

Wilfrid Laurier University

Scholars Commons @ Laurier

---

Theses and Dissertations (Comprehensive)

---

2009

## Neural and Behavioral Responses to the Use of Auditory Feedback in Vocal Control

Colin S. Haweo  
*Wilfrid Laurier University*

Follow this and additional works at: <https://scholars.wlu.ca/etd>



Part of the [Behavior and Behavior Mechanisms Commons](#)

---

### Recommended Citation

Haweo, Colin S., "Neural and Behavioral Responses to the Use of Auditory Feedback in Vocal Control" (2009). *Theses and Dissertations (Comprehensive)*. 913.  
<https://scholars.wlu.ca/etd/913>

This Thesis is brought to you for free and open access by Scholars Commons @ Laurier. It has been accepted for inclusion in Theses and Dissertations (Comprehensive) by an authorized administrator of Scholars Commons @ Laurier. For more information, please contact [scholarscommons@wlu.ca](mailto:scholarscommons@wlu.ca).



Library and  
Archives Canada

Bibliothèque et  
Archives Canada

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
*ISBN: 978-0-494-49982-5*  
*Our file* *Notre référence*  
*ISBN: 978-0-494-49982-5*

**NOTICE:**

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

**AVIS:**

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**



NEURAL AND BEHAVIORAL RESPONSES TO THE USE OF AUDITORY  
FEEDBACK IN VOCAL CONTROL

by

Colin S. Hawco

Bachelor of Science, Dalhousie University, 2001

THESIS

Submitted to the Department of Psychology

In partial fulfillment of the requirements for

Master of Science, Brain and Cognition

Wilfrid Laurier University

2009

© Colin S. Hawco 2009

**Author:**

**Colin S. Hawco**

**Cognitive Neuroscience of Communications Laboratory**

**And Language and Cognition Laboratory**

**Department of Psychology**

**Wilfrid Laurier University**

**Waterloo, Ontario, Canada**

**Supervisors:**

**Dr. Jeffery A. Jones**

**Cognitive Neuroscience of Communications Laboratory**

**and Centre for Cognitive Neuroscience**

**Department of Psychology**

**Wilfrid Laurier University**

**Waterloo, Ontario, Canada**

**Dr. Todd R. Ferretti**

**Language and Cognition Laboratory**

**And Centre for Cognitive Neuroscience**

**Department of Psychology**

**Wilfrid Laurier University**

**Waterloo, Ontario, Canada**

**Advisory Committee:**

**Dr. Philip Servos**

**Perception Laboratory**

**Centre for Cognitive Neuroscience**

**Department of Psychology**

**Wilfrid Laurier University**

**Waterloo, Ontario, Canada**

**Dr. Steven Perry**

**Department of Kinesiology**

**Wilfrid Laurier University**

**Waterloo, Ontario, Canada**

## Abstract

A large body of evidence suggests that the motor system maintains a forward model that predicts the sensory outcome of movements. When sensory feedback does not match the predicted consequences, a compensatory response corrects for the motor error and the forward model is updated to prevent future errors. Like other motor behaviours, vocalization relies on sensory feedback for the maintenance of forward models and to stabilize vocalizations.

Experiment 1 used event-related potentials (ERPs) to examine sensory processing of short feedback perturbations during an ongoing utterance. In one session, participants produced a vowel at an F0 of their own choosing. In another session, participants matched the F0 of a cue voice. An F0 perturbation of 0, 25, 50, 100, or 200 cents was introduced for 100 ms. A mismatch negativity (MMN) was observed. Differences between sessions were only found for 200 cents perturbations. Reduced compensation when speakers experienced the 200 cents perturbations suggests that this larger perturbation was perceived as externally generated. The presence of an MMN, and no earlier (N100) response suggests that the underlying sensory process used to identify and compensate for errors in mid-utterance may differ from feedback monitoring at utterance onset.

In Experiment 2, we used a frequency altered feedback (FAF) paradigm to study the role of auditory feedback in the control of vocal pitch (F0). We adapted participants to a one semitone shift and induced a perturbation by briefly removing the altered feedback. This was compared to a control block in which a 1 semitone perturbation was introduced into an unshifted trial, or trials were randomly shifted up 1 semitone, and a perturbation was introduced by removing the feedback alteration. The compensation response to mid-utterance perturbations was identical in all conditions, and was always smaller than the compensation to a shift at utterance onset. These results are explained by a change in the control strategy at utterance onset and mid-

utterance. At utterance onset, auditory feedback is compared to feedback predicted by a forward model to ensure the pitch goal is achieved. However, after utterance onset, the control strategy switches and stabilization is maintained by comparing feedback to previous F0 production.

Experiment 1 showed a MMN in response to a mid-utterance perturbation, which is distinct from the N100 found in previous studies that examined perturbations at utterance onset. This result suggests that there may be different underlying neurological mechanisms for the detection of perturbations at utterance onset and mid-utterance. Experiment 2 adds support for this idea by showing a difference in the compensation responses to mid-utterance and onset perturbations. We conclude that different mechanisms may be used to detect errors and compensate for these errors at utterance onset versus mid-utterance.

**Keywords:** Speech Production, Auditory Feedback, Vocalization, Event-Related Potentials, Sensory-motor Adaptation, Mismatch Negativity, Perturbation, Pitch-shift Reflex



## Acknowledgments

I would like to thank my advisors, Jeff Jones and Todd Ferretti for all for their help in completing my masters thesis, and for making it an enjoyable experience. Jeff was readily available whenever I needed some direction, and worked very hard to make sure that I stayed on the right track, and that I didn't get too frustrated when things didn't go my way. Todd offered great insight into my data, and invaluable advice, as well as making sure I could support my theoretical arguments with a sound foundation of data.

I would also like to thank the members of the Cognitive Neurosciences of Communications lab for their help and companionship. In particular, I'd like to thank Dwayne Keough for lots of discussion, and Farina Pinock and Danielle Culbert for their help and hard work collecting the data for Experiment 2. I'd also like to thank Michele Jarick for being a good friend, and offering advice and sometimes just a distraction, when it was needed.

Lastly but not lastly, I'd like to thank My wife Alexia, who left her home, and all her friends and family, to follow me here so I could pursue my goals and dream. Without her, I'd have never been able to finish my Masters. I owe her a debt I can never repay.

## Table of Contents

Abstract .....	iv
Acknowledgments .....	vi
Table of Contents .....	vii
List of Figures .....	viii
List of Abbreviations .....	ix
Chapter 1: General Introduction .....	9
Internal Models and Efference Copy .....	10
Vocal Adaptation .....	12
Compensation to Unanticipated Perturbations .....	14
Purpose of Thesis and Hypothesis .....	17
Chapter 2 Experiment 1 .....	19
Experiment 1 Background .....	20
Experiment 1 Methods .....	26
Experiment 1 Results .....	31
Experiment 1 Discussion .....	35
Chapter 3: Experiment 2 .....	45
Experiment 2 Background .....	46
Experiment 2 Methods .....	50
Experiment 2 Results .....	54
Experiment 2 Discussion .....	58
Experiment 2 Conclusions .....	65
Chapter 4: Conclusions .....	66
References .....	69

## List of Figures

Figure 2.1: Timeline for a single trial in Experiment 1. ....	28
Figure 2.2: Behavioral results for Experiment 1. ....	32
Figure 2.3: ERP results for absolute and relative sessions for all 5 shift conditions for Experiment 1. ....	33
Figure 2.4: Grand Average ERP results at Cz for Experiment 1. ....	34
Figure 3.1: Schematic representation of the four shifts presented in the different conditions for Experiment 2. ....	52
Figure 3.2: Median F0 at utterance onset (first 50 ms) and median F0 for the whole utterance for Experiment 2. ....	55
Figure 3.3: Compensation responses for the 3 perturbation conditions in Experiment 2. ....	56

## **List of Abbreviations**

ERP – Event-related potential

F0 – Fundamental Frequency (vocal pitch)

F1 – The first formant frequency in vowel production

FAF – Frequency altered feedback

fMRI – Functional Magnetic Resonance Imaging

MEG – Magnetoencephalography

MMN – Mismatch Negativity

PSR – Pitch-shift reflex

## Chapter 1

### General Introduction

During motor control, sensory feedback is used as the basis to correct online for errors of motor production and to adapt to novel environments and conditions, which might affect our motor output. In the case of reaching for an object, this feedback takes the form of kinesthetic feedback (an awareness of the location of the limb in space relative to the rest of the body, and the posture and muscle tension of that limb) and visual feedback, which allows a comparison of the current position of the limb to the desired final limb position. The use of feedback allows us to correct online for errors during movements, as we can observe errors during movement, and also to adjust future movements to prevent such errors from occurring again.

During speech, auditory feedback plays an important role in vocal control. Vocalization relies on auditory feedback for learning to correct for errors (Guenther, 2006) and maintaining verbal fluency. In post-lingual deafness, a situation in which auditory feedback is no longer available, a progressive decrease in vocal quality is observed, though these individuals still remain intelligible (Cowie & Douglas-Cowie, 1983). Auditory feedback is important for online correction of vocal productions (Larson, 1998; Burnett, Freedland, Larson, & Hain, 1998) and maintenance of stored motor commands for vocal production (Guenther, 2006; Jones & Munhall, 2000, 2005; Purcell & Munhall, 2006; Villacorta, Perkell, & Guenther, 2007).

While feedback plays an important role in monitoring motor production and the online detection of errors, feedforward control is used to initiate and plan actions. Feedforward control involves accessing and initiating stored motor plans relevant to the desired goal. Without a feedforward control system, every motor command would have to initiate with an arbitrary start point and adjust online for position and velocity effects to produce the desired motor act. Given

that feedback information is not available until some time after motor initiation (due to an inherent delay in information transfer), a reliance on online compensations would not result in the smooth and well controlled movements observed in all animals.

At present, studies of speech motor control have largely focused on behavioral studies, emphasizing changes to motor output following feedback manipulations, and a small number of imaging studies, which focus on the areas of the brain involved in motor control and feedback monitoring. While these studies have proven to be very informative, we still have a poor understanding of how the brain detects and responds to manipulations in auditory feedback. Also, many studies of speech and theories of speech motor control have the implicit assumption that all feedback errors are treated equally. However, this may not be the case. During motor production, feedback errors can be brief or prolonged, and can occur at utterance onset and during an ongoing utterance. The purpose of this thesis is to use event-related potential (ERP) and behavioral methods to study speech motor control, examining how the auditory system detects different feedback errors, and how the motor system responds to those errors. These results will be considered within a context of the theory of internal models, and feedback and feedforward control mechanisms.

### **Internal Models and Efference Copy**

The reliable and accurate performance of motor commands is a complex process which must not only select an appropriate set of muscle movements and contractions to reach the desired goal, but also account for context and environmental factors. For example, when we are lifting a heavy or light object, we might use the same motor configuration, but use significantly

more force for the heavy object than the light object. Alternately, we might adjust motor commands for different environmental contexts, such as walking through water, where we must account for the resistance of the water (which is modulated by the depth at our current location) to maintain a steady gait and balance.

Internal models are one system which has been proposed to explain how feedforward controllers might account for the wide array of motor commands and motor contexts encountered in our every day lives (Wolpert and Kawato, 1998). The theory of internal models proposed two types of controllers, inverse models and forward models, which are paired to produce motor commands. Forward models predict the upcoming state of the system based on current state and active motor command. Forward models have been proposed to be used in motor learning (Sutton and Barto, 1981; Jordan and Rumelhart, 1992). Inverse models, in contrast, select the appropriate motor command to achieve the desired motor output. Generally, feedback manipulations are presumed to act on the forward model, rather than the inverse model, and thus we will not further consider the inverse model within the current discussion.

The forward model includes an “efference” comparator, a copy of the motor command sent to the sensory cortex to allow a prediction of the sensory consequences (Nowak et al. 2007). When feedback does not match the efference copy during a motor command, it is registered as an error, and provokes some sort of compensation response. The efference serves the critical function of allowing us to differentiate between self-generated and externally-generated stimuli (Blakemore, Wolpert, & Frith, 2000). Several studies have shown evidence of suppression of early sensory responses during vocalizations. Suppression of activity in the middle and superior temporal gyri during vocalization has been found in electrode implantation studies in humans (Creutzfeldt, Ojemann, & Lettich, 1989), demonstrating a dampening of the auditory cortical

activity during vocal production. This dampening of activity in the auditory cortex is believed to be caused by the efference copy suppressing the sensory response to our own voice. A similar dampening occurs within the somatosensory system (Blakemore, Wolpert, & Frith, 2000).

### **Vocal Adaptation**

When a feedback alteration is introduced and left in place for a prolonged period, sensory-motor adaptation occurs as the system adjusts to compensate for this novel feedback context. Feedback alterations can take the form of mechanical loads on the system or a change in auditory feedback associated with vocalization (Jones & Munhall, 2000, 2005; Purcell & Munhall, 2006; Villacorta, Perkell, & Guenther, 2007). This adaptation represents a change in the forward model to maintain appropriate motor control and intelligibility when an error is detected between predicted and actual feedback.

Some studies have altered vowel formants during vowel production. This causes the produced vowel sound to be perceived as a different vowel. For example, Houde and Jordan (1998, 2002) examined sensorimotor representations for formants by shifting F1 and F2 for the vowel / $\epsilon$ / along the / $i$ / - / $\sigma$ / axis. Participants compensated for the feedback alterations by modifying their formant production. These modifications persisted when auditory feedback was removed, demonstrating adaptation occurred within the motor system. Similarly, Purcell and Munhall (2006) gradually shifted F1 during vowel production, and found a gradual return to baseline (de-adaptation) when feedback was abruptly returned to normal. The time of this de-adaptation response was not related to amount of time that the maximal feedback alteration was



maintained. Villacorta et al. (2007) modified F1 in consonant-vowel-consonant words, and likewise found a compensation response that persisted when feedback was removed. They also tested auditory discrimination, and found that participants with better F1 discrimination had a larger compensation response to the feedback alteration in F1, demonstrating a link between perceptual abilities and the motor response to a perceived error.

Manipulations of auditory feedback for fundamental frequency (F0), or vocal pitch, have also been used to study sensorimotor control during vocalizations. F0 is distinct from formant frequencies in that formants must be controlled within each vowel while F0 appears to be controlled suprasegmentally, at least for non-tone languages (Natke and Kalveram, 2001). Jones and Munhall (2000, 2002) slowly shifted F0 over many trials and observed a compensation response in the direction opposite the shift. When the feedback alteration was removed, they found prominent after-effects that suggested a re-mapping in the motor system for F0 control. Examining differences in singers and non-singers, Jones and Keough (2008) introduced an abrupt (as opposed to gradual) feedback change. They found non-singers adjusted to the feedback alteration almost immediately, while singers, who possess superior F0 control, were slower to modify their productions. Singers were slower to adapt and slower to de-adapt than non-singers, suggesting that their internal models were more entrenched than those of non-singers.

The response observed in adaptation studies has two potential components: a predictive adaptation response and an online compensation response. The predictive adaptation response changes the feed-forward motor plan *prior* to motor onset (and thus is present at the beginning of the motor action) and an online compensation response, in which the feedback alteration during the current motor command (e.g., the current utterance) is detected *after* motor onset, when the

sensory system has had time to detect the feedback alteration and send an error signal to the motor system, which engages in a compensatory response. For vocalization, this compensatory response takes approximately 100 to 200 ms (Burnett et al., 1998; Hain, et al., 2000). When examining F0 for a prolonged utterance (such as holding a vowel for 1 or 2 seconds), these responses can be separated by examining the F0 immediately after utterance onset (such as the first 50 ms of production, which is too early for feedback-based compensation to occur) and compare this to the final F0 attained later within the utterance (which is susceptible to an online feedback-driven compensation response).

### **Compensation to Unanticipated Perturbations**

During a sustained vocalization, F0 can be altered after utterance onset (Larson et al., 1998, Burnett et al., 1998). When speakers hear auditory feedback of their F0 suddenly shift in mid-utterance, a compensation response occurs in which the speaker shifts their voice in a direction opposite to that of the perturbation. This response has been termed the 'pitch shift reflex', or PSR, and begins approximately 100 to 150 ms after perturbation onset (Burnett et al., 1998; Hain, et al., 2000; Larson, Burnett, Bauer, Kiran, & Hain, 2001). The observed compensation is generally smaller than the feedback perturbation, except when very small perturbations are used (Larson et al. 2001; Lui & Larson, 2007). This compensation response is a closed-loop system, which relies on ongoing feedback.

Burnett et al. (1998) shifted F0 for periods varying from 100 to 500 ms, and found evidence that the compensation response is made up of two components, an early and a late

response. In a follow up study, Hain et al. (2000) used a 500 ms perturbation, and asked participants to compensate, not to compensate, or actively follow the pitch-shifted feedback when their voice was perturbed. They found that the early component of the pitch-shift reflex was automatic and not affected by task instructions while the late component was under volitional control.

Larson et al. (2001) introduced a feedback perturbation mid-utterance (onset trials), or introduced the perturbation prior to utterance onset, and then removed the perturbation mid-utterance (offset trials). They found that the compensation response was identical in either case. That is to say, introducing a feedback alteration produced the same compensation response as removing a feedback alteration. It is important to consider that participants produced at their habitual pitch (which may have varied substantially across trials), so the baseline F0 was not compared between control trials (where F0 was not shifted at onset) and experimental trials in which the perturbation was present at utterance onset. The results of this study suggest that the PSR is a voice stabilization response, and that the PSR is not using an absolute F0 reference. An absolute F0 reference, such as an efference copy sent from the motor system, should result in differences in compensation to a perturbation onset and a perturbation offset. However, as the baseline F0 is not known, it is possible that, in the case of the offset trials participants compensated at utterance onset, and the offset response observed was not a compensation, but the removal of a previous compensation response. If this was the case, it is possible an absolute F0 referent was being used.

Natke and Kalveram (2001) had participants vocalize a nonsense word (tatatas) with varying patterns of stress and/or long vowels. F0 of auditory feedback was shifted down on 20% of trials, for the whole utterance. They found compensation in response to the F0 shift, but that

this compensation was not observed in the first syllable. This suggests that F0 is controlled suprasegmentally, meaning the F0 compensation response is not meant to control F0 at the syllable level but to maintain overall F0 coherence and stability. The compensation response persists over time, and has an effect on unshifted utterances following a shifted utterance (Donath, Natke, & Kalveram, 2002).

Natke, Donath, and Kalveram (2003) compared singing to speaking a nonsense word. They found greater compensation in singing, suggesting F0 was more tightly controlled during singing than speech. This is in contrast to the results of Chen, Liu, Xu, and Larson (2007), who found a larger compensation response in speech than in vowel production. However, Chen et al. (2007) did not have participants match a target note, which may be distinct from singing (where the goal is to match a specific pitch), and used a question (“you know Nina?”) rather than a nonsense word as their speech stimuli. In the case of a question, with a rising F0, it is possible that F0 control is more tightly controlled than in declarative speech. Natke, Donath, and Kalveram (2003) noted that singing prolonged the period at which an F0 shift could affect a subsequent utterance (relative to speaking). They also found that a participant’s ability to match a target note was correlated with the amount of compensation, with participants who were better at hitting the target note having a larger compensation response. It is unclear if this is because of improved auditory perception, improved motor control, or both.

Compensation responses have also been observed when feedback amplitude (volume) is altered (Bauer, Mittal, Larson, & Hain, 2006). Bauer et al. (2006) shifted voice amplitude during vowel production during soft (low amplitude) or normal production. They found a compensation response similar to the PSR observed after F0 shifts, and that the amplitude compensation response was larger for soft productions. Larson, Sun, and Hain (2007) simultaneously shifted F0

and voice amplitude, and found that F0 and amplitude were controlled by separate but sometimes interacting mechanisms. Studies showing compensation responses following amplitude changes demonstrate that online closed-loop compensation responses are not unique to F0 control, but may generalize to all suprasegmental motor behaviors.

### **Purpose of Thesis and Hypothesis**

The purpose of this thesis was to examine differences in both the auditory response and motor output to perturbations of F0 at utterance onset or mid-utterance during sustained vowel productions. At present, it is largely assumed that feedback at utterance onset is treated the same as feedback during utterance maintenance (i.e., holding a steady F0 during a sustained vowel production). We also have a poor understanding in how errors in feedback at utterance onset and mid-utterance are detected by the brain. To examine these issues, we conducted 2 studies; one using ERPs to examine the sensory response to feedback manipulations, and a behavioral study to test for differences at how feedback is used at utterance onset and mid-utterance.

In Experiment 1, we introduced a random mid-utterance perturbation. We observed a compensation response, and a mismatch-negativity (MMN) in response to the perturbations. This is the first study to use ERPs to examine perturbations during an ongoing utterance. In addition, participants were instructed to either vocalize at a normal/habitual pitch, or to match a target note (which was close to the conversation pitch of most individuals of the participant's gender). This allowed us to compare both the compensation responses to absolute or relative pitch targets, as well as test for differences in the neural response.

Experiment 2 was a behavioral study incorporating aspects of both an adaptation and a random perturbation design. Participants completed 2 blocks. In one block, we adapted participants to a 100 cent upward shift. While the adaptation shift was still in place, we perturbed their voice by randomly removing the shift for a brief period within the middle of the utterance. In the other block, a 100 cent upwards perturbation was randomly introduced at utterance onset for some trials. A mid-utterance perturbation could then be introduced by briefly removing the onset perturbation. This was compared to perturbations introduced into an utterance that had not been shifted at onset. We found an identical compensation response to all mid-utterance perturbations (when the feedback shift was removed during either adaptation or the random onset trials, or when no onset-perturbations was present). In addition, the mid-utterance compensation was much smaller than the compensation at utterance onset.

Given the sensory responses observed in Experiments 1 (and their differences to ERP results reported for shifts at utterance onset) and the differences in the motor output to mid-utterance and onset perturbations in Experiment 2, we believe that different mechanisms are used to monitor feedback at utterance onset and mid-utterance. The difference in these mechanisms is likely related to changes in the goal at onset and during maintenance. When initiating an utterance, the goal is to match a desired F0 value as quickly and accurately as possible. Once this F0 goal is attained, the system shifts into a maintenance mode, which seeks to stabilize the current F0 rather than match a specific F0 target.

## Chapter 2

Experiment 1: Neural correlates of the detection of errors during online auditory feedback monitoring.

Accepted for Publication in *Psychophysiology* as:

Hawco, C., Jones, J.A., Ferretti, T.R. and Keough, D. (In Press). Neural correlates of the detection of errors during online auditory feedback monitoring. *Psychophysiology*

### Author Contributions:

The study was primarily designed by Colin S. Hawco, who also collected and analyzed all data.

Jeffery A. Jones and Todd R. Ferretti supervised the study, and were involved in discussions on the design and analysis, and interpretation of the data.

Dwayne Keough helped with discussions of the design and interpretation of the experiment.

## **Experiment 1 Background:**

The purpose of Experiment 1 is to examine the underlying sensory-neural mechanisms used in monitoring auditory feedback and detecting and correcting for alterations within that feedback. Control of F0, the fundamental frequency, or pitch, of the voice, is often studied using a frequency altered feedback (FAF) paradigm (e.g., Elman, 1981; Kawahara, 1995; Larson, 1998; Burnett et al., 1998; Natke, Donath, & Kalveram, 2003; Jones & Munhall, 2000, 2002, 2005). When speakers hear auditory feedback regarding their F0 suddenly shift in mid-utterance, a compensation response occurs in which the speaker shifts their voice in a direction opposite to that of the perturbation. This response has been termed the 'pitch-shift reflex', and begins approximately 100 to 150 ms after perturbation onset (Burnett et al., 1998; Hain, et al., 2000; Larson, Burnett, Bauer, Kiran, & Hain, 2001).

The pitch-shift reflex occurs as a result of changes in auditory feedback, demonstrating a closed-loop negative feedback system used in online correction of F0 while vocalizing. When unexpected deviations of F0 are heard, a compensatory response is initiated within the motor system to offset the error. The mechanism involved in the compensation process is hypothesized to compare auditory feedback to an 'efference copy' of the motor command that is sent to the auditory system. When the feedback does not match this efference copy, an error is determined to have occurred, and the motor system initiates a correction for the error (Guenther, 2006).

An alternative to the efference copy hypothesis is that speakers might use a variable referent rather than a fixed, internal referent (the efference copy). Larson et al. (2001) conducted an FAF study in which auditory feedback was altered shortly after the onset of an utterance (the onset condition), or the F0 shift was applied prior to utterance onset and removed soon after (the



offset condition). They found similar F0 trajectories for both conditions, suggesting that speakers were using their initial auditory feedback as a variable referent rather than comparing to an absolute fixed F0 goal. The authors suggest that both strategies, a variable referent or a fixed referent, may be used when appropriate.

Another way to examine this issue is to have speakers match a target pitch during vowel production. In that case, the target pitch, rather than the efference copy, could be used as the referent. When feedback matches the target, no adjustment is made, but when it does not match, adjustments are required to compensate for the error. Only a few FAF studies using short perturbations have used a target pitch. Liu and Larson (2007), for example, had speakers match an easy or difficult (high) target pitch. They found more compensation to the high target than to the easy target. Burnett and Larson (2002) had speakers match a target note which was either held steady or rose in pitch (glissandos) and found a compensation response during glissandos, but that the compensation response was smaller in glissandos than in steady-state production. The results of Liu and Larson as well as Burnett and Larson suggest task-dependent modulation of the compensation response, though no studies have compared target matching to vocalizing without a target.

Most FAF studies on the pitch shift response have focused on examining vocal changes in F0 in response to perturbations. These FAF studies have proven to be very informative, but they only allow us to observe the behavioral response to perturbations in auditory feedback. We can use this to infer the underlying mechanisms in feedback and feedforward control, but it is difficult to directly test these hypotheses. fMRI has been used to observe the brain regions involved in vocal control. Studies on perception of voices have consistently found activity in the superior temporal gyrus, and evidence of voice specific areas within this region, with activation

typically greater in the right hemisphere (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Belin, Zatorre, & Ahad, 2002; Fecteau, Armony, Joanette, & Belin, 2005).

Moreover, the right superior temporal gyrus has been found to be more specialized for pitch discrimination than the left (Zatorre, Belin, & Penhune, 2002). During vocalization, similar superior temporal activity has been observed. For example, Fu et al. (2006) conducted a study on verbal self-monitoring, in which speakers heard their own voice, their voice pitch-shifted by 2 semitones, or an alien voice. They found greater activation to alien voices than pitch-shifted voices, and greater activation to pitch-shifted than unshifted voices in the bilateral superior temporal cortices. Toyomura et al. (2007) compared trials in which speaker's voices were not modified to trials in which voices were briefly pitch-shifted up or down three times during a 3 second utterance and found increased activity in the right superior temporal gyrus when feedback was altered.

fMRI studies have provided many insights into the neural networks and mechanisms involved in vocal control, but the slow speed of acquisition makes it difficult to make temporal differentiations between activated areas to determine the timing of these responses. Event-related potentials (ERPs) and magnetoencephalography (MEG) provide us with a high temporal resolution of neural activity, at the cost of spatial resolution. ERPs and MEG can be used to examine sensory processes as they occur, and may be useful in studies of vocal control and perception. Several ERP and MEG studies have shown evidence of suppression of early sensory responses during vocalizations (Ford, Mathalon, Heinks, Kalba, & Roth, 2001; Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; Kudo et al., 2004), consistent with both the efference copy hypothesis of auditory cortex suppression during vocalization and with the results of Fu et al. (2006), who found reduced superior temporal activation during vocalizations.

Ford et al. (2001) found a reduction in the N100 (a sensory component related to perception of an auditory stimulus) during vocalization, with a smaller N100 during speech than when listening to speech playback. Kudo et al., (2004) found a reduced N100 to tones when participants were vocalizing (vocal feedback was not played back to them), but no reduction in other sensory components (the mismatch negativity, negative difference, or P300). The N100 and its magnetic equivalent, the M100, have been localized to the primary auditory cortex and surrounding areas (Hari et al., 1987; Pantev, Eulitz, Hampson, Ross, & Roberts, 1996; Krumbholz, Patterson, Seither-Preisler, Lammertmann, & Lutkenhoner, 2003; Ozaki et al., 2003).

Heinks-Maldonado et al. (2005) had participants vocalize an /a/ sound for 3 seconds, and altered their auditory feedback. They used the N100 to examine the sensory consequences of violations of the efference copy during vocalization and demonstrated that efference copy suppression was sensitive to changes in auditory feedback. During some trials, participants heard their unmodified feedback, while in other trials they heard their feedback perturbed downward by 2 semitones (200 cents), an alien voice (a male not familiar to the participant), or the alien voice shifted down by 2 semitones. They examined the N100 in response to the auditory feedback. They found a suppressed N100 in the unaltered feedback condition. During passive listening to the four feedback conditions without vocalizing, Heinks-Maldonado et al. (2005) found that the N100 response did not differ among conditions but was significantly larger than when participants were actively vocalizing. Heinks-Maldonado, Nagarajan, and Houde (2006) performed a replication of this study using MEG, and localized the response to the superior temporal cortex, showing larger areas of activation to the pitch-shifted and alien conditions than the unaltered voice when participants vocalized.

The ERP and MEG studies described above used perturbations that lasted for entire utterances (and were present at utterance onset). Grimm and Schröger (2005) conducted a study to examine the effects of modulating the frequency of an ongoing tone; similar to the way auditory feedback is altered in a transient perturbation FAF experiment. This experiment is the pure perceptual equivalent to a perturbation experiment using ERPs. Grimm and Schröger (2005) presented a 440 Hz tone for 1000 ms. During 12% of trials, there was a 50 ms period where the tone briefly rose to 480 Hz (approximately 150 cents) at various time intervals within the tone. They found a clear mismatch negativity (MMN) in response to the frequency modulation, with the MMN being smaller when the frequency modulation occurred later within the tone. The MMN is usually observed when a deviation is detected in an auditory stimulus, even when the stimuli are not attended to. Typical MMN studies utilize a stream of standard tones, with 10-20% of tones deviating from the standard in some way (by frequency, duration, or volume; Näätänen, Gaillard & Mäntysalo, 1978; Grimm & Schröger, 2005; and Näätänen, Paavilainen, Alho, Reinikainen & Sams, 1989, respectively).

The presence of an MMN is highly correlated with detection thresholds for deviations (Lang et al., 1995). MMN responses have also been found in response to deviant phonemic contrasts (Aaltonen, Tuomainen, Laine, & Niemi, 1993) and to arbitrary sound patterns, occurring when a change is detected in a predictable sound sequence (Tervaniemi, Maury, & Näätänen, 1994). The N100 and MMN overlap (with the MMN typically peaking 150-250 ms post-stimuli) but the MMN is believed to be distinct from the N100 (see Näätänen, Paavilainen, Rinne, & Alho, 2007, for review) that is related to detection of deviations rather than detection of the stimuli itself. The MMN is often seen in the absence of attention, and is produced by mechanisms that involuntarily direct attention to the deviant sound stimuli. This involuntary shift

in attention accompanied by the presence of an auditory MMN in an unattended auditory channel typically has an adverse effect on reaction time in the attended auditory channel (Schröger, 1996).

For the present study, we conducted an ERP experiment using an FAF paradigm to observe the sensory-neural components involved in the pitch-shift reflex, and to determine if we could observe ERP components in response to an F0 perturbation and how these components relate to the compensation observed in response to perturbations. This study bears certain similarities to that of Grimm and Schröger (2005), who altered ongoing tones. Instead of tones, we altered the participant's ongoing vocal auditory feedback. Participants vocalized for three seconds, hearing their auditory feedback perturbed by 0 (no perturbation), 25, 50, 100, or 200 cents for 100 ms during each utterance. The shift values were expected to lead to progressively larger compensation responses, up to some plateau. Participants produced vocalizations during two sessions: in one session, participants produced an F0 of their own choosing, while in the other they were instructed to match a specific target pitch. While vocalizing, a masking noise (multispeaker babble) was played to partially mask bone-conduction feedback through the chest and oral cavities.

When matching a target, a speaker is able to use the target note as a comparator, rather than the efference copy that is believed to be the standard comparator when vocalizing. Based on previous research (Grimm and Schröger, 2005; Lavikainen, Huottilainen, Ilmoniemi, Simola, & Näätänen, 1995), we hypothesized that an MMN might be observed in response to perturbations, and that the magnitude of this negative component would be related to the size of the pitch shift. Alternatively, we might observe an N100 response that is modulated by the size of the perturbation. It is unclear if an N100 can be observed to a change in an ongoing stimuli so soon

after stimulus onset. If there were differences between matching an external target or using an internal referent, we might see differences in the magnitudes of the ERP components in either the MMN or N100 time window. We also hope to lay groundwork for future studies examining the sensory aspects of auditory feedback.

### **Experiment 1 Methods**

*Participants:* Data were recorded from 21 participants. All participants reported no formal vocal training, did not speak a tonal language such as Mandarin or Cantonese, did not take any psychiatric medications, and were right-handed. Each participant took part in two separate EEG sessions on non-consecutive days, with each session lasting approximately 90 to 120 minutes. Seven participants were excluded from the final analysis (three because of excessive EEG artifacts, such as alpha activity and ocular artifacts, and four because they opted not to participate in a second session), leaving data from 14 participants (mean age 21.8, sd 2.4, 4 males). All participants read and signed an informed consent form, in accordance with the ethical policies of Wilfrid Laurier University.

*Behavioral Task:* Participants were seated in an electrically shielded room for the experiment and wore headphones with an attached boom microphone (Sennheiser HMD 280-13, Holte, Denmark). Participants were instructed to produce the vowel /a/ in a clear voice for 3 seconds following an auditory cue. They were instructed to vocalize in a loud voice (i.e., not shouting, but a loud clear voice such as might be used in a crowded room). Prior to the experiment,

participants practiced vocalizing such that they could hear their voice over the multi-speaker babble. Most participants heard their voice at approximately 85 to 95 dB SPL, depending on the volume at which they vocalized (with their voices being amplified by approximately 10 dB). Their vocalization was played back to them in real time via the headphones. Participants were informed that they would hear their voice shifted in pitch during the experiment.

The experiment was conducted in two sessions (a relative and absolute session) on non-consecutive days, with the order of sessions counterbalanced. Participants were instructed to begin vocalizing following a cue. In the relative session, the cue was a 1000 Hz tone, played for one second. Participants were instructed to begin vocalizing immediately after the end of the tone, at their normal pitch. Participants were instructed to maintain a consistent pitch and volume throughout the production. In the absolute session, the cue was another person's voice producing /a/ for 1 s. Male participants heard a male voice produce the note A3 (110 Hz) and female participants heard a female voice produce the note A4 (220 Hz). The cue was played at approximately 85 dB SPL. They were instructed to match the pitch of the cue voice to the best of their ability. The cues used in the absolute sessions were constructed by taking a male and female voice producing an /a/ sound at the desired target and digitally modifying them such that the pitch was exactly 110 Hz or 220 Hz, respectively. The F0 values of the cues are close to the natural pitch of most male and female speakers.

The cue was followed by multi-speaker babble, used as a masking noise to reduce the effects of bone-conducted feedback. The babble was played at 85 dB SPL and lasted 3 s. Participants were instructed to vocalize until the babble stopped, and thus vocalized for approximately 3 s. Between 1000 and 1200 ms following the end of the cue and onset of the babble sound, the participants' voice was perturbed upwards by either 0, 25, 50, 100, or 200

cents (the shift 0, shift 25, shift 50, shift 100, or shift 200 conditions, respectively), where 100 cents is equal to one semitone. The perturbations lasted 100 ms. The order of perturbations was randomized across trials. The babble was followed by a 3 s period of silence prior to the onset of the next cue (lasting 1 s). Thus, each trial lasted 7 s, with 3 s of vocalization and 4 s of rest. Details of a trial are shown in Figure 2.1. The experiment was divided into 5 blocks of 60 trials, with 12 of each size of perturbation in each block (for a total of 60 instances of each perturbation during the study). Each block lasted 7 minutes, and was followed by a brief break.

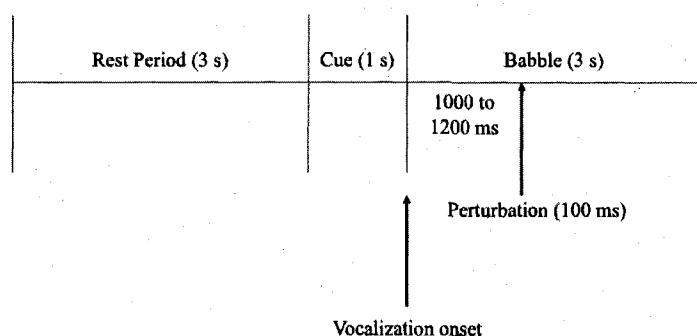


Figure 2.1: Timeline for a single trial. The cue was a 1000 Hz tone in the relative session, and a voice producing /a/ at A3 for males or A4 for females.

*Behavioral Recording and Analysis:* The auditory cue, multi-speaker babble, and shift onset were controlled by Max/MSP 4 (Cycling '74, San Francisco, CA), running on a Dell laptop. The sounds from the laptop were sent to a mixer (Mackie Oynx 1220, Loud Technologies, Woodinville, USA). As well, the participants' vocalizations were sent to the mixer and then sent to a digital signal processor (DSP; VoiceOne, T.C. Hellicon, Westlake Village, USA), which pitch-shifted the voice. The altered voice signal was returned to the mixer, mixed with the auditory signals from the laptop (cue and babble), and played back to the participant. The unaltered voice signal, as well as a tone that was triggered by the command to the DSP to pitch-shift, were digitally recorded (TASCAM HD-P2, Montebello, USA) at a sampling rate of 44.1 kHz.



Voice recordings were imported into a custom program written in Matlab (The Mathworks Inc, Natick, USA) that segmented recordings into separate trials. The trial data were then imported into Praat ([www.praat.com](http://www.praat.com)). The autocorrelation algorithm in Praat was used to calculate the F0 with a resolution of 5 ms. The F0 data were then imported into another Matlab program that time aligned each utterance with the perturbation onset.

The F0 data were converted into cents using the formula:

$$\text{cents} = 100 * (39.86 * \log_{10}(\text{F0}/\text{baseline}))$$

The baseline was defined as the mean F0 of the 50 data points (250 ms) preceding the onset of the perturbation. Any trial that did not have 500 ms of voice data prior to perturbation onset (because the participant was slow to begin production on that particular trial) was excluded, to ensure that the baseline represented steady-state F0 data, and was not affected by fluctuations present early in an utterance. Cents were calculated for 100 data points (500 ms) before and 200 data points (1000 ms) following the perturbation. Any trial that did not have 1000 ms of voice data following the perturbation was also excluded. On average, only 4.5% of trials were rejected for each participant. Data were smoothed with a 7-point moving window, and a linear detrend was performed on each trial to remove any drift in the vocalization (fitting a linear trend line to each trial, including the baseline segment, and removing that trend from the data). An average F0 trace was then constructed for each shift value for each participant.

For each of the averages of each shift value for each participant, the minimum value within the time window from 50 ms to 300 ms post-perturbation was calculated as a measure of the amplitude of the compensation response. Peak latency was calculated for the shift 25, shift 50, shift 100, and shift 200 conditions. No peak latency was calculated for the shift 0 condition, as there was no response in this condition due to the absence of a perturbation. Amplitudes were each analyzed with 2 X 5 (session by shift value) repeated measures ANOVA. Latency was

analyzed with a 2 X 4 (session by shift) ANOVA. All probability values were corrected for multiple degrees of freedom using Huynh-Feldt.

*EEG recording and Analysis:* Data from each session (relative and absolute) were recorded and analyzed separately. An electrode cap with 62 Ag/AgCl electrodes distributed across the scalp (60 channels plus reference at the vertex and a ground between Fz and Fpz) was fitted on to each participant. Data were recorded using a Neuroscan Synamps2 amplifier set at a bandpass of 0.01–250 Hz and digitized at 1000 Hz. Electrodes were also placed at the outer canthus of both eyes, and above and below the left eye to measure EOG artifacts. Electrode impedances below 5 K $\Omega$  were obtained for the majority of electrodes, with some impedances as high as 10 K $\Omega$ . After data acquisition, the EEG was rereferenced to the average of electrodes on each earlobe. Data were then epoched into segments from 100 ms before the onset of perturbation to 1000 ms after perturbation onset. The epoched data were baseline corrected, filtered with a band pass of 1 Hz (6 dB/octave) to 30 Hz (48 dB/octave) and any trials that exceeded 50 microvolts on any electrode were excluded from further analysis. On average, 22% of trials were rejected. Three participants were excluded due to high rejection rates (over 60% of trials), as described above.

For each participant, averaged waveforms were created for each shift type (i.e., shift 0, shift 25, shift 50, shift 100, and shift 200) for each electrode. Grand average waveforms were created for each of the 5 shift conditions in each of the two sessions by averaging the data from all participants for each electrode, resulting in 10 grand average files of 60 electrodes each. A subtraction waveform was created to visualize the MMN by subtracting the data from the shift 0 condition from all other conditions for both the relative and absolute session. For all average files for each participant, mean amplitudes were calculated for time windows from 50-150 ms (area

1), 151-250 ms (area 2), 251-350 ms (area 3), 351-450 ms (area 4), and 451-550 ms (area 5).

These windows were chosen based on a visual inspection of the data and on prior research (Grimm & Schröger, 2005). For each time window, a separate 2 X 5 X 60 repeated measures ANOVA (session by shift by electrode) was conducted, corrected for multiple degrees of freedom using Huynh-Feldt.

### **Experiment 1 Results**

*Behavioral Results:* Figure 2.2 shows plots of the F0 of the vocalizations (in cents). For peak latency of the compensation response, the 2 X 4 (session X shift) repeated measures ANOVA showed a significant main effect of shift,  $F(4,52) = 5.22$ ,  $p = 0.0048$ , but no main effect of session, and no shift by session interaction. Pairwise comparisons for the 4 shift values indicated no latency differences between shift 100 and shift 200. However, the shift 100 and shift 200 conditions had significantly shorter latencies than shift 25 and shift 50 conditions, which did not differ from each other. The 2 X 5 ANOVA for the amplitude of the compensation response showed a significant main effect of shift,  $F(4,52) = 16.37$ ,  $p < 0.0001$ , a marginal main effect of session,  $F(1,13) = 3.74$ ,  $p = 0.075$ , and no shift by session interaction. Pairwise comparisons for the 5 shift values revealed that all shift values significantly differed from all other shift values, except the shift 50 and shift 200 conditions, which were not significantly different. Of particular interest, the compensation response to the shift 200 condition was significantly smaller than that observed for the shift 100 condition ( $p < 0.0001$ ).

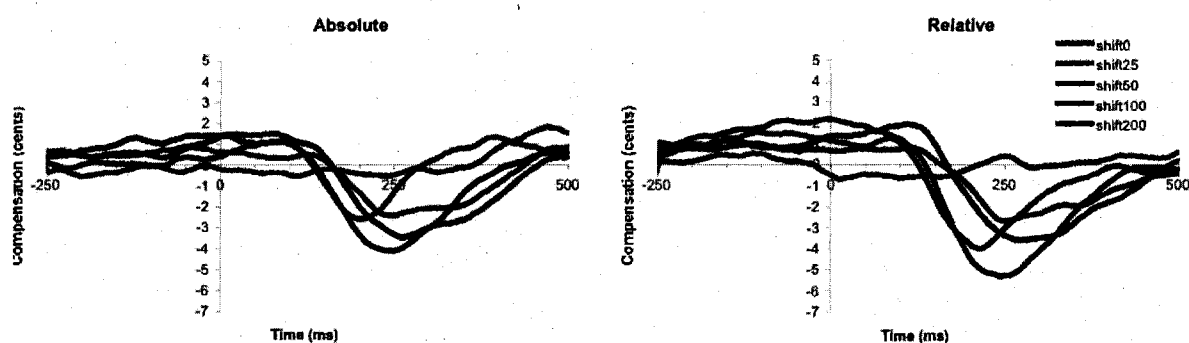


Figure 2.2: Averaged behavioral results from all participants for the absolute and relative sessions. Compensation shows how much the average of all participants altered their F0 (in cents) in response to the different perturbations. Time 0 represents the onset of the perturbation.

*EEG results:* ERP results for selected electrodes are detailed in Figure 2.3. Figure 2.4 shows the ERP results at Cz, as well as a comparison of the relative and absolute sessions for shift 100 and shift 200. The 2 X 5 X 60 (session by shift by electrode) ANOVA revealed no significant ERP changes in area 1 (50-150 ms). A prominent negative component was observed in area 2 (151-250 ms), with a main effect of shift,  $F(4,52) = 12.33$ ,  $p < 0.0001$ , and a shift by electrode interaction,  $F(236, 3068) = 3.8$ ,  $p = 0.0001$ . A linear trend analysis performed on the factor of shift revealed a significant linear trend ( $p = 0.0041$ ) indicating that the MMN response is larger with larger shift values. Planned comparisons were performed on the shift factor, revealing no difference between the shift 0 and shift 25 condition, but a significant difference between the shift 50 condition and shift 0 ( $p = 0.018$ ). The mean of the shift 100 and shift 200 conditions was also found to differ from the shift 0 condition ( $p = 0.0012$ ). Thus, although a negativity was present for perturbations that were 50 cents and larger, there was no negativity in area 2 when perturbations were 25 cents. The negativity response to the shift 200 condition was also found to be significantly larger than the negativity to the shift 100 condition ( $p = 0.017$ ).

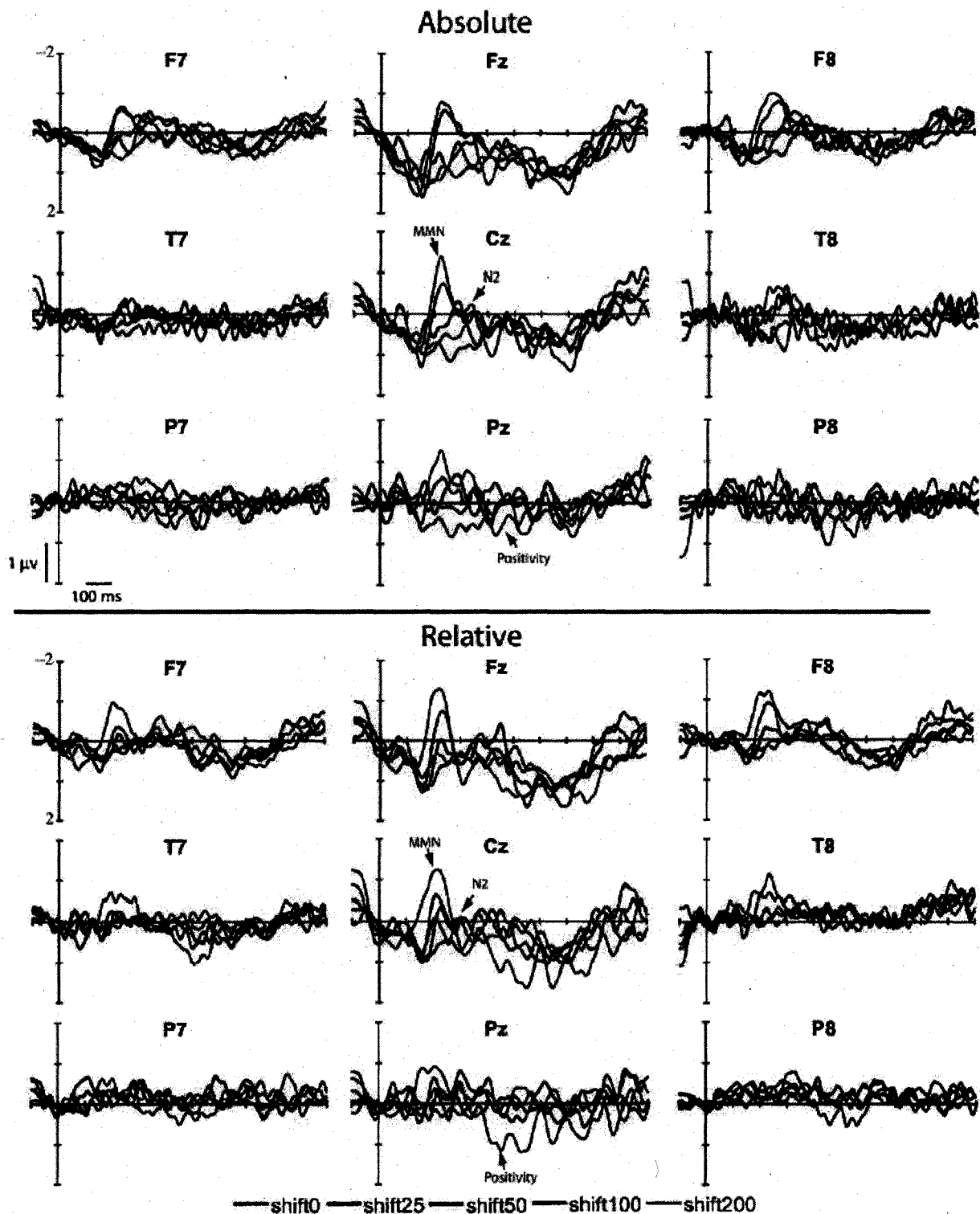


Figure 2.3: ERP results for absolute and relative sessions for all 5 shift conditions. Midline and lateral electrodes from the topographical analyses are shown. The MMN, N2, and late positivity responses are all indicated. Time 0 represents the onset of the perturbation.

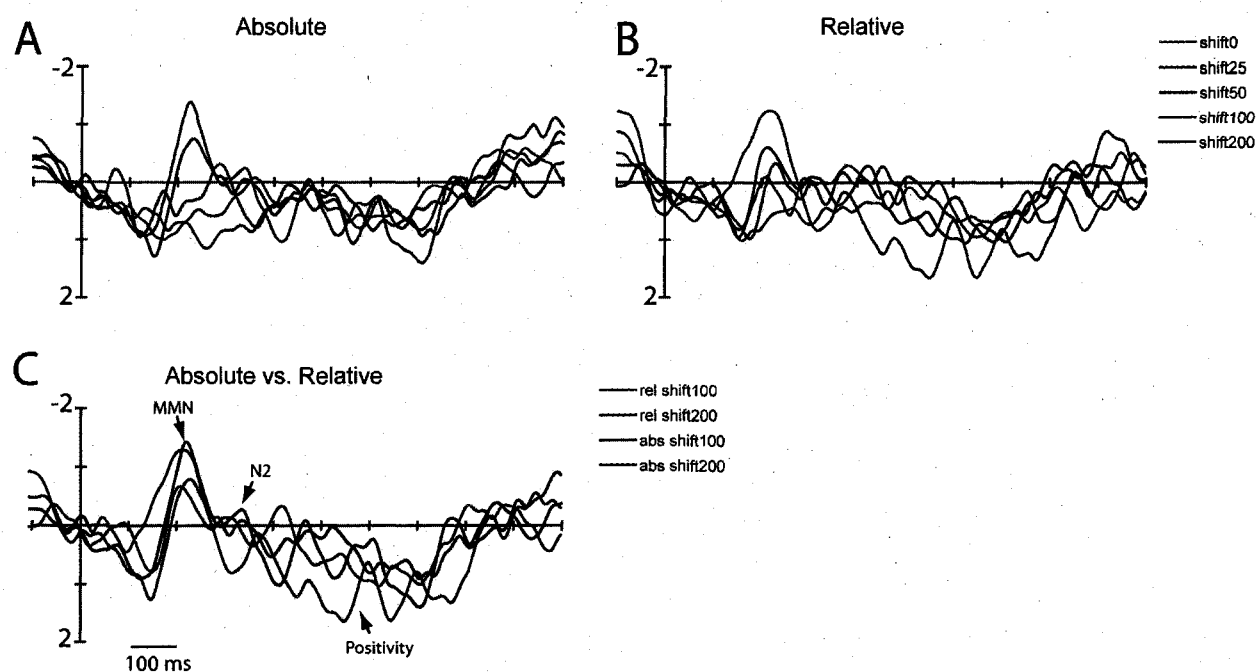


Figure 2.4: Grand Average ERP results at Cz, for a) the absolute session, b) the relative session, and c) comparing the grand averages for the absolute and relative sessions for the shift 200 and shift 100 conditions. No differences were observed between the relative and absolute sessions. The MMN, N2b, and late positivity are indicated by arrows. Time 0 represents the onset of the perturbation. Grand average waveforms represent an average across all participants.

Given the electrode by shift interaction, we performed a topographical analysis to determine the nature of this interaction, and to test for possible laterality effects (Figure 2.3). Frontal (F7, F3, Fz, F4, and F8), central (C7, C3, Cz, C4, and C8), and parietal (P7, P3, Pz, P4, and P8) electrodes were chosen for an anteriority factor, whereas left lateral (F7, T7, and P7), left medial (F3, C3, and P3), midline (Fz, Cz, and Pz), right medial (F4, C4, P4) and right lateral (F8, T8, and P8) electrodes were used as a laterality factor. A session by shift by anteriority by laterality (2 X 5 X 3 X 5) ANOVA was performed. A significant interaction in shift by anteriority,  $F(8,104) = 4.37$ ,  $p = 0.0096$ , indicated an anterior distribution to the response. Shift and laterality also produced a significant interaction,  $F(16,208) = 4.33$ ,  $p = 0.0002$ , which was

largely driven by a midline predominance in the response. Post-hoc tests revealed no hemispheric laterality. Changes in the topography with different perturbations were seen in a shift by anteriority by laterality interaction,  $F(32,416) = 1.72$ ,  $p = 0.036$ .

Examinations of area 3 (251-350 ms) found a significant response, with a significant main effect of shift,  $F(4,52) = 3.13$ ,  $p = 0.022$ . A significant linear trend was observed such that this response was increased with larger shift values ( $p = 0.027$ ).

No significant effects were found in area 4 (351-450 ms). A late positivity was observed in area 5 (451-550 ms) as a main effect of shift,  $F(4,52) = 4.47$ ,  $p = 0.0049$ , as well as a shift by electrode interaction,  $F(236, 2832) = 1.85$ ,  $p = 0.034$ . Planned contrasts revealed that this positivity was larger in the shift 200 condition than the shift 0 ( $p = 0.0037$ ) and the shift 100 condition ( $p = 0.009$ ), but that the shift 100 condition did not differ from the shift 0 condition ( $p = 0.46$ ), suggesting that this positivity only differed from the shift 0 condition in shift 200. The grand averages suggest that this positivity for shift 200 may be more prominent in the relative than the absolute condition (Figure 2.4), but once again a difference in relative and absolute sessions was not found in planned contrasts ( $p = 0.58$ ). A topographical analysis was performed, as described for area 2. The topographical analysis revealed a main effect of anteriority,  $F(2,26) = 4.30$ ,  $p = 0.034$ , though pairwise post-hoc comparisons did not reveal any specific differences. This positivity appears to be broadly distributed, as no other effects were observed in the regional analysis, with a somewhat frontal predominance.

## **Experiment 1 Discussion**

In this study, we demonstrated that ERPs represent an important tool in the search for the mechanisms underlying the pitch-shift reflex and monitoring of auditory feedback. We found a negativity in area 2 (from 151-250 ms post-perturbation), which we believe is an MMN, or an MMN-like component. The magnitude of this negative response increased with larger shifts (consistent with previous findings that the MMN increases with larger deviations in stimuli). We also observed an N2 response that was also modulated by shifts, and a late positivity that seems present mainly in the shift 200 condition. No systematic differences were found between the relative and absolute sessions. The observed MMN response may prove useful in future studies of preattentive sensory processing of auditory feedback. The MMN is generated by sources in the bilateral superior temporal lobes, believed to represent auditory processing, and a second frontal lobe generator associated with the attentional switch caused by an MMN (Näätänen et al., 1978; Giard, Perrin, Pernier & Bouchet, 1990; Rinne, Degerman & Alho, 2005). These sources for the MMN may overlap somewhat with the temporal and frontal activations observed in fMRI research on feedback alteration (Toyomura et al., 2007; Zarate & Zatorre, 2005, 2008).

Another potential interpretation of the present data is that the observed negativity may be a delayed N100 response to the perturbations. Our paradigm does not conform to a typical MMN study, in which the MMN is observed to rare stimuli within a train of standards. However, we do not believe that a delayed N100 response best describes our data for two reasons. First, the peak time of the MMN is over 200 ms post stimulus. It seems unlikely an N100 would be delayed by such a large amount. And second, the response follows patterns observed in the MMN literature, in that it is larger and earlier when the magnitude of the stimulus is greater. The presence of an N2b, in close proximity to the MMN response, may also argue against an N100 interpretation, as the MMN is often found in conjunction with the N2b. While our perturbations



were not rare stimuli, occurring in the majority of trials, they occurred randomly within an otherwise unaltered utterance. We therefore suggest that the pre-perturbation baseline represents the 'standard' used to evaluate the auditory stimuli (the participant's own auditory feedback) and form the appropriate memory representation. When the perturbation occurs, it represents a violation of the ongoing stimuli, producing an MMN. In the case of this study, the shift 0 condition represents a zero condition, in which no stimulus has been presented, and therefore the ERP findings for the shift 0 condition are equivalent to the baseline (pre-perturbation) period. The pitch manipulation performed in our study is very similar to those from Grimm and Schröger (2005), who also found a clear MMN to frequency changes in an ongoing sound. While we are characterizing the response as an MMN for purposes of discussion, we acknowledge that our MMN response may not be identical to the prototypical MMN observed in perceptual experiments using the standard MMN paradigm.

Interestingly, Lavikainen, et al. (1995) examined frequency modulation of a constant background tone, altering the tone for 100 ms approximately every 6 seconds. They found both an N100 and an MMN (at a similar latency to the negativity observed in the present study). We believe that, because the 'standard' tone in that study was played continuously, it took on the characteristic of background noise, and a change in that background was processed as a new stimulus. Similar to our study, there was no discrete 'standard' tone in Lavikainen et al. (1995) aside from the background tone. In the case of our study, the time difference between the onset of the auditory stimuli (the participant's auditory feedback) may have been too short to allow an additional N100 response. Instead, the perturbation was processed as a change in an ongoing and relatively novel stimuli, rather than as the onset of a discrete stimuli, resulting in a lack of an N100.

Interestingly, the shift 200 vocal compensation response was found to be smaller than the compensation in the shift 100 condition, while the MMN response was larger in shift 200 than shift 100. This smaller compensation has not been previously reported. Only two FAF studies with brief perturbations used stimulus magnitudes of 200 cents during vowel production. In Burnett et al. (1998), the duration of the perturbations was 500 ms. With a 500 ms perturbation, volitional responses are known to occur along with the automatic compensation responses (Hain et al., 2000). These volitional responses may have masked any differences in compensation responses between different shift magnitudes. Chen, Lui, Xu, and Larson (2007) found no differences between 50, 100, and 200 cents when participants vocalized /u/. We suggest the smaller compensation response we observed in shift 200 may be because the perturbation is large enough that the feedback alteration is recognized as externally generated by the auditory system. Presumably, such a large feedback error should be accompanied by a detectable kinesthetic error. The MMN amplitude may serve as a 'threshold', in which an MMN response exceeding a certain threshold must be accompanied by kinesthetic feedback to produce a normal compensation response. When such a large feedback alteration occurs without a detectable kinesthetic response, it is not likely that the error was internally generated, and thus a compensation response is not appropriate. If this is true, very large feedback alterations (such as 500 or 800 cents) should produce a small or no compensation response. This threshold may vary with a number of other experimental factors (volume of vocalizations, task instructions, etc).

The presence of a late positivity, which seems to only be distinct during a 200 cents perturbation, may also be consistent with the idea that the shift 200 condition is not seen as an internally generated error. This late positivity demonstrates enhanced and continued processing of the shift 200 condition, possibly because this large perturbation is attributed to an external,

rather than internal, source. Such continued processing shows that the shift 200 condition is viewed as distinct in some way from the other conditions.

The distinctiveness of the shift 200 conditions may have an important impact on future studies. Previous researchers have used a 200 cent shift as opposed to smaller shifts in ERP (Heinks-Maldonado et al., 2005), MEG (Heinks-Maldonado et al., 2006) and fMRI (Toyomura et al., 2007, Zarate & Zatorre, 2005, 2008) studies on auditory feedback. However, given the results of the present study, future imaging studies might wish to utilize a 100 cent shift, which maximizes the response without introducing complexities of continued processing not related to the use of feedback in vocalizations. Such extra processing might represent a significant confound in imaging studies trying to localize the regions involved in feedback monitoring, as the feedback change might no longer be recognized as a feedback error in the same way that a smaller error that was perceived as self-generated.

MEG and ERP studies examining N100 suppression related feedback perturbations have typically used 200 cent shifts (Heinks-Maldonado et al., 2005, 2006; Toyomura et al., 2007). These studies suggest a modulation of the auditory cortex by an efference copy comparator, both in that there is a reduction in activity in the auditory cortex when vocalizing compared to listening to recorded speech, and greater reduction to one's own unaltered voice than a pitch shifted voice. It might be interesting to see if such effects are present with a smaller perturbation (such as 50 cents), which is clearly detected as an error, but as a self-generated one. If the hypothesis that the suppression of the auditory cortex is related to the recognition of our own unaltered feedback as being internally generated (Heinks-Maldonado et al., 2005, 2006) is correct, we might expect to see an enhanced N100 (due to reduced suppression) in the presence

of large shifts (e.g. 200 cents), and no change in N100 suppression in the presence of small shifts (e.g. 50 cents).

The shift 25 condition is also interesting that no MMN was observed, while a compensation response was observed. The MMN is known to be highly correlated with detection thresholds of auditory stimuli (Lang et al., 1995, Näätänen & Alho, 1995), including in vowel discrimination (Aaltonen, et al., 1993). It is also known that compensation can occur without the need for perceptual awareness of a feedback alteration, though this threshold for awareness is generally less than 25 cents (Hafke, 2008). In the case of the present study, the multispeaker babble used as the masking may have caused participants to be perceptually unaware of the small 25 cent perturbations. White or pink masking noise may not have masked these small perturbations in the same way as multispeaker babble, which is constantly varying in frequency. Given the presence of an MMN response to perturbations, the MMN's known sensitivity to detection thresholds, and the fact that compensation can be observed when the perturbation is below the detection threshold, it is not surprising we would not see an MMN response to very small perturbations.

Though no site by electrode interaction in area 3 was observed, visual inspection of the ERP data suggests a posterior distribution to our N2 response, similar to the N2b response observed by Grimm and Schröger (2005) when participants attended to the frequency of stimuli with frequency modulations. We therefore believe that our N2 response is actually an N2b. Grimm and Schröger (2005) did not observe their N2b response when participants were instructed to attend to another dimension of the stimuli (duration), or when they ignored the stimuli. The presence of such a response in this study indicates that participants did in fact attend to and monitor their auditory feedback, given that the N2b is well known to be absent when

stimuli are not attended to (Näätänen, 1992). Visual inspection of the ERPs in the present study does not show a clear N2/N2b response in the relative condition, but there is a fairly clear N2/N2b in the absolute condition. If such a difference were to be observed, it would suggest greater feedback monitoring when matching a target. However, the high p-value ( $p = 0.48$ ) observed in this study suggests no systematic differences existed between the relative and absolute sessions, indicating that feedback monitoring was not driven by the task. A study with a larger number of trials at a large perturbation (100 cents) might allow an examination of differences among individuals in the amount of feedback monitoring as indexed by the N2b, and its possible effects on the compensation response.

The results of Heinks-Maldonado et al. (2005, 2006) show a clear early increase in N100 responses, starting at approximately 100 ms post-vocalization, whereas an N100 was not observed in the present study. The increase in the N100 responses, along with MEG studies showing a clear reduction in the M100 (the magnetic equivalent of the N100) activity when vocalizing as opposed to listening to speech (Pantev Eulitz, Hampson, Ross, & Roberts, 1996; Numminen & Curio 1999; Curio et al., 2000; Houde et al., 2002) give strong evidence of a feedforward efference copy which suppresses the auditory cortex during speech. Importantly, the observed effects in these studies occurred at utterance (and therefore motor command) onset.

Although no N100 was observed in our data, possibly because the perturbations occurred mid-utterance, we did expect to see an early negativity similar to that seen in Heinks-Maldonado et al. (2005, 2006) if an efference copy violation caused an increase in early auditory processing. The lack of a response raises several possibilities. Firstly, it may be that early auditory processing is not enhanced in the same fashion in mid-utterance as it is to a feedback alteration heard at the beginning of a vocalization. This would suggest that different mechanisms are

involved in vocalization onset and vocalization maintenance. Most discussions of feedback control do not differentiate between using feedback in vocalization onset and maintenance, so this issue should be examined in future studies.

The results of Kudo et al. (2004) illustrated that the MMN to tones was not reduced while an N100 response to tones was significantly reduced in the presence of an efference copy from vocalization. These results may suggest that the MMN response observed in the present research is not related to an efference copy comparison when evaluating auditory feedback. It is also possible that the early response occurs, but the low amplitude and poor signal to noise ratio of that response prevented us from detecting it. This possibility could be examined by a study using 100 cent shifts in mid-utterance with a large number of trials. However, given the current findings, we see no evidence that an efference copy comparator is being used in feedback monitoring during pitch maintenance.

Given the lack of an early response, and the characteristics of the MMN, we hypothesize that an efference copy comparator is not used in pitch maintenance during an ongoing utterance. It is not clear if normal speech would follow the same pattern of results when pitch was maintained, as the characteristics of the utterance are much more variable in speech, and each syllable might be treated as the initiation of a new motor plan (MacNeilage, 1998). The MMN is generally observed when an auditory stimulus is changed relative to a previous auditory stimulus, and relies on a memory trace for the preceding stimuli to form an expectation of incoming sensory stimuli. This contrasts strongly with the paradigm used in Heinks-Maldonado et al., (2005, 2006) where the feedback alterations are presented randomly, and so no stimuli trace can be formed to base a prediction of incoming stimuli. In our case, the pre-stimulus aspects of the vocalization allowed time for a memory trace to be formed.

In the present study, the beginning of the utterance could be used for trace formation. Given that the MMN requires the formation of a sensory trace of stimuli characteristics, and Heinks-Maldonado et al. (2005, 2006) observed an earlier N100 response rather than an MMN, we take this as evidence that pitch maintenance relies on an analysis of the signal characteristics of the ongoing utterance. When a non-volitional deviation in these characteristics occurs, it is classified as an error, and a compensatory response occurs. Because the efference copy comparator seems to interact with the auditory cortex at an earlier level (as it affects the N100, which occurs prior to the presently measured MMN), the activation of a new motor plan (i.e., a volitional change in pitch) may disrupt the sensory trace for the ongoing utterance, and thus stopping an MMN response and preventing us from compensating to an error-free volitional change in F0.

The hypothesis that we use our existing feedback as the referent in feedback monitoring of an unchanging utterance will require further testing. However, this hypothesis corresponds with the results of Larson et al. (2001), who found no differences in compensation when a perturbation was applied before an utterance and then removed during the utterance, or when the perturbation was turned on after utterance onset. In addition, sensorimotor adaptation studies using FAF often find complete compensation for the feedback alteration (Houde & Jordan, 1998) even for 100 cents shifts (Jones & Keough, in press). Mid-utterance perturbation studies, on the other hand, typically have found complete compensation for only very small shifts.

The results of the present study demonstrate the usefulness of ERPs to study processes of auditory feedback monitoring and their relationship to compensation responses and understanding the underlying sensory mechanisms of auditory feedback monitoring. The MMN gives an index of awareness of a feedback alteration, and as a window into the mechanism in

feedback monitoring and error detection. The N2b might be useful as an index into the amount of attentional resources being devoted to auditory feedback, and may be useful in differentiating people who attend to their feedback and those who do not, or in measuring the effects of shifting attention away from feedback by varying amounts. A wide array of studies present themselves, as much of the underlying mechanisms of feedback monitoring and compensation remain poorly understood at a neurological level.



### Chapter 3

Experiment 2: Vocal control at utterance onset and during F0 maintenance: Different mechanisms for different goals.

Submitted for Publication in Experimental Brain Research as:

Hawco, C., and Jones, J.A. (Submitted). Vocal control at utterance onset and during F0 maintenance: Different mechanisms for different goals. Brain Research

#### Author Contributions:

The study was primarily designed by Colin S. Hawco, who also analyzed all data.

Jeffery A. Jones supervised the study, and was involved in discussions on the design, analysis, and interpretation of the data.

## Experiment 2 Background

Studies of vocal control often use one of two FAF paradigms, studies of sensory-motor adaptation, or random perturbation studies. In studies of sensory-motor adaptation, a feedback alteration is introduced and left in place for over many trials. Over time, a remapping of the motor command for vocalization occurs, as participants adapt to the altered auditory feedback. When a brief perturbation is randomly introduced to an ongoing utterance, a transient compensation response is observed, where speakers adjust their F0 in the direction opposite the shift. This response has been termed the pitch-shift reflex (Burnett et al., 1998).

The pitch-shift reflex is distinct from the adaptation response, as the pitch-shift reflex reflects an online compensation process using ongoing auditory feedback when an unexpected feedback error is encountered, while adaptation represents an updating of the forward model in response to a predictable and constant feedback alteration. The results of Burnett et al. (1998) suggested that the pitch-shift reflex is made up of two components, an early and a late response. Hain et al. (2000) used a 500 ms perturbation, and asked participants to compensate, not to compensate, or actively follow the pitch-shifted feedback when their voice was perturbed. They found that the early component of the pitch-shift reflex was automatic and not affected by task instructions while the late component was under volitional control.

One important aspect of the pitch-shift reflex is that it is generally smaller than the perturbation, except for very small feedback alterations (Burnett et al., 1998; Larson et al. 2001; Liu and Larson, 2007). In contrast, responses to whole-utterance shifts are often close or equal in magnitude to the feedback alteration. Jones and Keough (2008) had a series of baseline trials, after which they exposed participants to a 100 cent (one semitone) feedback alteration. On the

first trial after feedback alteration, both singers and non-singers adjusted their F0 by more than 50 cents; in fact non-singers adjusted their F0 by approximately 70 cents. Moreover, non-singers achieved full compensation to the feedback alteration within 5 or 6 trials. This approximately 70 cent response is much larger than that typically observed in the pitch-shift reflex, where a 100 cent perturbation generally results in compensation responses of less than 50 cents, and sometimes as little as 9 cents (Liu and Larson, 2007).

It has been theorized that an efference copy comparator is used to monitor auditory feedback, and that this comparator plays an important role in motor learning (Nowak et al. 2007). The efference copy is sent from the motor system, by the forward model, to the auditory system so that sensory feedback can be predicted. That prediction is compared to real auditory feedback regarding the vocalization; when feedback does not match the prediction, an error is registered. This error is corrected online, and the forward model is subsequently modified. Some evidence for the use of an efference copy in monitoring auditory feedback has come from studies using event-related potentials (ERPs). Hienks-Maldonado et al. (2005) found the N100, an ERP component associated with sensory processing of a stimulus, was attenuated during vocalization. When feedback was altered, the N100 attenuation was reduced. This attenuation is believed to be driven by the efference copy. The auditory cortex is maximally attenuated when feedback matches the efference copy, allowing speakers to determine that what they are hearing is their own voice, and that no errors occurred during vocal production. Importantly, the N100 is only observed at utterance onset.

Larson et al. (2001) conducted a study in which they either altered feedback in mid-utterance, or altered feedback at utterance onset, and then removed the feedback alteration in mid-utterance. They found an identical mid-utterance compensation response in both conditions.

That is, the compensation response always stabilized production so that the current F0 value matched the F0 prior to the alteration, independent of whether the change detected was initiation of a feedback alteration or its removal. The authors suggest that this finding rules out the use of a fixed efference copy as the reference used in voice stabilization, as such a system should not stabilize the system to artificially altered feedback. However, participants were instructed to produce at a habitual F0, and the baseline F0 was not known. Thus, it is possible that participants responded to the initial feedback alteration with a compensation that was equal in magnitude to the compensation to the mid-utterance feedback alteration. When the feedback alteration was removed, they then simply stopped compensating and their F0s returned to their baseline value, resulting in a response equal to that observed when a perturbation was initiated mid-utterance. We have some evidence that responses to initial F0 shifts may be larger than those typically observed for mid-utterance F0 shifts (Jones and Keough, 2008), but that study used a fixed external reference in the form of a target note. Speakers attempting to match an external reference might produce larger compensation responses than speakers who are not required to achieve a specific pitch target.

Experiment 1 examined the response to a mid-utterance perturbation using ERPs. Rather than observing an early effect that replicated the N100 results of Hienks-Maldonado et al. (2005), Experiment 1 found a mismatch negativity (MMN), a later ERP component associated with a violation of a sensory memory trace. The MMN is generally observed when a memory trace is formed for a stimulus, or a stream of stimuli, and a stimulus is encountered which violates that memory trace (see Näätänen, et al., 2007, for review). Kudo et al. (2004) found that the N100 response to a tone was reduced during vocalization, while the MMN was unaffected, suggesting that the efference copy is related to the N100 but not to the MMN. The MMN

observed by Experiment 1 suggests that the error detected in a mid-utterance perturbation was compared to some echoic memory representation, such as the unshifted baseline within the utterance, rather than an efference copy.

In the present study, we compared the compensation response at utterance onset to that at mid-utterance, to determine if these responses were equivalent. Fifteen women heard a target female voice vocalizing the vowel /a/ at a specific frequency (D4, 296.33 Hz) and were asked to produce the vowel at the same pitch. In two blocks, speakers' auditory feedback was either shifted up by 1 semitone (100 cents) *mid-utterance*, or shifted up 1 semitone at *utterance onset*, and then perturbed by removing this feedback alteration. Specifically, in the control block, feedback was either shifted upward 100 cents in pitch mid-utterance (Control Perturb trials), or randomly shifted upward 100 cents at utterance onset with the feedback alteration either maintained throughout the entire utterance (Onset No-perturb trials) or perturbed by removing the feedback alteration for 500 ms (Onset Perturb trials). In a separate block of trials, we adapted participants to a feedback alteration by introducing a constant and predictable feedback alteration prior to utterance onset (100 cent shift up), and perturbed their voice by removing the feedback alteration during some trials (Adapt Perturb trials). The Onset Perturb and Adapt Perturb trials had identical feedback alterations (see Figure 3.1), although the Adapt trials were predictably shifted at utterance onset, while the Onset trials were randomly shifted within the control block. If the pitch-shift reflex is a pure maintenance response, and uses the current F0 as its reference (even when an absolute referent is available in the form of a target F0), we should observe a response of similar magnitude regardless whether the perturbation is the sudden onset of a feedback alteration in mid-utterance, or the removal of a feedback alteration that was introduced at utterance onset. If, on the other hand, an efference comparator is used, the size of the pitch-

shift response to a feedback alteration removal should match the size of a compensation response to the perturbation at utterance onset.

## **Experiment 2 Methods**

*Participants:* Fifteen female participants (aged 19 to 24) were recruited for this study. All reported that they had never received formal singing training, and were not practicing singers (e.g., in a choir). No participant spoke any tonal languages. All participants read and signed an informed consent form, in accordance with the ethical policies of Wilfrid Laurier University.

*Procedure:* Participants heard a target female voice vocalizing the vowel /a/ at a specific frequency (D4, 296.33 Hz) for 1 s. They were instructed to begin vocalizing for 3 s when the target voice finished, matching the pitch of the target. A 1000 Hz tone indicated when they should cease vocalizing. A loudness monitor in front of the participants allowed them to maintain a specific volume. Participants were instructed to vocalize at a volume of approximately 75 dB SPL. Auditory feedback was amplified and heard over the headphones at approximately 85 dB SPL.

The experiment was divided into two blocks, with a filler task between them. Each block had 140 trials and lasted approximately 18 min. In the adaptation block, participants heard their feedback unaltered for 20 trials. From trials 21 to 40, the F0 of their feedback was gradually increased by 5 cents/trial, until it had reached 100 cents at trial 40. This 100 cents upward feedback shift was maintained for the rest of the block (Adapt trials). On half of trials 61 to 140, participants' feedback was perturbed by removing the feedback alteration (i.e., returning their

feedback to its unaltered state of 0 cents, Adapt-Perturb trials). This effectively lowered the participant's feedback by 100 cents. The perturbation began between 1000 ms and 1800 ms after utterance onset and lasted for 500 ms.

During the control block, participants heard their unaltered feedback for the first 60 trials. In half of the trials from trial 61 to trial 140, a shift 100 cent upwards was pseudorandomly introduced *prior* to utterance onset (Onset trials), with no more than three successive Control or Onset trials. The Onset shift was equivalent to the shift used at onset in the adaptation block, though it was a randomly presented (i.e., unpredictable) as opposed to the predictable shift in the adaptation block. In half of the Control and Onset trials, a perturbation was induced between 1000 ms and 1800 ms after utterance onset for 500 ms. In the Onset condition, this perturbation was induced by removing the feedback alteration, while in the Control trials the perturbation consisted of a 100 cent downward shift. Therefore four conditions occurred during the control block: a Control No-perturb condition where participants heard unaltered feedback, a Control Perturb condition where participants heard their feedback suddenly shifted down mid-utterance, an Onset No-perturb condition where participants heard their voice shifted 100 cents up during their entire utterance, and an Onset Perturb condition where participants heard their voice shifted up 100 cents from the beginning of their utterance, but this alteration was turned off briefly mid-utterance. A schematic diagram of the F0 shifts in each condition in the control and adaptation block is shown in Figure 3.1.

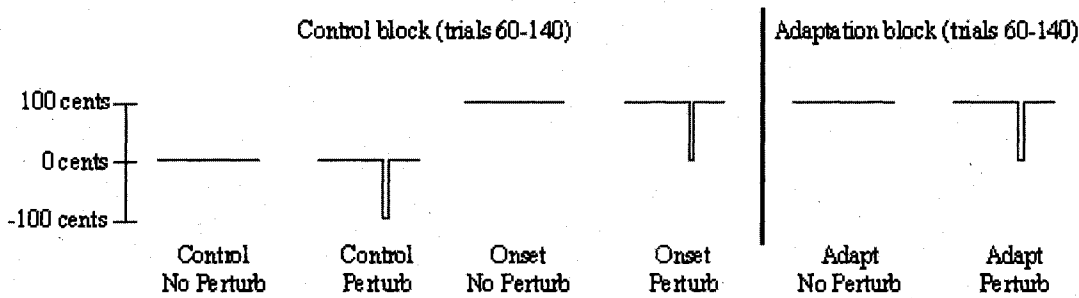


Figure 3.1. Schematic representation of the four shifts presented in the different conditions. No mid-utterance perturbations were presented in trials 1-60 in the control and adaptation blocks. The onset trials were similar to the adaptation trials except that the onset trials were presented randomly in the same block as the control trials.

A filler task was performed between blocks. Participants were asked to read 100 sentences (taken from Kalikow et al. 1977) silently, speak the sentence in a monotone voice, and then repeat the last word. This took approximately 15-20 minutes. This filler task served to remove any carry-over effects from the adaptation block, if the adaptation block occurred before the control block. During the filler task, participants heard their unaltered feedback. Participants performed the filler task even when the control block was presented first.

*Apparatus:* Participants sat in a double-walled sound attenuated booth (Industrial Acoustic Company, Model 1601-01), and wore headphones (Sennheiser HD 280 Pro) and a headset microphone (Countryman E6 Omni). Vocalizations were sent from the microphone to a mixer (Mackie Oynx 1220, Loud Technologies), which passed the voice signal to a digital signal processor (DSP) (VoiceOne, T.C. Hellicon). The DSP shifted the participant's voice and returned it to the mixer, where it was mixed with pink masking noise (70 dB SPL) and returned to the participant as auditory feedback. The unaltered voice signal was digitally recorded (TASCAM HD-P2) at a sampling rate of 44.1 kHz.



*Analysis:* Each trial onset was manually segmented and saved into a separate WAV file. F0 for each utterance was calculated using an autocorrelation algorithm included in the Praat program (Boersma, 2001), with a sampling rate of 5 ms.

To measure adaptation effects, the F0 for the entire utterance was converted into cents using the formula:

$$\text{cents} = 100 * (39.86 * \log_{10}(\text{F0}/\text{baseline}))$$

where the baseline was 296.33 Hz (the target pitch participants were instructed to match).

Adaptation was indexed by examining the F0 during the first 50 ms in each trial. The initial 50 ms of vocalization is driven by purely feed-forward controllers, as feedback is not available for at least 100 ms (Burnett et al., 1998). F0 data during later parts of the utterance can be influenced by both feed-forward controllers (because they control the initial F0 of the utterance) and feedback controllers, as F0 can be modified using auditory feedback during the later portions of vocalization (after 100 to 150 ms after utterance onset). If adaptation occurred, we expected to see a systematic shift in F0 at utterance initiation (the first 50 ms) across trials. If adaptation did not occur the initial F0 of each utterance would be equivalent.

The compensation response for the mid-utterance feedback alterations was also calculated. The F0 trajectories for each perturbation trial type were time aligned at the point of the perturbation and average waveforms were generated. The period between 500 ms before to 1000 ms after the perturbation was used to evaluate the compensation response. In order to compare the compensation responses across participants, each subject's averaged waveform was converted to cents and normalized such that the baseline period (the 250 ms preceding the perturbation) had a mean of 0. The magnitude of the compensation response was determined by calculating the area under the curve using the trapezoidal rule, for 2 time periods. The first

period, 100 to 250 ms post-perturbation was used to evaluate the automatic pitch-shift reflex, while the second time window, from 250-600 ms post perturbation, was used to evaluate later compensation responses that are subject to volitional control (Hain et al., 2000). An alpha of 0.05 was used for all statistical analysis in this study.

## **Experiment 2 Results**

Mean F0 across the entire vocalization, and median F0 for the initial 50 ms of adaptation trials and control trials are shown in Figure 3.2. Three participants were excluded from the analysis: two because they showed no evidence of adaptation effects and one participant was removed because they showed a following response, increasing their F0 throughout much of the adaptation block. To test for adaptation, we compared median 50 ms data from trials 11 to 20 (the last 10 trials prior to feedback alteration in the adaptation block) to trials 51 to 60 (the 10 utterances prior the onset of perturbations). These trials were compared between the adaptation and control blocks. A block (adaptation or control) by time (trials 11 to 20 or trials 51 to 60) ANOVA showed a main effect of block,  $F(1,11) = 9.7$ ,  $p = 0.0097$ , and a block by time interaction,  $F(1,11) = 4.8$ ,  $p = 0.049$ . Tukey's HSD post-hoc analysis of the 2-way interaction showed that trials 51 to 60 of the adaptation block differed from the other tested conditions, and no other conditions differed among themselves, suggesting a change in F0 onset brought on by adaptation.

Mean compensation responses are shown in Figure 3.3. To measure the size of the compensation responses, area under the curve of the normalized compensation response (where the baseline was normalized to zero, as described above) was calculated. Because we are

interested in comparing compensation responses in the presence of adaptation to compensation responses in the control and onset conditions, the three participants excluded from the adaptation analysis were also excluded from the compensation analysis. For the first time window (100 ms to 250 ms post-perturbation, evaluating the pitch-shift reflex) a perturbation (Perturb or No-perturb) by condition (Adapt, Control, or Onset) ANOVA was conducted. An effect of

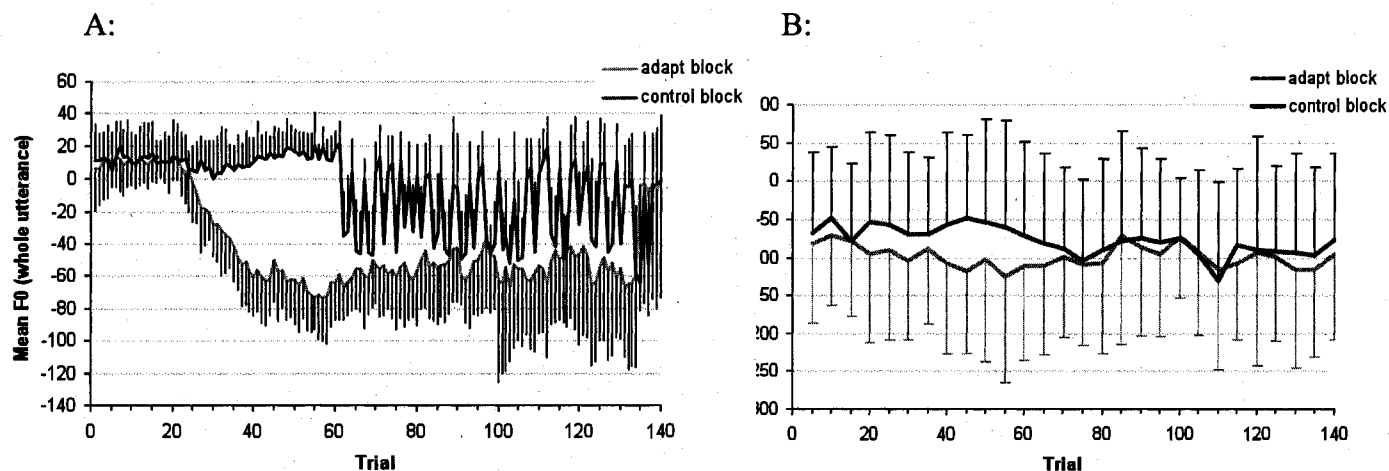


Figure 3.2. A) Mean of the median F0 of each utterance across trials for the control and adapt blocks. By trial 40, when auditory feedback F0 in the adaptation block has been increased by 100 cents, participants have reduced their F0 by approximately 70 to 80 cents to offset the feedback alteration. B) Mean of the median F0 for the first 50 ms of vocalization, averaged over blocks of 5 trials. The F0 for utterance initiation is lower in the adaptation block, indicating that participants adjusted their initial F0 in response to the change in feedback F0, suggesting adaptation occurred. Note that the data for trials 60 to 140 in the control block is affected by the presence of the onset trials, which have effects on both the mean F0 for the whole utterance, and the initial F0 of utterances following onset trials. Error bars represent 1 standard deviation.

perturbation was found,  $F(1,11) = 18.2$ ,  $p = 0.0013$ , indicating that participants responded to the feedback alteration, but no main effect of condition was observed,  $F(2,28) = 1.2$ ,  $p = 0.30$ , and there was no interaction, suggesting a similar magnitude to all compensation responses. The second time window was from 250 to 600 ms. Again, a main effect of perturbation was observed,  $F(1,11) = 27.1$ ,  $p = 0.00028$ . In addition, a main effect of condition,  $F(2,22) = 4.8$ ,  $p = 0.018$ , and an interaction existed,  $F(2,22) = 3.4$ ,  $p = 0.048$ , indicating that the late compensation response differed across conditions. Tukey's HSD post-hoc analysis of the interaction indicated

that when perturbations were present the adaptation and control conditions differed ( $p = 0.021$ ), but that the onset condition differed from neither the control ( $p = 0.89$ ) nor the adaptation condition ( $p = 0.18$ ).

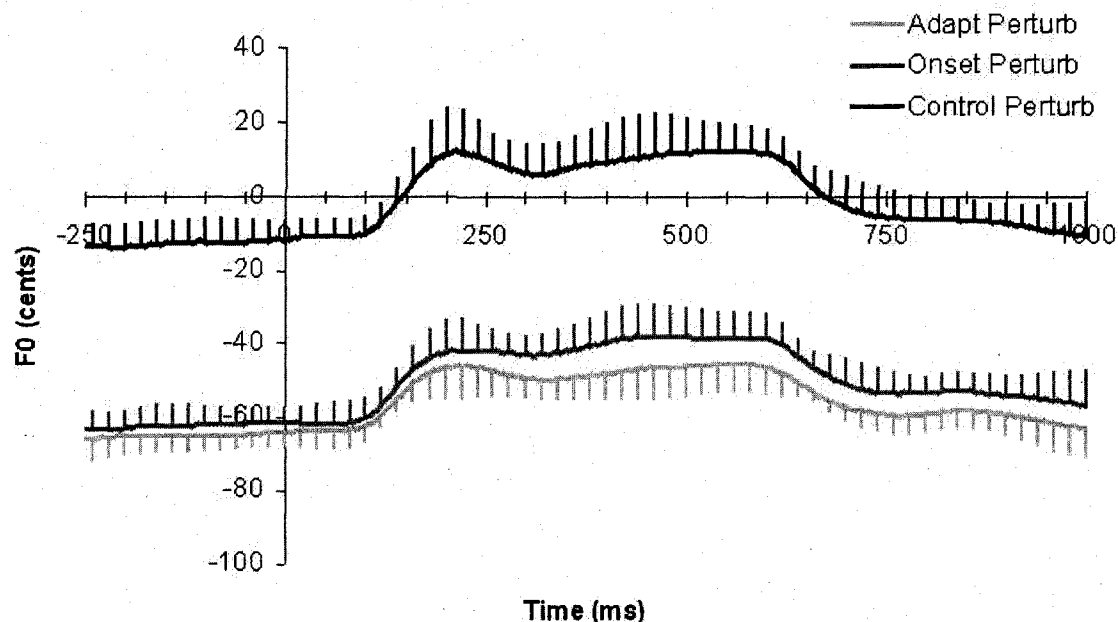


Figure 3.3. Compensation responses for the 3 perturbation conditions. Note that F0 in the onset condition was 50 cents below the F0 of the control condition before the onset of the mid-utterance perturbation, but the compensation to the mid-utterance perturbation in the onset condition is only approximately 17 cents. This shows that participants had a larger response to the perturbation at trial onset than the perturbation in mid-utterance. Error bars represent 1 standard deviation. Time 0 indicated the time of perturbation onset, which was 1000 to 1800 ms after utterance onset. Error bars indicate 1 standard deviation from the mean.

Visual inspection of the data (Figure 3.3) suggests that the F0 change in response to the feedback alteration at utterance onset is much larger than the compensation response to the mid-utterance perturbation. If the mid-utterance compensation responses were equal to the response to the onset, it would suggest that participants were returning their voice to pre-feedback shift baseline, and that the compensation response to the onset was the same as the response to the perturbation. If participants were returning to their baseline F0 when we removed the feedback

alteration in the onset trials (the perturbation), then the compensation response should reach a value equal to the control trials with no perturbations. Figure 3.3 shows that the compensation responses in the onset and adaptation trials were much smaller than the difference between the baseline, pre-perturbation values for Control and Onset or Adapt trials, suggesting that participants shifted their voices more for the initial feedback alteration (present before they began their utterance) than when the feedback alteration was removed. Note that the pre-perturbation baseline shown within Figure 3.3 represents F0 values within the middle of the utterance, not at utterance onset. We are therefore examining F0 when both predictive initial F0 changes due to adaptation effects and compensatory changes to F0 using auditory feedback after utterance onset have occurred. To test this, we compared the mean F0 values from 100 ms to 600 ms post-perturbation for the Onset Perturb and Adapt Perturb trials to the Control No-Perturb trials, which represents the baseline F0 value from which the other two conditions should differ. We found a significant difference,  $F(2,22) = 15.1$ ,  $p = 0.000072$ . Tukey's HSD post-hoc showed that the control data was significantly different than the adaptation ( $p = 0.00022$ ) and onset ( $p = 0.00096$ ) compensation responses. The adaptation and onset compensation responses were not found to be different. This demonstrates that the compensation to a feedback shift at utterance onset is larger than the compensation response to a mid-utterance perturbation.

It is possible that Onset trials may have an effect on subsequent trials. In order to assess single trial adaptation effects, we examined the first 50 ms of utterances of Control trials that *followed* either Onset trials or Control trials in the control block. Because it was possible to have 2 or 3 consecutive onset trials, we only examined Control trials that followed a single onset trial (i.e., onset trials that were preceded by a control trial) to prevent any confounds from adaptation occurring as the result of consecutive presentation of Onset trials. Likewise, we only examined

trials following control trials that were preceded by an Onset trial. Significant differences were observed between trials following onset trials and trials following control trials,  $F(1,11) = 7.7$ ,  $p = 0.017$ , with the  $F_0$  of the initial 50 ms being lower in trials following onset trials. This shows that the onset trials had an effect on the following trial, suggesting some rapid, single-trial adaptation occurred.

## **Experiment 2 Discussion**

The results of this study demonstrate that a mid-utterance perturbation results in an identical compensation response when the perturbation is the introduction of a feedback alteration, the removal of a randomly occurring feedback alteration, or the removal of a feedback alteration after adaptation has occurred. This is similar to the results of Larson et al. (2001), who found identical magnitudes of compensation when a feedback alteration was induced, or when an existing feedback alteration was removed. More importantly, it was found that the compensation response to a perturbation at utterance onset is much larger than the compensation response to a feedback alteration within an ongoing utterance. Overall, this pattern of results suggests differences in the mechanism used to evaluate  $F_0$  feedback at utterance onset and mid-utterance.

*Mid-utterance Compensation Responses:* We found similar compensation responses in all conditions for the early, automatic part of the mid-utterance compensation. While differences were found in the later phase of the mid-utterance compensation, these are less interesting as this part of the compensation response is subject to volitional control, and it is difficult to determine

what factors might contribute to the observed differences. Of particular interest is the Onset Perturb condition. In this condition, we initiated a feedback alteration at utterance onset, and removed it in mid-utterance for a 500 ms period. In this case, as in the adaptation trials, participants treated their unaltered feedback as an error and responded to it by compensating in the opposite direction. It is clear that this mid-utterance compensation response is not a 'switching-off' of the compensation to the perturbation at utterance onset, because if that were the case, we would have observed a much larger compensation in the Onset than Control conditions as F0 returned to the control-baseline values.

The mid-utterance compensation response may be driven by one of two possible mechanisms: comparison to a relative reference where the current F0 at the time of feedback alteration represents the goal, or to an absolute reference in which there is a specific, fixed F0 that represents the pitch goal. An efferent comparator is a form of internal absolute reference, in which feedback is compared to an efference copy of the motor command for a specific F0. In the current study, we introduced an absolute external reference by asking participants to match a specific pitch value. If participants were using an absolute referent, one of two possible things could occur; participants could maintain their F0 despite the perturbation, because the feedback alteration was induced by giving them their unaltered feedback (which by definition is not erroneous and thus might not cause a compensation response), or, more likely, they could show a larger response in the Onset condition (relative to the control) as they re-adjust their F0 back to the control baseline, removing their large compensation response to the perturbation at utterance onset. Larson et al. (2001) found that when speakers began an utterance under altered feedback and heard the alteration removed, they produced the same compensation response as when they heard their feedback altered part way through their utterance. The authors suggest that

participants used an internal, variable reference when no absolute reference was available. We found no differences in compensation responses even when we added an absolute external reference. This suggests that, when maintaining a steady F0, the reference is always internally based on the current, pre-shift F0. According to this hypothesis, the purpose of the compensation response is not to attain a specific pitch goal, but to adjust for unintentional fluctuations within F0 during an utterance.

Burnett and Larson (2002) found compensation to a perturbation during a changing utterance (glissandos). They suggest that the observed compensation discounts the possibility of the baseline, pre-shift F0 being used as the comparator. However, they also found smaller compensation for glissandos than for utterances with flat F0s. This might be consistent with the preceding F0 serving as the F0 reference considering the fact that they shifted the participants' voices down. If the pre-shift F0 is used to calculate the size of compensation, and the participants' voice increased in pitch over time, the compensation responses may have appeared smaller because the pre-shift F0 was lower than the F0 at the time the compensation was measured, about 150 ms after the shift. Another possibility is that, in the case of a changing baseline F0, the system is able to incorporate the velocity of the F0 change into compensation response, and thus compensate adequately during a changing utterance. Generally, the system responds to any change that is unintended, regardless of the context in which the motor command is being executed.

*Differences in comparators at onset or mid-utterance:* Figure 3.3 clearly shows that the compensation response to the feedback alteration at the beginning of the Onset and Adapt trials was larger than the compensation to the mid-utterance perturbation. It should be noted that the



pre-perturbation baseline for the Adapt trials may include changes in F0 from both feedforward adaptation and feedback based compensation, while the pre-perturbation baseline to the Onset trials represents a purely compensation based response, as the Onset trials are randomized and adaptation should not occur. The compensation response at utterance onset is much larger than the mid-utterance compensation. As discussed above, we have no evidence that an efference comparator is being used in the mid-utterance perturbations. However, several studies have suggested that efference copy plays a role at utterance onset (Curio et al., 2000, Houde et al., 2002, Hienks-Maldonado et al., 2005). For example, Hienks-Maldonado et al. (2005) found that N100 attenuation related to the efference copy was reduced when auditory feedback was altered, suggesting that the efference copy may be used as a means to detect errors within one's voice. This is in contrast to the results of Experiment 1, where we found an MMN, rather than an N100, to mid-utterance perturbations. The fact that Experiment 1 observed an MMN, and no early activity similar to the N100 effects from Hienks-Maldonado et al. (2005) suggests that feedback at utterance onset and during mid-utterance may be monitored using different mechanisms. However, it should be noted that Hienks-Maldonado et al. (2005) were interested in observing only the perception of the speaker's own voice, and not in linking that perception to vocal control.

At utterance onset, the goal is to match a specific and pre-planned F0, adjusting onset F0 to match a specific goal (e.g., a target note). Feedforward mechanisms must be used at utterance onset to hit an F0 target, or we would not observe a change in initial F0 following trials with altered feedback. In the present study, participants' initial utterance F0s were modified by repeated exposure to the feedback alteration indicating that the participants' feedforward plans were modified. However, this feedforward system was not highly accurate at attaining the

specific target immediately at utterance onset. Thus, feedback was also used to adjust for errors after utterance onset (initial F0 was lower than final F0, indicating a searching strategy). When feedback is used during utterance onset to reach a desired F0, an absolute reference must be used, as a variable reference would result in large and unpredictable errors. Indeed, it is difficult to imagine from where a variable reference would be derived at utterance onset, when no auditory feedback is available. Two possible absolute references exist at utterance onset during the task in the present experiments. The first is the efference copy, which is known to respond to changes in auditory feedback (Hienks-Maldonado et al., 2005). The second is a memory trace of the target note. It is difficult in the context of the current study to rule out the second possibility, though we believe that an efference comparator is more likely as it should exist during all vocalizations, including during normal speech, where an external target F0 does not generally exist. One way to test this may be to compare the compensation response for a feedback alteration at utterance onset with and without a target note. If a memory trace of the target note were used as the comparator, we would expect to see a larger compensation at utterance onset when a target is present than when one is not. It is also possible that both an efference comparator and a memory trace of the target note are being used in concert.

When the target F0 has been matched after utterance initiation, a goal switching takes place, in which the goal shifts from target matching (be it matching to an efference copy or to a memory trace of the target) to pitch maintenance. This maintenance mechanism drives the pitch-shift reflex, as described above. In other words, there is a switch from an efference or memory trace comparator to a current F0 comparator/voice stabilizer. When an alteration is encountered, the pitch-shift reflex serves to stabilize the voice to the current F0. We therefore arrive at a

separation of the mechanisms used at utterance onset to correct for feedback errors and those used after onset to stabilize F0 during an utterance.

Arm reaching studies have also suggested possible differences for specific aspects of a motor command. Dizio and Lackner (1995) exposed participants to Coriolis forces by placing them in a slowly rotating room. They had participants perform a reaching task in the dark (i.e., without visual feedback) using their dominant arm while the other arm remained stationary. When the Coriolis force was removed, the right arm, exposed to the Coriolis force, showed a curved trajectory that mirrored that of the Coriolis force. The non-exposed arm, in contrast, showed a linear trajectory, but had errors in their final position. This demonstrated a difference in end-point and trajectory adaptation and how it generalized between the limbs. Scheidt and Ghez (2007) conducted a study on differences in end-point and trajectory, and then made a simulation of their results. The simulation best matched their results when they added two sequential forward controllers; the first to initiate trajectory, and the second for control of final position. This difference in initial trajectory and final position may be somewhat analogous to the differences observed in the present study between utterance onset and F0 maintenance. In the contexts of the present study, we suggest that F0 maintenance uses distinct feedforward controllers from those used at utterance onset and during volitional changes in F0. Another possible analogue of the present study may be force exertion (such as pushing something), in which a fairly constant force or velocity is the goal. In such a case, differences might exist at the onset of the motor command (when the pushing movement is initiated and a desired velocity is reached), and during maintenance of the motor command.

*Single trial Adaptation effects:* We found that the onset trials caused a shift in the initial F0 of the following trials, even when an onset trial was preceded by a control trial, indicating some level of single-trial adaptation. This finding is consistent with the results of Donath et al. (2002) who found that compensation responses carried over within and between trials when speaking a nonsense word. Such single trial adaptation has also been found in arm reaching, with perturbations within a trial affecting subsequent trials (Thoroughman et al., 2007). The benefit of the current method is that we were able to use our analysis of the first 50 ms of the utterance to demonstrate that the single trial adaptation effects modified the forward motor plan used in the next utterance. Without an examination of the very beginning of a motor command, it is possible that any after-effects are caused by an early within-movement compensation for feedback alterations. In other words, we have demonstrated that any differences in trials following onset trials represents alterations of the feedforward motor plans.

A particularly important question to investigate is how much adaptation occurs between trials. While the forward model is modified by a single trial event, we do not yet know whether the magnitude of this modification is comparable to changes that occur during an adaptation trial. That is, the cumulative effect of repeated exposure has not been quantified. Even within the adaptation block, the change in initial F0 was much smaller than the feedback alteration and the final change in F0 (as measured by median F0 during the entire utterance). It is not clear if this is because our observed adaptation effects represent only a partial re-mapping of the forward model along with some additional compensation within each utterance, or because of physiological limitations within the system preventing a radical change in initial F0.

## **Experiment 2 Conclusions**

In this study, we have shown that the compensation response at utterance onset is larger than the compensation response to a perturbation mid-utterance. We suggest that our results are best explained by a change in the comparator used at utterance onset and mid-utterance. At utterance onset, an absolute comparator is used to match auditory feedback to the intended F0. After utterance onset, a goal change occurs to a stabilization mechanism, which uses the current F0 as the pitch goal (variable referent). This suggests that F0 is not universally controlled by an efference mechanism, but that different mechanisms for F0 control and stabilization can be used to suit different goals within different contexts. This may have implications for theories of motor control and the universality of motor control mechanisms under different goals and contexts.

## Chapter 4:

### Conclusions

In Experiment 1, we examined the ERP response to an unpredictable mid-utterance perturbation. We found an MMN or MMN-like response, which appeared to be distinct from the N100 observed in Heinks-Maldonado et al. (2005, 2006). This raises the possibility that the neural response to perturbations at utterance onset may be different to the neural response to unanticipated mid-utterance shifts. We also compared utterances using a relative (or habitual) pitch to utterances matching a target pitch, and found no differences in either the behavioral or ERP results.

Experiment 1 is important because it is the first study to use ERPs to examine the neural response to changes in a participant's ongoing vocalization. While there has been a flood of research using ERPs to examine neural responses to auditory stimuli, there has been relatively little study of the response to vocal production or to changes in ongoing auditory stimuli (self generated or external). Using the methods pioneered in Experiment 1, we will be able to use ERPs to examine auditory feedback in a variety of tasks. For example, the relationship between the MMN and a conscious perception of auditory deviations might allow us to determine a perceptual threshold for the detection of a perturbation, and how perception of the shift relates to the compensation response.

Based on the results of Experiment 1, we hypothesized that the response to a shift at utterance onset might be different to a mid-utterance perturbation. Experiment 2 was conducted to examine this issue. Participants F0 was shifted either at utterance onset (predictably in an adaptation block or unpredictably in a control block), and a perturbation was introduced by removing the shift, or an equivalent perturbation was introduced to an otherwise unshifted trial.

The compensation response to the mid-utterance perturbations was identical in all conditions, suggesting the pitch-shift reflex is a voice stabilization mechanism that attempts to maintain a constant F0 rather than match an F0 goal or target. Importantly, the compensation response to the onset perturbation was larger than the response to the mid-utterance perturbation, suggesting that there may be differences in the mechanism used in vocal control at utterance onset and after a stable F0 is achieved.

The differences in auditory and motor response to mid-utterance and onset perturbations needs to be considered in future studies of auditory feedback. This is true of both imaging studies of vocalization, which often use mid-utterance perturbations (Zarate and Zatorre, 2005; 2008; Toyomura et al., 2008), and models of vocal control. In addition, this difference in the use of feedback control at initiation of a movement and when that movement is maintained can be extended into other types of studies of motor control. Many studies of motor control use brief reaching movements with a very discrete start and end point, and a specific goal (e.g., Thoroughman et al., 2007; Scheidt & Ghez, 2007). The motor commands for such actions might differ significantly from actions with a prolonged goal, such as maintaining a constant force while pushing an object. When force is initially exerted, feedback may be used to establish the appropriate level of force to achieve the desired goal, and the system can then switch to a maintenance strategy similar to that used in maintaining a steady vocalization. If this is true, it has many implications for studies of motor control.

Our understanding of how the motor system works (both during vocalization and other motor tasks), and how the brain uses feedback to control and update actions, is still poorly understood. Both ERP and behavioral methods can be useful to help us understand perception of

feedback and its relationship to motor control, and may also be useful to understand how these systems are damaged in a wide variety of motor and perceptual disorders.



## References

- Aaltonen, O., Tuomainen, J., Laine, M., & Niemi, P. (1993). Cortical differences in tonal versus vowel processing as revealed by an ERP component called mismatch negativity (MMN). *Brain and Language, 44*, 139-152.
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Research: Cognitive Brain Research, 13(1)*, 17-26.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*, 309-312.
- Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *Neuroreport, 11*, R11-16
- Boersma P. (2001). Praat, a system for doing phonetics by computer. *Glott International, 5*, 341-345.
- Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America, 103(6)*, 3153-3161.
- Burnett, T. A., & Larson, C. R. (2002). Early pitch-shift response is active in both steady and dynamic voice pitch control. *Journal of the Acoustical Society of America, 112(3 Pt 1)*, 1058-1063.
- Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *Journal of the Acoustical Society of America, 121(2)*, 1157-1163.
- Cowie, R. I., & Douglas-Cowie, E. (1983). Speech production in profound post-lingual deafness. In M. E. Lutman & M. P. Haggard (Eds.), *Hearing Science and Hearing Disorders* (pp. 183-231). New York: Academic Press.
- Creutzfeldt, O., Ojemann, G., & Lettich, E. (1989). Neuronal activity in the human lateral temporal lobe: II. Responses to the subject's own voice. *Experimental Brain Research, 77*, 476-489.
- Curio G., Neuloh G., Numminen J., Jousmäki V., & Hari R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping, 9*, 183-191.
- Dizio P., Lackner J.R. (1995). Motor adaptation to Coriolis force perturbations of reaching movements: endpoint but not trajectory adaptation transfers to the nonexposed arm. *Journal of Neurophysiology, 74*, 1787-1792.

- Donath T.M., Natke U., & Kalveram K.T. (2002). Effects of frequency-shifted auditory feedback on voice F0 contours in syllables. *Journal of the Acoustical Society of America*, 111, 357-366.
- Elman, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *Journal of the Acoustical Society of America*, 70, 45-50.
- Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2005). Sensitivity to voice in human prefrontal cortex. *Journal of Neurophysiology*, 94, 2251-2254.
- Ford, J., Mathalon, D., Heinks, T., Kalba, S., & Roth, W. (2001). Neurophysiological evidence of corollary discharge dysfunctions in schizophrenia. *American Journal of Psychiatry*, 158, 2069-2071.
- Fu, C. H., Vythelingum, G. N., Brammer, M. J., Williams, S. C., Amaro, E. Jr., Andrew, C. M., Yaguez, L., van Haren, N. E., Matsumoto, K., & McGuire, P. K. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex*, 16, 969-977.
- Giard, M. H., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiology*, 27, 627-640.
- Grimm, S., & Schröger, E. (2005). Pre-attentive and attentive processing of temporal and frequency characteristics within long sounds. *Brain Research: Cognitive Brain Research*, 25(3), 711-721.
- Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, 39, 350-65.
- Hafke, H. Z. (2008). Nonconscious control of fundamental voice frequency. *Journal of the Acoustical Society of America*, 123, 273-278.
- 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, 133-141.
- Hari, R., Pelizzone, M., Makela, J., Hallstrom, J., Leinonen, L., & Lounasmaa, O. V. (1987). Neuromagnetic responses of the human auditory cortex to on- and offsets of noise bursts. *Audiology*, 26, 31-43.
- Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, 42, 180-190.
- Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport*, 17, 1375-1379.

- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279, 1213-1216.
- Houde J. F., Nagarajan S.S., Sekihara K., & Merzenich M.M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience*, 14, 1125-1138.
- Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech: I. Compensation and adaptation. *Journal of Speech, Language and Hearing Research*, 45, 295-310.
- Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research*. 190, 279-287.
- Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America*, 108, 1246-1251.
- Jones, J. A. & Munhall, K. G. (2002). The role of auditory feedback during phonation: Studies of Mandarin tone production. *Journal of Phonology*, 30, 303-320.
- Jones, J. A. & Munhall, K. G. (2005). Remapping auditory-motor representations in voice production. *Current Biology*, 15, 1768-1772.
- Kalikow D.N., Stevens K.N., Elliott L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, 61, 1337-1351.
- Kawahara, H. (1995). *Hearing Voice: Transformed auditory feedback effects on voice pitch control*. Paper presented at the 'Computational Auditory Scene Analysis' and International Joint Conference on Artificial Intelligence, Montreal, Quebec.
- Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lutkenhoner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, 13, 765-772.
- Kudo, N., Nakagome, K., Kasai, K., Araki, T., Fukuda, M., Kato, N., & Iwanami, A. (2004). Effects of corollary discharge on event-related potentials during selective attention task in healthy men and women. *Neuroscience Research*, 48, 59-64.
- Lang, H., Eerola, O., Korpilahti, P., Holopainen, I. E., Salo S. K., & Aaltonen, O. (1995). Practical issues in the clinical application of mismatch negativity. *Ear and Hearing*, 16, 118-130.
- Larson, C. R. (1998). Cross-modality influences in speech motor control: The use of pitch shifting for the study of F0 control. *Journal of Communication Disorders*, 31, 489-502.

- Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., & Hain, T. C. (2001). Comparison of voice F0 responses to pitch-shift onset and offset conditions. *Journal of the Acoustical Society of America*, 110, 2845-2848.
- Lavikainen J., Huotilainen M., Ilmoniemi R. J., Simola J. T., & Näätänen R. (1995). Pitch change of a continuous tone activates two distinct processes in human auditory cortex: a study with whole-head magnetometer. *Electroencephalography Clinical Neurophysiology*, 96, 93-96.
- 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, 3671-3677.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral Brain Science*, 21, 499-511.
- Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.
- Näätänen, R., & Alho, K. (1995). Mismatch negativity-a unique measure of sensory processing in audition. *International Journal of Neuroscience*, 80, 317-337.
- Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42, 313-329.
- Näätänen, R., Paavilainen, K., Alho, K., Reinikainen, K., & Sams, M. (1989). Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neuroscience Letters*, 98, 217-221.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118, 2544-2590.
- Natke, U., Donath, T. M., & Kalveram, K. T. (2003). Control of voice fundamental frequency in speaking versus singing. *Journal of the Acoustical Society of America*, 113, 1587-1593.
- Natke U., & Kalveram K.T. (2001). Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *Journal of Speech, Language and Hearing Research*, 44, 577-584.
- Nowak D.A., Topka H., Timmann D., Boecker H., & Hermsdörfer J. (2007). The role of the cerebellum for predictive control of grasping. *Cerebellum*, 6, 7-17.
- Numminen, J., & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters*, 272, 29-32.

- Ozaki, I., Suzuki, Y., Jin, C., Baba, M., Matsunaga, M., & Hashimoto, I. (2003). Dynamic movement of N100m dipoles in evoked magnetic field reflects sequential activation of isofrequency bands in human auditory cortex. *Clinical Neurophysiology*, *114*, 1681-1688.
- Pantev, C., Eulitz, C., Hampson, S., Ross, B., & Roberts, L. (1996). The auditory evoked "off" response: Sources and comparison with the "on" and the "sustained" responses. *Ear & Hearing*, *17*, 255-265.
- Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., & Guiod, P. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication*, *22*, 227-250.
- Purcell, D. W., Munhall, K. G. (2006). Compensation following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America*, *119*, 2288-2297.
- Rinne, T., Degerman A., & Alho, K. (2005). Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: An fMRI study. *NeuroImage*, *26*, 66-72.
- Scheidt R.A., & Ghez C. (2007). Separate adaptive mechanisms for controlling trajectory and final position in reaching. *Journal of Neurophysiology*, *98*, 3600-3613.
- Schröger, E. (1996). The influence of stimulus intensity and inter-stimulus interval on the detection of pitch and loudness changes, *Electroencephalography and Clinical Neurophysiology*, *100*, 517-526.
- Tervaniemi, M., Maury, S., & Näätänen, R. (1994). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport*, *5*, 844-846.
- Thoroughman K.A., Fine M.S., & Taylor J.A. (2007). Trial-by-trial motor adaptation: a window into elemental neural computation. *Progress in Brain Research*, *165*, 373-382.
- Thoroughman K.A., & Shadmehr R. (1999). Electromyographic correlates of learning an internal model of reaching movements. *Journal of Neuroscience*, *19*, 8573-8588.
- Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., & Kuriki, S. (2007). Neural correlates of auditory feedback control in human. *Neuroscience*, *146*, 499-503.
- Villacorta, V. M., Perkell J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations on vowel acoustics and its relation to perception, *Journal of the Acoustical Society of America*, *122*, 2306-2319.

- Wolpert D.M., Kawato M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11, 1317-1329.
- Zatorre, R. J., Belin P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Science*, 6(1), 37-46.
- Zarate, J. M., & Zatorre, R. J. (2005). Neural substrates governing audiovocal integration for vocal pitch regulation in singing. *Annals of the New York Academy of Science*, 1060, 404-408.
- Zarate, J. M., Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage*, 40, 1871-1887.