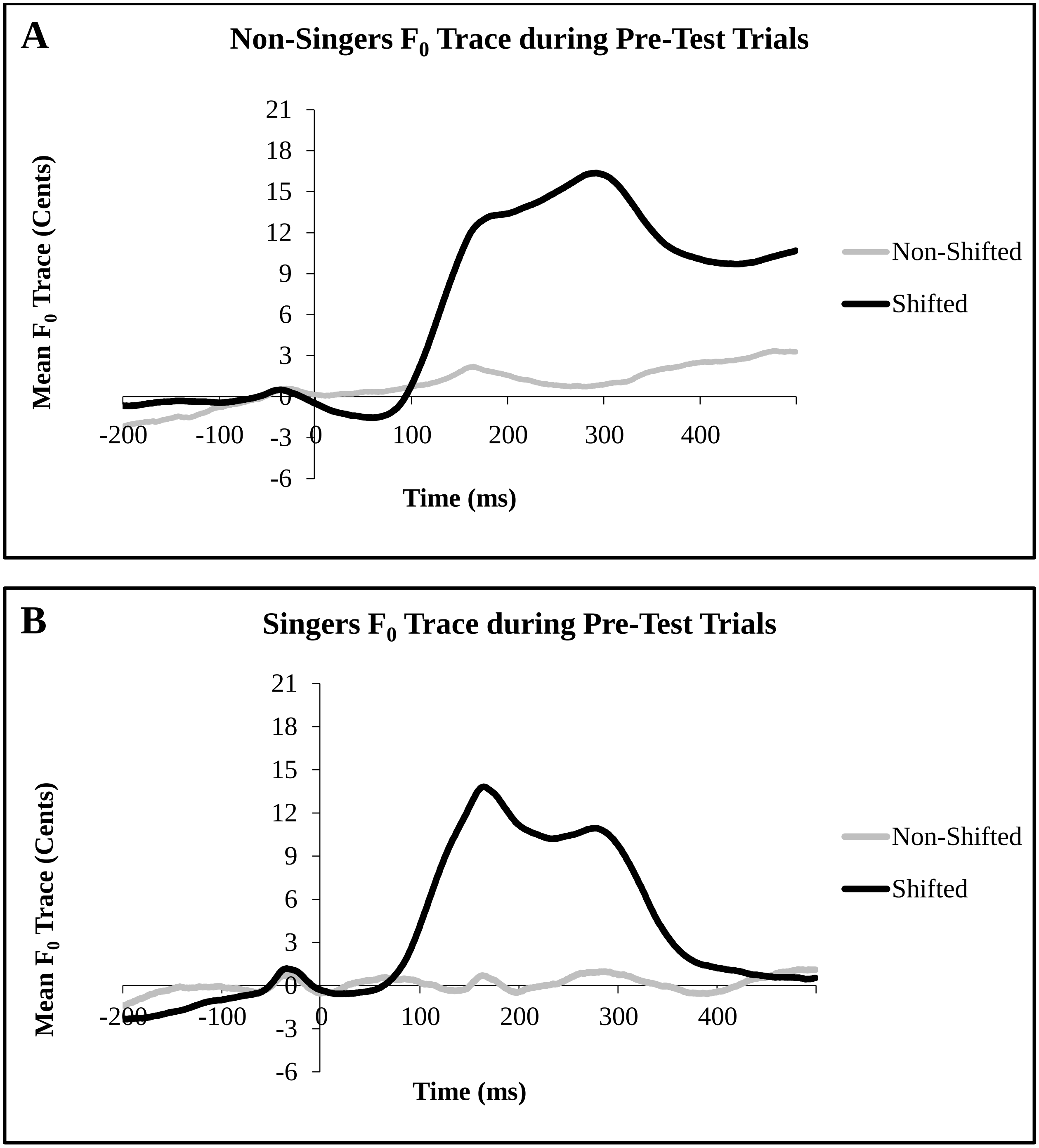

Figure 6. F_0 plots of the average vocalisation for non-singers (A) at the top and singers (B) below it in Experiment 1 at pre-test phase across conditions. Compensation is present when the shift is presented from 0 ms to 200 ms during the shifted trials (black) compared to the non-shifted trials (grey).

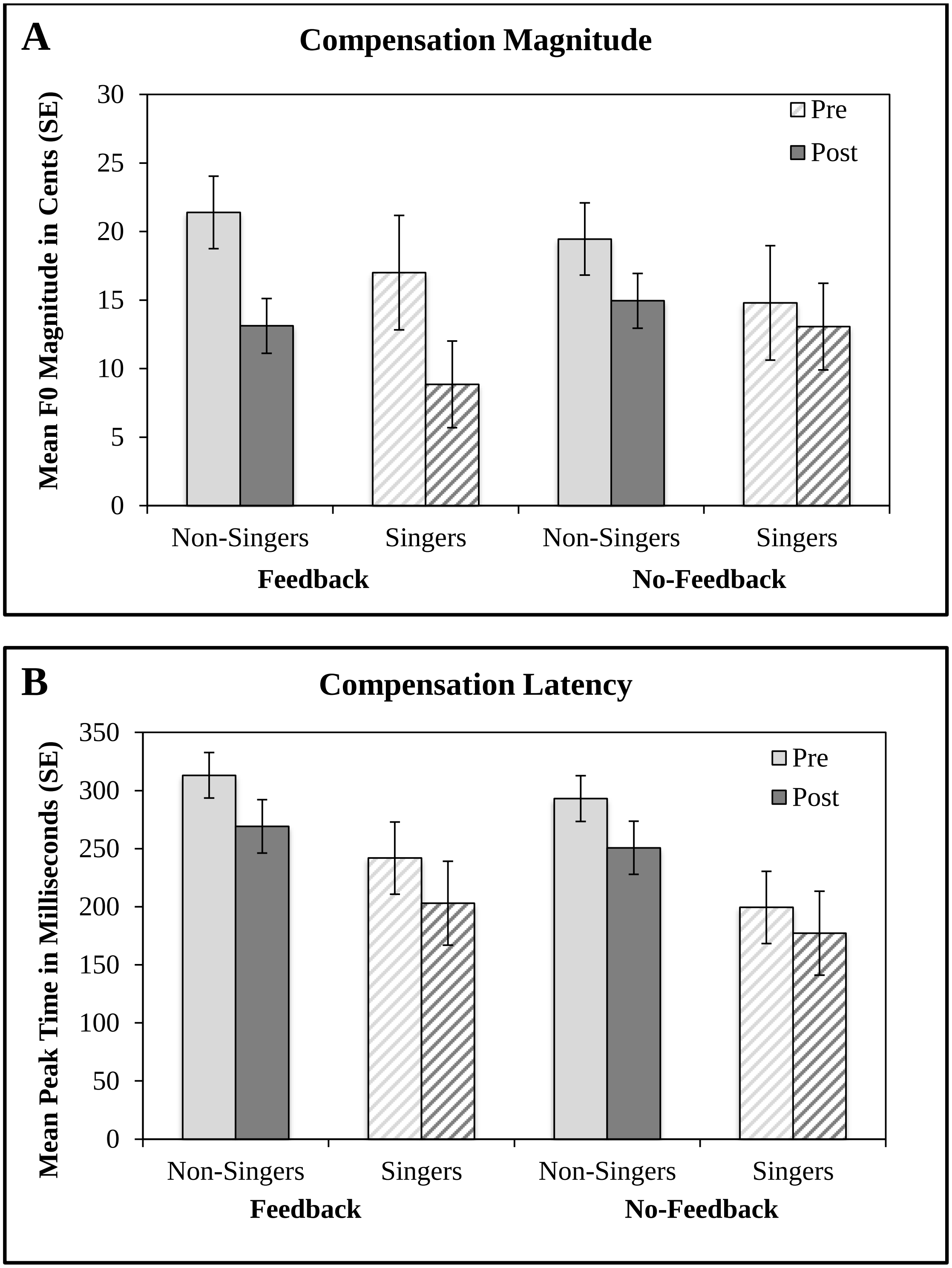

Figure 7. The differences between feedback and no-feedback training conditions in singers and non-singers at the pre-test (grey) and post-test (black) measures of (A) mean magnitude of compensation to FAF-perturbations and (B) mean compensation peak latency in Experiment 1.

Accuracy and variability. A three-way mixed ANOVA investigating the effect of the different feedback conditions on the change in accuracy, a main effect of experience reached significance $F(1, 52) = 9.667$, $p = .003$, $\eta^2 = .157$ indicating that singers were overall more accurate than non-singers. This is reflected in smaller F_0 deviation from the target note compared to the deviation of non-singers seen in Figure 8A. All other and main effects and interactions failed to reach significance ($p > .05$). Another three-way mixed ANOVA tested for the effect of the different feedback conditions on the change in vocal variability from the pre-test phase to the post-test phase (Figure 8B). All the main effects and interactions failed to reach significance ($p > .05$).

Correlations. Pearson's correlations were calculated for non-singers and singers between all four measures at pre-test. There was a significant positive correlation between the compensation magnitude measure and the measure of vocal variability only found in singers $r(14) = .563$, $p = .023$. Thus, singers with higher vocal variability were found to also have higher compensation magnitude for FAF-perturbations (see Table 1). All other correlations at pre-test did not reach significance ($p > .05$).

At post-test, Pearson's correlations were calculated to see if the relationships between the four measures changed after the training phase. There was a significant positive correlation between the measure of vocal variability and the measure of accuracy only found in non-singers who trained in the feedback condition $r(18) = .585$, $p = .007$. Thus, in non-singers who had higher accuracy (or lower deviations from the target note) it was found that their voices were also less variable (see Table 2). All other correlations at post-test did not reach significance in either condition ($p > .05$).

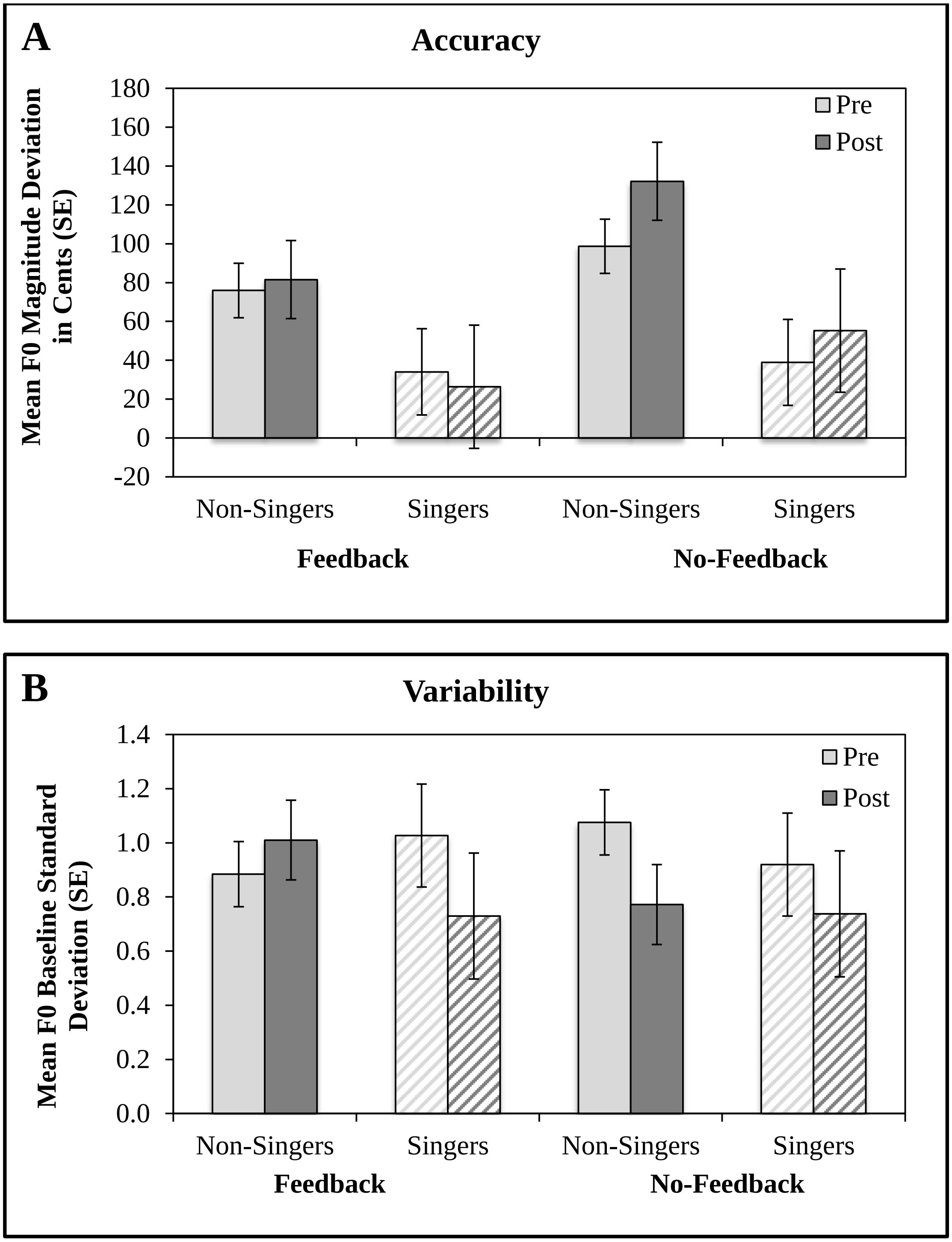

Figure 8. The differences between feedback and no-feedback training conditions in singers and non-singers at the pre-test (grey) and post-test (black) measures of (A) mean magnitude of accuracy deviation from the target note, and (B) mean baseline vocal variability in Experiment 1.

Table 1

Experiment 1 Correlations at Pre-test Across Conditions

	Non-Singers				Singers			
Measures	1	2	3	4	1	2	3	4
1. Compensation Magnitude	-				-			
2. Compensation Latency	-0.068	-			-0.194	-		
3. Accuracy	0.062	-0.173	-		-0.050	0.271	-	
4. Variability	0.114	0.069	0.275	-	0.563*	-0.200	0.095	-

Note. * $p < .05$ (2-tailed) ** $p < .001$ (2-tailed).

Table 2

Experiment 1 Correlations at Post-test in the Feedback and No-Feedback Conditions

	Non-Singers				Singers			
Feedback Condition	1	2	3	4	1	2	3	4
1. Compensation Magnitude	-				-			
2. Compensation Latency	0.232	-			-0.287	-		
3. Accuracy	0.101	0.173	-		-0.527	0.515	-	
4. Variability	0.090	0.393	0.585**	-	0.445	-0.188	-0.277	-
No-Feedback Condition								
1. Compensation Magnitude	-				-			
2. Compensation Latency	-0.278	-			-0.367	-		
3. Accuracy	-0.147	0.129	-		-0.328	0.531	-	
4. Variability	0.057	-0.135	0.179	-	0.455	-0.012	-0.009	-

Note. * $p < .05$ (2-tailed) ** $p < .001$ (2-tailed).

Training phase. A three-way mixed ANOVA examined the effect of the experimental condition and singing experience on the accuracy performance across the quartiles of the training phase. Mauchly's test indicated that the assumption of sphericity had been violated $\chi^2(5) = 31.994$, $p < .001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.738$). A main effect of experience was found $F(1, 52) = 4.923$, $p = .031$, $\eta2 = .086$ as singers performed generally more accurately than the non-singers regardless of condition. Furthermore, a main effect of condition was significant $F(1, 52) = 5.288$, $p = .026$, $\eta2 = .092$ such that those who had no-feedback performed on average worse than those with the feedback while training, shown in Figure 9. All other main effects and interactions failed to reach significance ($p > .05$).

Discussion

Together, the four measures used to indicate vocal control do not show significant changes between the pre-test and the post-test phases as a specific result of RTVF during the training phase. However, when examining each of the measures separately, training in general seems to have had an impact on vocal control. When looking at compensation magnitude, the results of the pre-test phase indicate that the PSR was consistently elicited when all participants were exposed to FAF-perturbations and consistently not elicited during the non-shifted trials. This study supports the large body of literature, which has established that people are able to quickly change their pitch when they perceive any error in their own feedback (Elman, 1981; Burnet et al., 1998; Bauer & Larson, 2003; Liu & Larson, 2007; Patel et al., 2013; Scheerer & Jones, 2012, 2014; Scheerer et al., 2013a, 2013b). At pre-test, there were no significant differences in the magnitude of compensation responses of singers and non-singers even though it was previously shown (Jones & Keough, 2008). The paradigm used by Jones and Keough

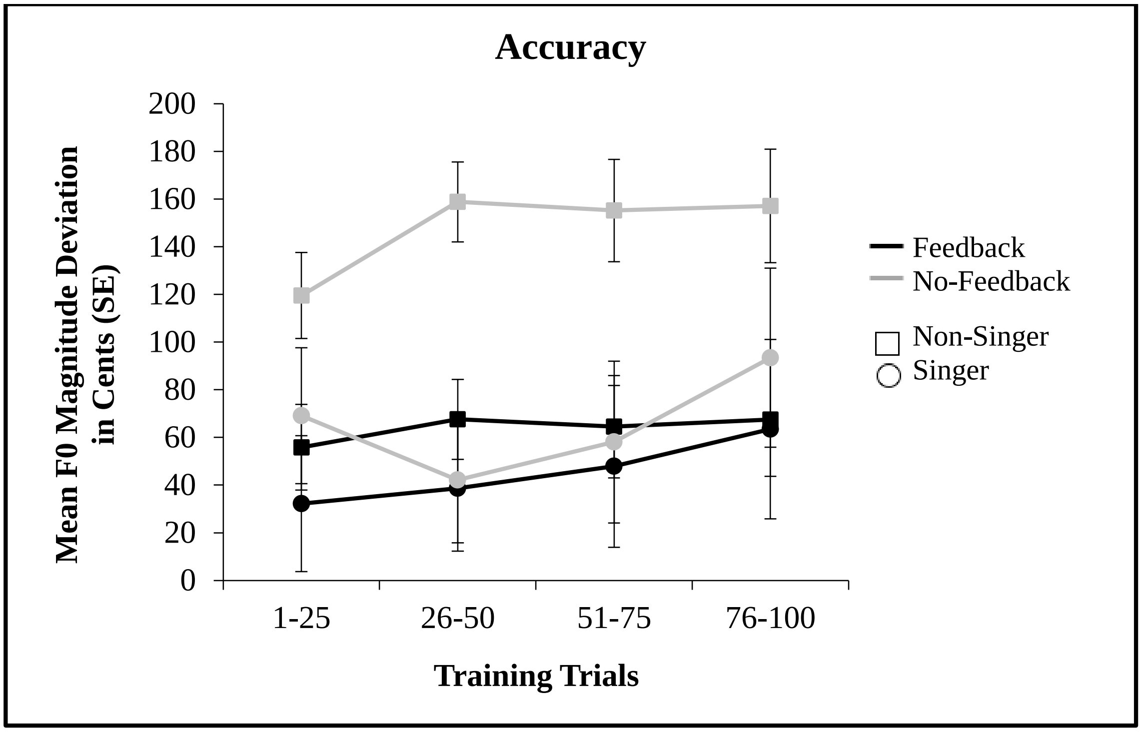

Figure 9. The plot of progression of the mean magnitude of accuracy deviation from the target note over course of the training phase of Experiment 1, divided into four quartiles. The differences between feedback condition (black) and the no-feedback condition (grey) are shown in singers (circles) and non-singers (squares).

(2008) was full utterance FAF compared to FAF-perturbations used in this experiment. FAF-perturbations seem to elicit similar compensation magnitudes among singers and non-singers. Another possible reason for not finding any significant differences between non-singers and singers is due to the lack of detailed training requirements at recruitment. Singers with any kind of formal training were permitted to participate in the study. When examining Figure 6, it is clear that singers and non- singers differ in their vocal response once the shift is removed. Although it was not tested for, it seems that the singers are able to return back to baseline while the non-singers are unable to return to baseline. This directly contradicts the findings of Jones and Keough (2008) though it may be because non-singers were more disrupted by the shift which resulted in worse vocal control after the shift. At post-test, the results show a general overall decrease in all participants' compensation magnitude regardless of condition. Although it was hypothesised that singers would be more resistant to errors and result in lower levels of compensation (Jones & Keough, 2008), there were no differences based on singing experience. Interestingly, this decrease in compensation magnitude was almost significantly different between those in the feedback condition compared to the no-feedback condition. Regardless of experience, the results hint towards an ameliorating effect of RTVF training compared to no-feedback training. Thus, it seems that it is not enough to simply vocalise any note, but in order to enhance vocal control with respect to compensation for errors in the voice, RTVF training may be beneficial to both non-singers and singers.

For the latency of the compensation response elicited in participants, the results contradict the literature and the expected outcome. First, it must be emphasised that the latency measure is not equal to time of the onset of the compensatory response, although the two are related. Rather, latency, as measured in this thesis, is equal to the time the compensatory response takes to reach its peak magnitude. Regardless of condition, all participants became faster at responding to their perceived errors at post-test. As previously suggested, more experienced singers tend to be more resistant, or rather less affected by errors heard in their own feedback. Therefore, their compensatory responses are usually slower than those of of less experienced singers (Grell et al., 2009). The results of this study actually contradict previous findings as earlier compensatory response times (lower latencies) were found among singers, compared to non-singers, regardless of the training condition they were assigned to. Thus, there were no results indicative of improvement due to RTVF training as initially hypothesised.

It may be hypothesised that the change in response to FAF across all conditions is the effect of the predictability of the shift. From debrief conversations with participants at the end of the study, it seemed that the three shifts in a trial were frequently anticipated. Not only did the shifted trials always have 3 shifts, but also they were all of the same magnitude. Scheerer and Jones (2014) and Burnett and her colleagues (2008) found that compensation magnitude decreased when FAF was predictable rather than unpredictable. They concluded that these findings reflect a change from feedback control to feedforward control because information from the feedback becomes consistently unreliable. However, similar to previous studies, this experiment reduced the effect of predictability by pseudo-randomly interrupting a sequence of shifted trials with ones that was not shifted at all. Furthermore, the time of shift onset for the three shifts in one trial differed randomly in an attempt to maintain unpredictability of the shift onset and the inter-stimulus interval time between shifts were also random. Thus, predictability is not a likely explanation for the decreased compensation magnitude or latency in this experiment. However, this reasoning would not explain change in the latency of responses to perturbations, because faster responses imply a faster feedback system to change the vocalisation.

A more probable cause of the diminished magnitude and increased latency in compensation responses is the effect of repeated exposure to the FAF-perturbation paradigm, or in other words habituation to the task. Habituation is when a naturally occurring behaviour, such as the PSR in this case, decreases or ceases to exist. In a preliminary study, DeMarco, Scheerer and Jones (2014) exposed participants to their F_0 shifted downward over multiple sessions on the same day, or over several days. They found that repeated exposure reduced behavioural F0 compensation (DeMarco et al., 2014). In the current study, participants complete 80 FAF trials in each test with 3 shifts in each trial, resulting in 480 FAF-perturbations per experimental session. Even though many trials are needed to reduce signal noise and compose smooth plots of vocalisations, the repetition of task may have resulted in task habituation.

Vocal variability did not significantly change between the two test phases, nor was it different as a result of the training condition to which participants were assigned. Despite Pfordresher and his colleagues' (2010) findings indicating that singers have less variability in their voices than non-singers, vocal variability did not differ based on the experience of participants recruited for this study. When examining the correlations between the measures at pre-test, it is interesting that singers in this study are the only ones to show a strong positive correlation between variability and compensation magnitude, as Scheerer and Jones (2012) found. When examining the correlations between the measures at post-test, another strong correlation between vocal variability and accuracy appears in the group of non-singers who trained using RTVF. This relationship also hints at the effect of RTVF on these vocal control measures as it may have polarised the group of non-singers. Those who sang fairly accurately also became less variable, and those who didn't sing accurately also became more variable.

As initially predicted, singers differed from non-singers with respect to accuracy at both tests simply based on their experience. Because singers had previous training, it was assumed that singers had previously acquired strong internal representations of pitches and stored them in memory. Although Wilson and her colleagues (2005) showed that singers do improve with RTVF, non-singers improved more when using a grid display similar to the one used in this study. Other training studies (Hoppe et al., 2006; Seashore & Jenner, 1910; Welch et al., 1989) found that visual training did reduce F_0 error, but this current study did not replicate these findings. Thus, the information on the display could not assist participants to improve the accuracy of their vocalisations significantly between the time of pre-test and post-test. By examining the accuracy measure over the course of the training phase, some reasons behind the lack of improvement can be surmised.

During the training phase, singers performed more accurately than non-singers. Furthermore, there was an effect of condition where participants who trained with RTVF actually performed more accurately than those without the RTVF. Although an interaction between condition and experience was not found, it is possible that the main explanation for this significantly elevated performance while using RTVF is because participants were reminded of their target note consistently before every trial. As seen in Figure 9, it appears that the non-singers in the feedback condition performed as accurately as the singers in the no-feedback condition. As a non-singer, just being reminded of the sound of the note as an external reference may lead to as good a performance as a singer without one. Although accuracy was enhanced during training, these improvements did not translate to learning to produce more accurate vocalisations at post-test.

Taken together, it is unclear whether RTVF training influenced changes in vocal control indicated by the four measures used in this thesis. However, it is clear that RTVF training did assist participants over the course of the training phase, encouraging the practice of accurate production and better vocal control. In the context of this study, the training implemented may not have been long enough for any learning to occur. Therefore, Experiment 2 of this thesis was designed to explore the effect of length of training on vocal control measures.

Chapter 6: Experiment 2

Since the role RTVF training plays in improving vocal control is unclear from the results in Experiment 1, this second experiment was created as an extension. The goal of this second study was to determine if the amount of RTVF training plays a mediating role in improving vocal control performance. In an attempt to influence the four different vocal control measures, the training session in Experiment 2 was designed to be two times longer than the training session in Experiment 1. Lengthening the training phase was intended to determine whether training time would increase the potential for vocal control improvements even after one session.

As in the first experiment, it was expected that after non-singers trained on one note using RTVF, they would approach the performance level of singers, and even more so because of the increased length of training compared to Experiment 1. As such, improvements in vocal control were predicted from the pre-test phase to the post-test phase. More specifically, as found in previous studies, improvements were expected to be expressed as lower (Jones & Keough, 2008) and slower (Grell, et al., 2009) compensation responses to detected vocal errors during FAF in the feedback condition compared to the no-feedback condition. Again, it was expected that vocal accuracy would increase (or deviations from the target note decrease; Wilson et al., 2005), and vocal variability would decrease (Pfordresher et al., 2010) in non-singers in the feedback condition compared to the no-feedback condition.

During the training phase, previous training studies (Welch et al., 1989; Wilson et. al, 2005) found a benefit of RTVF compared to no visual feedback. In Experiment 1, improved singing accuracy was found while training with RTVF as an aid; however, this did not transfer into post-test improvement of accuracy. This experiment attempted to investigate whether longer RTVF training would be required to influence vocal control significantly enough to reflect as improvements. The results of this study were compared to the results of Experiment 1 in order to investigate any effects of longer RTVF training on compensation responses, pitch-matching accuracy and vocal variability.

Method

Participants. Forty participants between the ages of 18 and 28 years with a mean age of *M* = 21.20 (*SD* = 2.66) were recruited to participate in the study; 15 males and 25 females. All participants reported no formal vocal training, did not speak a tonal language and were right handed. Prior to participating in the study, participants gave written informed consent and upon completion of the study, all participants received either course credit or financial compensation for their involvement. The procedures of this study complied with the ethical standards of Wilfrid Laurier University Research Ethics Committee.

Apparatus. All components of the equipment and program were exactly the same as in Experiment 1 with one exception. Participants were given noise-cancelling headphones (Sennheiser HD 280 Pro) and a wraparound microphone (AkG C 420) that was maintained at a fixed distance of approximately 3 cm from their mouth.

Procedure. The procedures of this study were exactly the same as in Experiment 1, with two exceptions. First, both the pre FAF test and the post FAF test were reduced from 100 trials to 50 trials where 10 were pseudo-randomly unaltered while the remaining 40 trials had pitch shifts. Second, the training phase was elongated from 100 trials to 200 trials. Due to the proportional adjustments to trials, participants produced the same number of vocalisations (300) as in Experiment 1.

Design. This experiment was a mixed design with one within-subjects factor (Test Phase) and one between-subjects factor (Condition). Both factors had two levels: Test Phase (pre-test and post-test), and Condition (feedback, no-feedback). The same four measurements that were taken in Experiment 1 were measured during the two test phases of this experiment.

Analysis. The analyses of this study follow those of Experiment 1. Further analyses comparing this group of non-singers to the non-singers from Experiment 1 were conducted to explore the effects of the length of training on vocal control.

Results

Pre-test PSR. A paired samples t-test was conducted in order to examine the presence of the PSR in non-singers' vocal responses at pre-test prior to any conditions of training. There was a significant difference found between peak F_0 of the non-shifted trials compared to the shifted trials such that $t(39) = -7.990$, $p < .001$. The average peak F_0 of the vocalisations during the shifted trials was significantly higher than the F_0 of the vocalisations during the non-shifted trials. A graphical representation of the average F_0 of all shifted trials compared to all non-shifted trials is shown in Figure 10.

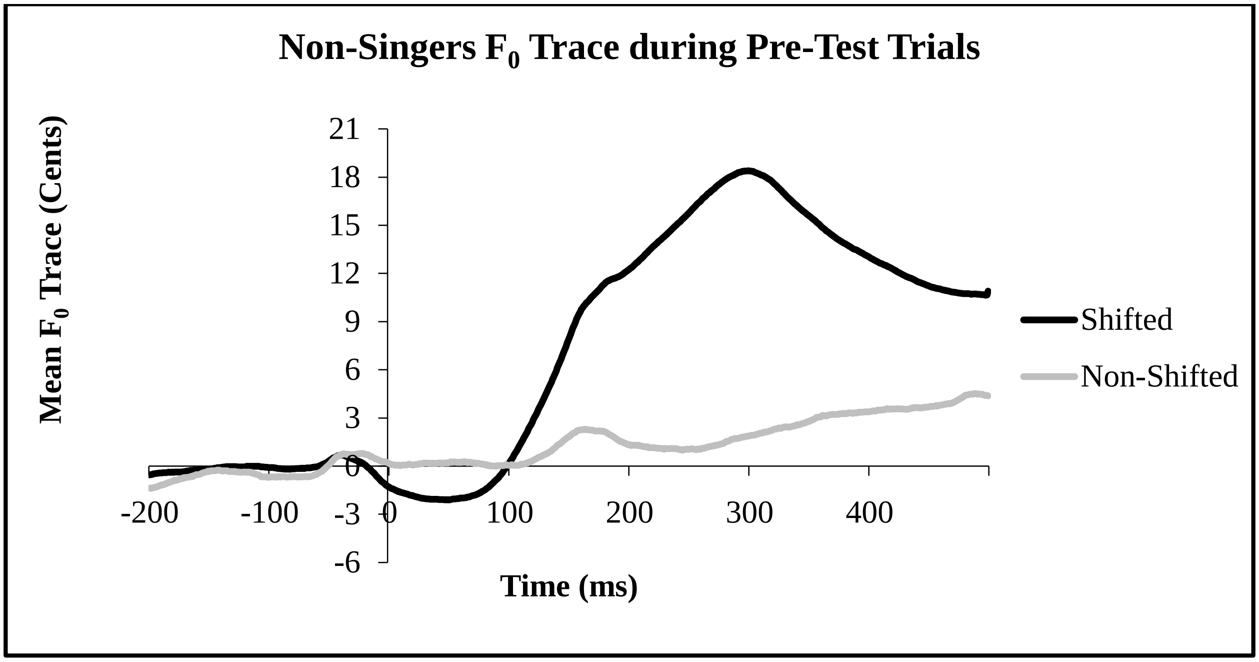

Figure 10. F_0 plots of the average vocalisation for non-singers in Experiment 2 at pre-test phase across conditions. Compensation is present when the shift is presented from 0 ms to 200 ms during the shifted trials (black) compared to the non-shifted trials (grey).

Test Phases.

Compensation and latency. A two-way mixed ANOVA investigating the effect of the two different feedback conditions on the change in peak compensation magnitude from the pre-test phase to the post-test phase detected a significant main effect of test phase $F(1, 38) = 27.029$, $p < .001$, $\eta^2 = .416$. Under both feedback and no-feedback conditions, the compensation magnitude of the post-test phase was less than the pre-test phase (Figure 11A). The effect of condition and the interaction effect of condition and test phase failed to reach significance ($p > .05$). Another two-way mixed ANOVA tested for the effect of the different feedback conditions on the change in latency of peak compensation from pre-test phase to the post-test phase. A significant effect of test phase was found $F(1, 38) = 4.732$, $p = .036$, $\eta^2 = .111$. Thus, regardless of condition, participants compensated sooner at post-test than at pre-test (Figure 11B).

Accuracy and variability. A two-way mixed ANOVA investigating the effect of the two different feedback conditions on the change in accuracy magnitude from the pre-test phase to the post-test phase (Figure 12A) detected no significant main effects or interactions ($p > .05$). With respect to the measures of vocal variability (Figure 12B), the two-way mixed ANOVA indicated no differences between conditions, nor test phases ($p > .05$).

Correlations. Pearson's correlations were calculated for non-singers between all four measures at pre-test. There was a significant positive correlation between the compensation magnitude measure and the measure of vocal variability $r(38) = .653$, p $< .001$. Thus, participants with higher vocal variability were found to also have higher compensation for FAF-perturbations (see Table 3). All other correlations at pre-test did not reach significance ($p > .05$). At post-test, Pearson's correlations were calculated to see if the relationships between the four

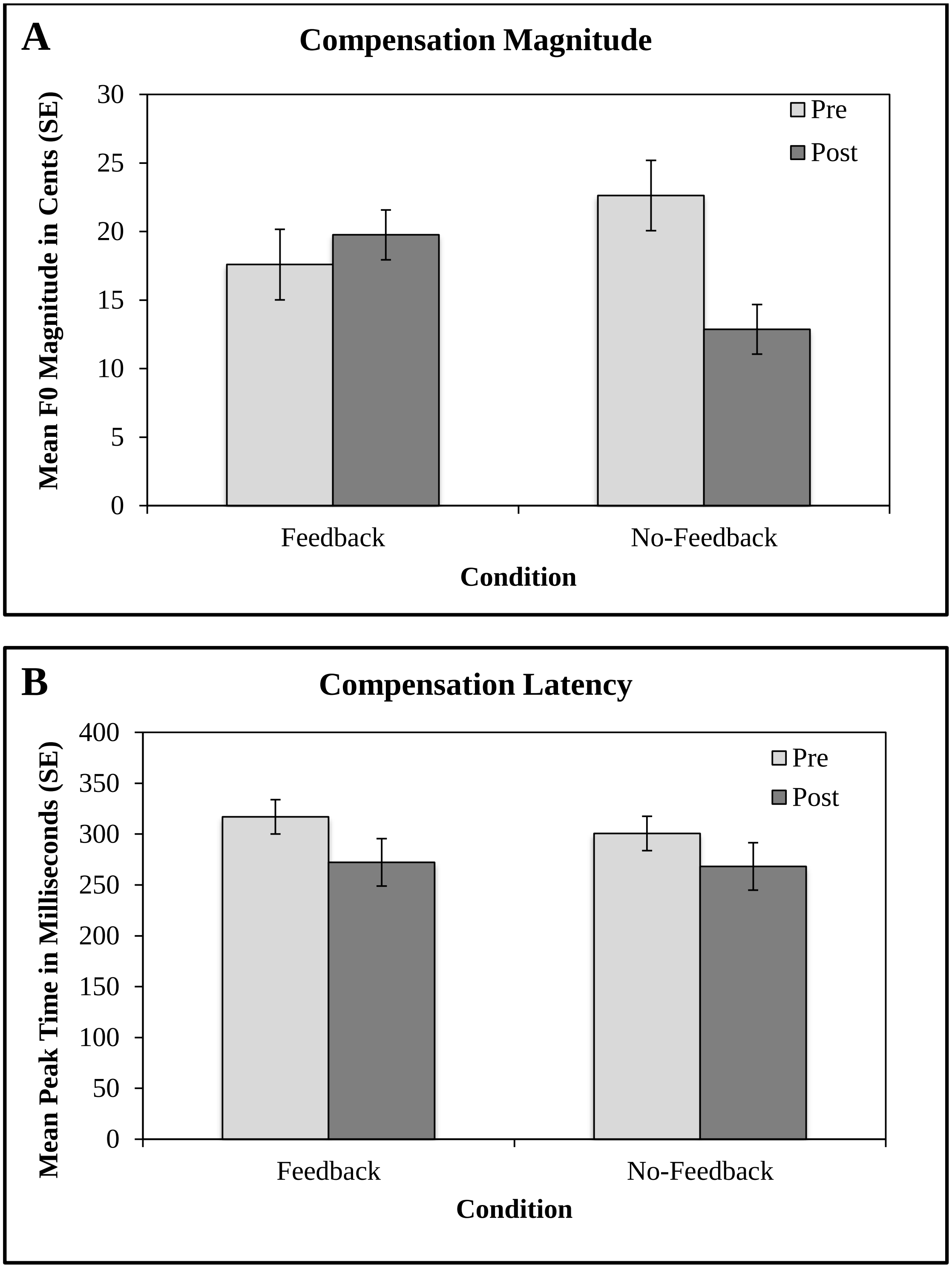

Figure 11. The differences between feedback and no-feedback training conditions in non-singers at the pre-test (grey) and post-test (black) measures of (A) mean magnitude of compensation to FAF-perturbations and (B) mean compensation peak latency in Experiment 2.

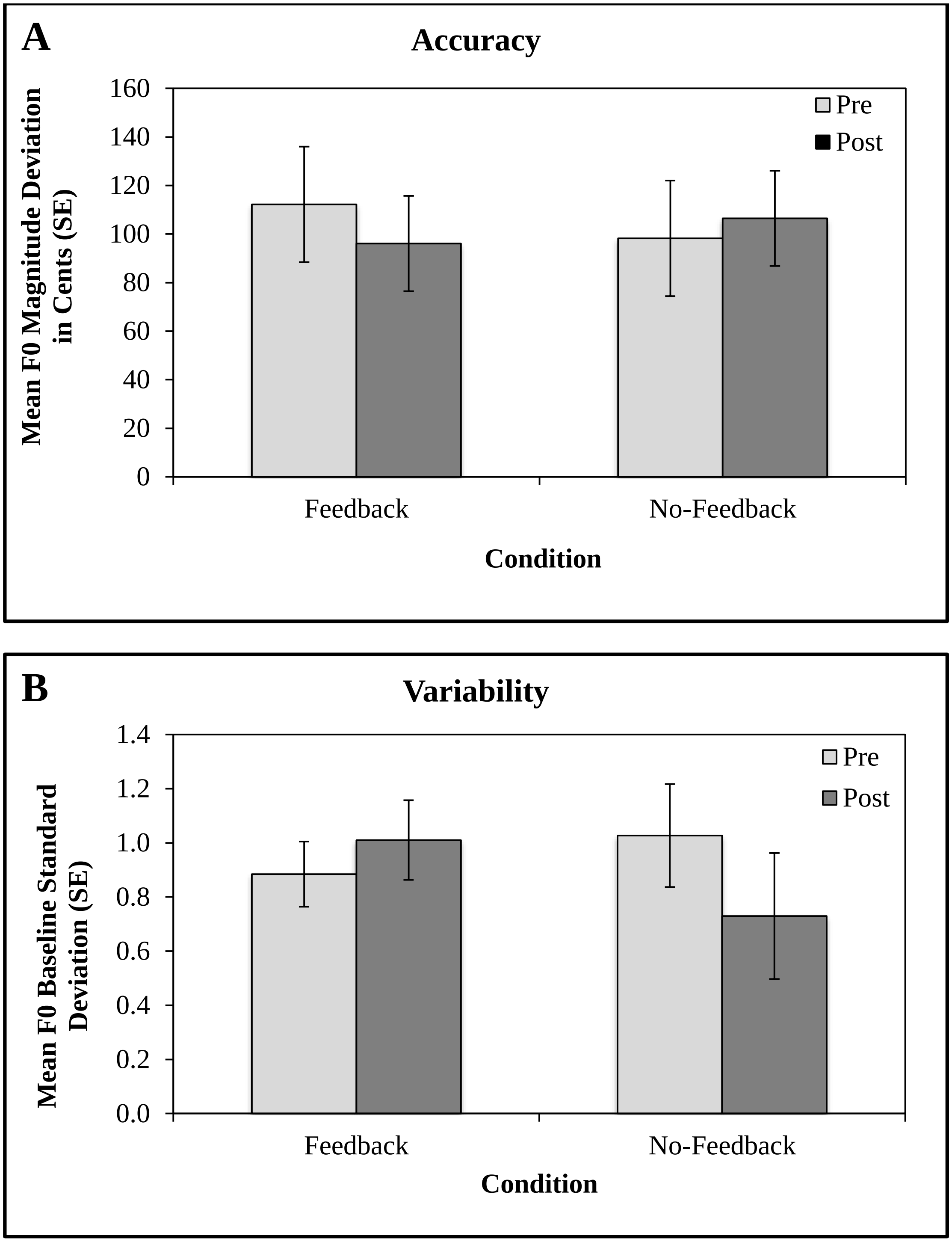

Figure 12. The differences between feedback and no-feedback training conditions in non-singers at the pre-test (grey) and post-test (black) measures of (A) mean magnitude of accuracy deviation from the target note, and (B) mean baseline vocal variability in Experiment 2.

Table 3

Experiment 2 Correlations at Pre-test Across Conditions

Measures	1	2	3	4
1. Compensation Magnitude	-			
2. Compensation Latency	-0.047	-		
3. Accuracy	0.034	-0.027	-	
4. Variability	0.653**	0.012	-0.091	-

Note. * $p < .05$ (2-tailed) ** $p < .001$ (2-tailed).

measures changed after the training phase. As shown in Table 4, none the correlations between the measures at post-test reached significance in either condition ($p > .05$).

Training phase. A two-way mixed ANOVA examined the effect of the experimental condition on the accuracy performance across the training phase trials. Mauchly's test indicated that the assumption of sphericity had been violated $\chi^2(27) = 121.681$, $p < .001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.450$). A main effect of training trials was found $F(3.153, 119.801) = 3.030$, $p = .030$, $\eta^2 = .074$ where accuracy performance followed a quadratic trend $F(1, 38) = 6.872$, $p = .013$, $\eta^2 = .153$ regardless of the training condition. Furthermore, a main effect of condition was significant $F(1, 38) = 9.818$, $p = .003$, $\eta^2 = .205$ such that those that had no-feedback performed on average less accurately than those with the feedback while training, shown in Figure 13. All other main effects and interactions failed to reach significance ($p > .05$).

Comparison to experiment 1 non-singers. When comparing the non-singers from both experiments, a three-way mixed ANOVA considering the effects of training length, condition and test phase on peak compensation magnitude detected a main effect of test phase $F(1, 76) = 47.027$, $p < .001$, $\eta^2 = .382$. This indicates that, overall compensation diminished from the time of the pre-test to the time of the post-test across all the different groups. All other main effects and interactions failed to reach significance ($p > .05$). For the measure of latency, a three-way mixed ANOVA detected a main effect of test phase $F(1, 76) = 13.395$, $p < .001$, $\eta^2 = .150$. Therefore, regardless of training length or condition, compensation responses happened sooner at post-test compared to at pre-test. All other interactions and main effects failed to reach significance ($p > .05$). When comparing both experiments with respect to accuracy and vocal variability measures, all main effects and interactions failed to reach significance ($p > .05$).

Table 4

Experiment 2 Correlations at Post-test in the Feedback and No-Feedback Conditions

Feedback Condition	1	2	3	4
1. Compensation Magnitude	-			
2. Compensation Latency	0.256	-		
3. Accuracy	0.263	-0.416	-	
4. Variability	0.397	0.279	-0.114	-
No-Feedback Condition				
1. Compensation Magnitude	-			
2. Compensation Latency	-0.019	-		
3. Accuracy	-0.112	0.205	-	
4. Variability	0.385	0.011	0.228	-

Note. * p < .05 (2-tailed) ** p < .001 (2-tailed).

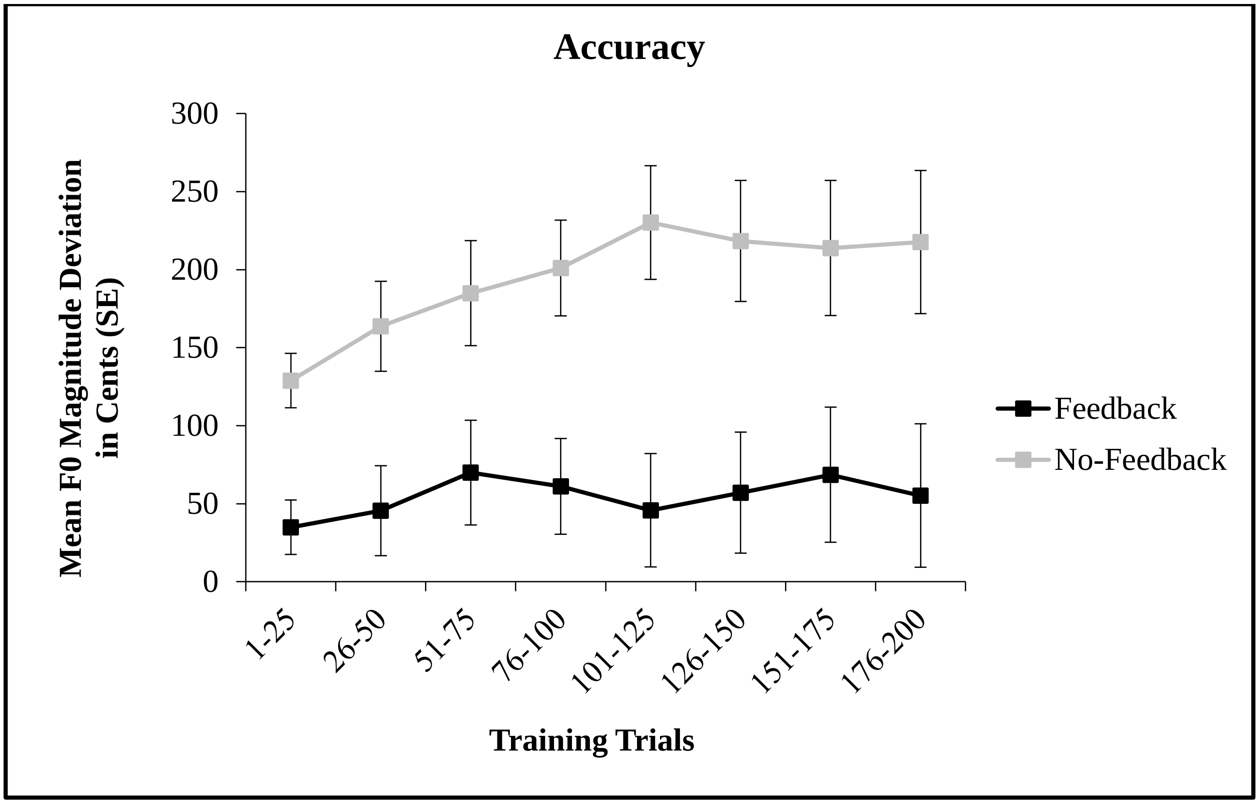

Figure 13. The plot of progression of the mean magnitude of accuracy deviation from the target note over course of the training phase of Experiment 2. The differences between the feedback condition (black) and the no-feedback condition (grey) performance are shown.

Discussion

Taken together, the effect of RTVF training on the four different measures of vocal control failed to reach significance in non-singers. Similar to the findings of the Experiment 1, there were no effects of RTVF training in non-singers of this experiment on the four measures taken. As was seen with non-singers in Experiment 1, compensations for errors perceived in the voice were consistently elicited at the pre-test. Compensation responses were generally smaller in magnitude and latency (faster) at post-test compared to at pre-test regardless of condition. In this experiment, the number of FAF-perturbations trials during the test phases were reduced by one half, which should reduce any effect of task habituation, as participants would be exposed to 150 fewer perturbations. This did not seem to reduce compensation magnitude or latency when compared to Experiment 1. Thus, it remains a possibility that participants in Experiment 2 had compensatory responses to errors due to their exposure to the FAF task during the pre-test as previously found (DeMarco et al., 2014).

Vocal variability at pre-test was correlated with compensation magnitude as in Experiment 1 and by Scheerer and Jones (2012). Replicating the findings of Experiment 1, vocal variability did not change significantly from pre-test to post-test. In both experiments, participants vocalised a total of 300 times, however, both Figure 8B and Figure 12B suggest that the RTVF training condition caused an increase in variability. It is possible that during the RTVF training, participants were trying different methods to accurately produce the notes. They might have purposefully varied their pitch multiple times during the training phase, which would result in more variable overall F_0.

Finally, in this experiment, non-singers' pitch-matching accuracy was unaffected by RTVF training as no expected differences between the pre-test and post-test were found.

Although most of the correlations between all four measures became stronger in non-singers after feedback training in the post-test of Experiment 1, this was not the same in Experiment 2. No correlations between compensation magnitude, latency, vocal accuracy, and variability were found at post-test, regardless of condition.

During the extended RTVF training designed in this experiment, it appears that the non-singers used the information provided as an aid to generally vocalise more accurately as seen in Figure 13. It was interesting that the training trials followed a quadratic trend for accuracy measured as deviation in cents from the target note. Over the course of the training phase, participants seemed to perform most accurately at first, then worsen around the middle, and finally accuracy began to increase and approach the level of accuracy observed at the beginning of the training phase. This pattern can be understood when exploring the potential pattern of motivation, which was not incorporated in the design. It is possible that at first, participants were enthusiastic about the training and put in a significant amount of effort. As they progressed into the training phase, they might have found the task quite mundane regardless of the number of breaks they were permitted to take. Finally, when approaching the end of the training trials count, participants realised that they were almost done, and so, they may have regained focus and motivation during the last few trials to finish strongly.

General Discussion

This thesis aimed to explore the effect of RTVF training on vocal control that has been shown in the previous singing training literature (Hoppe et al., 2006; Seashore & Jenner, 1910; Welch et al., 1989; Wilson et al., 2005). This thesis also explored the progression of performance over the course of the training phase and how the length of training impacted the amount of vocal control exerted while singing a target note of the participant’s choice. Two studies were conducted where participants were tested on FAF-perturbation trials at the pre-test prior to training. Then participants moved on to the training phase where they had to practice that target note a number of times, and they were randomly assigned to either the RTVF training condition or the no-feedback condition. Finally, participants were tested again using the FAF-perturbation paradigm and measures of vocal control were taken and compared to pre-test ability. Progression of accuracy over the course of training phase was examined.

This thesis is also one of the first attempts at integrating concepts and measures used in the speech literature to concepts and measures from the singing literature that are reflective of different aspects of vocal control. More specifically, the measures of compensatory magnitude and latency reflect vocal control in response to perceived errors in the voice (Grell et al., 2009; Jones & Keough, 2008). Pitch-matching accuracy has been used as a measure to reflect the feedforward control loop relying on categorical mental representations of pitches stored in memory, activated when vocalisations are initiated (Wilson et al., 2005). Finally, the measure of vocal variability reflects vocal control from the perspective of sensorimotor translation and motor execution (Pfordresher et al., 2010).

Vocal Control in Response to Errors in Feedback

Experiments 1 and 2 found that participants have generally smaller magnitudes and latencies of compensation to errors in the post-test than in the pre-test. As previously discussed, it is possible this reduced response may be due to the participants' ability to predict the FAF-perturbations of the same magnitude and length throughout the test phase. The other possible explanation for a diminishing response to error previously discussed was the effect of repeated exposure, where the participant is habituated to the error, and so, the more frequent the FAF perturbation, the less effective errors are at eliciting the PSR. However, the second study not only increased the number of training trials but also reduced the number of test trials, which reduced the amount of exposure participants had to FAF-perturbation. When comparing the non-singers from the first and second experiments, no differences between the compensation responses were found, so it is unlikely that repeated exposure to FAF caused the reduction in compensation magnitude or compensation latency.

Although not significant, there were different effects of RTVF training on compensation magnitude in Experiment 1 as both singers and non-singers in the feedback condition became less reactive to FAF-perturbations at post-test compared to at pre-test. As hypothesised, this change in compensation magnitude may be due to the change in the weightings of the feedforward loop and the feedback loop such that the reliance on the feedforward loop control increases, while the reliance on the feedback loop control decreases as participants become more confident of their practiced pitch. However, this change in the weighting of the feedforward and the feedback loop control was not reflected in the measure of compensation latency, which according to Scheerer and Jones (2012), should be correlated to the measure of compensation magnitude. Furthermore, the effect of the length of training on compensation magnitude was not significant in Experiment 2; therefore, it is unclear whether the hypothesised change in the vocal control system did occur.

Vocal Control at the Initiation of the Vocalisation

Although the RTVF training program was focused on participants' efforts to aim for a target pitch, measures of accuracy at the onset of the vocalisations did not change from pre-test to post-test. Over the course of the training phase, however, using the RTVF training program as an aid improved performance compared to the no-feedback condition. This is contrary to some of the studies in the literature that found that while using an aid, singing performance is actually diminished (Wilson et al., 2005). These previous studies have attempted to explain the worse accuracy during the training using theories of cognitive load, suggesting that the added information available in the visual stimuli causes attention to become divided between the visual representation and the vocal task (Wilson et al., 2005). The results of Experiments 1 and 2 do not support these theories because in both conditions participants were given similar visual stimuli (although in one condition, the stimulus was less meaningful; see Figure 4).

Following a singing symposium attended by a few leading singing researchers, a first attempt was made to develop a battery of tasks which standardise a baseline measure of singing accuracy (Demorest, et al., 2015). This battery is referred to as the Seattle Singing Accuracy Protocol, or SSAP. The tasks are designed to provide a baseline for any study of singing that could be used to compare the performance of one study population directly to the performance of populations from other studies across different ages and levels of training. Although this endeavour was a small step, it is one of the first attempts to unify the singing literature and promote proper replications in future studies. Thus, measuring performance accuracy of participants using SSAP during the test phases may have been a better indicator of vocal accuracy than a simple deviation from target calculation.

Vocal Control in Variability of Production

Neither study showed any effects of RTVF training on vocal variability. The only exception to this was the strong relationship between vocal variability and accuracy, which developed only in non-singers after training with RTVF in the first study. As a positive correlation, individuals with lower deviations from the target pitch (higher accuracy) also had less variability in their voices. This relationship aligns with the categorisation of different types of singers based on their accuracy and variability as suggested by Pfordresher and colleagues

(2010). For example, Pfordresher et al. suggest that poor singers are inaccurate and very variable, while good singers are inaccurate but not variable, and then finally trained singers, who are accurate and not variable. Experiments 1 and 2 did not divide participants on a continuum from poor-pitched singers to good singers beyond asking what their previous singing training was. A clearer picture of vocal control measures, and their interaction with RTVF training, can be obtained in future studies that investigate the relationship between the level of experience or training of the participants to compensation magnitude, compensation latency, vocal accuracy and vocal variability.

Training Programs

Many RTVF training programs have been developed for singing training research, and many of them have customizable features to present KR to the user in many different ways (Hoppe et al., 2006). Furthermore, some of these programs have previously shown successful improvement of vocal control in singers and non-singers, even immediately after one training session (Wilson et al., 2005). Other studies have examined different types of RTVF displays and their effect on vocal accuracy (Callaghan et al., 2004; Welch et al., 2004; Wilson et al., 2005), such as a piano display instead of a graphical one in future studies.

The studies in this thesis are the first studies that use an RTVF training program which provides the learner with KR in the form of real-time percent evaluation of performance, in addition to the graphical representation of pitch found in other training programs. It would be interesting and useful to separate the different components in a RTVF display and test their effect on measures of vocal control other than accuracy, as was attempted by this thesis. For instance, one of the KR used in this thesis was a target accuracy range of ±30 cent. Thus, it is possible that vocal control gains were limited to that range as participants could settle for 100%

performance rating if they stayed just under +30 cents and just above -30 cents from the target. In future studies, RTVF training using a stricter target range may yield greater improvements in the internal representations of the voice.

With that in mind, another explanation of the results obtained in these studies arises. As better vocal control is characterised by the stronger contribution of the feedforward loop as seen in singers (Keough & Jones, 2008), the RTVF training session may have decreased the participants' dependency on their feedforward loop and increased the contribution of the feedback loop. As participants were instructed to watch the visual representation of their vocal pitch presented in real time, their attention was drawn to the shape and location of the line on the screen. Whenever that line deviated from their target range, they reacted in a way that brought it back to the target range, similar to the PSR. This highlighted attention may have led to more focus on the feedback. While the results do not indicate that vocal control diminished significantly, the use of RTVF may have hindered any improvements or fine-tuning of the vocal representations used in the feedforward loop.

Although adaptation, a form of learning, can occur quickly, as seen in the literature (e.g. Hawco & Jones, 2009) RTVF training may not specifically result in the immediate vocal improvement that Wilson and her colleagues (2005) suggest. Given that Experiment 2 tested the effect of more training, it is possible that more than 200 trials of practice in one session are needed to elicit improvements in all aspects of vocal control. However, more practice in one session is not feasible because of vocal fatigue. Thus, it is possible that either multiple RTVF training sessions or more time between the training and the test would allow for memory consolidation (i.e. sleep) and lead to reliable improvements in vocal control.

As Guenther's DIVA model is used to explain vocal control mechanisms involved in speech acquisition and speech production (2006), Pfordresher and colleagues' singing accuracy model is used to explain vocal control mechanisms involved in singing performance (2015), and Welch's training model is used to investigate vocal control improvements through the RTVF singing training regime (2005), there is neither integration nor collaboration between them. This may be due to the notion that speech and song are two different cognitive processes, although because they use the same biological instrument and similar mechanisms, it would benefit both fields to collaborate and come together to create a complete model of the mechanisms that support communication.

Limitations and Future Directions

As mentioned throughout the thesis, there were a few limitations to the studies presented. The KR granted to the user may have differential effects on the outcome of vocal training dependent on whether the KR is meaningful or not (Welch, 1985a). Therefore, the lack of practice trials or demonstrations in using the RTVF training program in Experiments 1 and 2 may have resulted in inappropriate use of the RTVF tool. Future studies may assist the participants in their understanding of the tool thoroughly and grant them a few tries to practice in order to make it more meaningful and useful to them (Callaghan et al., 2004).

Another limitation to the experiments conducted in this thesis was the potential habituation to the FAF perturbation paradigm. As shift onset was random, predictability was avoided at a micro level, though at a macro level, especially after the first shift, participants may have anticipated the two shifts after it as they were in the same direction and of the same magnitude for the whole experiment; thus reducing their compensatory responses. As previous studies have done (Scheerer & Jones, 2014), future studies may use the training paradigm used in this study with FAF testing phases with different directions and magnitudes of shifts to decrease predictability of FAF errors at test.

For the singers that were recruited for Experiment 1, there was no minimum amount of training nor a minimum quality of training required to participate (i.e. Royal Conservatory Training); this may have led to an overestimation of their abilities and incorrect categorisation thereafter. Thus, it may be necessary to analyse the singers' abilities on a continuum and compare their levels of vocal control. A longitudinal study may also accomplish this by measuring the abilities of a singer over the course of their training. Furthermore, it may be important to see whether musicians differ on some aspects of vocal control such as compensation to FAF responses. Due to their heightened ability to translate their perceptual information and compare it to their categorical representation of pitches, it is possible that musicians are better at error detection and correction than non-singers. They will have stronger representations of pitch to compare their feedback to, and be able to adjust their voice accordingly. On the other hand, because musicians do not necessarily have the same training of the vocal articulators as trained singers do, it is possible that musicians are worse than singers not only at correcting for errors but also how fast they do so. Thus, recruiting participants on a continuum of perception and production abilities may provide a clearer picture of the variations in vocal control.

Conclusion

As the vocal production literature examined in this thesis suggests, and the mere existence of vocal teachers, singing training should enhance vocal performance - just as with other motor skill training. Furthermore, despite the evidence from previous studies that suggested RTVF training does result in enhanced vocal accuracy, this series of studies did not replicate these results. Of the literature examined, none of the previous studies have examined a complete picture of vocal control including feedforward internal representations and feedback error monitoring. In an attempt to evaluate changes in vocal control by including measures of vocal variability and compensatory PSR responses of magnitude and latency, this thesis found no consistent immediate effects of short-term RTVF training irrespective of vocal experience or length of the single training session. Although RTVF during training has been found helpful as an aid to increase performance accuracy compared to having no feedback, this did not translate into robust improvements at test. Future studies should examine the learning and memory pathways involved in singing training with the knowledge we have today to design better training regimes, which specifically improve vocal control.

Appendix A

Language Questionnaire

Date of Birth: __ __ / __ __ / __ __ (MM/DD/YY) *Age:* ______ years

Gender: Male Female Other

Current Year of Education (Gr 1 to 1st Year =13): ________________________

1. What is your mother tongue (the first language you learned)?
2. What other languages do you know?
3. What is your best language for speaking?
4. What is your best language for writing?
5. What language(s) did your family speak at home?
6. In what city (and country) were you born?
7. How long did you live in the city that you were born?
8. In what city did you go to elementary school?
9. In what city did you go to high school?
10. How many years have you lived in Canada?

Handedness Questionnaire

Instructions: Think carefully about each of the following tasks and indicate by circling, whether you use your left hand, right hand or either hand.

1. Which hand do you use to hold scissors?	Left	Either	Right
2. With which hand do you draw?	Left	Either	Right
3. With which hand do you screw the top off a bottle?	Left	Either	Right
4. With which hand do you deal cards?	Left	Either	Right
5. Which hand do you use to hold a toothbrush?	Left	Either	Right
6. With which hand do you use a bottle opener?	Left	Either	Right
7. With which hand do you throw a ball away?	Left	Either	Right
8. Which hand do you use to hold a hammer?	Left	Either	Right
9. With which hand do you thread a needle?	Left	Either	Right
10. With which hand do you hold a racket when playing tennis?	Left	Either	Right
11. With which hand do you open the lid of a small box?	Left	Either	Right
12. With which hand do you turn a key?	Left	Either	Right
13. With which hand do you cut a cord with a knife?	Left	Either	Right
14. With which hand do you stir with a spoon?	Left	Either	Right
15. With which hand do you use an eraser on paper?	Left	Either	Right
16. With which hand do you strike a match?	Left	Either	Right
17. With which hand do you write?	Left	Either	Right

Music Experience Questionnaire

1. Do you have any formal musical training (vocal or instrumental)? No Yes
    i. What you have been trained in:____
    ii. How old you were when you received this training:________________________
    iii. How many years have you studied:______________________________________
2. How many members of your family sing to you when you were a child?

    ___ Number of People
3. Was choral or individual singing encouraged in your childhood environment?

    No Yes
4. How often did singing occur in your childhood environment?

    Never Rarely Sometimes Often Always
5. How often did you hear music in your childhood environment?

    Never Rarely Sometimes Often Always
6. Were any of your family members particularly fond of music?

    ___ Number of People
7. Were musical instruments played in your childhood environment?

    No Yes
8. Types of musical education (e.g., private, group, self-taught, conservatory examinations).

    ___ Number of Types
9. Number of instruments played

    ___ Number of Instruments
10. Years of training on primary instrument

    ___ Number of Years

11. At the peak of your interest, how many hours per week did you play/practice this instrument?

___ Number of Hours

12. Regarding your peak of interest (10), how long did you maintain this peak?

___ Number of Years

13. Given the opportunity, my interest in participating in future musical instruction is:

Non-Existent Low Neutral High Very High

14. I sing in private (e.g., in my car, in the shower, in my environment)

Never Rarely Sometimes Often Always

15. I sing in public (as part of a group or solo: e.g., a choir, carols, a sing-a-long, with friends)

Never Rarely Sometimes Often Always

16. How often do you purposely listen to music, as opposed to music in your environment that you had no part in choosing, e.g., music in stores, elevators, and restaurants?

Never Rarely Sometimes Often Always

17. When you listen to music, how difficult is it to hear the difference between the notes?

Very Difficult Difficult Indifferent Easy Very Easy

18. How difficult do you find singing in general?

Very Difficult Difficult Indifferent Easy Very Easy

19. Rate your ability to memorize a short song.

Non-existent Poor Average Good Excellent

20. I find it hard/easy to repeat a tune someone else has recently sung to me.

Very Difficult Difficult Indifferent Easy Very Easy

21. If I imagine the tune Happy Birthday, I can hear the melody in my head.

Not able Inaccurately able Average Able Accurately able

22. When music is being played in my environment (e.g., on the radio, in a store, on TV), I can recognize familiar songs by the first two or three notes.

Never Rarely Sometimes Often Always

23. I find it hard/easy match the notes and to sing or hum along with my favourite recorded music.

Very Difficult Difficult Indifferent Easy Very Easy

24. Singing a note to match one played on the piano is a task I find:

Very Difficult Difficult Indifferent Easy Very Easy

25. If someone played two notes on the piano, separately, and asked me which was higher in pitch, I would find this task:

Very Difficult Difficult Indifferent Easy Very Easy

26. When I sing, I can tell when I'm out of tune.

Not able Inaccurately able Average Able Accurately able

27. When I sing, I perform best when I am:

Solo In a Group

28. How often do you get a tune stuck in your head?

Never Rarely Sometimes Often Always

Appendix B

Experimental Instructions:

Welcome to Singing Training study in Dr. Jeffery Jones' Lab.

Step 1 – Consent Form: Please read the consent form and sign and date at the bottom.

Step 2 – Questionnaires: Just to get to know you better, please fill out the questionnaire!

Step 3 – Getting prepared: Please put on the headphones in front of you, the researcher will come and adjust them. Once on, please do not take them off unless the researcher says so!

Step 4 – Setting up: Do you know what is your singing range? Are you a Soprano, Alto, Tenor, or Bass? If you don't know we can find out. Please sing AAHHH to the most comfortable note.

Step 5 – Pre-test Phase: You will see on the screen a red square and you will hear the note we picked. That note is now your target note; the note you will need to keep matching. This red square will turn into a green square, which means GO! So as soon as you see it, we want you to start singing the target note for as long as the green square is up (which is about 5 seconds, and one breath's worth of an AAHHH. Then the red square will re-appear and you will be reminded of your target note and that's when you should listen carefully to it while you have about a 5 second break to catch your breath. You will do this about 100 times *(50 times for Experiment 2)* and take breaks after blocks of 25 trials to give your voice a break. That is also when you can drink some water in the cup, which is provided to you. (In between blocks: This is one of your breaks, if you need to take a sip, now is the time).

Step 6 – Training Phase *(Feedback Condition)*: In front of you now is the training program you will be using to practice and get better at that target note we chose. You must follow the steps provided on the screen for the duration of this phase. First, you must press the button to remind you of the target note you are practicing. Make sure you press it at least once every time before you start to vocalise. When you are ready, you will press the red record button and begin to sing an AAHHH sound to that note. The record button will turn off by itself, after you have vocalised for about 5 seconds as you've been doing before. While you are vocalising, you will see your voice being plotted on the graph in blue in real-time. As you can see the center of the graph shows the line of what the perfect note would be, but we give you some leeway and so you must aim for the white range around the note for an acceptable accuracy. Once you are finished your vocalisation, you will see a percent performance rating based on how long your pitch remained in the acceptable accuracy range. You have to keep doing this until the counter at the top reaches 100 trials *(200 trials for Experiment 2)* in hopes of improving your pitch accuracy. Feel free to take a few moments break in between trials and a sip if you feel your voice is exhausted.

Step 6 – Training Phase *(No-Feedback Condition)*: In front of you now is the training program you will be using to practice and get better at that target note we chose. You must follow the steps provided on the screen for the duration of this phase. To start, you will press the red record button and begin to sing an AAHHH sound at the target note you remember we chose before. The record button will turn off by itself, after you have vocalised for about 5 seconds as you've been doing before. While you are vocalising, you will see your voice being plotted <u>as a straight line</u> on the screen. You have to keep doing this until the counter at the top reaches 100 trials *(200 trials for Experiment 2)* in hopes of improving your pitch accuracy. Feel free to take a few moments break in between trials and a sip if you feel your voice is exhausted.

Step 7 – Post-test Phase: Repeat Step 5.

Step 8 – Disassembling: Please take off the headphones now.

Step 9 – Debrief: Congratulations! You successfully finished the study! Do you have any questions? Do you know what the study's about? *(Discuss).* Thank you very much for participating!

References

Apfelstadt, H. (1984). Effects of melodic perception instruction on pitch discrimination and vocal accuracy of kindergarten children. *Journal of Research in Music Education*, *32*(1), 15-24.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia. *Brain*, *125*(2), 238-251.

Bradshaw, E., & McHenry, M. A. (2005). Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *Journal of Voice*, *19*(3), 431-439.

Bauer, J. J., & Larson, C. R. (2003). Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique. *The Journal of the Acoustical Society of America*, *114*(2), 1048-1054.

Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America*, *119*(4), 2363-2371.

Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology*, *5*, 69.

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, *103*(6), 3153-3161.

Burnett, T. A., McCurdy, K. E., & Bright, J. C. (2008). Reflexive and volitional voice fundamental frequency responses to an anticipated feedback pitch error. *Experimental brain research, 191*(3), 341-351.

DeMarco, A. R., Scheerer, N. E., & Jones, J. A. (2014, December). Investigating the Role of Repeated Exposure to Alterations of Vocal Auditory Feedback. In *Canadian Journal of*

Experimental Psychology (Vol. 68, No. 4, pp. 261-261). 141 Laurier Ave West, Suite 702, Ottawa, Ontario K1P 5J3, Canada: Canadian Psychological Association.

Elman, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *The Journal of the Acoustical Society of America*, *70*(1), 45-50.

Callaghan, J. (2000). Singing and voice science. San Diego, California: Singular Pub. *Group*.

Callaghan, J., Thorpe, W., & van Doorn, J. (2004, April). The science of singing and seeing. In *Proceedings of Conference of Interdisciplinary Musicology, Graz, Austria*.

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

Christiner, M., & Reiterer, S. M. (2013). Song and speech: examining the link between singing talent and speech imitation ability. *Frontiers in Psychology*, *4*, 874

Cuddy, L. L., Balkwill, L. L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare. *Annals of the New York Academy of Sciences*, *1060*(1), 311-324.

Dalla Bella, S., Berkowska, M., & Sowiński, J. (2011). Disorders of pitch production in tone deafness. *Frontiers in Psychology*, *2*, 164

Dalla Bella, S., Giguère, J. F., & Peretz, I. (2007). Singing proficiency in the general population. *The Journal of the Acoustical Society of America*, *121*(2), 1182-1189.

Dalla Bella, S., Giguère, J. F., & Peretz, I. (2009). Singing in congenital amusia. *The Journal of the Acoustical Society of America*, *126*(1), 414-424.

Demorest, S. M., & Pfordresher, P. Q. (2015). Singing accuracy development from K-adult: A comparative study. *Music Perception: An Interdisciplinary Journal*, *32*(3), 293-302.

Demorest, S. M., Pfordresher, P. Q., Bella, S. D., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015). Methodological perspectives on singing accuracy: an introduction to the special issue on singing accuracy (Part 2). *Music Perception: An Interdisciplinary Journal*, *32*(3), 266-271.

Dennis, C. C. (1975). The conditioning of a pitch response using uncertain singers. *Research in music behavior: Modifying music behavior in the classroom*, 139-150.

Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., & Wicha, N. Y. (2011). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Frontiers in Psychology*, *2*, 146

Granot, R. Y., Israel-Kolatt, R., Gilboa, A., & Kolatt, T. (2013). Accuracy of pitch matching significantly improved by live voice model. *Journal of Voice*, *27*(3), 13-20.

Grell, A., Sundberg, J., Ternström, S., Ptok, M., & Altenmüller, E. (2009). Rapid pitch correction in choir singers. *The Journal of the Acoustical Society of America*, *126*(1), 407-413.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological cybernetics*, *72*(1), 43-53.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological review*, *102*(3), 594.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of communication disorders*, *39*(5), 350-365.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and language*, *96*(3), 280-301.

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of neurolinguistics*, *25*(5), 408-422.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., & Kenney, M. K. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Experimental Brain Research*, *130*(2), 133-141.

Harries, M. L., Griffith, M., Walker, J., & Hawkins, S. (1996). Changes in the male voice during puberty: Speaking and singing voice parameters. *Logopedics Phonatrics Vocology*, *21*(2), 95-100.

Hawco, C. S., & Jones, J. A. (2009). Control of vocalization at utterance onset and mid-utterance: Different mechanisms for different goals. *Brain research*, *1276*, 131-139.

Hawco, C. S., & Jones, J. A. (2010). Multiple instances of vocal sensorimotor adaptation to frequency-altered feedback within a single experimental session. *The Journal of the Acoustical Society of America*, *127*(1), 13-18.

Heinks-Maldonado, T. H., & Houde, J. F. (2005). Compensatory responses to brief perturbations of speech amplitude. *Acoustics Research Letters Online*, *6*(3), 131-137.

Hoppe, D., Sadakata, M., & Desain, P. (2006). Development of real-time visual feedback assistance in singing training: a review. *Journal of computer assisted learning*, *22*(4), 308-316.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213-1216.

Howard, D. M., & Welch, G. F. (1989). Microcomputer-based singing ability assessment and development. *Applied Acoustics*, *27*(2), 89-102.

Howell, P., & Sackin, S. (2002). Timing interference to speech in altered listening conditions. *The Journal of the Acoustical Society of America*, *111*(6), 2842-2852.

Hutchins, S., & Moreno, S. (2013). The Linked Dual Representation model of vocal perception and production. *Frontiers in Psychology*, *4*, 825.

Hutchins, S., Zarate, J. M., Zatorre, R. J., & Peretz, I. (2010). An acoustical study of vocal pitch matching in congenital amusia. *The Journal of the Acoustical Society of America*, *127*(1), 504-512.

Hutchins, S. M., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology: General*, *141*(1), 76.

Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental brain research*, *190*(3), 279-287.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, *108*(3), 1246-1251.

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: studies of Mandarin tone production. *Journal of Phonetics*, *30*(3), 303-320.

Jones, J. A., & Munhall, K. G. (2005). Remapping auditory-motor representations in voice production. *Current Biology*, *15*(19), 1768-1772.

Kennell, R. (2013). Toward a methodology of vocal pedagogy research. *The phenomenon of singing*, *1*, 129-137.

Keough, D., Hawco, C., & Jones, J. A. (2013). Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. *BMC neuroscience*, *14*(1), 25.

Keough, D., & Jones, J. A. (2009). The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing. *The Journal of the Acoustical Society of America*, *126*(2), 837-846.

Kramer, S. J. (1985). *The effects of two different music programs on third and fourth grade children's ability to match pitches vocally* (Doctoral dissertation, Rutgers, The State University of New Jersey).

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*(3), 346-374.

Larson, C. R., Altman, K. W., Liu, H., & Hain, T. C. (2008). Interactions between auditory and somatosensory feedback for voice F 0 control. *Experimental Brain Research*, *187*(4), 613-621.

Lévêque, Y., Giovanni, A., & Schön, D. (2012). Pitch-matching in poor singers: human model advantage. *Journal of Voice*, *26*(3), 293-298.

Liu, H., & Larson, C. R. (2007). Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. *Journal of the Acoustical Society of America*, *122*(6), 3671-3677.

Merrill, J. (2013). *Song and speech perception: Evidence from fMRI, lesion studies and musical disorder* (Doctoral dissertation, Max Planck Institute for Human Cognitive and Brain Sciences Leipzig).

Mürbe, D., Pabst, F., Hofmann, G., & Sundberg, J. (2002). Significance of auditory and kinesthetic feedback to singers' pitch control. *Journal of Voice*, *16*(1), 44-51.

Natke, U., Donath, T. M., & Kalveram, K. T. (2003). Control of voice fundamental frequency in speaking versus singing. *The Journal of the Acoustical Society of America*, *113*(3), 1587-1593.

Özdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage*, *33*(2), 628-635.

Patel, S., Nishimura, C., & Larson, C. (2013). Effects of vocal training on voluntary responses to pitch-shifted voice auditory feedback. *The Journal of the Acoustical Society of America, 133*(5), 3339-3339.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature neuroscience*, *6*(7), 688-691.

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., Paquette, S. & Bouchard, B. (2008). On-line identification of congenital amusia. *Music Perception: An Interdisciplinary Journal*, *25*(4), 331-343.

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, *25*(5), 382-407.

Pfordresher, P. Q., & Beasley, R. T. (2014). Making and monitoring errors based on altered auditory feedback. *Frontiers in Psychology*, *5*, 914.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of 'tone-deafness'. *Music Perception*, *25(2)*, 95–115.

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, perception, & psychophysics*, *71*(6), 1385-1398.

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *The Journal of the Acoustical Society of America*, *128*(4), 2182-2190.

Pfordresher, P. Q., Demorest, S. M., Bella, S. D., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015). Theoretical perspectives on singing accuracy: an introduction to the special issue on singing accuracy (Part 1). *Music Perception: An Interdisciplinary Journal*, *32*(3), 227-231.

Pfordresher, P. Q., & Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive psychology*, *70*, 31-57.

Pfordresher, P., & Palmer, C. (2002). Effects of delayed auditory feedback on timing of music performance. *Psychological Research*, *66*(1), 71-79.

Purcell, D. W., & Munhall, K. G. (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120*(2), 966-977.

Rossiter, D., & Howard, D. M. (1996). ALBERT: a real-time visual feedback computer tool for professional vocal development. *Journal of voice: official journal of the Voice Foundation*, *10*(4), 321-336.

Saitou, T., & Goto, M. (2009, September). Acoustic and perceptual effects of vocal training in amateur male singing. In *INTERSPEECH* (pp. 832-835).

Scheerer, N. E., Behich, J., Liu, H., & Jones, J. A. (2013a). ERP correlates of the magnitude of pitch errors detected in the human voice. *Neuroscience*, *240*, 176-185.

Scheerer, N. E., & Jones, J. A. (2012). The relationship between vocal accuracy and variability to the level of compensation to altered auditory feedback. *Neuroscience letters*, *529*(2), 128-132.

Scheerer, N. E., & Jones, J. A. (2014). The predictability of frequency-altered auditory feedback changes the weighting of feedback and feedforward input for speech motor control. *European Journal of Neuroscience*, *40*(12), 3793-3806.

Scheerer, N. E., Liu, H., & Jones, J. A. (2013b). The developmental trajectory of vocal and event-related potential responses to frequency-altered auditory feedback. *European Journal of Neuroscience*, *38*(8), 3189-3200.

Seashore, C. E., & Jenner, E. A. (1910). Training the voice by the aid of the eye in singing. *Journal of Educational Psychology*, *1*(6), 311-320.

Siupsinskiene, N., & Lycke, H. (2011). Effects of vocal training on singing and speaking voice characteristics in vocally healthy adults and children based on choral and nonchoral data. *Journal of voice*, *25*(4), 177-189.

Smith, R. B. (1963). The effect of group vocal training on the singing ability of nursery school children. *Journal of Research in Music Education*, 137-141.

Stegemöller, E. L., Skoe, E., Nicol, T., Warrier, C. M., & Kraus, N. (2008). Music training and vocal production of speech and song.

Sundberg, J. (1987). The Science of the Singing Voice. DeKalb, IL: Northern Illinois University Press.

Sundberg, J. (2001). Level and center frequency of the singer's formant. *Journal of voice*, *15*(2), 176-186.

Sundberg, J. (2007). The human voice in speech and singing. In *Springer handbook of acoustics* (pp. 669-712). Springer New York.

Titze, I. R., & Martin, D. W. (1998). Principles of voice production. *The Journal of the Acoustical Society of America*, *104*(3), 1148-1148.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, *39*(3), 1429-1443.

Tsang, C. D., Friendly, R. H., & Trainor, L. J. (2011). Singing development as a sensorimotor interaction problem. *Psychomusicology: Music, Mind and Brain*, *21*(1-2), 31.

Wakefield, G. H. (2003). Vocal pedagogy and pedagogical voices. Proceedings of the 9th International Conference on Auditory Display (ICAD2003), Boston, MA, July 7-9, 2003. Eds. Eoin Brazil and Barbara Shinn-Cunningham. International Community for Auditory Display, 2003.

Wallin, N. L., & Merker, B. (2001). *The origins of music*. MIT press.

Watts, C., Moore, R., & McCaghren, K. (2005). The relationship between vocal pitch-matching skills and pitch discrimination skills in untrained accurate and inaccurate singers. *Journal of Voice*, *19*(4), 534-543.

Watts, C., Murphy, J., & Barnes-Burroughs, K. (2003). Pitch matching accuracy of trained singers, untrained subjects with talented singing voices, and untrained subjects with nontalented singing voices in conditions of varying feedback. *Journal of Voice*, *17*(2), 185-194.

Welch, Graham F. "Vocal range and poor pitch singing." *Psychology of Music* (1979).

Welch, G. F. (1985a). A schema theory of how children learn to sing in tune. *Psychology of music*, *13*(1), 3-18.

Welch, G. F. (1985b). Variability of practice and knowledge of results as factors in learning to sing in tune. *Bulletin of the Council for Research in Music Education*, 238-247.

Welch, G. F. (2005). Singing as communication. *Musical communication*, 239-259.

Welch, G. F., Himonides, E., Howard, D. M., & Brereton, J. (2004, April). VOXed: Technology as a meaningful teaching aid in the singing studio. In *Proceedings of the conference on interdisciplinary musicology*.

Welch, G. F., Howard, D. M., Himonides, E., & Brereton, J. (2005). Real-time feedback in the singing studio: an innovatory action-research project using new voice technology. *Music Education Research*, *7*(2), 225-249.

Welch, G. F., Howard, D. M., & Rush, C. (1989). Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychology of Music*, *17*(2), 146-157.

Wilson, P. H., Lee, K., Callaghan, J., & Thorpe, C. W. (2007, August). Learning to sing in tune: Does real-time visual feedback help. In CIM07: 3rd Conference on Interdisciplinary Musicology, Tallinn, Estonia (Vol. 2, No. 1, p. 1519).

Wulf, G., & Prinz, W. (2001). Directing attention to movement effects enhances learning: A review. *Psychonomic bulletin & review*, *8*(4), 648-660.

Zendel, B. R., Lagrois, M. É., Robitaille, N., & Peretz, I. (2015). Attending to Pitch Information Inhibits Processing of Pitch Information: The Curious Case of Amusia. *The Journal of Neuroscience*, *35*(9), 3815-3824.