2016

An Examination of the Factors that Dictate the Relative Weighting of Feedback and Feedforward Input for Speech Motor Control

Nichole E. Scheerer
Wilfrid Laurier University, nikkischeerer@gmail.com

Recommended Citation

Scheerer, Nichole E., "An Examination of the Factors that Dictate the Relative Weighting of Feedback and Feedforward Input for Speech Motor Control" (2016). *Theses and Dissertations (Comprehensive)*. 1802.
http://scholars.wlu.ca/etd/1802

An Examination of the Factors that Dictate the Relative Weighting of Feedback and Feedforward Input for Speech Motor Control

by

Nichole Elizabeth Scheerer

Submitted to the Department of Psychology

in partial fulfilment of the requirements for

Doctor of Philosophy in Psychology: Cognitive Neuroscience

Wilfrid Laurier University

Declaration of Co-Authorship/Previous Publication

Study 1:

Scheerer, N.E, Jacobson, D.S., & Jones, J.A. (In Prep). The Role of Auditory Feedback for Speech Motor Control in Toddlers.

Study 2:

Published as:

Scheerer, N., Liu, H., & Jones, J.A. (2013). The Developmental Trajectory of Vocal and ERP Responses to Frequency Altered Auditory Feedback. European Journal of Neuroscience, 38(8), 3189-3200.

Study 3:

Submitted as:

Scheerer, N.E., Jacobson, D.S., & Jones, J.A. (2015). Sensorimotor Learning in Children and Adults: Exposure to Frequency-Altered Auditory Feedback during Speech Production. Neuroscience. Accepted: November 18, 2015.

Study 4:

Published as:

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

Study 5:

Scheerer, N.E., & Jones, J.A. (In Prep). The relationship between vocal variability and sensorimotor learning.

Study 6:

Published as:

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

Abstract

Speech is arguably the most important form of human communication. Fluent speech production relies on auditory feedback for the planning, execution, and monitoring of speech movements. Auditory feedback is particularly important during the acquisition of speech, however, it has been suggested that over time speakers rely less on auditory feedback as they develop robust sensorimotor representations that allow speech motor commands to be executed in a feedforward manner. The studies reported in this thesis recorded speaker's vocal and neural responses to altered auditory feedback in order to explore the factors that dictate the relative importance of auditory feedback for speech motor control. More specifically, studies 1 through 3 examined how the role of auditory feedback changes throughout development, while studies 4 and 5 examined the relationship between vocal variability and auditory feedback control, and lastly study 6 looked at how the predictability of auditory feedback errors influences the role of auditory feedback for speech motor control. Results of the first study demonstrated that toddlers use auditory feedback to regulate their speech motor commands, supporting the long held notion that auditory feedback is important during the acquisition of speech. While mapping out the developmental trajectory of vocal and event related potential responses to altered auditory feedback, the second study demonstrated that vocal variability, rather than age, best predicts responses to altered auditory feedback. Importantly, this suggests that the maturation of the speech motor control system is not strictly dependent on age. The third study in this thesis demonstrated that children and adults show similar rates of sensorimotor adaptation, suggesting that once speech is acquired, speakers are proficient at using sensory information to modify the planning of future speech motor commands. However, since adults produced larger compensatory responses, these results also suggested that adults are more proficient at comparing incoming auditory feedback with the feedback predicted by their sensorimotor representations, as a result of possessing more precisely mapped sensorimotor representations. The results of studies four and five demonstrated that vocal variability can be used to predict the size of compensatory responses and sensorimotor adaptation to changes in one's auditory feedback, respectively. Furthermore, these studies demonstrated that increased variability was related to increased auditory feedback control of speech. Finally, the sixth study in this thesis demonstrated that experimentally induced predictability and variability can be used to induce increases in feedforward and auditory feedback control, respectively. In conclusion, the results reported in this thesis demonstrate that age and vocal variability, both naturally occurring and experimentally induced, are important determinants of the role of auditory feedback in speech motor control.

Acknowledgements

This document, as well as the enclosed work, would not have been possible without the guidance and support of many people.

First and foremost, I would like to thank my supervisor Jeff, as before I met you I did not know graduate school was a thing. Not only did you encourage me to pursue a graduate career, but you have provided me with endless opportunities over the past 6 years. I hope you take pride in what I have accomplished, as my achievements would not have been possible with your unwavering trust and support. I consider myself extremely fortunate to have met you.

Throughout my graduate career I have also had the privilege of working along side many extraordinary faculty members and students. I would especially like to thank Dr. Diano Marrone, Dr. Philip Servos, and Dr. Bob Gebotys for the opportunities they provided me to further develop my research and teaching skills. Also, a special thanks to Doreen Weise for all of your advice and support — both personally and academically. To all of my colleagues, particularly Anu Tumber, Arden DeMarco, Dani Jacobson, Vivian Cheng, and Zeynep Barlas thank you for putting up with, and helping me to control my many neurotic behaviours. Also, to all of the undergraduate volunteers, thesis, directed study, and summer students, particularly my toddler whisperers, Andrea Mohr, Dani Jacobson, and Michelle Lewis, thank you for all of your help — many of these projects would not have been possible without you.

Last, but certainly not least, I would like to thank my family and friends. Mom and Dad, you raised me to be hardworking and independent, and it is those skills that have allowed me to excel over the past few years. Josh and Amy, you have both been extremely encouraging throughout this whole process and never let me believe I could do anything but succeed. To all my friends that have cheered me on throughout this process, particularly Amanda, Danielle, Laura, and Marissa, I am extremely lucky to have you in my life. Although these past few years have not always been easy, having a great family and friends to fall back on definitely kept me going. By whisking me away on vacations, joining me at Blue Jays games, or simply listening to me vent about how Jeff is the worst, you all helped pave the way to this finish line.

Table of Contents

List of Figures

Study 1 — The Role of Auditory Feedback for Speech Motor Control in Toddlers

Study 2 — The Developmental Trajectory of Vocal and ERP Responses to Frequency Altered Feedback

Study 3 — Sensorimotor Learning in Children and Adults: Exposure to Frequency-Altered Auditory Feedback during Speech Production

Study 4 — The Relationship between Vocal Accuracy and Variability to the Level of Compensation to Altered Auditory Feedback

Study 5 — The Relationship between Vocal Variability and Sensorimotor Learning

Study 6 — The Predictability of Frequency-Altered Auditory Feedback Changes the Weighting of Feedback and Feedforward Input for Speech Motor Control

List of Abbreviations

AAF - Altered Auditory Feedback

AEP - Auditory Evoked Potential

ANOVA - Analysis of Variance

ASD - Autism Spectrum Disorder

DIVA - Directions into the velocity of the articulators

DSP - Digital Signal Processor

DTI - Diffusion Tensor Imaging

EEG - Electroencephalography

ERP - Event Related Potential

F0 - Fundamental Frequency

FAF - Frequency Altered Feedback

fMRI - Functional Magnetic Resonance Imaging

LSD - Least Significant Difference

MEG - Magnetoencephalography

PET - Positron Emission Tomography

pIFG - Posterior Inferior Frontal Gyrus

RM - Repeated Measures

SD - Standard Deviation

SFC - State Feedback Control

TD - Typically developed

TMS - Transcranial Magnetic Stimulation

vPMC - Ventral Premotor Cortex

General Introduction

Speech is arguably the most important form of human communication. Speech allows complex ideas and emotions to be conveyed between individuals, and is a skill that is unique to the human species (Rauschecker & Scott, 2009; Kingyon et al., 2015). Since the functional goal of speech production is the transfer of information, speech production must be carefully regulated to ensure its intelligibility is maintained (Burnett, Senner, & Larson, 1997; Perkell, 2012). The regulation of speech production requires the precise coordination of the respiratory, laryngeal, and supralaryngeal (articulatory) systems, as speech would not be possible without air passing through the glottis, vibration of the vocal folds, and the resonance characteristics imposed by movements of the articulators, respectively (Jurgens, 2002, 2009; Zarate, 2013). In order to regulate the activity of the muscles that make up the speech motor control system, the central nervous system relies on peripheral sensory feedback, such as auditory and somatosensory feedback for the planning, execution, and monitoring of speech movements (Guenther, 2006). Regulation of the speech motor control system is particularly challenging during development, as the structures involved in speech production undergo dramatic changes (Guenther, 1994; Callan, Kent, Guenther, & Vorperian, 2000). For this reason, it is likely that sensory feedback plays a critical role in maintaining the intelligibility of speech throughout development (Callan et al., 2000). However, once the speech motor control system is fully developed, and the structures involved in speech production have stabilized, the role of sensory feedback for speech motor control may change.

The Role of Auditory Feedback for Speech Motor Control

It has been argued that auditory feedback, or the sound of one's voice, is critical for the acquisition of speech (Oller & Eilers, 1988; Hickok, Houde, & Rong, 2011). Support for this argument comes from examining individuals with congenital deafness, as without auditory feedback these individuals fail to acquire fluent speech (Smith, 1975; Svirsky, Teoh, & Neuburger, 2004). However, individuals who become deafened post-lingually are generally able to maintain relatively fluent speech, despite the loss of auditory feedback, suggesting that the role of auditory feedback for speech motor control changes throughout one's lifespan (Cowie, Douglas-Cowie, & Kerr, 1982; Goehl & Kaufman, 1984; Cowie & Douglas-Cowie, 1992).

Models of Speech Motor Control

Fluent speech production depends on the precise control of the musculature of the respiratory, laryngeal, and articulatory systems (Tourville & Guenther, 2011; Guenther & Vladusich, 2012). For this reason, speech production models are generally built upon more general motor control principles. In the motor control literature, movement regulation is often described as being under feedback or feedforward control (Max, Guenther, Gracco, & Ghosh, 2004). Under feedback control, motor commands are generated by constantly monitoring sensory information regarding effector positions and comparing this sensory information to the target position of the effector (Max et al., 2004). Based on this information, motor commands are then executed in parallel with the movement (Max et al., 2004). On the other hand, under feedforward control, motor commands are generated prior to the onset of a movement, and are executed by the musculature without alteration (Max et al., 2004). Since speech production occurs rapidly, and the processing of peripheral sensory feedback is not instantaneous, a speech motor control system relying solely on feedback control would result in disfluent speech (Perkell, Matthies, Lane, & Guenther, 1997; Perkell, 2012). However, a speech motor control system operating under strict feedforward control would also be implausible, as it would result in a rigid control system incapable of adjusting to fluctuating environmental conditions and developmental changes to the respiratory, laryngeal, and articulatory systems, amongst other things (Perkell et al., 1997). For this reason, the two most widely accepted models of speech motor control, the directions into the velocities of the articulators model (DIVA; Guenther, 2006; Villacorta, Perkell, & Guenther, 2007; Tourville, Reilly, & Guenther, 2008), and the state feedback control model (SFC; Hickok et al., 2011; Houde & Nagarajan, 2011; Houde & Chang, 2015), suggest that fluent speech production relies on an integrated control system that combines both feedback and feedforward control.

Directions into the Velocities of the Articulators (DIVA)

The DIVA model is a neurocomputational model that describes the roles of various brain regions in the acquisition and production of speech (Guenther, 1994, 1995, 2006; Guenther, Ghosh, & Tourville, 2006; Guenther, Hampson, & Johnson, 1998; Tourville et al., 2011; Guenther & Vladusich, 2012). According to the DIVA model, the production of a speech sound begins with the activation of a small population of neurons, or a neural representation, associated with the to-be-produced sound (Tourville et al., 2011). These neural representations have been suggested to reside in the left ventral premotor cortex (vPMC), as well as the posterior inferior frontal gyrus (pIFG; Guenther & Vladusich, 2012). Following the activation of a neural representation, both the feedback and feedforward control systems send motor commands to the primary motor cortex (Tourville et al., 2011; Guenther & Vladusich, 2012). The motor commands executed by the feedforward system travel from the neural representation to articulatory control units in the cerebellum, before arriving at the primary motor cortex. Whereas, the motor commands executed by the feedback system pass through auditory and somatosensory feedback control subsystems, before reaching the primary motor cortex. These sensory feedback control subsystems are proposed to play a critical role in the sensorimotor transformations required for the execution and maintenance of sensorimotor representations (Tourville et al., 2011; Guenther & Vladusich, 2012).

During development, the DIVA model suggests that random articulatory movements, or babbling, leads to the detection of novel speech sounds (Guenther, 2006). These novel speech sounds activate neurons in the vPMC and pIFC that were previously unused. This activation creates an auditory target, represented as a time-varying acoustic signal, for the novel sound (Guenther, 2006; Tourville et al., 2011; Guenther & Vladusich, 2012). This auditory target is then encoded in synaptic projections from the newly formed neural representation to the auditory brain regions including Heschl's gyrus, the posterior superior temporal gyrus, the posterior superior temporal sulcus, and the planum temporale. Importantly, this auditory target encodes the allowable variability in the acoustic signal, which is crucial for later error monitoring and correction during the production of the speech sound (Tourville et al., 2011; Guenther & Vladusich, 2012). Once the neural representation for the auditory target is established, it can then be activated for production of the associated speech sound (Guenther, 2006). During production, activation of the neural representation results in projections to both the feedforward and feedback control systems. The projection to the feedforward system is in the form of a time-varying signal that encodes the articulator velocities for the production of the auditory target, and is sent to the articulatory control regions of the motor cortex. On the other hand, the projection to the feedback system contains the time-varying auditory expectations associated with the activated neural representation. These projections to the feedback system result in inhibition of the auditory regions dedicated to processing the expected auditory feedback for the auditory target. At the same time, these auditory regions are also receiving the auditory feedback resulting from the articulatory movements. Since the auditory regions are receiving inhibitory input from the activated neural representation, and excitatory input resulting from the articulatory movements, the net activity in the auditory regions represents the size of the auditory error. This auditory error, calculated as the difference between the expected and actual sensory consequences of the activated neural representation, is then transformed into a feedback-based articulator velocity command using an auditory-to-motor transformation. This feedback-based articulator velocity command is sent to the articulatory control regions of the motor cortex to modify the ongoing movement. This corrective command is also incorporated into the feedforward command associated with the activated neural representation, such that future activation of the representation results in more accurate articulatory movements (Guenther, 2006; Tourville et al., 2011; Guenther & Vladusich, 2012). Importantly, the DIVA model predicts that eventually the feedforward command will be sufficient for the production of the auditory target (Guenther, 2006). Specifically, once the feedforward command is well tuned, the actual sensory consequences of the articulatory movements will match the auditory predictions from the neural representation, making feedback based movement corrections unnecessary. Following repeated production of a speech sound, a somatosensory target for the sound is also learned (Guenther, 2006). This somatosensory target encodes the expected tactile and proprioceptive sensations associated with the production of the sound, and allows somatosensory feedback based error correction to occur in a manner similar to that described for auditory feedback (Guenther, 2006; Tourville et al., 2011; Guenther & Vladusich, 2012).

State Feedback Control (SFC)

The state feedback control model (Hickok et al., 2011; Hickok, 2012; Houde and Chang, 2015) is a speech motor control model that is derived from optimal feedback control models of motor behaviour (c.f. Todorov, 2004). The SFC model states that when a speech sound is produced, a motor controller sends a motor command to the vocal tract in a ‘feedforward’ manner, while also sending an efference copy of the motor command to an internal model of the vocal tract. This efference copy of the motor command contains an estimate of the current dynamic state of the vocal tract, which is based on previously learned mappings between speech motor commands and their resultant sensory consequences. This state estimate allows for both internal and external feedback control. Internal feedback control uses the state estimate to predict the sensory consequences of the motor command prior to the arrival of peripheral sensory feedback. If the sensory consequences predicted based on the state estimate differ from the intended speech target, corrective feedback can be sent to the motor controller so that the motor behaviour is modified before the erroneous peripheral sensory feedback arrives. Similarly, external feedback control allows peripheral sensory feedback to be compared with the state estimate, and deviations from this state estimate can then be used to generate an error correction signal that is sent to the motor controller to modify the ongoing movements. The external feedback system also sends a copy of the corrective command to the internal model to update the state estimate, so that future motor commands are more accurate. While the conceptualization of speech motor control by the SFC model is quite similar to that of the DIVA model, the SFC model emphasizes the role of state feedback (internal feedback) over peripheral sensory feedback (external feedback) for the regulation of ongoing speech movements (Hickok et al., 2011; Hickok, 2012).

Both the DIVA and SFC models of speech motor control emphasize the role of auditory feedback for creating and maintaining representations that support speech motor control. However, both the DIVA and SFC models also recognize that online error correction using peripheral sensory feedback is impractical, as peripheral sensory feedback is unreliable as it can be noisy, delayed, and intermittently absent (Guenther, 2006; Purcell & Munhall, 2006a; Tourville et al., 2008; Hickok et al., 2011; Hickok, 2012). Hence, it has been suggested that as development progresses, the speech motor control system relies more heavily on feedforward commands (Bailly, 1997; Guenther & Bohland, 2002, Max et al., 2004; Guenther, 2006; Villacorta et al., 2007; Tourville et al., 2008; Perkell, 2012; Beal, Quraan, Cheyne, Taylor, Gracco, DeNil, 2011; Kim & Max, 2014). Feedforward control permits speech motor commands to be executed by stored sensorimotor representations that encode the relationship between the motor commands responsible for speech, and the sensory consequences of these commands (Tourville et al., 2008). Since these feedforward commands are read out from memory, the delays associated with the processing of peripheral sensory feedback are avoided (Houde, Nagarajan, & Sekihara, 2002; Jones & Munhall, 2002; Guenther et al., 2006). Despite this transition to feedforward control, auditory feedback continues to play an important role in maintaining the precision and accuracy of the mapping of the sensorimotor representations that drive feedforward control (Perkell et al., 1997; Civier, Tasko & Guenther, 2010). While the role of auditory feedback in fluent speech production has been suggested to change over time, it is currently unclear which factors dictate the relative weighting of feedback and feedforward input for speech motor control (Kim & Max, 2014).

Frequency Altered Feedback

Much of the empirical evidence that supports the involvement of both feedback and feedforward control systems in fluent speech production comes from experimental paradigms that utilize frequency altered feedback (FAF; Elman, 1981; Burnett et al., 1997; Burnett, Freedland, Larson, & Hain, 1998; Burnett & Larson, 2002; Jones & Munhall, 2005; Jones & Keough, 2008; Hawco & Jones, 2009; Keough & Jones, 2009, 2011; Liu, Chen, Larson, Huang, & Liu, 2010a, Liu, Russo, & Larson, 2010b; Scheerer et al., 2013a; Tumber, Scheerer, & Jones, 2014). As part of these paradigms, speakers are exposed to changes to their auditory feedback, and any changes in the speaker's fundamental frequency (F0) resulting from the manipulation are measured. Previous research has shown that speakers tend to compensate for the FAF manipulation by shifting their voice in the opposite direction of the change. For example, in response to an increase in the vocal pitch of a speaker's auditory feedback, the speaker tends to respond by decreasing their vocal pitch (Burnett et al., 1997, 1998, 2002; Jones & Munhall, 2005; Jones & Keough, 2008; Hawco & Jones, 2009; Keough & Jones, 2009, 2011; Scheerer et al., 2013a; Tumber et al., 2014). By exposing speakers to unexpected FAF perturbations to their ongoing speech, it is possible to assess how the speakers monitor and correct for errors in their ongoing speech. On the other hand, by exposing speakers to persistent predictable FAF, in addition to assessing how speakers monitor and correct for errors in their ongoing speech, it is also possible to assess if, and how, speakers use the deviant auditory feedback to plan future speech motor commands (Villacorta et al., 2007). For this reason, the FAF perturbation paradigm is useful for identifying the role of the feedback control system in the online control of speech motor commands, while the FAF adaptation paradigm can be used to assess both feedback and feedforward control simultaneously. Furthermore, since the synthetic manipulations imposed on a speaker's auditory feedback during FAF tasks mimic naturally occurring sensory events that may arise as a result of developmental changes or environmental fluctuations, amongst other things, FAF provides an effective means for revealing important properties of the speech motor control system (Bauer & Larson, 2003).

Frequency Altered Feedback (FAF) Perturbation Paradigm

The FAF perturbation paradigm is most commonly used to assess responses to changes to the F0 of a speaker's auditory feedback. As part of this paradigm speakers are exposed to brief unpredictable perturbations to the F0 of their auditory feedback (Burnett et al., 1997, 1998, 2002; Jones & Munhall, 2005; Liu et al., 2010a, 2010b; Scheerer et al., 2013a; Tumber et al., 2014). These brief perturbations have been shown to consistently elicit a rapid compensatory response, with a latency of approximately 100-150 ms (Burnett et al., 1997, 1998), that has been termed the 'pitch-shift reflex' (Larson, Carrell, Senner, Burnett, & Nichols, 1995; Burnett et al., 1998, 2002; Bauer et al., 2003; Bauer, Mittal, Larson, & Hain, 2006; Liu et al., 2010a; Liu, Meshman, Behroozmand, & Larson, 2011). Due to its reflexive nature, this response has been suggested to play a role in stabilizing voice F0 around a desired target (Hain et al., 2000; Natke, Donath, & Kalveram, 2003, Bauer et al., 2006; Hawco & Jones, 2009). Using the FAF perturbation paradigm, researchers have been able to assess how various manipulations to the F0 of a speaker's auditory feedback influences the pitch shift reflex. Valuable information about the feedback control of speech motor control has come from studies that have varied task instruction (Hain et al., 2000), perturbation magnitude (Scheerer et al., 2013a), and perturbation onset (Hawco & Jones, 2009).

i. Task Instruction

Hain and colleagues (2000) exposed speakers to FAF perturbations and asked them to respond by either opposing the direction of the pitch shift stimulus, following the direction of the pitch shift stimulus, or ignoring the pitch shift stimulus, in attempt to investigate the extent to which the pitch shift reflex is under voluntary control. The results of this study revealed that speakers made both an early and a late response to the pitch shift. The early response had a shorter latency and smaller magnitude than the late response, and was almost always compensatory regardless of the task instructions. On the other hand, the late response was larger and was almost always produced in the direction specified by the task instructions. These results suggest that changes in voice F0 elicit an automatic compensatory response that may serve to stabilize F0 against internal perturbations. However, these changes in voice F0 are also regulated by slower more voluntary response that may serve to maintain F0 at a desired level.

ii. Perturbation Magnitude

A consistent finding in the FAF literature is that compensatory responses to FAF are only a fraction of the size of the imposed feedback perturbation (Burnett et al., 1997, 1998; Donath, Natke, & Kalveram, 2002; Bauer & Larson, 2003; Natke et al., 2003; Chen et al., 2007). In order to systematically investigate the relationship between FAF perturbation magnitudes and compensatory responses to FAF, Scheerer and colleagues (2013a) parametrically manipulated the magnitude of perturbations to speaker's auditory feedback and recorded the size of the speaker's compensatory responses. The results of this study indicated that smaller magnitude feedback perturbations (50-250 cents, 100 cents = 1 semitone) elicited a small but reliable response that was similar across the varying perturbation magnitudes. On the other hand, larger feedback perturbations (> 300 cents) resulted in relatively smaller vocal response magnitudes. These results suggest that smaller magnitude feedback perturbations elicit one size of response, while larger magnitude feedback perturbations result in relatively smaller vocal response magnitudes. These results support the notion that the reflexive response to FAF is a limited response intended to stabilize vocal F0 (Burnett et al., 1998; Hain et al., 2000; Bauer et al., 2006). In addition, the drop in magnitude of speaker's responses for relatively large perturbation magnitudes, may indicate that speakers ignore auditory feedback if it is dramatically different from the sound in which they are trying to produce (Burnett et al., 1997; Scheerer et al., 2013a). This control strategy is optimal, as large feedback errors likely fall outside of the range of errors normally experienced by the speaker, thus they are unlikely to represent true production errors (Burnett et al., 1997; Sober & Brainard, 2012; Scheerer et al., 2013a; Kuebrich & Sober, 2015).

iii. Perturbation Onset

Hawco and Jones (2009) investigated whether the timing of the onset of a perturbation influenced responses to that perturbation. In this study, speakers were adapted to a 100 cent shift in their auditory feedback, which was then briefly removed to induce an auditory feedback perturbation. In a control condition, a 100 cent perturbation was introduced midway through an unaltered vocalization, or the vocalization was randomly shifted upwards 100 cents, and then a perturbation was introduced by briefly removing the feedback alteration. Speakers produced similar sized compensatory responses to all mid-vocalization auditory feedback perturbations, regardless of whether the perturbation was introducing a pitch shift, or removing a pitch shift. However, compensatory responses to pitch changes at voice onset in both the adaptation and control conditions were much larger than those induced mid-vocalization. These results suggest that auditory feedback control operates differently at voice onset and midway through a vocalization. Based on these results the authors suggest that at vocalization onset a static reference, such as a sensorimotor representation or a target note, is used to match auditory feedback to the intended F0. On the other hand, after vocalization onset a more dynamic reference, such as the current F0, is used to match auditory feedback to the target F0.

Assessing responses to brief FAF perturbations of varying magnitudes, onsets, and under various task instructions has provided valuable information regarding the auditory feedback control of speech motor control. Generally speaking, these results suggest that individuals automatically compensate to deviations in their auditory feedback in order to stabilize their voice F0 (Burnett et al., 1998; Hain et al., 2000; Bauer et al., 2006). Furthermore, the size and the conditions under which these responses are produced suggest that the reflexive responses to changes in auditory feedback are part of a fine tuning mechanism, rather than a system geared to make elaborate adjustments to vocal F0 (Burnett et al., 1998; Bauer et al., 2006). The limited nature of these reflexive responses seems optimal for speech motor control, as it allows cognitive and voluntary mechanisms of voice control to dominate and prevents excessive feedback control, which may result in unwanted fluctuations of voice F0 with nearby environmental sounds, other speaker’s voices, etc. (Burnett et al., 1998; Bauer et al., 2006).

Frequency Altered Feedback Adaptation Paradigm

While the FAF perturbation paradigm is useful for assessing online control of voice F0, it does not allow researchers to assess the role of auditory feedback in motor planning. For this reason, researchers often utilize the FAF adaptation paradigm. As part of this paradigm, speakers are exposed to persistent and predictable changes to aspects of their auditory feedback, such as its F0 (Jones & Keough, 2008, Keough & Jones, 2009, 2011; Hawco & Jones, 2010; Keough, Hawco, & Jones, 2013) or its formant frequencies (Houde & Jordan, 1998, 2002; Purcell & Munhall, 2006a; Villacorta et al., 2007; Cai et al., 2012). Much like the FAF perturbations, these manipulations tend to elicit compensatory vocal responses, allowing the role of auditory feedback for ongoing vocal control to be assessed. In addition, persistent exposure to these FAF manipulations results in aftereffects, or a persistence of the compensatory response after the FAF manipulation has been removed (Jones & Keough, 2008; Hawco & Jones, 2009). These aftereffects suggest that the deviant auditory feedback was used to update the mapping of the speaker's sensorimotor representations, so that subsequent vocal productions more closely resemble the characteristics of the speaker's unaltered voice (Jones & Keough, 2008; Hawco & Jones, 2009). For this reason, these aftereffects can be used to assess the influence of auditory feedback on motor planning.

i. Fundamental Frequency Manipulations

When F0 is manipulated as part of the FAF adaptation paradigm, speakers produce vocalizations that are initially unaltered (the baseline phase), followed by vocalizations where the researcher alters the F0 of the speaker's vocalization by shifting it upwards or downwards (the shifted phase) either gradually (Keough & Jones, 2009, 2011; Keough, Hawco, & Jones, 2013) or abruptly (Jones & Keough, 2009, Hawco & Jones, 2010). During the shifted phase the speaker's vocalizations are altered prior to voice onset, and this alteration is maintained for several trials, which results in persistent and predictable exposure to FAF (Keough & Jones, 2009, 2011; Hawco & Jones, 2010; Keough et al., 2013). Following the shifted phase, the F0 manipulation is abruptly removed (the test phase), and speaker's once again produce vocalizations while listening to their unaltered auditory feedback (Keough & Jones, 2009, 2011; Hawco & Jones, 2010; Keough et al., 2013).

To assess speaker's responses during the FAF adaptation paradigm researchers generally focus on two different measures, one that assesses the speaker's F0 at the beginning of their vocalization, and one that assesses the speaker's F0 across their entire vocalization (Keough & Jones, 2009, 2011; Hawco & Jones, 2010; Keough et al., 2013). Measuring the speaker's F0 at the beginning of their vocalization allows sensorimotor learning to be assessed (Keough & Jones, 2009, 2011; Hawco & Jones, 2010; Keough et al., 2013). Since auditory feedback is delayed by cortical processing, the F0 at the beginning of a speaker's vocalization can be used to index feedforward control, or the extent to which the deviant auditory feedback is used for the planning of subsequent speech motor commands (Keough & Jones, 2009, 2011; Hawco & Jones, 2010; Keough, et al., 2013). On the other hand, by looking at the speaker's F0 across each vocalization, it is possible to index the extent to which the deviant auditory feedback modulated the ongoing vocalization (Keough & Jones, 2009, 2011; Hawco & Jones, 2010; Keough et al., 2013).

Results of previous studies utilizing the FAF adaptation paradigm have demonstrated that the F0 at the beginning of a speaker's vocalization gradually changes in the direction opposite to that of the FAF manipulation (Keough & Jones, 2009, 2011; Hawco & Jones, 2010; Keough et al., 2013). This demonstrates that the speaker is using the deviant auditory feedback to modify the subsequent speech motor commands so that their auditory feedback is more in line with their intended speech target. Once the FAF manipulation is removed, the compensatory response persists, further demonstrating that the deviant auditory feedback was used to modify the speaker's sensorimotor representation, and thus their subsequent speech motor commands (Keough & Jones, 2009, 2011; Hawco & Jones, 2010; Keough et al., 2013). Similarly, the speaker's F0 across the entire vocalization also changes in the direction opposite to the FAF manipulation (Keough & Jones, 2009, 2011; Hawco & Jones, 2010; Keough et al., 2013). Unlike the compensatory responses recorded following FAF perturbations, the magnitude of the compensatory response to these full vocalization shifts is quite large, often approximating the size of the FAF manipulation (Keough & Jones, 2009, 2011; Hawco & Jones, 2010; Keough et al., 2013). Thus, the compensatory responses produced in response to perturbation and adaptation F0 manipulations appear to be controlled in different manners.

i. Formant Frequency Manipulations

Similarly, when formant frequencies are manipulated as part of a FAF adaptation paradigm, speakers produce consonant-vowel-consonant (CVC) word(s) that are initially unaltered (the baseline phase), followed by CVC word(s) where the researcher alters the formant frequencies of the speaker's vocalization (the ramp phase) by gradually shifting the formant frequencies of the vowel upwards or downwards (Houde & Jordan, 1998, 2002; Purcell & Munhall, 2006a; Villacorta et al., 2007; Cai et al., 2012). Once the formant manipulation reaches its maximum value, the manipulation is held for a series of trials (the hold phase). Lastly, the speaker once again produces CVC word(s) while no longer exposed to the manipulated auditory feedback (the test phase), rather they hear their own unaltered auditory feedback (Houde & Jordan, 1998, 2002; Purcell & Munhall, 2006a; Villacorta et al., 2007).

During the ramp phase of the FAF adaptation paradigm, the formant frequency manipulations cause the production of a vowel sound, for example /ɛ/, to sound like another vowel sound, for example /i/, which induces a compensatory response where the speaker produces a vowel sound that more closely resembles the intended vowel sound (Houde &

Jordan, 1998, 2002; Purcell & Munhall, 2006a; Villacorta et al., 2007). Once the speaker's auditory feedback is returned to normal, the compensatory response persists, indicating the deviant auditory feedback was used to update the sensorimotor representation, thus future motor acts are influenced by exposure to the deviant auditory feedback (Purcell & Munhall, 2006a).

Assessing responses to both manipulations of fundamental and formant frequencies using the FAF adaptation paradigm provides further support for the notion that feedback and feedforward control systems regulate speech motor control. Online compensatory responses during the shifted phase of these experiments demonstrate that auditory feedback is being used to modify ongoing vocalizations in a feedback-like manner. On the other hand, the aftereffects measured when speakers auditory feedback is returned to normal, following exposure to the deviant auditory feedback, demonstrates that the deviant auditory feedback is used to update the speaker's sensorimotor representations. Thus, the fact that the deviant auditory feedback not only influences ongoing vocalizations, but also influences subsequent vocalizations, provides support for the notion that speech is regulated by both feedback and feedforward control systems.

Importantly, while both compensatory responses and aftereffects are observed following manipulations to F0 and formant frequencies using the adaptation paradigm, these responses differ in some respects. For example, when the F0 of a speaker's auditory feedback is manipulated, the size of the compensatory vocal response approximates the size of the F0 manipulation (Keough & Jones, 2009, 2011; Hawco & Jones, 2009, 2010; Keough et al., 2013). However, when the formant frequencies of a speaker's auditory feedback are manipulated, the compensatory vocal responses are often incomplete, and only account for a portion of the frequency manipulation (Purcell and Munhall, 2006a). These differences highlight the fact that not all of a sounds spectral characteristics are equally sensitive to sensory perturbations

(Smotherman, 2007). F0 is considered a postural setting of speech (Perkell et al., 1997), and is regulated by varying muscle tension in the larynx (Smotherman, 2007). This regulation of F0 helps to maintain the intelligibility of speech sounds (Perkell et al., 1997). On the other hand, formant frequencies are considered part of the phonemic speech settings (Perkell et al., 1997), and are regulated by changing the shape of the oral and nasal cavities (Smotherman, 2007). This regulation helps to maintain phonemic distinctions (Perkell et al., 1997). Interestingly, following a change in hearing status, F0 values (Cowie and Douglas-Cowie, 1992; Perkell et al., 1997, 2010), but not formant frequencies are found to change rapidly (Perkell et al., 1997). Since F0 values are found to change more rapidly following hearing loss, and manipulations of a speaker's F0 produces larger compensatory responses, relative to formant frequencies, these results suggest that F0 and formant frequencies may be regulated in different manners. For this reason, it is important to consider manipulations to F0 and formant frequencies separately.

Electroencephalography (EEG)

Assessing changes to speakers' vocal output following exposure to deviant auditory feedback provides a useful means for evaluating the role of auditory feedback in speech motor control. However, by monitoring auditory-cortical processing using EEG responses it is also possible to identify the neural correlates of speech motor control. Many neural imaging and recording techniques, including diffusion tensor imaging (DTI; Sommer, Koch, Paulus, Weiller, & Buchel, 2002; Peeva et al., 2013), functional magnetic resonance imaging (fMRI; Christoffels, Formisano, & Schiller, 2007; Tourville, Reilly, & Guenther, 2008; Beal et al., 2010, 2011; Golfinopoulos et al., 2011; Behroozmand et al., 2015), magnetoencephalography (MEG; Houde, Nagarajan, Sekihara, & Merzenich, 2002; Heinks-Maldonado, Nagarajan, & Houde, 2006; Aliu, Houde, and Nagarajan, 2009; Ventura, Nagarajan, & Houde, 2009; Niziolek, Nagarajan, & Houde, 2013; Tian & Poeppel, 2014), positron emission tomography (PET; Fox et al., 1996, 2000; Braun, Varga, Stager, Schulz, & Selbie, 1997), single-cell recordings (Eliades, 2002; Eliades & Wang, 2005, 2008; Greenlee et al., 2011; Kingyon et al., 2015), and transcranial magnetic stimulation (TMS; Grabski et al., 2013; Mottonen, van de Ven, & Watkins, 2014), have provided valuable information to aid in our understanding of the neural processes involved in speech motor control. However, since fluent speech production is a highly rapid motor behaviour, the superior temporal resolution that EEG affords (Jeste & Nelson III, 2008) makes it ideal for monitoring the neural correlates of speech production.

EEG is one of the oldest and most widely used experimental techniques for recording electrical activity from the brain (Buzsaki, Anastassiou, and Koch, 2012a). The electrical activity recorded from scalp EEG electrodes provides an index of the local field potential integrated over a small area (Buzsaki et al., 2012a). Due to the distorting and attenuating effects of the soft and hard tissues between the current source and the recording electrodes, EEG signals are not directly related to the firing patterns of individuals neurons (Buzsaki et al., 2012a). Rather, the recorded local field potential reflects the combined activity of all ionic processes, from fast action potentials to slow fluctuations in glia (Buzsaki et al., 2012a).

A typical EEG session yields a continuous recording of oscillating neural activity, and the frequency and amplitude of these oscillations can be used to infer information about a subject's behavioural state (Roach & Mathalon, 2008; Lapenta & Boggio, 2014). However, in order to take advantage of the superior temporal resolution of EEG, event-related potentials (ERPs) are often extracted from the continuous recording. ERPs are an average of many EEG waves that are time locked to a specific stimulus or event. Accordingly, ERPs represent transient changes in neural activity in response to the presentation of a stimulus or event (Friedman, Cycowicz, & Gaeta, 2001; Jeste & Nelson III, 2008). ERPs are generally classified by their latency with respect to the stimulus, with early components beginning 50 ms post stimulus onset, and late components occurring as late as 600 - 1000 ms post stimulus onset (Jeste & Nelson III, 2008).

Early ERP components are thought to reflect the basic sensory processing of stimuli, while later components are thought to reflect the perceptual and cognitive processing of stimuli (Picton & Hillyard 1988; Banaschewski & Brandeis 2007; Jeste & Nelson III, 2008).

Speech Induced Suppression

One of the first neurophysiological phenomena found to be associated with auditory feedback processing during speech was ‘speech induced suppression’ (Ford & Mathalon, 2005; Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; Heinks-Maldonado et al., 2006; Beal et al., 2010; Sitek et al., 2013; Wang, Mathalon, Roach, Reilly, & Keedy, 2014; Houde & Chang, 2015). Sensory feedback regarding motor movements, such as auditory feedback resulting from the production of speech, is processed differently than sensory information generated by an external source (Heinks-Maldonado et al., 2005, 2006). According to forward modelling theories, or “re-afference hypotheses,” when a motor command is generated an efference copy of that motor command is sent to sensory areas of the brain, which allows the sensory consequences of the motor command to be predicted (Blakemore, Wolpert, & Frith, 2000; Heinks-Maldonado et al., 2005, 2006; Sitek et al., 2013). When the actual sensory consequences of the motor command are processed, a comparison of the predicted and perceived sensory feedback occurs (Blakemore et al., 2000; Heinks-Maldonado et al., 2005, 2006). If the perceived sensory feedback matches the sensory feedback predicted by the efference copy, a cancellation occurs (Weiskrantz et al., 1971; Blakemore et al., 1998; Blakemore Wolpert, & Frith, 2000; Heinks-Maldonado et al., 2005, 2006). This cancellation results in suppressed neural activation during the perception of self generated sensory feedback, relative to externally generated sensory feedback (Weiskrantz, Elliott, & Darlington, 1971; Blakemore, Rees, & Frith, 1998; Blakemore et al., 2000; Heinks-Maldonado et al., 2005, 2006; Sitek et al., 2013).

The term ‘speech induced suppression’ was coined following a series of experiments that demonstrated that self-produced speech results in a dampened sensory experience relative to the perception of externally generated speech, or the playback of self-generated speech (Heinks-Maldonado et al., 2005, 2006). For example, Heinks-Maldonado and colleagues conducted an EEG experiment (2005) as well as a MEG experiment (2006) where speakers were asked to produce vowel sounds while they listened to real-time playback of their unaltered voice, a pitch-shifted version of their voice, or an alien voice substituted for their own voice. In addition, the speakers also participated in a listening phase, where they heard recordings of their unaltered voice, a pitch-shifted version of their voice, and an alien voice played back to them. Examination of the auditory N1 event-related potential (ERP; 2005) and its magnetic equivalent the M1 (2006) revealed that speakers perception of their own unaltered voice resulted in a dampened sensory experience, or smaller N1/M1 amplitudes, relative to the N1/M1s elicited by the playback of their own unaltered voice, as well as the pitch-shifted and alien voice substituted versions of their voice in both the production and playback conditions. Together these results suggest that the neural processing of auditory feedback resulting from self-produced speech is suppressed (Heinks-Maldonado et al., 2005, 2006). Furthermore, the fact that this suppression was specific to the perception of self-produced unaltered speech, and not speech that was being altered in real-time, suggests that this suppression is not the result of a general dampening of all incoming auditory information during speech production, but rather a highly specific mechanism for processing the auditory consequences of self-produced speech (Heinks-Maldonado et al., 2005, 2006; Chen, Chen, Liu, Huang, & Liu, 2012a).

The results of these speech induced suppression studies support the forward or ‘re-afference’ hypotheses that suggest that when a motor command is executed, a copy of this motor command is sent to sensory cortical areas in preparation for the upcoming sensory feedback (Blakemore et al., 1998; Blakemore et al., 2000; Heinks-Maldonado et al., 2005, 2006). The results of these speech induced suppression studies are also in accordance with the DIVA and SFC speech motor control models, which both suggest that when speech motor commands are executed in a feedforward manner, a copy of the motor command is sent to sensory cortical areas for comparison with the resultant sensory consequence of the motor movements. This speech induced suppression is thought to play an important role in speakers' abilities to detect speech production errors, as the comparison between predicted and perceived auditory feedback allows deviations from the intended speech motor act to be readily detected. This rapid detection in turn facilitates rapid error correction by the feedback control system, thus promoting fluent speech production. In addition, since the sensory feedback resulting from self produced actions is predictable, cortical activation only increases following unpredictable sensory stimulation. This means that instead of fully processing our own speech while talking, processing resources are freed up in order to process potentially important external events (Wang et al., 2014). As a result, externally generated sounds, which are unpredictable, are easily distinguished from self produced sounds. Since externally generated sounds may be indicative of potential threats to the speaker, this rapid discrimination of externally generated sounds may promote additional evaluation of the stimuli in order to determine if a response is necessary (Blakemore et al., 2000; Heinks-Maldonado et al., 2005, 2006).

Event-Related Potentials (ERPs)

In addition to the use of ERPs in investigating speech induced suppression, ERPs are often utilized to investigate the neural correlates of auditory feedback based error correction during ongoing speech. For example, when speakers are exposed to FAF, the deviant auditory feedback reliably elicits the P1-N1-P2 ERP complex (Behroozmand, Karvelis, Liu, & Larson, 2009; Behroozmand, Korzyukov, & Larson, 2011a; Behroozmand, Korzyukov, & Larson, 2012;

Chen et al., 2012; Korzyukov, Karvelis, Behroozmand, & Larson, 2012a; Liu et al., 2013; Scheerer et al., 2013a; Tumber et al., 2014).

i. The P100 (P1)

The P1 is a positive deflection that occurs approximately 50 to 100 ms after the onset of FAF (Scheerer et al., 2013a; Tumber et al., 2014). Intracerebral recordings from auditory cortical areas in humans indicate that the auditory P1 response represents activity in the lateral portion of Heschl's gyrus, or the secondary auditory cortex (Ponton, Eggermont, Kwong, & Don, 2000; Moore, 2002). Based on its origin in secondary cortex, it has been suggested that the auditory P1 represents a relatively low level of auditory processing (Moore, 2002). Accordingly, previous FAF studies suggest that changes in the amplitude of the P1 component reflect the basic detection of deviant auditory feedback, as its amplitude increases in an all-or-nothing manner in responses to deviant auditory feedback (Scheerer et al., 2013a). In other words, these results suggest that although the auditory P1 is sensitive to FAF, its amplitude does not reflect specific properties of the eliciting stimulus.

ii. The N100 (N1)

The N1 is a negative deflection that occurs approximately 100 to 200 ms after the onset of FAF (Korzyukov et al., 2012a; Scheerer et al., 2013a; Tumber et al., 2014), and has been suggested to be the most robust and reproducible auditory component across subjects (Beal et al., 2010). The N1 is maximally distributed over fronto-central scalp regions (Picton, 1974) and is thought to be generated in the primary and secondary auditory cortices (Sitek et al., 2013). The auditory N1 can reflect specific stimulus attributes, such as sound intensity (Sitek et al., 2013) or stimulus frequency (Butler, 1968), and may also provide an indication of how the auditory stimulus is being perceptually evaluated (Picton, 1974), as well as the amount of resources dedicated to processing the stimulus (Sitek et al., 2013). As previously mentioned, the N1 has been found to be maximally attenuated during the perception of one's own unaltered speech, relative to FAF (Heinks-Maldonado et al., 2005, 2006; Sitek et al., 2013), alien speech (Heinks-Maldonado et al., 2005, 2006), and the playback of self-produced speech (Houde et al., 2002; Beal et al., 2011). This neural modulation has been suggested to reflect suppression in the auditory cortex during the perception of one's own unaltered speech, as a result of a match between perceived and predicted sensory feedback. In comparison, increases in activation when sensory feedback is incongruent with the predicted feedback is thought to reflect a violation of the prediction created by the feedforward system (Heinks-Maldonado et al., 2006; Beal et al., 2011).

iii. The P200 (P2)

The P2 is a positive deflection that occurs approximately 200 to 300 ms after the onset of FAF (Korzyukov et al., 2012a; Scheerer et al., 2013a; Tumber et al., 2014). Unlike the P1 and N1 components, the P2 is not generated in temporal cortex, rather the P2 component is thought to reflect auditory driven output of the mesencephalic reticular activating system (Ponton et al., 2000). Similar to the N1, the P2 is maximally distributed over fronto-central scalp regions, and is thought to reflect stimulus specific attributes, in addition to providing an indication of how the stimulus is being perceptually evaluated (Picton, 1974). Accordingly, FAF studies have shown that the amplitude of the P2 ERP component increases linearly with increasing feedback perturbation magnitudes (Behroozmand et al., 2009; Scheerer et al., 2013a). As the size of the P2 response varies with the size of the feedback manipulation, it has been suggested that the P2 component indexes the size of the mismatch between the perceived and predicted auditory feedback (Scheerer et al., 2013a).

Together, these results suggest that the P1, N1, and P2 neural responses are valuable for assessing feedback and feedforward input to the speech motor control system.

Current Studies

The neurocomputational models of speech motor control, such as the DIVA (Guenther, 1994, 1995, 2006; Guenther, et al., 1998, 2006; Tourville & Guenther, 2011; Guenther & Vladusich, 2012) and SFC models (Hickok et al, 2011; Hickok, 2012; Houde & Chang, 2015), as well as experimental evidence (e.g. Jones & Munhall, 2005; Jones & Keough, 2008; Hawco & Jones, 2009; Keough & Jones, 2009, 2011; Scheerer et al., 2013a) have provided adequate support for the notion that both feedback and feedforward control systems support speech motor control. Early in speech acquisition, feedback control plays an important role in mapping speech sounds onto the appropriate speech motor commands (Guenther, 2006; Callan et al., 2000). However, once these mappings are well learned feedback control becomes less crucial, as the information provided by sensory feedback becomes redundant. In addition, since there are delays involved in processing peripheral sensory feedback, strict feedback control has the potential to induce speech disfluencies. For this reason, it has been suggested that speech motor control transitions to a more feedforward dominant system (Guenther, 2006; Beal et al., 2010). Since feedforward commands are executed from stored sensorimotor representations, the delays associated with processing peripheral sensory feedback are avoided. However, it is important to note that although speech motor control transitions to a more feedforward dependent system, the feedback control system still plays an important role in error correction as well as the maintenance of the sensorimotor representations supporting the feedforward control system (Guenther, 2006; Beal et al., 2010). While it is clear that the relative weighting of feedback and feedforward control changes throughout development, it is currently unclear which factors dictate this relative weighting (Kim & Max, 2014). With this in mind, the purpose of this thesis was to explore the factors that dictate the relative weighting of feedback and feedforward input for speech motor control using both the FAF perturbation and FAF adaptation paradigms, as well as ERP responses. More specifically, experiments 1 through 3 examined how the role of auditory feedback changes throughout development, while experiments 4 and 5 examined the relationship between vocal variability and feedback control, and experiment 6 looked at how the predictability of auditory feedback errors influences the weighting of feedback control and feedforward control.

Study 1: The Role of Auditory Feedback for Speech Motor Control in Toddlers

Scheerer, N.E, Jacobson, D.S., & Jones, J.A. (In Prep). The Role of Auditory Feedback for Speech Motor Control in Toddlers.

MacDonald, Johnson, Forsythe, Plante, and Munhall (2012) reported that toddlers aged 2-3 years old do not use auditory feedback to compensate for formant frequency manipulations imposed during ongoing speech. Despite these findings, observation of individuals suffering from congenital deafness would argue that auditory feedback is necessary for speech development (Svirsky et al., 2004). For this reason, we exposed toddlers aged 2-3 years old to brief unpredictable perturbations of the F0 of their auditory feedback in order to assess whether these changes in their auditory feedback would elicit compensatory responses. If the toddlers produced compensatory responses following changes in their auditory feedback, these results would provide evidence to support the notion that auditory feedback is important for the regulation of voice F0 early in development.

Abstract

Children maintain relatively fluent speech despite the dramatic changes that occur to the shape, size, and musculature of their articulators as they develop (Guenther, 1994). Behavioural, clinical, and neurophysiological evidence demonstrates that auditory feedback aids in the acquisition and lifelong maintenance of the sensorimotor mechanisms that underlie vocal motor control in both humans and songbirds (Guenther, 1994, Perkell et al., 1997, Doupe & Kuhl, 1999, Callan et al., 2000, Sakata & Brainard, 2008, Civier et al., 2010; Kelly & Sober, 2014). Surprisingly, a recent study by MacDonald and colleagues reported that in toddlers, the speech motor control system may “suppress” the influence of auditory feedback since exposure to altered auditory feedback regarding their formant frequencies did not lead to reciprocal modifications to their speech, counter to what was observed for older children and adults. Specifically, the authors argued that a social context may be necessary to promote auditory feedback based error correction in toddlers. Although plausible, as social reinforcement certainly plays a role during speech acquisition (Goldstein, King, & West, 2003; Kuhl, 2004), a lack of sensory feedback based error correction before 4 years of age is not parsimonious with current theories of vocal motor control (Guenther, 1994; Callan et al., 2000). Here, we exposed toddlers to brief perturbations to their auditory feedback regarding the pitch of their ongoing vocalizations. Toddlers compensated for the auditory feedback manipulations by rapidly shifting their vocal pitch in the opposite direction of the manipulation, producing significantly different responses to upward and downward perturbations. These data represent the first empirical demonstration that toddlers integrate auditory feedback into their ongoing vocal motor control. Furthermore, our findings suggest that toddlers may be more sensitive to changes to the postural properties of their auditory feedback, such as fundamental frequency, relative to the phonemic properties, such as formant frequencies.

Introduction

Throughout development children maintain relatively fluent speech despite dramatic changes to the shape, size, and musculature of the articulators involved in speech production (Guenther, 1994; Callan et al., 2000). The vocal tract of a child is not simply a smaller version of an adult vocal tract, as there are many differences in the structures supporting speech production in children relative to adults (Kent & Vorperian, 1995; Kent, 1999; Callan et al., 2000). In both humans and songbirds, auditory feedback has been suggested to not only aid in the acquisition of vocal motor control, but also act as an adaptive signal to guide movements of the articulators during development (Guenther, 1994; Perkell et al., 1997; Doupe & Kuhl, 1999; Callan et al., 2000; Sakata & Brainard, 2008; Civier et al., 2010; Kelly & Sober, 2014).

The importance of auditory feedback for monitoring and correcting ongoing vocalizations has been demonstrated by utilizing the frequency altered feedback (FAF) paradigm to synthetically alter speakers' auditory feedback (Burnett et al., 1997; Burnett et al., 1998; Scheerer & Jones, 2012; Scheerer, et al., 2013a; Scheerer & Jones, 2014). When a speaker's auditory feedback is manipulated by changing properties such as the fundamental frequency (F0; Burnett et al., 1997; Burnett et al., 1998; Scheerer & Jones, 2012; Scheerer et al., 2013a; Scheerer & Jones, 2014), or the formant frequencies (Houde & Jordan, 1998, 2002; Purcell & Munhall, 2006a; Villacorta et al., 2007; Cai et al., 2012), the speaker reflexively responds by opposing the manipulation. These compensatory responses demonstrate that when a speaker detects changes in their auditory feedback, they use information from the deviant auditory feedback to modify their ongoing vocalization.

Scheerer and colleagues (2013b) investigated the developmental trajectory of the speech motor control system by exposing children and adults to changes in their F0 using the FAF paradigm. The results of that study indicated that children as young as 4 years old produce compensatory responses to changes in their voice F0, consistent with the notion that auditory feedback is important for monitoring and correcting for speech production errors. Surprisingly, MacDonald and colleagues (2012) reported that children and adults, but not toddlers, compensate for manipulations of the formant frequencies of their auditory feedback. The results of that study led to the suggestion that the auditory feedback component of the speech motor control system may be ‘suppressed’ in toddlers. However, given the dramatic changes occurring to the acoustic properties of the vocal tract during development (Callan et al., 2000), it is hard to imagine how toddlers can preserve speech fluency without the aid of auditory feedback.

Auditory feedback is important for maintaining both the phonemic and postural settings of speech, which ensure that phonemic distinctions are preserved and that speech remains intelligible in dynamic acoustic environments, respectively (Perkell et al., 1997). F0, a property of auditory feedback related to postural control, has been shown to be more sensitive to the loss of auditory feedback, relative to formant frequency, a property of phonemic control (Perkell et al., 1997). Since the phonemic and postural settings of speech appear to be regulated independently, it is also possible that these speech properties develop at different rates.

Early in development, caregivers communicate with infants using infant directed speech (IDS), which is characterized by exaggerated prosodic features including longer pauses, a slower tempo, more prosodic repetitions, a higher mean F0, and a wider F0 range (Fernald, 1993; Cooper & Aslin, 1994; Kuhl, 1994; Saint-Georges et al., 2013). These prosodic features vary according to the caregivers intentions, and have been shown to provide an effective way to communicate affect with prelinguistic children (Fernald, 1993; Saint-Georges et al., 2013). For example, Fernald (1993) demonstrated that even when hearing IDS in another language, infants smile more to positive expressions of IDS, and express negative affect in response to expressions of disapproval. Since recognizing changes in F0 provides young children with a means for comprehending the affect and intentions of caregivers, prior to becoming linguistically proficient (Saint-Georges et al., 2013), IDS may increase the saliency of the prosodic features of speech such as F0. On the other hand, since non-prosodic features of the speech signal, such as formant frequencies, are less beneficial for speech comprehension in prelinguistic children (Saint-Georges et al., 2013), these features may be less salient. Thus it is possible that exposure to IDS may facilitate infants' ability to process prosodic aspects of speech, which may later manifest itself in an increased proficiency at monitoring and correcting for prosodic changes in their own speech.

In the current study, we utilized the FAF paradigm to manipulate the F0 of toddlers auditory feedback in order to investigate whether toddlers are capable of extracting information from their auditory feedback in order to modify their speech motor commands during ongoing speech, or whether the auditory feedback component of the speech motor control system is in fact 'suppressed'. Although previous research suggests that toddlers do not produce compensatory responses to manipulations of the formant frequencies of their auditory feedback (MacDonald et al., 2012), formant frequency and F0 control have been shown to be regulated independently (Perkell et al., 1997). In addition, research on IDS indicates that prosodic aspects of speech, such as F0, play a crucial role in communication with prelinguistic children. For this reason, we expected that toddlers would produce compensatory responses to the F0 manipulations, demonstrating the role of auditory feedback for the ongoing monitoring and maintenance of the toddlers' speech motor control. Furthermore, coupled with the findings of MacDonald and colleagues (2012), these results would suggest toddlers are more proficient at regulating the postural aspects of their speech, such as F0, relative to the phonemic aspects, such as formant frequencies.

Methods

i. Participants

Twenty-five toddlers between the ages of 2 and 3 years (M = 36.15 months, SD = 6.05 months; 11 females) participated in this study. However, five of the toddlers did not complete the study, as they refused to produce the vocalizations. The remaining 20 toddlers were divided into two groups: younger toddlers who were between 24 and 35 months (n = 11, M = 31.55 months, SD = 3.64 months) and older toddlers who were between 40 and 46 months (n = 9, M = 41.78 months, SD = 2.33 months). All participants were Canadian-English speakers who did not speak a tonal language, and had no formal vocal training. Informed consent was obtained from all participants, as well as from a parent or guardian of each participant. All procedures were approved by the Wilfrid Laurier Research Ethics Board and were in accordance with the World Medical Association 2013 Declaration of Helsinki.

ii. Procedure

Participants were seated in front of a computer monitor that displayed an interactive farm game. As part of the game, a pre-recorded voice produced a ‘baa’ sound and the toddler was asked to mimic the ‘baa’ sound in order to progress through the game. This process was repeated until approximately 100 vocalizations were completed.

Participants were asked to vocalize at a loud, but comfortable volume. Vocalizations were played back to the participants in real time via headphones. Since the participants were quite young, sound level was not precisely monitored. However, the experimenter provided online feedback throughout the experiment if the participant was vocalizing too loudly or too softly to help regulate the participants’ volume. This volume monitoring technique was deemed adequate as previous research has shown that the relative loudness of auditory feedback has no influence on the amplitude or latency of responses to FAF (Burnett et al., 1998). Participants were also told that during the experiment their voice might sound different, but they should ignore these differences and continue to produce their vocalizations at a consistent pitch. The experiment consisted of four blocks of 25 trials, and lasted approximately 20 minutes. During each vocalization the participant's voice was perturbed one semitone (100 cents) upwards or downwards three times. Each perturbation lasted for 200 ms, with an interstimulus interval of 900-1100 ms (see Figure 1). The perturbation direction was held constant in each block, but alternated across blocks. The block orders were also counterbalanced so that half of the participants were initially exposed to the upward shift, while the other half were initially exposed to the downward shift.

iii. Apparatus

Participants were seated in front of a computer monitor and fitted with a headset-microphone (Sennheiser HMD 280-13 Pro, Sennheiser Electronics, Germany). Presentation of the visual stimuli were controlled by powerpoint, while the onset and offset of the auditory feedback perturbations were controlled by Max/MSP (v.5.0, Cycling '74). During the experiment vocalizations were sent to a Mac Mini (Apple, Cupertino, CA) where they were mixed using Studio One Software (Presonus, Baton Rouge, LA) and then sent to a digital signal processor (VoiceOne, T.C. Hellicon, Victoria, BC), which allowed the pitch of the participant's voice to be altered. This process introduced approximately a 10 millisecond delay in the feedback signal, which was then presented back to the participant as auditory feedback. Both the altered and the unaltered voice signals were digitally recorded (Presonus FireStudio Project, Baton Rouge, LA) at a sampling rate of 44.01 kHz for later analysis.

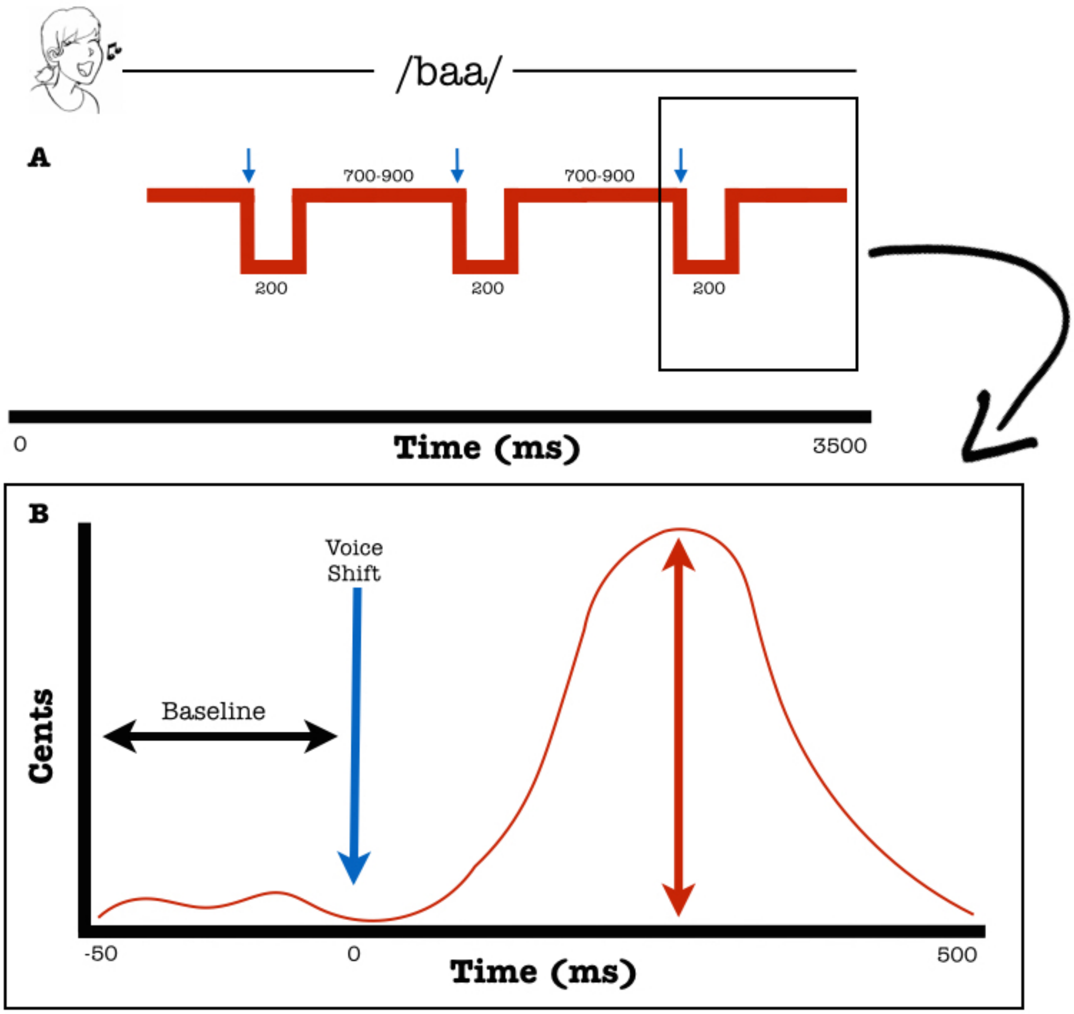

Figure 1: Experimental Paradigm

A - Participants vocalized a /baa/ sound while their F0 was perturbed either upwards or downwards one semitone (+/- 100 cents) three times per vocalization for 200 ms, with an interstimulus interval of 900-1100 ms. B - Each vocalization was segmented relative to the voice perturbation (blue arrow), with the 50 ms prior to the voice perturbation serving as a baseline to normalize the vocalization. The response magnitude was calculated by determining the maximum deviation from the baseline in the 500 ms following the voice perturbation (red arrow).

iv. Data Analysis

The digital recording of the vocalizations was segmented into separate utterances and F0 values were calculated for each utterance. Utterances were then segmented on the basis of the pitch-shift onset (see Figure 1). F0 values for each of the three segments were normalized to the baseline (50 ms of speech prior to the onset of the pitch shift) by converting Hertz values to cents using the following formula:

$$\boxed{\text{Cents} = 100\ (12\ \text{log2}\ F/B)}$$

In this formula, F is the F0 value in Hertz and B is the mean frequency of the baseline period. Cents values were calculated for 100 ms before and 500 ms after the pitch shift. An averaged F0 trace was constructed for each shift magnitude (-100 cents, +100 cents), for each participant. The magnitude of each vocal response, as well as the participants' vocal variability (SD of the baseline period), were evaluated. The magnitude of each participant's compensatory response was determined by finding the point at which the averaged F0 trace deviated maximally from the baseline mean, while the SD of the baseline period was calculated on the basis averaged F0 of the 50 ms of unaltered voice prior to the pitch shift.

Results

Averaged F0 traces were created for each participant and averaged separately for the younger toddlers and older toddlers (see Figure 2). Both the younger and older toddlers compensated for the auditory feedback perturbations by changing their vocal pitch in the opposite direction of the manipulation. To confirm these observations, response magnitudes were subjected to a repeated measures analysis of variance (RM-ANOVA) with perturbation direction (upward, downward) and age (younger toddler, older toddler) as factors. A main effect of perturbation direction, $F_{1,18} = 43.522$, $P < .001$, confirmed that participants compensated for the perturbations by increasing their F0 in response to downward perturbations ($M = 12.23$ cents, $SD = 14.76$), and decreasing their F0 in response to upward perturbations ($M = -25.07$ cents, $SD = 19.19$). However, the main effect of age, $F1,18 = .387$, $P = .541$, and the interaction between perturbation direction and age, $F_{1,18} = .017$, $P = .897$, were not significant.

Since a relationship between the age and vocal variability has previously been identified (Scheerer & Jones, 2012), a RM-ANOVA was also conducted to investigate the effect of perturbation direction (upward, downward) and age (younger toddler, older toddler) on vocal variability. This analysis revealed a main effect of age, $F_{1,18} = 4.60$, $P = .046$, as the younger toddlers ($M = 2.14$ cents, $SD = 1.19$) were more variable than the older toddlers ($M = 1.32$ cents, $SD = 0.99$). However, the main effect of perturbation direction, $F_{1,18} = 1.716$, $P = .207$, and the interaction between perturbation direction and age, $F_{1,18} = .277$, $P = .605$, were not significant.

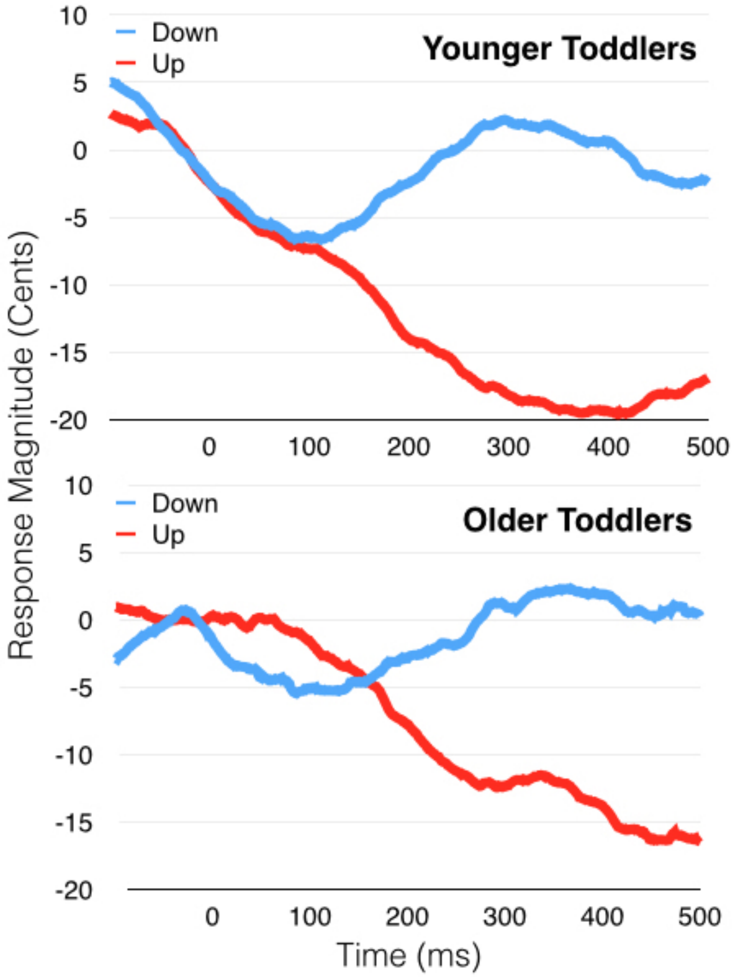

Figure 2: Averaged F0 Trace

Averaged F0 trace for both the upward perturbation (red) and downward perturbation (blue) conditions plotted separately for younger toddlers (top) and older toddlers (bottom).

Discussion

The aim of the current study was to investigate whether toddlers are capable of extracting information from their auditory feedback in order to regulate their speech motor control system during ongoing speech. The results indicate that toddlers produced compensatory responses to brief perturbations of the F0 of their auditory feedback. As the responses produced following upward perturbations were significantly different than those produced following downward perturbations, these findings indicate that toddlers can in fact extract information from their auditory feedback to modify their ongoing vocalizations.

MacDonald and colleagues (MacDonald et al., 2012) reported that children and adults compensate for alterations of the frequency of the first and second formants of their auditory feedback, but toddlers do not. Based on these findings, the authors concluded that the auditory feedback component of speech motor control system may be suppressed in toddlers, or may not develop until closer to the age of 4. The results of the current study argue against these conclusions, as they clearly demonstrate that toddlers under the age of 4 are actively using auditory feedback to monitor and correct for errors in their ongoing speech. However, it is important to note that MacDonald and colleagues (MacDonald et al., 2012) manipulated the formant frequencies of the toddlers' auditory feedback, while the current study manipulated the F0 of the toddlers' auditory feedback. Auditory feedback plays an important role in maintaining the phonemic settings of speech, in order to ensure phonemic distinctions are preserved (Perkell et al., 1997). However, auditory feedback is also important for maintaining the postural settings of speech, which ensure that speech remains intelligible in dynamic acoustic environments (Perkell et al., 1997). Formant frequency, a property of auditory feedback related to phonemic control, has been shown to be less sensitive to the loss of auditory feedback, relative to F0, a property of postural control (Perkell et al., 1997). This difference may explain why toddlers were able to modify their F0 using information from the deviant auditory feedback in the current study, but were unaffected by the formant frequency manipulations imposed by MacDonald and colleagues (MacDonald et al., 2012). Importantly, these differences suggest that toddlers may be more proficient at regulating the postural properties of their speech, relative to the phonemic settings, using auditory feedback.

Toddlers increased proficiency at regulating the postural properties of speech, specifically F0, may also be a consequence of exposure to infant directed speech (IDS). Early in development, caregivers communicate with infants using IDS, and this IDS has been shown to provide an effective way to communicate affect with prelinguistic children (Fernald, 1993, Saint-Georges et al., 2013). Since recognizing changes in F0 provides young children with a means for comprehending the affect and intentions of caregivers, prior to becoming linguistically proficient (Saint-Georges et al., 2013), IDS may increase the saliency of the prosodic features of speech such as F0. On the other hand, since non-prosodic features of the speech signal, such as formant frequencies, are less beneficial for speech comprehension in prelinguistic children (Saint-Georges et al., 2013), these features may be less salient. For this reason, we suggest that exposure to IDS may facilitate infants' ability to process prosodic aspects of speech, which later manifests itself as an increased proficiency at monitoring and correcting for prosodic changes in their own speech.

Although the younger and older toddlers in this study produced similar sized compensatory responses to the FAF, the younger toddlers were found to be more variable than the older toddlers. This finding is in line with previous studies that have demonstrated that vocal variability decreases with age (Scheerer et al., 2013b). Scheerer and colleagues (2012, 2013b, 2014) have suggested that vocal variability is related to the proficiency of speech motor control. As vocal stability improves through vocal training or development, speakers become less dependent on auditory feedback for closed-loop speech motor control (Scheerer & Jones, 2012; Scheerer et al., 2013b, 2014). Instead, speakers transition to a predominantly open-loop speech motor control system, where speech motor commands are generated from stored representations (Perkell et al., 1997; Scheerer & Jones, 2012; Chen et al., 2013, Max et al., 2004; Guenther, 2006; Tourville et al., 2008; Civier et al., 2010). The decreased variability with age observed in this study, may reflect a gradual maturation of the speech motor control system throughout development.

In summary, the current results demonstrate that toddlers as young as 2 years old produce compensatory responses to changes in their auditory feedback. These results provide empirical support for the long held assumption that auditory feedback functions as an adaptive signal to guide movements of the articulators in order to preserve speech fluency while the articulators are restructured during development. In addition, these results suggest that toddlers may be more sensitive to changes to the postural properties of their auditory feedback, such as fundamental frequency, relative to the phonemic properties, such a formant frequencies.
Risk factors for speech disorders often appear early in childhood (Schneider, Zumtobel, Prettenhofer, Aichstill, & Jocher, 2010). Understanding the normal development of auditory feedback control of speech can aid in identifying irregularities in the vocal system, which may serve as precursors to future pathologies. The future application of this research to the diagnosis of disorders with known disturbances in vocal control, such as Autism Spectrum Disorders, which are lacking viable diagnostic markers, may prove invaluable.

Study 2: The Developmental Trajectory of Vocal and ERP Responses to Frequency Altered Feedback

Published As:

Scheerer, N., Liu, H., & Jones, J.A. (2013). The Developmental Trajectory of Vocal and ERP Responses to Frequency Altered Auditory Feedback. European Journal of Neuroscience, 38(8), 3189-3200.

Neurocomputational models of speech suggest that speech motor control develops gradually as the acoustics of speech are mapped onto the positions and movements of the articulators (Guenther, 2006). In this ERP study, children and adults produced vocalizations while exposed to FAF. Behaviourally, vocal response magnitudes, vocal response latencies, and vocal variability were measured to assess whether the role of auditory feedback for ongoing speech differs across different age groups and genders. In addition, P1-N1-P2 ERP responses were recorded in order to assess whether the neural systems that integrate auditory feedback during vocal motor control undergo changes with age and gender. The results of this study will help to map out the developmental trajectory of auditory feedback control of speech.

Abstract

Speech motor control develops gradually as the acoustics of speech are mapped onto the positions and movements of the articulators. In this event-related potential (ERP) study, children and adults aged 4–30 years produced vocalizations while exposed to frequency-altered feedback. Vocal pitch variability and the latency of vocal responses were found to differ as a function of age. ERP responses indexed by the P1–N1–P2 complex were also modulated as a function of age. P1 amplitudes decreased with age, whereas N1 and P2 amplitudes increased with age. In addition, a correlation between vocal variability and N1 amplitudes was found, suggesting a complex interaction between behavioural and neurological responses to frequency-altered feedback. These results suggest that the neural systems that integrate auditory feedback during vocal motor control undergo robust changes with age and physiological development.

Introduction

The ability to produce speech develops gradually as speech sounds are mapped onto the positions and movements of the articulators. This sensorimotor mapping begins as infants explore auditory and somatosensory space by babbling (de Boisson-Bardies, 1999; Civier et al.,

2010) and it remains plastic throughout development to accommodate growth of the articulators and vocal folds, increases in musculature, and changes in lung capacity (McAllister et al., 1993; Callan et al., 2000; Guenther, 2006). Although proprioception plays a role in speech motor control (Larson, Altman, Liu, & Hain, 2008), it is clear that the development and maintenance of fluent speech is largely dependent on auditory feedback (Lombard, 1911; Siegel, Pick, Olsen, & Sawin, 1976; Burnett & Larson, 2002; Jones & Munhall, 2002; Chen et al., 2012b; MacDonald et al., 2012). The importance of auditory feedback for maintaining accurate vocal output has been demonstrated by studies that have used the frequency-altered feedback (FAF) paradigm. The FAF paradigm involves shifting the fundamental frequency (F0) of an individual's vocalizations and instantaneously presenting the altered auditory feedback (AAF) back to the individual through headphones (Elman, 1981; Burnett et al., 1997). This paradigm consistently elicits a response termed the pitch-shift reflex (Burnett et al., 1998; Jones & Munhall, 2002; Natke et al., 2003). The pitch-shift reflex is the tendency of speakers to respond to the FAF by shifting their vocal pitch in the opposite direction of the alteration (Burnett et al., 1998). This response has been suggested to play a role in stabilizing voice F0 around a desired target (Hain et al., 2000; Natke et al., 2003; Hawco & Jones, 2009).

Although the role of auditory feedback during ongoing speech production has been extensively studied in the adult population, attempts to investigate the developmental trajectory of sensorimotor control of speech in children have been limited by narrow age ranges and contradictory results. For example, Liu and colleagues (2010b) investigated vocal responses to FAF in a group of English-speaking children aged between 7 and 12 years. The authors reported that the children produced longer-latency vocal responses to the FAF relative to the young adults (Liu et al., 2010b). A follow-up study testing Mandarin-speaking children aged between 7 and 12 years, showed that Chinese children produced larger-magnitude and longer-latency voice F0 responses than young adults (Liu et al., 2010a). Although these studies provide evidence of developmental changes in the feedback control of vocal production, they fail to portray a clear picture of the developmental trajectory of speech motor control, owing to the limited age ranges. Furthermore, the neural activity related to these audio-vocal responses throughout development is currently unclear.

Auditory cortical processing is often indexed with the use of electroencephalography (EEG) and magnetoencephalography. It has been suggested that the M50, the magnetic equivalent of the P1 event-related potential (ERP) component, is the most prominent and reliable indicator of auditory processing in children, whereas the M100 component, the magnetic equivalent of the N1 component, is the most robust component in adults (Bruneau & Gomot, 1998; Oram Cardy, Ferrari, Flagg, Roberts, & Roberts, 2004; Kotecha et al., 2009; Beal et al., 2011). Ponton & Eggermont (2007) suggested that the P1 component develops early in childhood, reaches its maximum in toddlers, and declines shortly thereafter. On the other hand, the N1 component is generally found to be smaller in children, and increases in amplitude into adulthood (Oram Cardy et al., 2004). The magnitude of the P2 component is generally unaffected by age.

Previous research has shown that, when adults are exposed to FAF, the P1, N1 and P2 components increase in amplitude, reflecting a mismatch between expected and perceived auditory feedback (Behroozmand et al., 2009; Scheerer et al., 2013a). The P1 component has been found to increase in an all-or-nothing manner in response to FAF, suggesting that it is sensitive to deviant auditory feedback in a nonspecific manner (Scheerer et al., 2013a). The N1 ERP component is more sensitive to the size of the deviant feedback, as it shows small increases in response to small feedback perturbations, and larger increases in response to larger feedback perturbations. It has been suggested that deviant, but still physiologically feasible feedback perturbations elicit small increases in N1 amplitudes, whereas larger feedback perturbations that are more obviously externally generated elicit relatively larger N1 responses. For this reason, it has been proposed that the N1 component plays a role in classifying auditory feedback as internally or externally produced. Similarly, P2 amplitudes have been found to increase in a stepwise manner in response to increasingly large feedback perturbations. For this reason, it has been suggested that the P2 component indexes the size of the mismatch between perceived and expected auditory feedback (Scheerer et al., 2013a). Although these neural changes in response to FAF have been well studied in the adult population, few studies have investigated neural responses to FAF in children.

Although children's auditory responses to FAF have rarely been reported, more generalized auditory cortical responses to stimuli such as pure tones have been well documented. As positive peaks tend to be more robust and consistent in younger subjects, the P1 component is the most frequently reported auditory evoked potential (AEP; Ceponiene, Rinne, & Nantanen, 2002). The results of developmental AEP studies are unanimous in showing that P1 amplitudes decrease with age (Ponton et al., 2000; Ponton, Eggermont, Khosla, Kwong, & Don, 2002; Ponton & Eggermont, 2001, 2007; Ceponiene et al., 2002; Oram Cardy et al., 2004; Sussman, Steinschneider, Gumenyuk, Grushko, & Lawson, 2008), whereas P1 latencies become shorter with age (Ponton et al., 2000, 2002; Ceponiene et al., 2002; Gage et al., 2003; Oram Cardy et al., 2004; Beal et al., 2011). Reports regarding the N1 component have been far less consistent. It has been suggested that the superimposition of the strong P1 component leads to difficulty in detecting a robust N1 component (Ponton et al., 2000, 2002). However, in studies where the N1 component has been detected, some have reported that the N1 component increases in amplitude with age (Goodin, Squires, Henderson, & Starr, 1978; Allison, Hume, Wood, & Goff, 1984; Polich, Ladish, & Burns, 1990; Bruneau, Roux, Guerin, Barthelemy, & Lelord, 1997; Cunningham, Nicol, Zecker, & Kraus, 2000; Oram Cardy et al., 2004), whereas others have found that it decreases with age (Ceponiene et al., 2002). Bruneau et al. (1997) reported that the N1 component decreases in amplitude with age until the age of 8 years, at which point it begins to resemble the characteristic adult-like form. Although there is some debate about the developmental pattern of N1 amplitudes, latency trends have been more consistent. Similarly to P1 latency, N1 latency has been found to decrease with age (Goodin et al., 1978; Allison et al., 1984; Polich et al., 1990; Paetau, Ahonen, Salonen, & Sams, 1995; Tonnquist-Uhlen, Borg, & Spens, 1995; Bruneau et al., 1997; Ponton et al., 2000; Gage et al., 2003; Oram Cardy et al., 2004). Although less frequently reported, the P2 component has also been found to vary throughout development. Some studies have reported increases in P2 amplitude with age (Kraus et al., 1993; Johnstone, Barry, Anderson, & Coyle, 1996; Oades, Dittman-Balear, & Zerbin, 1997), others have reported decreases in amplitude with age (Ponton et al., 2000), and still others have reported no systematic change in P2 amplitude as a function of age (Tonnquist-Uhlen, 1996; Ceponiene et al., 2002). Reports of P2 latency, however, have reliably reported a lack of change as a function of age (Johnstone et al., 1996; Tonnquist-Uhlen, 1996; Ponton et al., 2000).

Whereas the AEP literature is useful for determining the developmental course of auditory cortical processing, a specific investigation of auditory cortical responses to AAF would help to identify the neural changes accompanying speech development. To date, there has only been one study investigating the neural responses of children to FAF. Liu et al. (2013) investigated the modulation of the P1–N1–P2 complex in response to FAF in 10–15-year-old Mandarin-speaking children, relative to young adults (18–30 years). The results indicated that P1 amplitudes decreased with age, whereas N1 and P2 amplitudes were modulated by an interaction between sex and age. Response latencies for the P1 and N1 components decreased with age, whereas P2 latencies were not modulated by age (Liu et al., 2013). Although these results suggest a developmental sex-specific change in neural responses to FAF, the small range of ages tested fails to adequately depict the developmental trajectory of neural responses to FAF. In addition, the fact that a Mandarin-speaking population was used in this study limits the generalizability of these results. It has been shown that both vocal and cortical responses to FAF are sensitive to language experience (Chen et al., 2012b), which suggests that an English-speaking population may show different patterns of neural activation in response to FAF. Given that the precursors of many speech-related disorders are already evident during childhood, establishing an understanding of the normal developmental trajectory of responses to AAF is necessary (Schneider et al., 2010). The aim of the current study was to do exactly that – investigate the developmental trajectory of the audio-vocal system, both behaviourally and neurologically. During this study, Canadian English-speaking individuals ranging in age from 4 to 30 years participated in a FAF paradigm in order to investigate the use of auditory feedback during vocalization. In addition, P1–N1–P2 ERP responses were recorded in order to enable comparison of the neural activity in response to FAF throughout development. As developmental changes occur at different rates for males and females, we also investigated sex-related differences in the development of speech motor control.

Successful motor learning relies on the ability to form internal models that predict the sensory consequences of motor commands (Miall & Wolpert, 1996; Shadmehr, Smith, & Krakauer, 2010). Accordingly, both empirical research and computational models of speech motor control (e.g. directions into velocities of articulators model; Guenther, 2006) suggest that fluent speech motor control results from the combined effort of feedback and feedforward control systems. Feedforward control relies on sensorimotor representations, or internal models, which contain an estimate of the current dynamic state of the vocal tract, based on incoming motor commands as well as previously learned mappings between articulatory movements and the resultant sensory consequences of these movements. This feedforward control system regulates speech motor control by providing internal feedback to speech motor control planning systems, while avoiding the delays associated with peripheral feedback (Miall & Wolpert, 1996; Guenther, 2006). On the other hand, feedback control relies on auditory feedback for detection and correction of overt production errors. Although feedback control is crucial for learning the mapping between articulator movements and the sensory consequences of the movements, it has been suggested that as speech development progresses a transition from feedback control to feedforward control takes place (Bailly, 1997; Guenther & Bohland, 2002; Schmidt & Lee, 2005; Guenther, 2006; Civier et al., 2010; Beal et al., 2011).

On the basis of the previous empirical and modelling research, we expected that the magnitude of compensation to the FAF would decrease with age, reflecting increasingly proficient speech motor control, and a transition from feedback to feedforward control. We also predicted that behavioural response latencies would decrease with age, reflecting increasingly efficient speech motor control. Neurologically, we hypothesized that P1 amplitudes would decrease with age, as shown in previous research (Liu et al., 2013). On the other hand, we expected N1 and P2 amplitudes to increase with age. We believed that larger N1 and P2 amplitudes with age would reflect the development of increasingly stable sensorimotor representations (Goodin et al., 1978; Allison et al., 1984; Polich et al., 1990; Bruneau et al., 1997; Cunningham et al., 2000; Oram Cardy et al., 2004; Scheerer et al., 2013a). As the sensorimotor representation becomes more stable, the evaluation of the mismatch between perceived and expected feedback should likewise be more precise, and thus the amplitude of the N1 and P2 components should increase. Finally, we predicted that P1, N1 and P2 latencies would decrease with age, reflecting the increased synaptic efficiency that occurs with development (Tonnquist-Uhlen et al., 1995; Sharma, Kraus, McGee, & Nicol, 1997; Albrecht, Suchodoletz, & Uwer, 2000; Devous, Altuna, Ngai, Chiu, & Harris, 2006).

Methods

i. Participants

One hundred participants were recruited. Participants were divided into five age groups: 4–6 years [mean, 5.88; SD, 0.60], 7–10 years (mean, 8.47; SD, 1.26), 11–13 years (mean, 11.88; SD, 0.69), 14–17 years (mean, 16.37; SD, 0.89), and 18–30 years (mean, 22.72; SD, 3.03), with 10 males and 10 females in each group. These age groups were chosen in order to allow vocal and neural responses to FAF to be compared throughout development. All participants were native Canadian English speakers, did not speak a tonal language, reported no formal vocal training, and were right-handed. Prior to testing, participants were screened with a diagnostic audiometer (Midimate 602; Madsen Electronics, Minneapolis, MN, USA) to ensure that hearing thresholds were ≤ 20 dB HL at 250, 500, 1000, 2000, 4000 and 8000 Hz. All participants received financial compensation or course credit for participation in this study. Informed consent was obtained from all participants, as well as from a parent or guardian of participants under the age of 18 years, in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All procedures were approved by the Wilfrid Laurier University Research Ethics Board.

ii. Apparatus

Participants sat in an electrically shielded booth (Raymond EMC, Ottawa, Ontario, Canada), wearing a HydroCel GSN 64 1.0 Cap (Electrical Geodesics, Eugene, OR, USA) and a headset microphone (Countryman Isomax E6 Omnidirectional Microphone). During the experiment, vocalizations were sent to a mixer (Mackie Oynx 1220; Loud Technologies, Woodinville, WA, USA) and pitch-shifted by a digital signal processor (VoiceOne; T. C. Hellicon, Victoria, BC, Canada). The pitch-shifted vocalization was then presented back to the participant as auditory feedback through insert headphones (ER- 3; Etymotic, Elk Grove Village, IL, USA).

The unaltered voice signal was digitally recorded (HD-P2; TASCAM, Montebello, CA, USA) at a sampling rate of 44.1 kHz for later analysis.

iii. Procedure

Participants sat in front of a computer screen that displayed an interactive space game. Participants were instructed that when a red box with the word ‘activate’ appeared they should keep their gaze fixed on the box, while vocalizing the sound /a/. They were further instructed to stop vocalizing when the red box disappeared. This process was repeated until 100 vocalizations were completed.

Participants were asked to vocalize at a loud, but comfortable, volume. Vocalizations were played back to the participants in real time via headphones. Owing to the young age of many of the participants, sound level was not precisely monitored. However, previous research has shown that the relative loudness of auditory feedback has no influence on the amplitude or latency of responses to FAF (Burnett et al., 1998). Despite this fact, the experimenter monitored the volume of participants’ vocalizations and provided feedback if individuals were too loud or too soft. All participants were instructed to refrain from blinking and making extraneous movements while vocalizing.

During 50 of the 100 vocalizations produced, the participant’s voice was unaltered. Although the voice was unaltered, three samples were randomly taken per utterance, with an intersample interval of 1000–1200 ms. During the other 50 vocalizations the participant’s voice was shifted –100 cents (down one semitone), three times per vocalization, with an interstimulus interval of 1000–1200 ms. Each perturbation had a fixed duration of 200 ms. This resulted in vocalizations between 4 and 5 s in length. Altered and unaltered trials were pseudo-randomly presented throughout the experiment.

iv. Behavioural Analysis

The digital recording of the vocalizations was segmented into separate utterances, and F0 values were calculated for each utterance. Utterances were then segmented on the basis of the pitch-shift onset. F0 values for each of the three segments were normalized to the baseline period (the segment 200 ms prior to the onset of the pitch-shift) by converting Hertz values to cents with the following formula:

$$\boxed{\text{Cents} = 100\ (12\ \text{log2}\ F/B)}$$

In this formula, F is the F0 value in Hertz, and B is the mean frequency of the baseline period.

Cent values were calculated for 200 ms before and 1000 ms after the pitch-shift. An averaged F0 trace was constructed for each shift magnitude, 0 and -100, for each participant. The magnitude and latency of vocal responses, as well as participants' vocal variability (SD of the baseline period), were evaluated. The magnitude of each participant's compensatory response was determined by finding the point at which the averaged F0 trace deviated maximally from the baseline mean, and the latency was calculated as the time at which this maximal deviation occurred. The SD of the baseline period was calculated on the basis of the 200 ms of unaltered voice prior to the pitch-shift.

v. EEG Recording and Analysis

Electroencephalography data were recorded from 64 scalp electrodes referenced to an electrode at the vertex (Cz) by use of a HydroCel GSN 64 1.0 Cap, and amplified with a Net Amps 300 system (Electrical Geodesics), with a fixed anti-aliasing filter with a frequency of 6 kHz. Data were bandpass-filtered (1–30 Hz) and digitized (12-bit precision) at 1000 samples per second. Electrode impedances were adjusted and maintained below 50 kΩ (Ferree, Luu, Russell, & Tucker, 2001). After data acquisition, the EEG signals were re-referenced to the average voltage across all sites. The data were then epoched into segments from 100 ms before to 500 ms after perturbation onset. Data were analyzed offline for movement artifacts, and any segment with voltage values exceeding 55 µV of the moving average over an 80-ms span were rejected. Additional visual inspection of all data ensured that artifacts were being adequately detected. On the basis of the rejection criterion and visual inspection, participants with > 50% of their trials rejected in a single condition were not included in further analyses. This resulted in 47 rejected participants, most of whom were young children who were either unable to perform the task, or had excessive rejected trials because of ocular and movement artifacts.

Five electrodes were included in the analysis, and divided for anteriority [Fz (frontal), Cz (medial), and Pz (posterior)] and laterality [C3 (left) and C4 (right)] analyses. These electrodes were chosen in order to allow for comparison with previous FAF (Liu et al., 2013) and AEP (Bruneau et al., 1997) research.

For each participant, averaged waveforms were created for the unaltered and the -100 cent shift conditions for each electrode. Grand-averaged waveforms were created for both conditions by averaging the data from all participants for each electrode, and then performing baseline correction. For all averaged files for each participant, the amplitude and latency were calculated for the P1–N1–P2 complex. These components were extracted at time windows of 50–100 ms, 100–200 ms, and 200–300 ms, respectively. These windows were chosen on the basis of a visual inspection of the data.

vi. Statistical Analysis

In order to ensure that the recorded responses were the result of the FAF manipulation, preliminary repeated measures (RM)-ANOVAs were conducted to compare vocal response magnitude and ERP amplitudes in the -100 cent condition relative to the 0 cent (unaltered) condition. As expected, a main effect of shift magnitude indicated that responses were larger in the -100 cent condition than in the 0 cent (unaltered) condition for all measures. For this reason, subsequent analyses included only FAF trials. In addition, the 0 cent (unaltered) condition was not included in latency comparisons, as data were randomly sampled from vocalizations in the 0 cent condition, so latency values have no true reference.

For FAF trials, ANOVAs were conducted comparing the vocal responses across age groups and sex, and ERP responses across age groups, sex, and electrode site. Regression analyses were also conducted to investigate relationships between the behavioural and ERP data. In cases where Mauchley's assumption of sphericity was violated, the Greenhouse–Geisser correction was used (Greenhouse & Geisser, 1959). However, for ease of interpretation, original degrees of freedom were reported.

Results

i. Behavioural Results

Response Magnitude

A two-way ANOVA was conducted to compare the effects of age and sex on vocal response magnitude. The main effects of age, $F_{4,90} = 1.845$, $P = 0.127$, (see Figure 3) and sex, $F_{1,90} = 1.849$, $P = 0.177$, and the interaction between age and sex, $F_{4,90} = 1.383$, $P = 0.246$, failed to reach significance.

Response Latency

A two-way ANOVA was conducted to investigate the effects of age and sex on vocal response latency. The results indicated a main effect of age, $F_{4,90} = 3.024$, $P = 0.022$. Least significant difference (LSD) comparisons indicated that 4–6-year-olds ($P = 0.004$), 7–10- year-olds ($P = 0.045$) and 11–13-year-olds ($P = 0.003$) had significantly longer response latencies than 18–30-year-olds. In addition, the interaction between age and sex, $F_{4,90} = 2.602$, $P = 0.041$, was significant, as males and females had similar responses, with the exception of the 14–17-year-old males, who had significantly faster responses than the females ($P = 0.005$; see Figure 4). However, the main effect of sex, $F_{1,90} = 0.263$, $P = 0.609$, failed to reach significance.

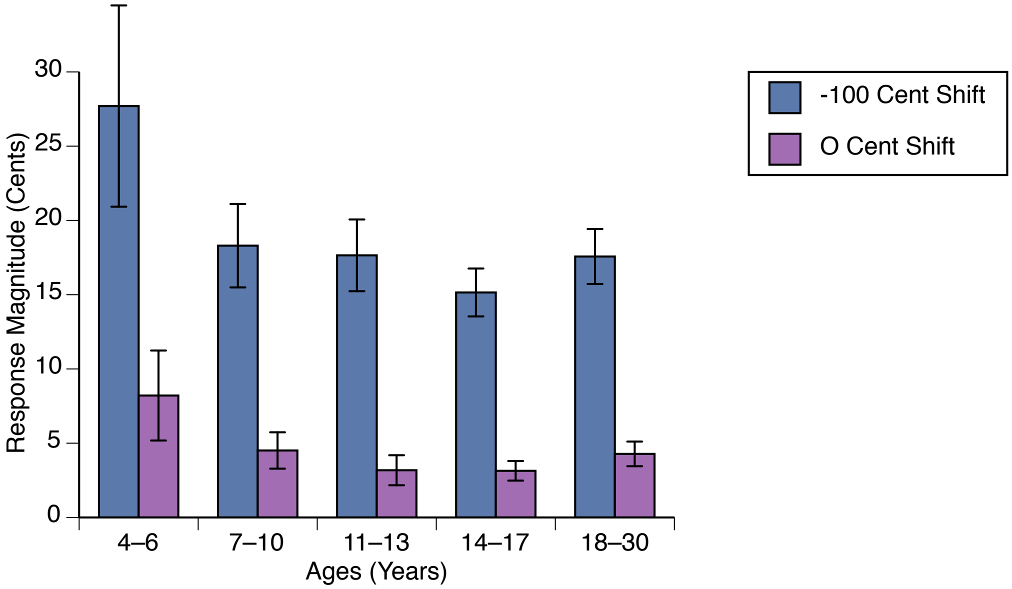

Figure 3: Vocal Response Magnitudes (mean and standard error) in the -100 cent (blue) and 0 cent (purple) conditions as a function of age group.

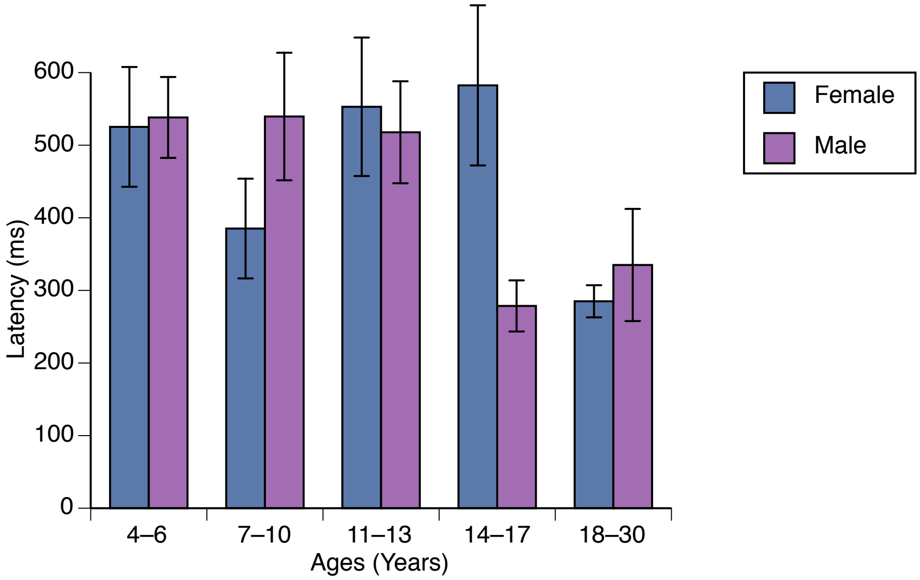

Figure 4: Vocal response latencies (mean and standard error) in the -100 cent condition as a function of age and sex.

Vocal Variability

A two-way ANOVA was conducted to compare the effects of age and sex on vocal variability (as indexed by baseline SD). The main effect of age, $F_{4,90} = 2.626$, $P = 0.040$, was significant (see Figure 5). LSD comparisons indicated that 4–6-year-olds were significantly more variable than 18–30-year-olds ($P = 0.002$). The main effect of sex, $F_{1,90} = 0.754$, $P = 0.388$, and all interactions failed to reach significance.

In order to determine whether baseline variability predicted the magnitude of the responses to FAF, a correlational analysis was conducted. As baseline variability did not differ as a function of the shift magnitude, $F_{1,90} = 0.036$, $P = 0.851$, baseline variability values were collapsed across the -100 cent shift and unaltered feedback conditions. The results indicated a significant correlation, $r = 0.628$, $P_{\text{(two-tailed)}} < 0.001$, $n = 100$, between baseline variability and the magnitude of responses to FAF. As vocal variability was found to vary as a function of age ($P = 0.040$), a regression analysis was performed to determine whether age and vocal variability could account for a significant proportion of the variance in vocal response magnitude. The results indicated that vocal variability, $b = 21.013$, $t_{97} = 7.683$, $P < 0.001$, but not age, $b = 0.074$, $t_{97} = 0.336$, $P = 0.738$, accounted for a significant proportion of the variance in vocal response magnitude. Together, both age and vocal variability accounted for a significant proportion of the variance, $R^2 = 0.395$, $F_{2,97} = 31.632$, $P < 0.001$, in vocal response magnitude.

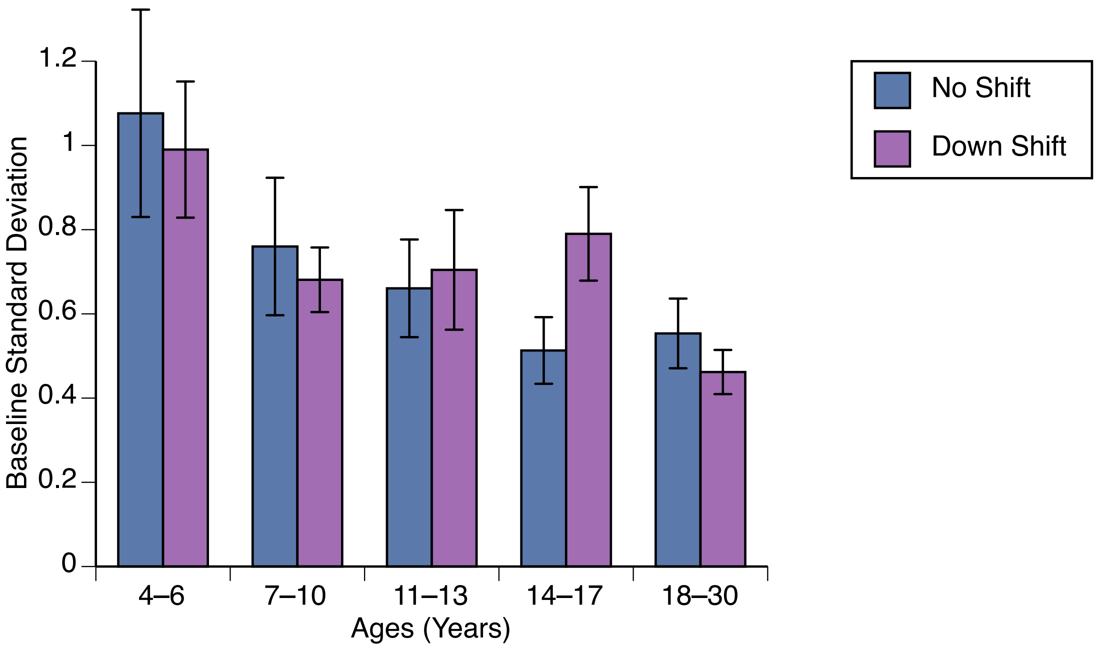

Figure 5: Vocal pitch (F0) variability as indexed by baseline standard deviation (mean and standard error) in the -100 cent (purple) and 0 cent (blue) conditions as a function of age group.

ii. ERP Results

P1 Amplitude Anteriority

A RM-ANOVA was conducted to investigate the effects of age and sex on P1 amplitudes across three midline electrodes. The results indicated a significant main effect of anteriority, $F_{2,180} = 53.168$, $P < 0.001$, as P1 amplitudes were largest at the anterior site, followed by the medial and then the posterior site (all three sites significantly different at $P < 0.001$). The main effect of age was also significant, $F_{4,90} = 8.844$, $P < 0.001$ (see Figure 6), with LSD comparisons indicating that 4–6-year-olds had significantly larger P1 amplitudes than 11–13-year-olds ($P = 0.001$), 14–17-year-olds ($P < 0.001$), and 18–30-year-olds ($P < 0.001$), and that 7–10-year-olds had significantly larger P1 amplitudes than 14–17-year-olds ($P = 0.002$) and 18–30-year-olds ($P = 0.003$). The interaction between sex and age was significant, $F_{4,90} = 3.012$, $P = 0.022$, as male 4–6-year-olds had larger P1 amplitudes than females ($P = 0.035$), whereas 7–10-year-old females had larger P1 amplitudes than males ($P = 0.010$). The main effect of sex, $F_{1,90} = 0.192$, $P = 0.662$, and all other interactions, failed to reach significance.

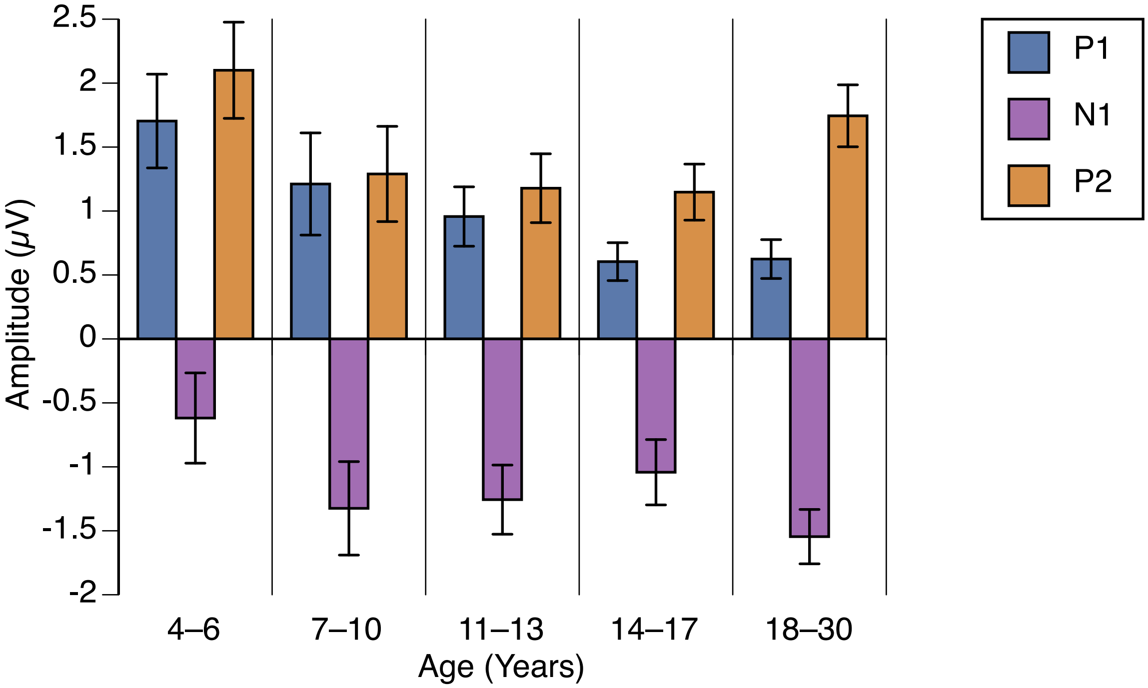

Figure 6: P1 (blue), N1 (purple), and P2 (orange) amplitudes (mean and standard error) collapsed across all three electrode sites (Fz, Cz, and Pz) as a function of age.

P1 Amplitude Laterality

A RM-ANOVA was conducted to investigate the effects of age and sex on P1 amplitudes in the right and left hemispheres. A significant main effect of sex, $F_{1,90} = 13.051$, $P < 0.001$, was found, as males had larger P1 amplitudes than females. The main effect of age was also significant, $F_{4,90} = 4.297$, $P = 0.003$. Least significant difference comparisons indicated that 4–6-year-olds had significantly larger P1 amplitudes than 14–17-year-olds ($P = 0.003$) and 18–30 year-olds ($P = 0.001$), 7–10-year-olds had significantly larger P1 amplitudes than 18–30-year-olds ($P = 0.016$), and 11–13-year-olds had significantly larger P1 amplitudes than 18–30-year-olds ($P = 0.046$). The main effect of laterality, $F_{1,90} = 0.001$, $P = 0.972$, and all other interactions, failed to reach significance.

P1 Latency Anteriority

A RM-ANOVA was conducted to investigate the effects of sex and age on P1 latency across three midline electrodes. A main effect of anteriority was found, $F_{2,180} = 67.272$, $P < 0.001$, as P1s peaked more rapidly at the posterior sites, with latency increasing from posterior to anterior electrodes on the scalp ($P < 0.01$ for all comparisons). There was also an anteriority x age interaction, $F_{8,180} = 4.217$, $P < 0.001$, as changes in latency across the three midline sites were robust in the 4–6-year-olds, 7–10-year-olds, and 11–13-year-olds, whereas they were more moderate in the 14–17- year-olds and 18–30-year-olds. Similarly, a main effect of age was also found, $F_{4,90} = 2.745$, $P = 0.033$. LSD comparisons indicated that 18–30-year-olds had significantly faster P1s than 4–6-year-olds ($P = 0.003$), 7–10-year-olds ($P = 0.031$), 11–13-year-olds ($P = 0.033$), and 14–17-year-olds ($P = 0.015$). Finally, the main effect of sex was significant, $F_{1,90} = 8.317$, $P = 0.005$, as females had faster P1s than males. All other interactions failed to reach significance.

P1 Latency Laterality

A RM-ANOVA was conducted to investigate the effects of sex and age on P1 latency in the right and left hemispheres. Although the main effect of laterality, $F_{1,90} = 0.592$, $P = 0.444$, failed to reach significance, main effects of sex, $F_{1,90} = 16.333$, $P < 0.001$, and age, $F_{4,90} = 4.769$, $P = 0.002$, were found. LSD comparisons indicated that females had faster P1s than males ($P = 0.002$). Furthermore, LSD comparisons indicated that 4–6-year-olds had significantly slower P1s than 14–17-year-olds ($P = 0.002$) and 18– 30-year-olds ($P < 0.001$), 7–10-year-olds had significantly slower P1s than 14–17-year-olds ($P = 0.039$) and 18–30-year-olds ($P = 0.009$), and 11–13-year-olds had significantly slower P1s than 18–30-year-olds ($P = 0.014$). All interactions failed to reach significance.

N1 Amplitude Anteriority

A RM-ANOVA was conducted to investigate the effects of sex and age on N1 amplitudes over three midline sites. A main effect of anteriority, $F_{2,180} = 21.740$, $P < 0.001$, was found (see Figure 6 and 8), as N1 amplitudes were significantly larger at the posterior site than at the medial site ($P = 0.012$) and anterior site ($P < 0.001$), which were also significantly different from one another ($P < 0.001$). The interaction between anteriority and age was significant, $F_{8,180} = 7.679$, $P = 0.001$, as the younger age groups (4–6, 7–10 and 11–13 years) had robust N1 amplitude differences across the midline electrode sites, with the largest amplitude occurring at the posterior electrode, whereas the older age groups (14–17 and 18–30 years) had more moderate N1 amplitude differences across the midline electrode sites, with the largest N1 amplitude occurring at the medial electrode site (see Figure 7). Accordingly, the main effect of age, $F_{4,90} = 5.128$, P = 0.001, was also significant (see Figure 6), with 4–6-year-olds having significantly smaller amplitudes than 7–10-year-olds ($P = 0.002$), 11–13-year-olds ($P = 0.005$), and 18–30-year-olds ($P < 0.001$). In addition, 14–17-year-olds had significantly smaller amplitudes than 18–30-year-olds ($P = 0.024$). The main effect of sex, $F_{1,90} = 8.178$, $P = 0.005$, was also significant, as females had significantly larger amplitudes than males. All other interactions failed to reach significance.

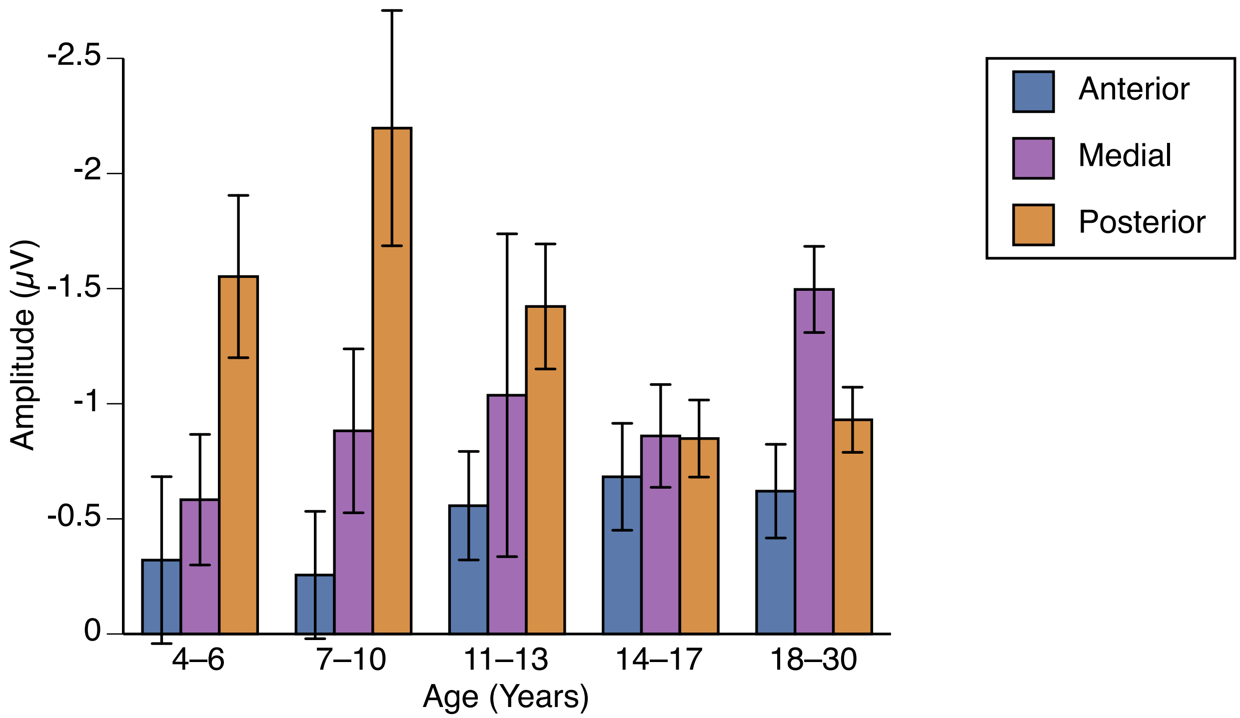

Figure 7: N1 Amplitudes (mean and standard error) as a function of age and electrode site (anterior, Fz; medial, Cz; posterior, Pz).

N1 Amplitude Laterality

A RM-ANOVA was conducted to investigate the effects of sex and age on N1 amplitudes in the right and left hemispheres. The main effects of laterality, $F_{1,90} = 0.092$, $P = 0.762$, sex, $F_{1,90} = 0.779$, $P = 0.380$, and age, $F_{4,90} = 1.881$, $P = 0.121$, and all interactions, failed to reach significance.

N1 Latency Anteriority

A RM-ANOVA was conducted to investigate the effects of sex and age on N1 latency across three midline electrodes. A significant main effect of anteriority was found, $F_{2,180} = 58.111$, $P < 0.001$, as N1s were faster at the posterior site than at the anterior site ($P < 0.001$) and medial site ($P < 0.001$). A main effect of sex was also found, $F_{1,90} = 9.650$, $P = 0.003$, as females had faster N1s than males. Finally, a main effect of age was found, $F_{4,90} = 2.819$, $P = 0.030$. LSD comparisons indicated that 18–30-year-olds had significantly faster N1s than 4–6-year-olds ($P = 0.013$), 7–10-year-olds ($P = 0.002$), and 11–13-year-olds ($P = 0.029$). All interactions failed to reach significance.

N1 Latency Laterality

A RM-ANOVA was conducted to investigate the effects of sex and age on N1 latency in the right and left hemispheres. A significant main effect of age was found, $F_{4,90} = 9.711$, $P < 0.001$. LSD comparisons indicated that 14–17-year-olds had significantly faster N1s than 4–6-year-olds ($P = 0.001$), 7–10-year-olds ($P < 0.001$), and 11–13-year-olds ($P = 0.010$), whereas 18–30-year-olds also had significantly faster N1s than 4–6-year-olds ($P < 0.001$), 7–10-year-olds ($P < 0.001$), and 11–13-year-olds ($P = 0.001$). The interaction between age and laterality was also significant, as 4–6-year-olds showed faster N1s in the right hemisphere, whereas 11–13-year-olds and 14–17-year-olds showed faster N1s in the left hemisphere. Finally, a significant main effect of sex was found, $F_{1,90} = 9.680$, $P = 0.002$, as females had faster N1s than males. The main effect of laterality, $F_{1,90} = 1.607$, $P = 0.208$, failed to reach significance.

P2 Amplitude Anteriority

A RM-ANOVA was conducted to investigate the effects of sex and age on P2 amplitudes across three midline sites. A significant main effect of anteriority was found, $F_{2,180} = 8.045$, $P = 0.001$, as P2 amplitudes were smaller at the posterior site relative to the medial site, ($P = 0.004$), and anterior site ($P = 0.002$). The interaction between anteriority and age was also significant, $F_{8,180} = 2.727$, $P = 0.007$. Correspondingly, the main effect of age was significant, $F_{4,90} = 5.521$, $P = 0.001$ (see Figure 6). LSD comparisons indicated that 4–6-year-olds had significantly larger P2 amplitudes than 7–10-year-olds ($P = 0.002$), 11–13-year-olds ($P < 0.001$), and 14–17-year-olds ($P < 0.001$), while 18–30-year-olds had significantly larger P2 amplitudes than 11–13-year-olds ($P = 0.024$) and 14–17-year-olds ($P = 0.018$). The main effect of sex, $F_{1,90} = 0.832$, $P = 0.507$, and all other interactions, failed to reach significance.

P2 Amplitude Laterality

A RM-ANOVA was conducted to investigate the effects of sex and age on P2 amplitudes in the right and left hemispheres. A significant main effect of age was found, $F_{4,90} = 3.575$, $P = 0.009$, as 7–10-year-olds and 11–13-year-olds had significantly smaller P2 amplitudes than all other age groups ($P < 0.05$). However, the main effects of laterality, $F_{1,90} = 0.351$, $P = 0.555$, and sex, $F_{1,90} = 3.353$, $P = 0.070$, and all other interactions, failed to reach significance.

P2 Latency Anteriority

A RM-ANOVA was conducted to investigate the effects of sex and age on P2 latency across three midline electrodes. A significant main effect of anteriority was found, $F_{2,180} = 4.043$, $P = 0.022$, as latencies were significantly longer at the anterior site relative to the medial site ($P = 0.006$) and posterior site ($P = 0.042$). A main effect of sex was also found, $F_{1,90} = 4.617$, $P = 0.034$, as females had faster P2s than males. The main effect of age, $F_{4,90} = 0.210$, $P = 0.932$, and all interactions, failed to reach significance.

P2 Latency Laterality

A RM-ANOVA was conducted to investigate the effects of sex and age on P2 latency in the right and left hemispheres. A significant main effect of laterality was found, $F_{1,90} = 6.278$, $P = 0.014$, as P2 latency was shorter in the left hemisphere than in the right hemisphere. The main effects of sex, $F_{1,90} = 0.208$, $P = 0.650$, and age, $F_{4,90} = 1.194$, $P = 0.319$, and all interactions, failed to reach significance.

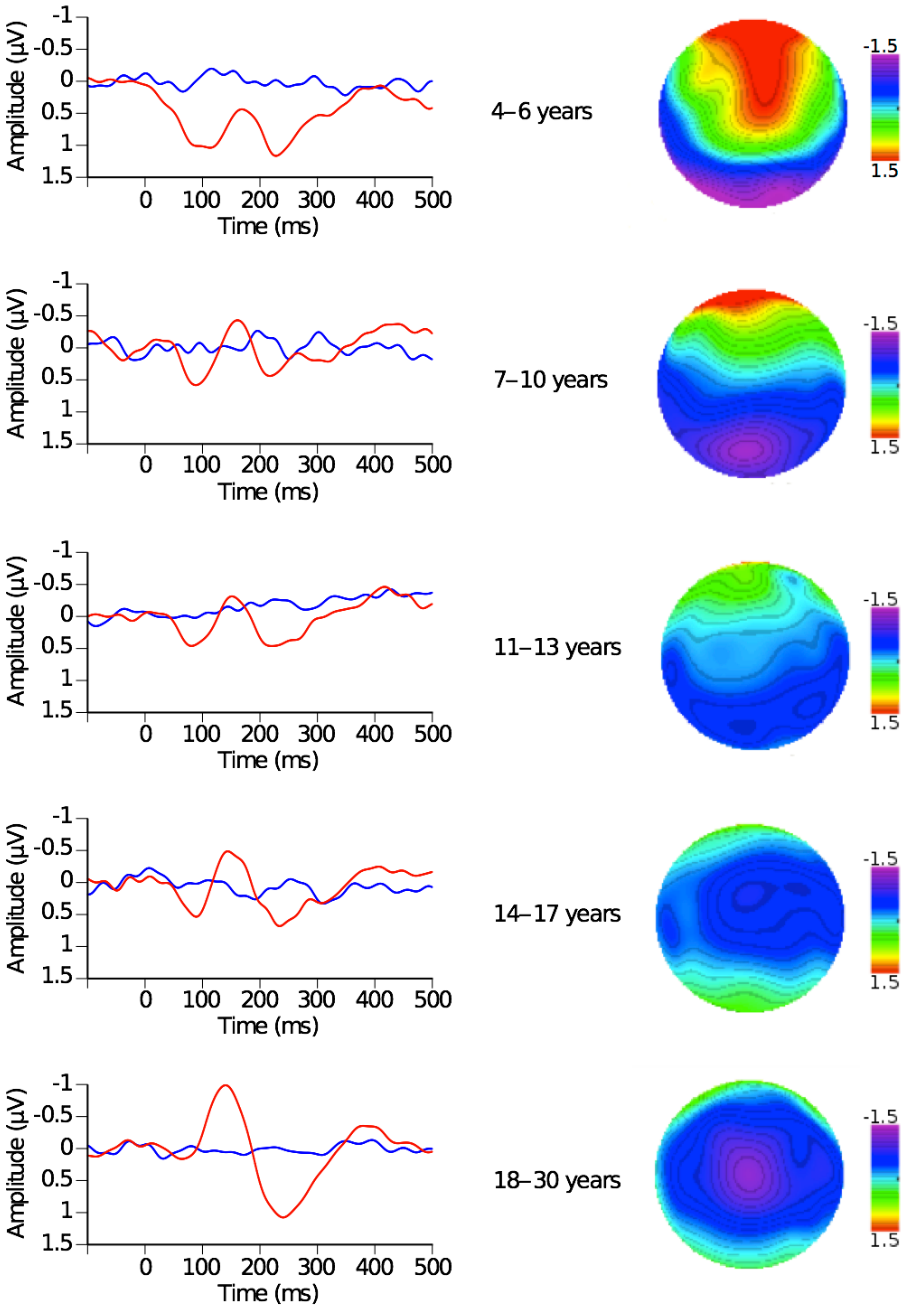

Figure 8: Left column: ERP waveforms averaged across three electrode sites (Fz, Cz, and Pz) for each age group. The blue lines depict ERPs during the 0 cent condition, and the red lines depict ERPs during the -100 cent condition. Time zero represents the point at which the feedback perturbation was initiated. Right column: topographical scalp distribution of ERPs in the -100 cent condition between 100 and 200 ms following the feedback perturbation.

iii. Regression Analyses

A regression analysis was conducted to determine whether the variance in P1 amplitudes and latencies could be accounted for by age, vocal response magnitude, and vocal variability. Age, $b = -0.496$, $t_{96} = -5.351$, $P < 0.001$, significantly predicted P1 amplitude; however, vocal response magnitude, $b = -0.147$, $t_{96} = -1.290$, $P = 0.200$, and vocal variability, $b = 0.011$, $t_{96} = 0.096$, $P = 0.924$, did not. Together, age, vocal response magnitude and vocal variability explained a significant proportion of the variance in P1 amplitudes, $R^2 = 0.245$, $F_{3,96} = 10.363$, $P < 0.001$. Similarly, age, $b = -0.349$, $t_{96} = -3.610$, $P < 0.001$, but not vocal response magnitude, $b = -0.239$, $t_{96} = -2.009$, $P = 0.051$, or vocal variability, $b = 0.217$, $t_{96} = 1.771$, $P = 0.080$, significantly predicted P1 latencies. Together, age, vocal response magnitude and vocal variability explained a significant proportion of the variance in P1 latencies, $R^2 = 0.177$, $F_{3,96} = 6.890$, $P < 0.001$.

The regression analysis of N1 amplitudes revealed that age, $b = -0.234$, $t_{96} = -2.325$, $P = 0.022$, and vocal variability, $b = 0.260$, $t_{96} = 2.054$, $P = 0.043$, were both significant predictors of N1 amplitude; however, vocal response magnitude was not, $b = -0.141$, $t_{96} = -1.145$, $P = 0.255$. Together, age, vocal variability and vocal response magnitude explained a significant proportion of the variance in N1 amplitudes, $R^2 = 0.120$, $F_{3,96} = 4.354$, $P < 0.001$. Analysis of N1 latencies indicated that age, $b = -0.488$, $t_{96} = -5.306$, $P < 0.001$, but not vocal response magnitude, $b = -0.026$, $t_{96} = -0.226$, $P = 0.821$, or vocal variability, $b = 0.101$, $t_{96} = 0.891$, $P = 0.375$, significantly predicted N1 latency. Together, age, vocal variability and vocal response magnitude explained a significant proportion of the variance in N1 latency, $R^2 = 0.255$, $F_{3,96} = 10.949$, $P < 0.001$.

The regression analysis of P2 amplitudes revealed that neither age, $b = 0.090$, $t_{96} = 0.855$, $P = 0.395$, vocal response magnitude, $b = -0.130$, $t_{96} = -0.999$, $P = 0.320$, nor vocal variability, $b = 0.181$, $t_{96} = 1.358$, $P = 0.178$, accounted for a significant proportion of the variance in P2 amplitudes. Accordingly, overall age, vocal variability and response magnitude did not significantly correlate with P2 amplitudes, $R^2 = 0.023$, $F_{3,96} = 0.744$, $P = 0.528$. This trend continued for P2 latency, as neither age, $b = -0.074$, $t_{96} = -0.670$, $P = 0.486$, vocal response magnitude, $b = -0.061$, $t_{96} = -0.467$, $P = 0.642$, nor vocal variability, $b = 0.005$, $t_{96} = 0.039$, $P = 0.969$, accounted for a significant proportion of the variance in P2 latency. Together, as expected, age, vocal response magnitude and vocal variability failed to account for a significant proportion of the variance in P2 latency, $R^2 = 0.008$, $F_{3,96} = 0.245$, $P = 0.865$.

Discussion

In this study, behavioural and neurological responses to FAF were examined to investigate developmental changes in the processing of auditory feedback. Overall, the behavioural results indicated that vocal responses to FAF were modulated as a function of age, as both response latency and vocal variability decreased with age. In addition, a strong correlation between vocal variability and response magnitude further highlighted the modulatory effects of vocal skill on vocal responses to FAF. Neurological changes to FAF were also observed, as the P1–N1–P2 complex varied as a function of age. Developmental changes in the P1 component were indexed by a decrease in amplitude and latency with age. The N1 component was found to increase in amplitude and decrease in latency with age, whereas the P2 component was found to increase in amplitude with age. Besides the overall effect of age, age by electrode site interactions also existed, which showed changes in the location of the peak amplitudes and latencies across the age groups tested. In addition to developmentally related changes, the P1–N1–P2 complex was found to differ as a function of sex, as females showed faster responses across all three components. Although, as hypothesized, age was shown to modulate a number of behavioural and neurological responses independently, regression analyses also highlighted relationships between the N1 component and both age and vocal variability, suggesting a complex relationship between behavioural and neurological responses to FAF.

To date, only one other study has investigated both behavioural and neurological responses to FAF in children. As in the current study, Liu et al. (2013) found that vocal response magnitude did not differ as a function of age. However, they did find a main effect of sex, suggesting that males produced larger responses than females, a result that was not confirmed by the results of the current study. In line with the current study, Liu et al. (2013) also found that vocal response latency varied as a function of age. However, their results did not show that females produce more rapid vocal responses than males, as reported here. When considering the ERP results of Liu et al. (2013), we found that both P1 amplitude and latency, as well as N1 latency, were modulated by age in the same fashion as reported here. In addition, P2 amplitude modulation showed a similar trend as in Liu et al. (2013), who found that P2 amplitudes increased for older girls relative to younger girls, whereas the current study found that P2 amplitudes increased with age, independently of sex. However, both N1 and P2 latency, as well as the sex-related differences reported by Liu et al. (2013), are incongruent with the results found here. Although Liu et al. (2013) used a very similar paradigm, the current study investigated responses to FAF across a larger range of ages. The current study investigated FAF responses in 4–30-year-olds, whereas the study by Liu et al. (2013) looked at children and adults between the ages of 10 and 25 years. As many of the age-related differences found in this study were the result of comparisons between children under the age of 10 years, and older children and adults, the differences found between these two studies may have resulted from the ages of the participants. Another confounding factor is the language experience of the participants. The participants in the Liu et al. (2013) study were Mandarin-speaking children and adults. Previous behavioural and neurophysiological studies have shown that vocal and neural responses to FAF are language-specific (Liu et al., 2010a,b; Chen et al., 2012b), and tonal language speakers, such as Mandarin speakers, respond to FAF faster than non-tonal language speakers, such as English speakers (Chen et al., 2007; Liu et al., 2010a,b). These findings suggest that the differences found between the current study and that of Liu et al. (2013) may have also arisen as a result of differences in the way in which tonal and non-tonal language speakers process FAF. Further investigation is required to determine whether the neural activation in response to FAF differs between tonal and non-tonal language speakers.

Although children's neurological responses to FAF have been rarely reported, there is an abundance of literature reporting on the developmental trajectory of AEPs. AEP research has consistently shown that P1 amplitudes decrease with age (Nelson, Hall III, & Jacobson, 1997; Ponton et al., 2000, 2002; Ponton & Eggermont, 2001, 2007; Ceponiene et al., 2002; Oram Cardy et al., 2004; Sussman et al., 2008), whereas P1 latencies become shorter with age (Ponton et al., 2000, 2002; Ceponiene et al., 2002; Gage et al., 2003; Oram Cardy et al., 2004; Beal et al., 2011). Both of these trends are in line with the P1 results reported in this study, as well as the FAF study by Liu et al. (2013). The N1 latency results obtained in this study are also in line with previous AEP research, as it has been well documented that N1 latencies decrease with age (Goodin et al., 1978; Allison et al., 1984; Polich et al., 1990; Paetau et al., 1995; Tonnquist-Uhlen et al., 1995; Bruneau et al., 1997; Ponton et al., 2000; Gage et al., 2003; Oram Cardy et al., 2004). However, the AEP literature is far less clear regarding N1 amplitude. Many AEP studies have found that N1 amplitudes increase with age (Goodin et al., 1978; Allison et al., 1984; Polich et al., 1990; Bruneau et al., 1997; Cunningham et al., 2000; Oram Cardy et al., 2004), as we found here. However, other studies have reported that N1 amplitudes decrease with age (Ceponiene et al., 2002) or are absent in younger children (Ponton et al., 2000, 2002), making it difficult to compare the current N1 amplitude results with previous AEP studies. Although the pattern is more complicated, the results of this study suggest that P2 amplitude increases with age, which is consistent with the previous AEP literature (Kraus et al., 1993;

Johnstone et al., 1996; Oades et al., 1997). Also in line with previous AEP research, P2 latencies were not found to vary systematically as a function of age (Johnstone et al., 1996; Tonnquist-Uhlen, 1996; Ponton et al., 2000). The strong similarities between the P1–N1–P2 values found in this study and previous AEP studies suggest that many of the ERP differences found in this study result from generalized changes in auditory cortical processing throughout development.

It has commonly been suggested in AEP research that the P1 component is the childhood correlate of the adult N1 component (Sharma et al., 1997; Ceponiene et al., 2002). The P1 component has been suggested to reflect the analysis of the basic sound features of a stimulus, whereas the adult N1 wave is thought to reflect a higher-level analysis that is integrated with ongoing mental activity (Ceponiene et al., 2002). This is in line with the results of the current study, as it was found that P1 amplitudes decreased with age, whereas N1 amplitudes increased with age. Furthermore, the P1 component was not found to vary systematically with any of the behavioural responses, suggesting that P1 amplitude modulation was the result of developmental changes in auditory cortical processing. On the other hand, N1 amplitude was found to correlate with vocal variability. Previous work with adults has shown modulation of the N1 component by FAF (Heinks-Maldonado et al., 2005; Behroozmand & Larson, 2011; Scheerer et al., 2013a). It is likely, then, that the N1 responses in the current study were likewise altered by vocal motor input, and were not purely the result of developmental changes in auditory cortical processing. Like the P1 component, P2 amplitudes were not found to vary systematically with any of the behavioural measures. Previous FAF research has suggested that the P2 component indexes the size of the mismatch between perceived and expected feedback. Accordingly, the trend for an increase in P2 amplitudes with age may reflect the fact that, with age, individuals develop more accurate sensorimotor representations, which result in better detection of the deviation between perceived and expected feedback. However, as the 4–6-year-old group also showed significantly larger P2 amplitudes, and previous FAF studies have utilized multiple perturbation magnitudes, further investigation is required to elucidate the relationship between age and P2 amplitude modulation. In addition to amplitude differences, both P1 and N1 latencies were found to differ as a function of age. It has been suggested that latency changes reflect changes in synaptic density and efficacy in the auditory cortex (Eggermont, 1989; Tonnquist-Uhlen, 1996; Albrecht et al., 2000). As developmental changes in the vocal tract and articulators of adults are infrequent, one would expect the synaptic connections in the auditory cortex, as well as the pathways between speech-related brain regions, to remain relatively consistent. As a result of this consistent processing, long term potentiation is likely to create synaptically efficient connections within the auditory cortex and between speech-related brain regions, thus increasing the speed of processing (Devous et al., 2006). As a result of this increased processing speed, one would expect responses to also be more rapid, accounting for the decreased latencies found here.

From the results of previous studies, it was expected that the magnitude of the vocal responses to FAF would decrease with age, reflecting increased speech motor control, and a transition from feedback to feedforward control. However, there are discrepancies in the literature as to whether response magnitudes increase with age (Liu et al., 2010b; MacDonald et al., 2012), decrease with age (Liu et al., 2010a), or are unaffected by age (Liu et al., 2013). The results of this study suggest that response magnitudes do not vary as a function of age, but that vocal variability does. Interestingly, vocal variability differed across the age groups, and also correlated with response magnitude and N1 amplitude. When they were combined in a regression analysis, it was found that age and vocal variability accounted for a significant proportion of the variance in response magnitudes and in N1 amplitudes. These results suggest that, although age can have a modulatory effect on response magnitudes, vocal variability is a more reliable predictor of response magnitudes.

Previous research has also shown a significant correlation between vocal variability and vocal response magnitude (Scheerer & Jones, 2012). It has been suggested that vocal variability is related to the proficiency of speech motor control (Scheerer & Jones, 2012). When vocal stability improves through vocal training or development, vocal variability decreases, and speech motor control transitions from a feedback-biased control system to a feedforward-biased control system (Schmidt & Lee, 2005; Civier et al., 2010). This transition from a feedback-biased control system to a feedforward-biased control system is hypothesized to be driven by the stability of the sensorimotor representation (Scheerer & Jones, 2012). When the mapping between the motor commands responsible for speech production and their sensory consequences is weak, auditory feedback is required to monitor and retune the sensorimotor representation (Tourville et al., 2008). However, when the sensorimotor representation becomes more reliable at predicting the sensory consequences of the upcoming articulations, the additional information provided by auditory feedback becomes redundant, and a transition from feedback to feedforward control occurs. However, strict feedforward control is not ideal, as auditory feedback would no longer be available to retune the sensorimotor representations (Tourville et al., 2008). In the case of development, auditory feedback is particularly important, as the size and shape of the vocal tract and articulators change throughout development, so retuning of the sensorimotor representation is necessary (MacDonald et al., 2012). Strict feedback control is also not ideal, as it introduces time lags and disfluencies (Civier et al., 2010). Instead, current speech production models suggest that speech production involves a combination of feedback and feedforward control (Guenther, 2006; Hickok et al., 2011).

Although it does not appear that age was directly related to changes in vocal response magnitude, vocal response latency was found to vary with age. This result is consistent with previous research investigating the processing of FAF in children (Liu et al., 2010a,b, 2013). This decrease in latency with age may simply reflect neurophysiological maturation (Rojas et al., 1998; Kotecha et al., 2009). As discussed with regard to ERP latencies, it has been suggested that decreases in neural response latencies reflect changes in synaptic density and efficacy in the auditory cortex (Eggermont, 1989). It is likely that these changes in efficacy also influence behavioural response latencies, as increased synaptic efficacy promotes faster neural processing, which should result in more rapid behavioural responses. Ultimately, this result suggests that vocal response latency can be used as an index of the maturation of the audio-vocal system.

A regression analysis performed to disentangle the relationship between vocal and neurological responses indicated that age and vocal variability accounted for a significant proportion of the variance in N1 amplitude. More specifically, as vocal variability increased, N1 amplitude decreased, or became more positive. To our knowledge, this is the first study to report a relationship between N1 amplitude and vocal variability. Previous research (Scheerer et al., 2013a) has shown that N1 amplitude increases in an all-or-nothing manner in response to auditory feedback that is deviant, but is still considered to be self-produced. However, when deviant feedback becomes large enough to be classified as externally produced, N1 amplitudes increase further (Scheerer et al., 2013a). We suggest that N1 amplitude is modulated by the proficiency of speech motor control, or the weighting of the feedback vs. the feedforward control system. As mentioned previously, the transition from feedback control to feedforward control is driven by the stability of the sensorimotor representation (Schmidt & Lee, 2005; Civier et al., 2010). In accordance with the results found here, we suggest that individuals who weigh the feedback control system more heavily, as a result of poor speech motor control, are likely to accept a larger range of deviant feedback as internally produced. As a result, larger feedback perturbations are required to elicit larger N1 responses. On the other hand, individuals who weigh the feedforward control system more heavily, as a result of stable speech motor control, are more likely to classify the deviant feedback as externally produced, thus resulting in larger

N1 amplitudes in response to relatively smaller feedback perturbations. This would explain why, in the current study, individuals with increased vocal variability produced smaller N1 amplitudes in response to FAF, whereas individuals with more stable vocal output produced larger N1 amplitudes. This relationship between N1 amplitude and vocal variability highlights the complex interaction between vocal and neurological responses to FAF.

Although the focus of this study was to investigate the developmental trajectory of behavioural and neurological responses to FAF, sex and ERP distribution across the scalp were also investigated. The results indicated that females produced faster responses both behaviourally and neurologically, as both vocal response and P1– N1–P2 latencies were shorter for females than for males. Chen et al. (2010) previously reported shorter vocal response latencies in response to FAF in females than in males. Shorter ERP latencies in females than in males have also previously been reported in both FAF (Liu et al., 2013) and AEP (Allison et al., 1984; Nanova, Lyamova, Hadjigeorgieva, Kolev, & Yordanova, 2008) studies. Although it is also unclear why females produce faster ERP responses, Nanova and colleagues (2011) have suggested that females have stronger functional synchronization of oscillatory responses during stimulus processing, resulting in shorter-latency ERPs. Despite the general trend for females to respond faster than males, 14–17-year-old males were found to have significantly shorter vocal response latencies than 14–17-year-old females. As the vocal response latencies recorded for 14–17-year-old males are similar to those reported for 18–30-year-olds, these results suggest that the vocal motor control system may mature faster in males than in females. However, further investigation is required to elucidate the cause of the sex-related latency differences in these age groups. In addition to sex related differences in latency, females were found to have larger N1 amplitudes. Further investigation is required to clarify sex-related differences in the processing of FAF.

In addition to sex, behavioural and neurological responses to FAF also varied as a function of electrode site. Both amplitudes and latencies of the P1–N1–P2 complex varied as a function of anteriority. ERP latencies were shortest at the posterior electrode site relative to the medial and anterior sites, which is consistent with previous AEP studies (Bruneau et al., 1997; Ponton et al., 2000). Also in accordance with previous AEP (Nelson et al., 1997; Ponton et al., 2000) and FAF (Liu et al., 2013) studies, P1 amplitudes were found to be largest at the anterior site. Similarly, P2 amplitudes were largest at the anterior site, which is also in accordance with previous FAF studies (Korzyukov et al., 2012; Liu et al., 2012). In contrast with AEP studies, which have consistently reported the largest N1 amplitudes at anterior sites (Bruneau et al., 1997; Ponton et al., 2000; Nanova et al., 2008), the results of this study showed the largest N1 amplitudes at the posterior site in children (4–13 years) and at the medial site in young adults (14–30 years). Previous FAF studies have also reported N1 amplitudes to be maximal at medial sites in adults (Korzyukov et al., 2012a; Scheerer et al., 2013a). Although the site of maximal N1 amplitude found in this study is inconsistent with previous AEP research, N1 amplitude was found to correlate with vocal variability, suggesting that N1 amplitude was not only modulated by basic auditory processing, but was also influenced by the motor aspect involved in producing a response to FAF. With regards to laterality, P2 latency was found to differ as a function of laterality, with more rapid P2s occurring in the left hemisphere. Once again, the congruency of the P1 and P2 activation patterns with the AEP literature suggests that the results found here were predominantly attributable to changes in basic auditory processing with age. However, the incongruency of the N1 amplitude results with the AEP literature suggests that N1 amplitudes were further modulated by the FAF manipulation.

Conclusion

The age-related differences in both behavioural and neurological responses to FAF found in this study suggest that children and adults process auditory feedback in different ways. This trend demonstrates that the auditory feedback system undergoes robust changes with age and physiological development. Although many of the responses to FAF reported here appear to be the result of developmentally related changes in auditory processing, the relationship between vocal variability and N1 amplitude suggests that a complex interaction between vocal and neural responses to FAF exists. Furthermore, the finding that vocal variability, rather than age, significantly predicts the magnitude of the vocal response to FAF suggests that audio-vocal proficiency is independent of age. Although this study has begun to uncover the developmental trajectory of the audio-vocal system, further work is required to adequately describe the influence of age on the processing of FAF. Studies utilizing multiple shift magnitudes would help to more effectively delineate the effects of FAF from developmental changes in auditory processing. In addition, performing a similar study with the addition of a playback condition would help to separate the influence of the auditory system from that of the auditory and motor systems working in concert. Finally, the new sex-based and electrode site-based trends reported here warrant more extensive investigation.

Study 3: Sensorimotor Learning in Children and Adults: Exposure to Frequency-Altered Auditory Feedback during Speech Production.

Submitted As:

Scheerer, N.E., Jacobson, D.S., & Jones, J.A. (2015). Sensorimotor Learning in Children and Adults: Exposure to Frequency-Altered Auditory Feedback during Speech Production. Neuroscience. Accepted: November 18, 2015.

Speech is acquired early in childhood, with many children speaking fluently by the age of 3 (Ingram, 1989; Singleton & Ryan, 2004). However, the speech motor control system continues to develop for many years after the initial acquisition of speech. Throughout development the shape, size, and muscle innervation of the articulators involved in speech production undergo dramatic changes, which influence the acoustic properties of the vocal tract (Callan et al., 2000). Despite these changes, the speech motor control system accomplishes the same functional goals, allowing children to maintain relatively fluent speech. In addition to facilitating the acquisition of speech, auditory feedback may also function as an adaptive signal to guide movements of the articulators in order to preserve speech fluency while the articulators are restructured during development (Callan et al., 2000). Although auditory feedback plays an important role in the acquisition of fluent speech, this role may change once speech is acquired and individuals are no longer experiencing persistent developmental changes to their speech motor control system. For this reason, this study was designed to investigate if younger children who are still experiencing ongoing changes to their speech motor control system, demonstrate different rates of sensorimotor learning following persistent exposure to deviant auditory feedback, relative to young adults. In order to investigate this research question, young children aged 5-8 years, and young adults aged 18-30 years produced vocalizations as part of a FAF adaptation paradigm. The results of this study will help to uncover how development influences the plasticity of sensorimotor representations, and in turn modulates feedforward control.

Abstract

Auditory feedback plays an important role in the acquisition of fluent speech; however, this role may change once speech is acquired and individuals no longer experience persistent developmental changes to the brain and vocal tract. For this reason, we investigated whether the role of auditory feedback in sensorimotor learning differs across children and adult speakers. Participants produced vocalizations while they heard their vocal pitch predictably or unpredictably shifted downward one semitone. The participants' vocal pitches were measured at the beginning of each vocalization, before auditory feedback was available, to assess the extent to which the deviant auditory feedback modified subsequent speech motor commands. Sensorimotor learning was observed in both children and adults, with participants' initial vocal pitch increasing following trials where they were exposed to predictable, but not unpredictable, frequency-altered feedback. Participants' vocal pitch was also measured across each vocalization, to index the extent to which the deviant auditory feedback was used to modify ongoing vocalizations. While both children and adults were found to increase their vocal pitch following predictable and unpredictable changes to their auditory feedback, adults produced larger compensatory responses. The results of the current study demonstrate that both children and adults rapidly integrate information derived from their auditory feedback to modify subsequent speech motor commands. However, these results also demonstrate that children and adults differ in their ability to use auditory feedback to generate compensatory vocal responses during ongoing vocalizations. Since vocal variability also differed across the children and adult groups, these results also suggest that compensatory vocal responses to frequency altered feedback manipulations initiated at vocalization onset may be modulated by vocal variability.

Introduction

Speech is acquired early in childhood, with many children speaking fluently by the age of 3 (Ingram, 1989; Kuhl, 1994; Singleton & Ryan, 2004). However, the speech motor control system continues to develop for many years after the initial acquisition of speech. Throughout development the shape, size, and musculature supporting the articulators involved in speech production undergo dramatic changes, which influence the acoustic properties of the vocal tract (Guenther, 1994; Callan et al., 2000, Max et al., 2004). Despite these changes, the speech motor control system accomplishes the same functional goals, allowing children to maintain relatively fluent speech (Guenther, 1994). Investigation of individuals suffering from congenital deafness provides support for the notion that auditory feedback is crucial during the acquisition of speech (Oller & Eilers, 1988), as without prosthetic devices such as cochlear implants, these individuals often fail to acquire fluent speech (Svirsky et al., 2004). In addition to facilitating the acquisition of speech, auditory feedback may also function as an adaptive signal that guides the movements of the articulators in order to preserve speech fluency while the articulators are restructured during development (Guenther, 1994; Callan et al., 2000). Although auditory feedback plays an important role in the acquisition of fluent speech, this role may change once speech is acquired and individuals no longer experience persistent developmental changes to their brain and vocal tract.

Speech production is rapid, and the processing of sensory feedback is not instantaneous, thus a speech motor control system that relies solely on auditory feedback would result in dysfluent speech (Perkell et al., 1997; Max et al., 2004). For this reason, experimental evidence (Perkell et al., 1997; Scheerer & Jones, 2012; Chen et al., 2013), and neurocomputational models of speech motor control, such as the "directions into the velocities of articulators" model (DIVA; Max et al., 2004; Guenther, 2006; Tourville et al., 2008; Civier et al., 2010), suggest that fluent speech production relies on the combined effort of a feedback and a feedforward control system. While these researchers acknowledge the fundamental role of auditory feedback in mapping the relationship between the motor commands responsible for speech, and their sensory consequences, it has been suggested that as development progresses, the speech motor control system relies more heavily on feedforward commands. Under feedforward control, speech motor commands are executed by stored sensorimotor representations that encode the relationship between the motor commands responsible for speech and the sensory consequences of these commands. Since these feedforward commands are read out from memory, the delays associated with the processing of peripheral sensory feedback are avoided (Houde et al., 2002; Guenther et al., 2006; Scheerer & Jones 2012, 2014). Despite this transition to feedforward control, auditory feedback continues to play an important role in maintaining the precision and accuracy of the mapping of the sensorimotor representations that drive feedforward control (Perkell et al., 1997; Civier et al., 2010). While the role of auditory feedback in fluent speech production has been suggested to change over time, it is currently unclear how these changes influence the role of auditory feedback in sensorimotor learning.

Speech is often produced in dynamic and unpredictable environments, so even after the speech motor control system is fully developed, adult speakers must use sensory information to calibrate their speech motor commands. Since adult speech is thought to occur in a primarily feedforward manner, it is unclear if this decreased reliance on auditory feedback influences the rate of sensorimotor learning in adults. The role of auditory feedback in the sensorimotor learning of speech motor commands is often investigated using the frequency-altered-feedback (FAF) adaptation paradigm (Jones & Keough, 2008; Hawco & Jones, 2009, 2010; Keough & Jones, 2009, 2011; Keough, Hawco, & Jones, 2013; Scheerer & Jones, submitted; Scheerer, Tumber, & Jones, submitted). As part of this paradigm, participants are exposed to predictable changes to their auditory feedback (e.g., their vocal pitch is decreased), followed by a sudden removal of the manipulation (Jones & Keough, 2008; Hawco & Jones, 2009, 2010; Keough & Jones, 2009, 2011; Keough, Hawco, & Jones, 2013; Scheerer & Jones, submitted; Scheerer et al., submitted). Previous research has shown that individuals tend to compensate for the FAF manipulation, by shifting their voice in the opposite direction of the change (e.g., they increase their vocal pitch; Jones & Keough, 2008; Hawco & Jones, 2009, 2010; Keough & Jones, 2009, 2011; Keough, Hawco, & Jones, 2013; Scheerer & Jones, submitted; Scheerer et al., submitted). These compensatory responses can be used to index the extent to which auditory feedback is used to modify ongoing vocalizations. In addition, participants often show after-effects, that is their compensatory response persists for some time after the FAF manipulation is removed (Jones & Keough, 2008; Scheerer et al., 2015; Scheerer & Jones, In Prep). These after-effects suggest that the deviant auditory feedback was used to update the mapping of the speaker's sensorimotor representations, so that subsequent vocal productions more closely resembled their unaltered voice.

Jones and Keough (2008) used this FAF adaptation paradigm to investigate sensorimotor learning in trained singers and non-singers. During this study a group of singers, and a group of non-singers, produced vocalizations while exposed to a persistent -100 cent (one semitone decrease) manipulation of their fundamental frequency (F0; perceived as vocal pitch). The results of this study indicated that relative to singers, non-singers produced larger compensatory responses to the FAF. On the other hand, singers demonstrated larger after-effects, which were not observed in the non-singers. Based on this pattern of results, it was suggested that as a result of their extensive training, singers rely less on auditory feedback and more on their precisely tuned feedforward system. This decreased reliance on auditory feedback reduced the singers' susceptibility to the feedback manipulation, resulting in smaller compensatory responses. However, their increased reliance on feedforward control appears to have facilitated sensorimotor learning. Based on the results of this study, we hypothesized that as a result of differences in the weighting of feedback and feedforward control, children and adults may demonstrate different amounts of sensorimotor adaptation.

Previous research has demonstrated that children and adults, but not toddlers, show sensorimotor adaptation to changes in formant frequencies during vocalization (MacDonald et al., 2012). In addition, the results of that study indicated that toddlers, young children, and adults demonstrated differences in vocal variability, with variability decreasing with age (MacDonald et al., 2012). However, it is unclear if the same pattern would be observed in response to changes in F0. As demonstrated by Perkell and colleagues (1997) when investigating patients with cochlear implants, auditory feedback has two important roles. Auditory feedback is important for maintaining phonemic settings, which ensure phonemic distinctions are preserved, but it is also important for maintaining postural settings, which ensure that speech remains intelligible in changing acoustic environments. Formant frequencies, a property of auditory feedback related to phonemic control, were shown to be less affected by hearing loss in patients with cochlear implants, while F0, a property of auditory feedback related to postural control, changed rapidly following changes in hearing (Perkell et al., 1997). Since formant frequencies and F0 have been shown to be differentially influenced by hearing loss (Perkell et al., 1997) it is plausible that they also develop at different rates.

When considering other postural settings of speech such as vocal intensity and vocal timing, the results of studies investigating the control of these settings throughout development have been mixed. For example, when exposed to changes in sidetone amplification, younger children produced smaller changes in vocal intensity relative to older children (Siegel et al., 1976). Yet when exposed to increases in background noise, both children and adults demonstrated the ‘Lombard Effect’, increasing their vocal intensity in the presence of background noise (Siegel et al., 1976). Thus it is unclear if age influences the regulation of vocal intensity. Similarly, studies utilizing delayed auditory feedback (DAF) in participants of various ages have also demonstrated mixed results. Chase and colleagues (1961) reported that DAF impairs the speech of young children less than older children, while MacKay (1968) reported that young children are more sensitive to DAF than adults. Thus, when considering the developmental pattern of various speech parameters such as formant frequencies, vocal intensity, and vocal timing, there is no clear developmental trend across these parameters. For this reason, it is possible that each of these speech parameters are controlled by separate mechanisms, necessitating the specific investigation of F0 control throughout development.

One reason to suspect that F0 control develops earlier than other speech parameters such as formant frequencies is its importance for infant directed speech. Early in development caregivers use infant directed speech, characterized by exaggerated prosodic features such as F0, in order to communicate their intentions and affect (Fernald, 1993; Cooper & Aslin, 1994; Kuhl, 1994; Saint-Georges et al., 2013). Since changes in F0 provide an effective means for communication in pre-linguistic children, infant directed speech may promote the rapid development of F0 processing. On the other hand, since phonemic aspects of the speech signal, such as formant frequencies, are less beneficial for speech comprehension in pre-linguistic children (Saint-Georges et al., 2013), the processing of these features may develop more gradually. That being said, it is reasonable to expect that sensorimotor adaptation to changes in formant frequencies and F0 may be differentially influenced by age.

Although it is unclear if children and adults show the same pattern of sensorimotor adaptation in response to persistent changes in the F0 of their auditory feedback, Scheerer and colleagues (2013b) examined the vocal responses of children and adults between the ages of 4-30 years when exposed to brief unpredictable changes in the F0 of their auditory feedback. The results of that study indicated that vocal variability, rather than age, predicted the size of the compensatory response to the FAF. Since the children in that study were also found to be more variable than the adults (Scheerer et al., 2013b), these results suggest that sensorimotor adaptation may differ between children and adults, with vocal variability acting as a mediating factor.

In line with the findings reported by Scheerer and colleagues (2013b), previous research has also demonstrated that increased vocal variability leads to larger compensatory responses to brief unpredictable (Scheerer & Jones, 2012) changes to the F0 of adults' auditory feedback during ongoing vocalizations. Based on the results of these studies it has been suggested that increased stability of the speech motor control system results in increased feedforward control, and smaller compensatory responses to FAF. On the other hand, more variable speech motor control results in increased reliance on feedback control, making these individuals more susceptible to deviant auditory feedback, and thus resulting in larger compensatory responses to FAF. However, it is currently unclear if younger children, who are still experiencing ongoing changes to their speech motor control system, will demonstrate different rates of sensorimotor adaptation following persistent exposure to deviant auditory feedback. To investigate this research question, we exposed children and adults to both predictable and unpredictable changes to their auditory feedback. The predictable changes to the children and adults' auditory feedback allowed the extent to which auditory feedback is used for sensorimotor learning and motor planning, or feedforward control, to be assessed. On the other hand, the unpredictable auditory feedback manipulation allowed the extent to which auditory feedback is used for ongoing motor control, or feedback control, to be assessed. We expected that since children's speech motor control systems have yet to reach maturity, children would demonstrate an increased reliance on feedback control, manifesting in larger compensatory responses to the deviant auditory feedback following predictable and unpredictable auditory feedback manipulations, relative to adults. On the other hand, since adults are no longer experiencing persistent developmental changes, we expected that they would demonstrate an increased reliance on feedforward control, which would result in larger after-effects following exposure to the predictable FAF. Investigation of the pattern of sensorimotor adaptation following exposure to persistent changes to the F0 of children's and adults' auditory feedback will help to further our understanding of how the role of auditory feedback in speech motor control changes throughout the lifespan.

Methods

i. Participants

Twenty-five children between the ages of 3 and 8 years (M = 6.1, SD = 1.7; 6 females) and 25 young adults between the ages of 19 and 27 years (M = 21.8, SD = 2.4; 16 females) participated in this study. All participants were Canadian-English speakers who did not speak a tonal language, and had no formal vocal training. Participants received course credit, or financial compensation for participating in this study. Informed consent was obtained from all participants, as well as from a parent or guardian of participants under the age of 18 years. All procedures were approved by the Wilfrid Laurier Research Ethics Board and were in accordance with the World Medical Association 2013 Declaration of Helsinki.

ii. Procedure

Participants were seated in front of a computer monitor that displayed an interactive space game. Participants were instructed to start vocalizing an /a/ sound each time a red box with the word 'activate' appeared on the screen, and to stop vocalizing when the red box disappeared. This process was repeated until 100 vocalizations were completed.

Participants were asked to vocalize at a loud, but comfortable volume. Vocalizations were played back to the participants in real time via headphones. Since the child participants were quite young, sound level was not precisely monitored. However, the experimenter provided online feedback throughout the experiment if the participant was vocalizing too loudly or too softly to help regulate the participants' volume. This volume monitoring technique was deemed adequate as previous research has shown that the relative loudness of auditory feedback has no influence on the amplitude or latency of responses to FAF (Burnett et al., 1998). Participants were also told that during the experiment their voice might sound different, but they should ignore these differences and continue to produce their vocalizations at a consistent pitch.

The experiment consisted of four blocks of 25 trials, and lasted approximately 20 minutes. The first and third blocks were identical, and the second and fourth blocks were also identical (see Figure 9). During the first and third blocks, the first five vocalizations were left unaltered, while the last 20 vocalizations were pitch shifted downwards 100 cents (1 semitone) for the entire utterance. During the second and fourth blocks, the first five vocalizations were left unaltered, while half of the last 20 vocalizations were randomly shifted downwards 100 cents (1 semitone) for the entire utterance, and the other half were left unaltered.

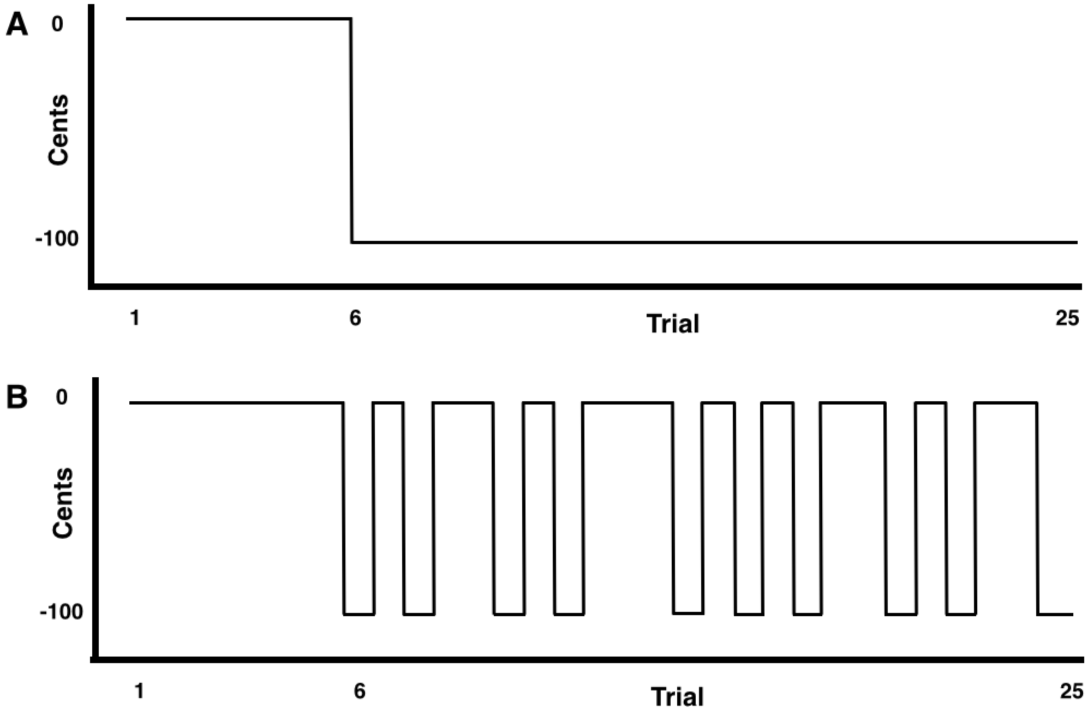

Figure 9: Each experimental block contained 25 trials. (A) Blocks 1 and 3 were identical. The first 5 vocalizations were left unaltered, while the last 20 vocalizations were pitch shifted downwards 100 cents (1 semitone) for the entire utterance. (B) Blocks 2 and 4 were also identical. The first 5 vocalizations were left unaltered, while half of the last 20 vocalizations were randomly shifted downwards 100 cents (1 semitone) for the entire utterance, and the other half were left unaltered.

iii. Apparatus

Participants were seated in a double-walled sound attenuated booth (Industrial Acoustics Company, Model 1601-01) and fitted with a headset-microphone (Sennheiser HMD 280-13 Pro, Sennheiser Electronics, Germany). Presentation of the visual stimuli was controlled by Max/MSP (v.5.0, Cycling ‘74). During the experiment vocalizations were sent to a mixer (Mackie Oynx 1640, Loud Technologies, Woodinville, WA), followed by a digital signal processor (VoiceOne, T.C. Hellicon, Victoria, BC), which allow the pitch of the participant’s voice to be altered. This process introduced approximately a 10 ms delay in the feedback signal, which was then presented back to the participant as auditory feedback. The onset and offset of the pitch alterations were controlled manually by the experimenter between the experimental phases. Both the altered and the unaltered voice signals were digitally recorded (TASCAM HD-P2, Montebello, CA) at a sampling rate of 44.01 Hz for later analysis.

iv. Data Analysis

The digital recording of the vocalizations was segmented into separate utterances and F0 values were calculated for each utterance using the SWIPE algorithm (Camacho and Harris, 2008). F0 values were normalized to the participants habitual F0, by converting Hertz values to cents using the following formula:

$$\boxed{\text{Cents} = 100\ (12 \log2\ F/B)}$$

In this formula, F is the F0 value in Hertz and B is the participant's habitual F0. Participants habitual F0 was estimated separately for each block, and for each measure (median 50 and median 1500) by calculating the average median 50 (the median F0 of the first 50 ms of the vocalization) and median 1500 (the median F0 of the first 1500 ms of the vocalization) of the participants' vocalizations using trials 2-5 of blocks 1 and 3 (the last four trials before the feedback manipulation was introduced). Only trials from block 1 and 3 were used to measure these habitual F0 measures, as these baseline trials were not immediately following the persistent F0 manipulation, making them less prone to contamination as a result of after-effects of the manipulations in the previous block. Similarly, the first trial of blocks 1 and 3 were not included in the baseline measure to allow for acclimation to the experimental setup and wash-out from the previous block, respectively.

For each vocalization the median F0 of the first 50 ms (the 'median 50' value) and median F0 of the first 1500 ms (the 'median 1500' value) were calculated. Since neural processing delays prevent the auditory feedback resulting from a vocal motor command from being processed for at least 100 ms (Hawco & Jones, 2009; Keough & Jones, 2009; Keough, Hawco, & Jones, 2013), the median 50 value provides an index of the F0 at which the vocalization was initiated, before auditory feedback was available. On the other hand, median 1500 values provide an index of the F0 of the vocalization once auditory feedback becomes available. In addition, in order to index the normal amount of variability in each participant's habitual F0, the standard deviation of the median 1500 of trial 5, the last trial before the feedback manipulations was introduced, was calculated and then averaged over these trials from each block (4 trials total).

To investigate changes in speech motor control following persistent and predictable changes in the pitch of one's auditory feedback, commonly referred to as sensorimotor adaptation, four groups of experimental trials were used for statistical analysis: baseline (the 2nd-5th trial of blocks 1 and 3), shift-start (the first five shifted trials of blocks 1 and 3), shift-end (the last five trials of blocks 1 and 3), and post-shift (the first 5 trials of blocks 2 and 4). The size of each participant's compensatory response at the end of the shift-phase was calculated by subtracting the average of the baseline trials from the average of the shift-end trials for median

1500. Lastly, to assess whether the compensatory behaviour persisted after the participant's auditory feedback was returned to normal, an after-effect value was calculated by subtracting the average of the baseline trials from the average of the post-shift trials for median 50. Following calculation of the compensatory response, it became clear that five children and five adults followed the direction of the manipulation, rather than compensating for the manipulation, as their compensation response was a negative value. Since the aim of this experiment was to understand the typical pattern of sensorimotor adaptation and compensatory responses across children and adults, these participants were excluded as their responses represented statistical outliers, and thus were atypical. For this reason, data from a total of 20 children and 20 young adults were subjected to statistical analyses.

To investigate changes in speech motor control as a result of exposure to unpredictable changes in the pitch of one's auditory feedback, two groups of experimental trials were utilized for statistical analysis: shifted trials (10 shifted trials from each of blocks 2 and 4), and unshifted trials (10 unshifted trials from each of blocks 2 and 4). As the first 5 trials of both blocks 2 and 4 were used to assess the after-effects resulting from the persistent exposure to FAF in blocks 1 and 3, these trials were obtained by taking the last 20 trials from each of blocks 1 and 3.

v. Statistical Analysis

Since the F0 of the baseline phase was used to normalize the F0 values for all trials, the mean and standard deviation of the baseline phase became zero. For this reason, to investigate sensorimotor adaptation, a series of one-sample t-tests were conducted to compare both median 50 and median 1500 values across the trial types (shift-start, shift-end, and post-shift) for both experimental groups (children and adults), relative to the baseline. In order to control for multiple comparisons, a Bonferroni adjusted alpha level of 0.0167 was utilized to assess statistical significance. Independent samples t-tests were also conducted to compare compensatory responses and after-effects across the children and adult groups.

To investigate compensatory responses to the unpredictable pitch shifts in block 2 and 4, a RM-ANOVA was performed to assess the influence of trial type (shifted and unshifted) and experimental group (children and adults) on median 50 and median 1500 values. In all instances where the sphericity assumption was violated, the Greenhouse-Geisser correction was applied (Greenhouse & Geisser, 1959). However, for ease of interpretation original degrees of freedom are reported.

Vocal variability differences were assessed by comparing baseline standard deviation across the children and adult groups using an independent samples t-test. Pearson correlations were also calculated to investigate the relationship between habitual vocal variability and both the size of the compensatory responses (median 1500) and the after-effects (median 50).

Results

i. Sensorimotor Adaptation (Blocks 1 and 3)

Median 50

One-sample t-tests were conducted on median 50 values to assess whether the children and adults' F0 values deviated from their habitual F0 throughout the sensorimotor adaptation blocks (see Figure 10). The children's median 50 values were significantly larger than their baseline F0 at shift-start, $t_{19} = 3.337$, $P = .003$, shift-end, $t_{19} = 3.360$, $P = .003$, and post-shift, $t_{19} = 2.758$, $P = .013$. Similarly, the adults' median 50 values were significantly larger than their baseline F0 at shift-start, $t_{19} = 3.522$, $P = .002$, shift-end, $t_{19} = 4.032$, $P = .001$, and post-shift, $t_{19} = 2.901$, $P = .009$. Independent samples t-tests revealed that the magnitude of the after-effects, $t_{38} = -.100$, $P = .921$, did not differ across the experimental groups.

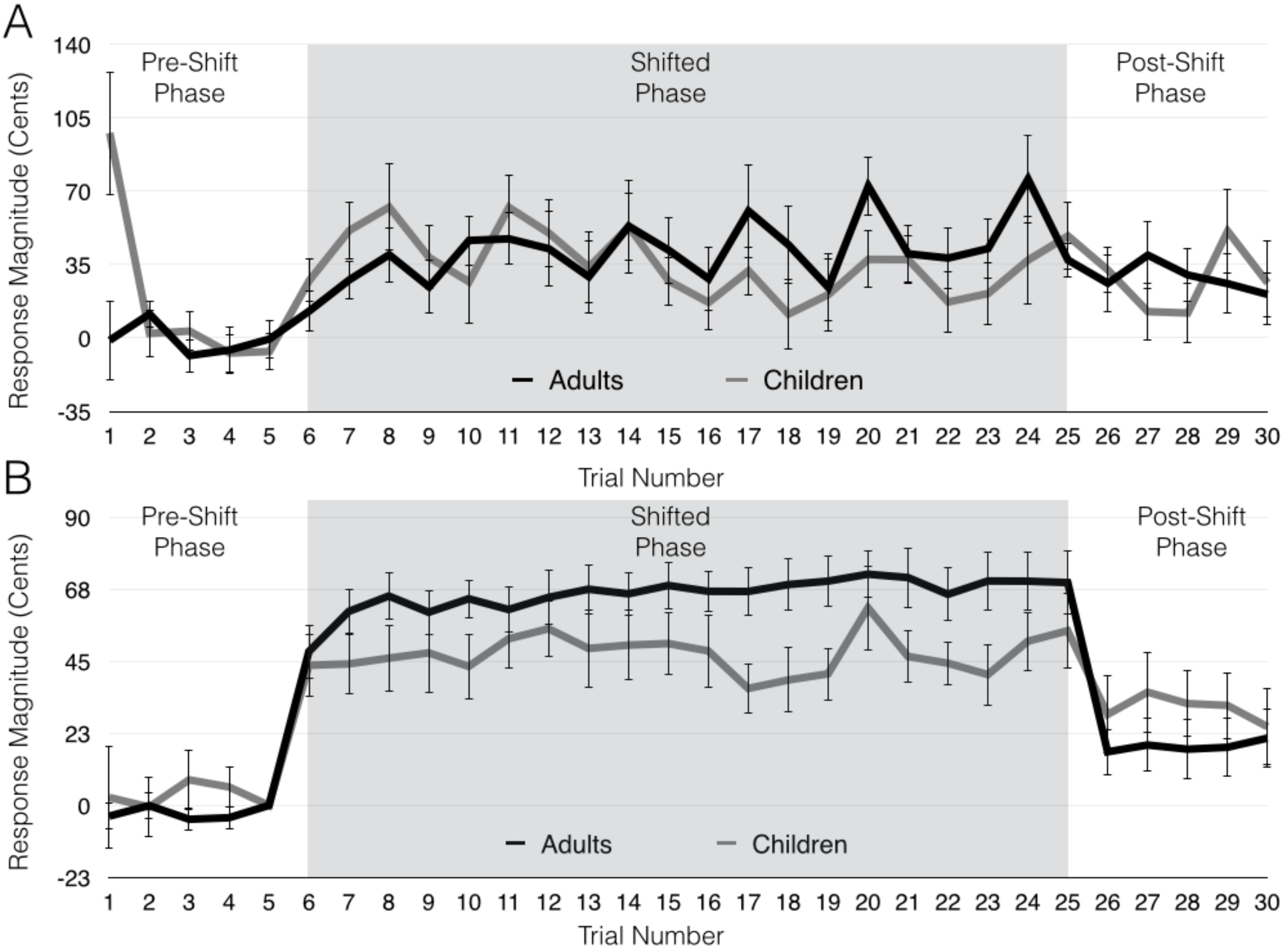

Figure 10: (A) Median 50 (top) and (B) Median 1500 (bottom) responses for each trial, averaged across block 1 and 3, for children (grey) and adults (black). The grey area represents the portion of the experiment where participants voice was shifted downwards 1 semitone (100 cents). Error bars represent the standard error of the mean.

Median 1500

One-sample t-tests were also conducted on median 1500 values to assess whether the children and adults' F0 values deviated from their habitual F0 throughout the sensorimotor adaptation blocks (see Figure 10). The children's median 1500 values were significantly larger than their baseline F0 at shift-start, $t_{19} = 3.895$, $P = .001$, shift-end, $t_{19} = 5.498$, $P < .001$, and post-shift, $t_{19} = 2.636$, $P = .0162$. Similarly, the adults' median 1500 values were significantly larger than their baseline F0 at shift-start, $t_{19} = 10.009$, $P < .001$, shift-end, $t_{19} = 8.789$, $P < .001$, and post-shift, $t_{19} = 2.867$, $P = .010$. An independent samples t-tests comparing the size of compensatory responses across children and adults was significant, $t_{38} = 2.366$, $P = .023$, as adults produced larger responses than children.

ii. Compensatory Responses (Blocks 2 and 4)

Median 50

A RM-ANOVA was conducted to investigate the effect of trial group (shifted and unshifted) and experimental group (children and adults) on median 50 values. The main effect of trial group, $F_{1, 38} = .549$, $P = .463$, the main effect of experimental group, $F_{1, 38} = .001$, $P = .971$, and the interaction between trial group and experimental group, $F_{1, 38} = .205$, $P = .205$, were not significant (see Figure 11).

Median 1500

The RM-ANOVA conducted to investigate the effect of trial group (shifted and unshifted) and experimental group (children and adults) on median 1500 values, revealed a main effect of trial group, $F_{1, 38} = 116.884$, $P < .001$, as overall median 1500 values were larger for shifted trials compared to unshifted trials (see Figure 3). The interaction between trial group and experimental group was marginally significant, $F_{1, 38} = 3.343$, $P = .075$, as both groups produced similar sized responses during the unshifted trials, while adults produced larger responses than the children during the shifted trials (see Figure 11). However, the main effect of experimental group, $F_{1, 38} = .613$, $P = .438$, was not significant.

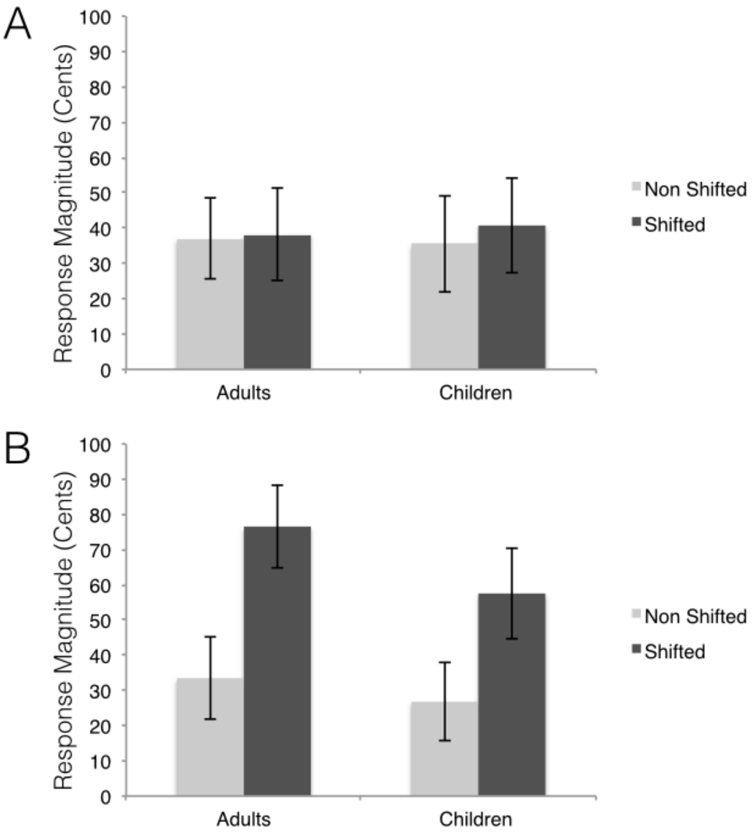

Figure 11: Median 50 (top) and Median 1500 (bottom) responses for the shifted and unshifted trials, averaged across block 2 and 4, for children (right) and adults (left).

iii. Vocal Variability

The independent samples t-test revealed that vocal variability differed across the experimental groups (children and adults), t_{38} = -2.064, *P* = .046, as the children (*M* = 41.21, *SD* = 13.42) were more variable than the adults (*M* = 30.93, *SD* = 17.79). Correlational analyses indicated that vocal variability did not correlate significantly with the size of compensatory responses in the sensorimotor adaptation blocks (median 1500: r_{38} = -.179, *P* = .268), the compensatory responses to the shifted trials in the unpredictable shift blocks (median 1500: r_{38}

=.029 *P* = .860), or the size of the after-effects on the sensorimotor adaptation blocks (median 50: r_{38} = -.202, *P* = .210).

Discussion

The aim of the current study was to investigate whether younger children who are still experiencing ongoing changes to their speech motor control system demonstrate different rates of sensorimotor adaptation following persistent exposure to deviant auditory feedback. During this study young children and adults were exposed to persistent predictable changes to their auditory feedback (blocks 1 and 3), as well as unpredictable changes to their auditory feedback (blocks 2 and 4). To assess sensorimotor learning the F0 of the first 50 ms of each participant's vocalizations, or the median 50, was measured. Since auditory feedback is delayed by cortical processing, the F0 at the beginning of a participant's vocalization was used to index feedforward control, or the extent to which deviant auditory feedback was incorporated into the planning of subsequent vocalizations (Keough & Jones, 2008, 2011; Hawco et al., 2009; Keough, Hawco, & Jones, 2013). On the other hand, by measuring participants' F0 across a larger portion of the vocalization, the median 1500, or the extent to which deviant auditory feedback was used to modify ongoing vocalizations was assessed (Keough & Jones, 2008, 2011; Hawco et al., 2009; Keough, Hawco, & Jones, 2013).

Sensorimotor learning was demonstrated in this study, as median 50 values were found to increase after persistent and predictable exposure to the FAF (blocks 1 and 3), but median 50 values were not found to differ when participants were exposed to the unpredictable FAF manipulation (blocks 2 and 4). More specifically, in the sensorimotor adaptation blocks, median 50 values were larger in the shift-start, shift-end, and post-shift phases, relative to the baseline phase. Changes in the median 50 values following predictable exposure to FAF indicate that the way in which speech motor commands were executed by the feedforward controller changed as a result of the exposure to FAF. Furthermore, the fact that median 50 values in the post-shift phase were larger than those recorded in the baseline phase, suggests that the effects of sensorimotor learning persisted even when the FAF manipulation was removed. On the other hand, the fact that median 50 values did not differ across the shifted and non-shifted trials in the unpredictable FAF manipulation blocks, demonstrates that it is the persistent and predictable nature of the FAF manipulation that drives sensorimotor learning.

Although sensorimotor learning was demonstrated across the experimental phases, sensorimotor adaptation was not found to vary across the two experimental groups. This result suggests that children and adults show similar rates of sensorimotor adaptation following persistent exposure to deviant auditory feedback. Also, since children and adults produced similarly sized after-effects, these results suggest that both children and adults used the deviant auditory feedback for motor planning. In line with the current findings, MacDonald and colleagues (2012) found that young children and adults, but not toddlers, demonstrated sensorimotor adaptation to changes in formant frequencies while vocalizing. Together these results suggest that once fluent speech is acquired, generally around the age of 3 (Kuhl, 1994), speakers are proficient at using sensory information to modify the planning of future motor commands. While we hypothesized that an increased reliance on auditory feedback for speech motor control in children may influence their rate of sensorimotor learning and thus their use of auditory feedback for motor planning, this is not the case. Neurocomputational models of speech motor control suggest that as speech development progresses children transition from a primarily feedback driven speech motor control system, to a primarily feedforward driven speech motor control system, but adequate feedforward control is necessary even at a young age to ensure that speech is produced fluently (Max et al., 2004; Guenther, 2006; Tourville et al., 2008; Civier et al., 2010). If the speech motor control system relied too heavily on feedback control, the time lags between the execution of the motor commands and the processing of their sensory consequences would result in an unstable speech motor control system. As a result of this instability speakers would produce disfluent speech (Max et. al, 2004). So although previous research has suggested that children have an increased reliance on auditory feedback (Scheerer et al., 2013b), if this is the case, it does not appear that increased weighting of the feedback control system influences sensorimotor learning.

In addition to investigating sensorimotor learning, the current study also investigated the role of auditory feedback for the regulation of the F0 of ongoing vocalizations in young children and adults. Following persistent exposure to FAF (blocks 1 and 3), both the young children and adults were found to increase their F0, as their median 1500 values were larger in the shift-start, shift-end, and post-shift phases, relative to the baseline phase. This increase in median 1500 across the experimental phases indicates that participants were increasing their F0 in order to compensate for the auditory feedback manipulation. Interestingly, adults were found to produce larger compensatory responses than young children, as there was a significant difference between young children's and adults' median 1500 values at shift-end. Similarly, following unpredictable exposure to FAF (blocks 2 and 4) both young children and adults were found to produce larger compensatory responses during shifted trials, relative to non-shifted trials. In addition, adults were found to produce larger compensatory responses during the shifted trials, relative to the young children. The finding that in both the predictable and unpredictable FAF conditions adults demonstrated larger changes in their median 1500 values was unexpected.

Previous studies have shown that when exposed to brief changes to the F0 of their auditory feedback, children and adults produce similar sized responses (Liu et al., 2013; Scheerer et al., 2013b). However, in the current study participants' vocalizations were manipulated at vocalization onset, with the manipulation persisting for the duration of the vocalization. The earlier studies that have compared the compensatory responses to FAF in children and adults used brief 200 ms perturbations, with the onset of the perturbation always occurring after the onset of the vocalization (Liu et al., 2013; Scheerer et al., 2013b). These methodological differences may explain the observed findings. Specifically, Hawco and Jones (2009) demonstrated that compensatory responses to FAF manipulations at voice onset are much larger than those introduced mid-utterance. For this reason, it was proposed that different control strategies are used at voice onset, relative to mid-utterance. Hawco and Jones suggested that at utterance onset auditory feedback is compared to the feedback predicted by the sensorimotor representation in order to ensure the intended F0 is produced. On the other hand, after utterance onset auditory feedback is compared to the F0 of the auditory feedback received prior to the manipulation, in attempt to stabilize the F0 of the vocalization (Hawco & Jones, 2009). In light of these previous findings, we suggest that although children and adults may not differ in their reflexive responses to brief perturbations in their auditory feedback (Liu et al., 2013; Scheerer et al., 2013b), adults may be more proficient at comparing incoming auditory feedback with the feedback predicted by their sensorimotor representations. We propose that this increased proficiency may be the result of adults possessing more precisely mapped sensorimotor representations. Since adults are no longer experiencing developmental changes to their speech motor control system, it is likely that their speech motor control system is more finely tuned, resulting in more consistent sensory predictions. On the other hand, since children are still experiencing developmental changes, it is likely that their speech motor control system is more plastic, resulting in more variable sensory predictions. For this reason, we suggest that adults may produce larger compensatory responses, as they are better at resolving the discrepancy between the deviant auditory feedback and their sensory prediction because their sensory prediction is more precise.

Another possible explanation for the differences in the size of the compensatory responses produced by the children and adults is attention. Scheerer and colleagues (submitted) reported that while participating in a FAF adaptation study, participants produced smaller compensatory responses when their attention was divided. For this reason, it was suggested that divided attention interferes with the use of auditory feedback for the regulation of ongoing speech. The current experiment utilized an interactive space game to prompt both children and adults to produce the required vocalizations. As it is likely that relative to the adults, the children were more captivated by the game, it is possible that less attentional resources were available for processing their auditory feedback, resulting in smaller compensatory responses to the deviant auditory feedback. However, it is important to note that Scheerer and colleagues (submitted) also found that divided attention diminished sensorimotor learning. Since compensatory responses, but not sensorimotor adaptation, differed across children and adults, it is likely that the observed differences in compensatory response magnitudes reflect differences in speech motor control across children and adults.

While adults produced larger compensatory responses to the FAF manipulations, children were found to produce more variable vocalizations. This finding is in line with previous studies that have reported that children demonstrate more vocal variability (MacDonald et al., 2012; Scheerer et al., 2013b). However, unlike previous studies that have reported a relationship between vocal variability and the size of compensatory responses to FAF (Scheerer & Jones, 2012; Scheerer et al., 2013b), vocal variability was not found to vary with the size of the compensatory responses produced in this experiment. Interestingly, although previous studies have shown that vocal variability correlates with the size of the compensatory responses produced following exposure to brief mid-utterance auditory feedback perturbations (Scheerer & Jones, 2012; Scheerer et al., 2013b), this finding has not be replicated in studies using FAF manipulations imposed at utterance onset. As previously discussed, compensatory responses to FAF manipulations imposed at vocal onset have been suggested to be planned by comparing the deviant auditory feedback to the feedback predicted by the speaker’s sensorimotor representation. Individuals who demonstrate greater vocal variability have been argued to have less precisely mapped sensorimotor representations (Scheerer & Jones, 2012; Scheerer et al., 2013b). Although the measures were not found to correlate, as a group children produced more variable vocal responses, but smaller compensatory responses. Perhaps then as a result of increased vocal variability, children have less precisely mapped sensorimotor representations, which decrease their proficiency in planning compensatory responses to FAF manipulations imposed at vocalization onset.

The aim of the current study was to investigate whether the role of auditory feedback for sensorimotor learning changes with development. The results of the current study demonstrate that children and adults can both rapidly integrate information derived from their auditory feedback to modify subsequent speech motor commands. Since even young children were able to extract information from their auditory feedback to modify future speech motor acts, these results provide support for the notion that auditory feedback functions as an adaptive signal that guides movements of the articulators in order to preserve speech fluency while the articulators are being restructured during development (Callan et al., 2000). However, these results also demonstrate that children and adults differ in their ability to use auditory feedback to generate compensatory vocal responses. Since vocal variability also differed across the children and adult groups, it is possible that the size of compensatory vocal responses to FAF manipulations initiated at vocalization onset may be related to vocal variability. In conclusion, although both children and adults were able to extract information from their auditory feedback to regulate their speech motor control systems, differences in the size of the compensatory responses produced by children and adults suggests that the ability to incorporate information derived from auditory feedback to modify ongoing vocalizations may change with age.

Study 4: The Relationship Between Vocal Accuracy and Variability to the Level of Compensation to Altered Auditory Feedback.

Published As:

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

Jones and Keough (2008) demonstrated that non-singers produced larger compensatory responses to FAF relative to trained singers. This finding suggests that superior vocal control may make individuals less susceptible to deviant auditory feedback. Since even without training the degree of vocal control varies naturally amongst non-singers, we aimed to investigate whether there is a relationship between vocal variability and the size of the compensatory response elicited by FAF. In order to examine this research question, we had non-trained singers produce vocalizations while we exposed them to unpredictable changes in their auditory feedback. It was expected that the results of this study would demonstrate that increased vocal variability is related to larger compensatory responses to FAF. If the expected pattern of results is obtained, it will suggest that vocal variability increases the weighting of the feedback control system, making these individuals more susceptible to deviant auditory feedback.

Abstract

Auditory feedback plays an important role in monitoring vocal output and determining when adjustments are necessary. In this study a group of untrained singers participated in a frequency altered feedback experiment to examine if accuracy at matching a note could predict the amount of compensation to auditory feedback that was shifted in frequency. Participants

were presented with a target note and instructed to match the note in pitch and duration. Following the onset of the participants' vocalizations their vocal pitch was shifted down one semitone at a random time during their utterance. This altered auditory feedback was instantaneously presented back to them through headphones. Results indicated that note matching accuracy did not correlate with compensation magnitude, however, a significant correlation was found between baseline variability and compensation magnitude. These results suggest that individuals with a more stable baseline fundamental frequency rely more on feedforward control mechanisms than individuals with more variable vocal productions. This increased weighting of feedforward control means they are less sensitive to mismatches between their intended vocal production and auditory feedback.

Introduction

The act of singing involves monitoring one's fundamental frequency (F0) in order to produce and maintain an intended vocal pitch (Elliott & Niemoeller, 1970). Particular pitch targets are produced and maintained through adjustments of the positioning of laryngeal structures, as well as changes in subglottal air pressure (Watts, Murphy, & Barnes-Burroughs, 2003). Several decades of research indicates that auditory feedback plays an important role in monitoring vocal output and determining when adjustments are necessary. For example, research has shown that when an individual's auditory feedback is altered it influences the quality of their vocal productions. When auditory feedback is delayed (Yates, 1963), masked (Bauer et al., 2006), or formant frequencies (Houde & Jordan, 1998; Purcell & Munhall, 2006b) or F0 are shifted (Elman, 1981; Burnett et al., 1998; Jones & Munhall, 2000), individuals tend to compensate for the manipulation by adjusting their vocal output. Despite this fact, the ability of singers to produce accurate pitches when reliable auditory feedback is masked, as in the case of individuals singing in a choir or in the presence of a band, suggests that auditory feedback is only partially responsible for monitoring voice F0.

According to the DIVA (directions into velocities of articulators) model, a neural network model capable of simulating speech, fluent speech is the result of a combination of feedback and feedforward control systems (Guenther, 2006). The feedforward control system relies on stored motor commands that contain detailed instructions on how to move the articulators, which are read out directly from memory. On the other hand, the feedback component relies on auditory feedback for detection and correction of production errors. Since strict feedforward control results in difficulties correcting for errors in auditory feedback, and strict feedback control introduces time lags and disfluencies of speech, the DIVA model relies on a combination of feedback and feedforward commands (Civier et al., 2010). Although the DIVA model has led to the successful simulation of speech, many questions regarding how auditory feedback contributes to successful speech production in humans remain unanswered.

In order to assess the importance of auditory feedback during vocal production, the frequency altered feedback (FAF) paradigm is often utilized. FAF has been consistently shown to elicit the pitch shift reflex (PSR) [e.g., Elman, 1981; Burnett et al., 1998; Jones & Munhall, 2000, 2002, 2005; Natke et al., 2003), which is the tendency for speakers to unconsciously respond to perturbations in their vocal output by shifting their F0 in the direction opposite of the perturbation (Burnett et al., 1998). Although this compensatory response has been reliably shown, it is a fraction of the size of the imposed perturbation, except when very small perturbations are used (Liu & Larson, 2007).

Although many FAF studies focus on perturbations to auditory feedback while speaking, results obtained in tasks requiring participants to sing have shown similar compensation patterns (e.g., Burnett et al., 1997; Natke et al., 2003). Interestingly, it has also been shown that perturbing the vocal output of trained singers leads to smaller compensatory responses than when testing untrained singers (Jones & Keough, 2008). This smaller response in singers may be the result of a decreased reliance on auditory feedback. It is suggested that auditory feedback is important when trying to produce a particular note, but over time as an individual has more experience producing the note, a sensorimotor representation encoding the relationship between the motor commands responsible for producing the note, and the sensory consequences of those motor commands is created (Keough & Jones, 2009). Subsequent attempts to produce the same note can then be executed by virtue of this established sensorimotor representation, independent of auditory feedback (Proteau, Marteniuk, Girouard, & Dugas, 1987; Proteau, Marteniuk, & Levesque, 1992; Keough & Jones, 2009).

Jones and Keough (2008) investigated whether or not trained singers rely more on sensorimotor representations than untrained singers while participating in a FAF study. They found that not only did singers compensate less when exposed to FAF, but singers also showed large aftereffects when their feedback was returned to normal. These results suggest that trained singers rely more on sensorimotor representations, that guide feedforward control, than untrained singers do.

Natke et al. (2003) investigated note matching accuracy and degree of compensation to FAF in German speaking participants. They found that when participants spoke a nonsense word and their voice was perturbed, there was no correlation between accuracy and compensation magnitude. On the other hand, when participants sang a nonsense word and their voice was perturbed, there was a significant inverse correlation between accuracy (measured as the absolute deviation from the target F0) and compensation magnitude. These results counter those of Jones and Keough (2008) who found that trained singers, who as a result of training or natural talent were more accurate than untrained singers, compensated less than the untrained singers in the study.

In order to further investigate the relationship between singing accuracy and compensation magnitude, we had a group of untrained singers participate in an FAF study and asked them to sing a target note. We aimed to determine if accuracy at matching a note could predict the degree of compensation to FAF in the auditory feedback of an untrained population. These results may help to further disentangle the relationship between singing ability and reliance on auditory feedback versus feedforward control.

Methods

i. Participants

Forty-nine female participants (aged 17–36, mean age 20.7, SD 3.1) were recruited for participation in this study. All participants reported no formal vocal training, did not speak a tonal language, and were right-handed. All participants received financial compensation for participation in this study. Informed consent was obtained from each participant, in accordance with the ethical policies at Wilfrid Laurier University.

ii. Apparatus

Participants were seated in an electrically shielded booth and fitted with Etymotic ER-3 insert headphones (Etymotic Research, Elk Grove Village, IL) and a headset microphone (Countryman Isomax E6 Omnidirectional Microphone). Presentation of the visual stimuli, target notes, as well as the shift onsets were controlled by Max/MSP 4 (Cycling ‘74, San Francisco, CA). The target notes were processed by combining three sine wave tones to create a harmonic tone using the program Praat (Boersma, 2001). The three target notes, 261.63 (C4), 277.18 (C#4), or 293.66 (D4) Hz, were chosen as they fall within the physiological frequency range of a healthy female’s voice (Moore, Keaton, & Watts, 2007). During the experiment vocalizations were sent to a mixer (Mackie Oynx 1220, Loud Technologies, Woodinville, WA), followed by a digital signal processor (DSP; VoiceOne, T.C. Hellicon, Victoria, BC), which shifted the pitch of the participant's voice. This pitch-shifted vocalization was then presented back to the participant as auditory feedback. The unaltered voice signal and the tone that was triggered by the command to the DSP were digitally recorded (TASCAM HD-P2, Montebello, CA) at a sampling rate of 44.1 Hz for later analysis.

iii. Procedure

Participants were seated in front of a computer screen displaying a small box in the centre of the screen that alternated in colour from red to green. When the box was red participants were presented with a target tone (C4, C#4, or D4) for 2000 ms. Following the presentation of the target tone, the participants were told that the box would turn green, and this was their cue to begin matching the target tone in pitch and duration by vocalizing the vowel sound /a/. Participants were instructed to vocalize at a loud, but comfortable amplitude. Vocalizations were played back to the participants in real time via headphones. Participants were told that during the experiment their feedback might be altered, but they should ignore any alterations and continue trying to match the target note in pitch and duration.

The experiment consisted of one block of 150 trials, and lasted approximately 20 min. Each target note was presented a total of 50 times, and the order of the note presentation was varied pseudorandomly throughout the experiment. During each trial the pitch of the participant's voice was shifted downwards 100 cents (1 semitone). The pitch shift occurred at a random time between 500 ms and 1500 ms after utterance onset and lasted for the duration of the utterance.

iv. Data Analysis

The digital recording of the vocalizations was segmented into separate utterances and F0 values were calculated for each utterance using the SWIPE algorithm (Camacho & Harris, 2008). F0 values were normalized to the target tone (C4, C#4, or D4) by converting Hz values to cents using the following formula:

$$\boxed{\text{Cents} = 100\ (12 \log_2 F/B)}$$

In the formula, F is the F0 value in Hz and B is frequency of the target tone participants were to vocalize (261.63, 277.18, or 293.66).

Cents values were calculated for 200 ms before the pitch shift (the baseline period), and 500 ms after the pitch shift. An average F0 trace was constructed for each of the three notes for each participant. For each average F0 trace the mean baseline value was calculated, or the average of the first 200 ms of the vocalization. Since cents values were normalized to the target tone, a cents value of 0 indicated a perfect match to the target tone. In order to index the normal amount of variability in each participant’s vocalizations, the standard deviation of the baseline period was also calculated. The amplitude of the compensation response was also determined by finding the point at which the participant’s average F0 trace deviated maximally from the baseline mean. Participants who followed the direction of the FAF manipulation rather than compensating for it (n = 10) were excluded from the analyses. Thus in total 39 subjects were included in the statistical analyses.

A repeated measures analysis of variance was performed for each of the three measures: accuracy, compensation magnitude, and baseline variability, in order to determine if responses differed significantly across notes. In instances where the assumption of sphericity was violated, the Greenhouse–Geisser correction was applied and adjusted values are reported

(Greenhouse & Geisser, 1959). Pearson correlations were calculated to investigate the relationship between the accuracy at matching the target notes and the degree of compensation to the FAF. Pearson correlations were also calculated to investigate the relationship between the variability of the participants' baseline periods and the magnitude of compensation to the FAF.

Results

In order to determine if accuracy at matching a note could predict the degree of compensation to the FAF, the absolute deviation from the target F0 and the magnitude of compensation were calculated for each participant for each note. Since absolute deviation from the target F0, $F_{1.49,56.50}=2.178$, $P=0.135$, and the magnitude of compensation, $F_{1.39,52.74}=1.135$, $P=0.311$, did not significantly differ as a function of note, values were collapsed across all three notes for the correlation analysis. No significant correlation was found, $r = -0.132$, $P_{(2\text{-tailed})} = 0.424$, $n = 39$) between degree of compensation and absolute deviation from the target F0. Magnitude of compensation and absolute deviation from the target F0 are plotted against each other in Figure 12.

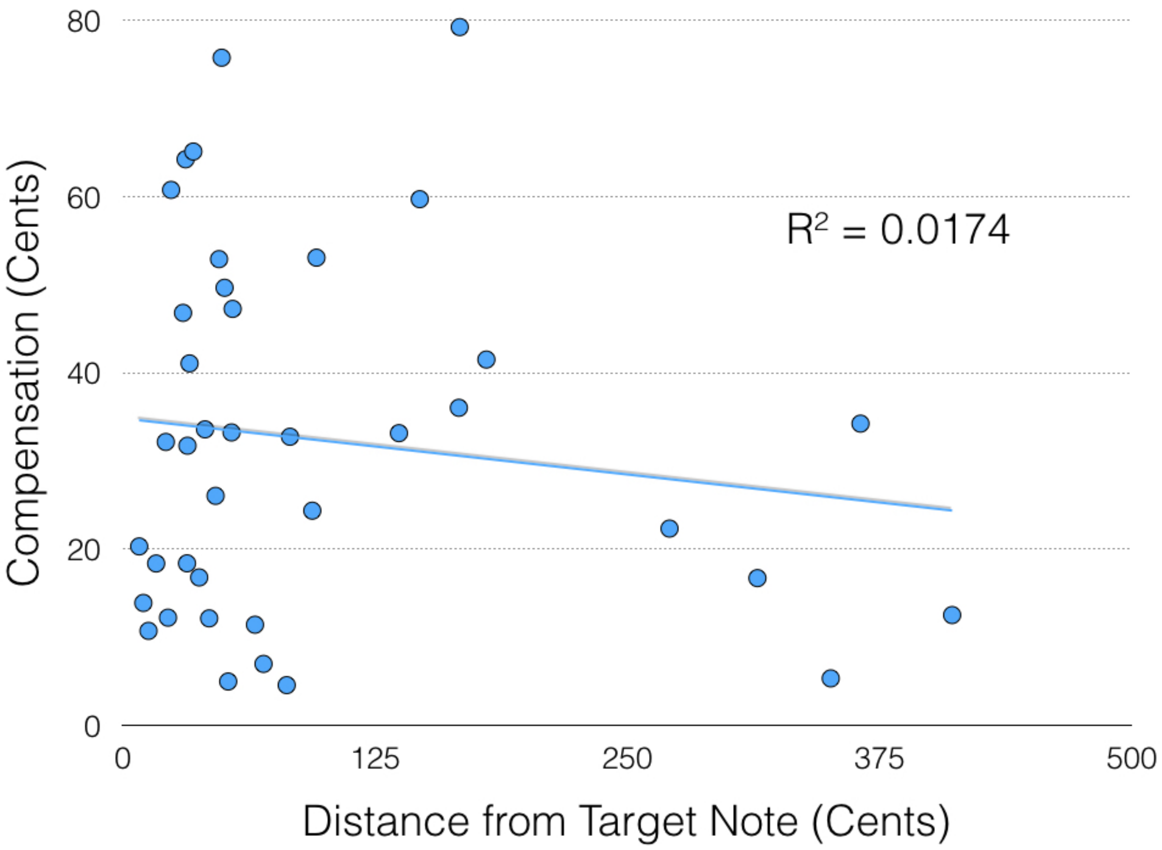

Figure 12: Relationship between Accuracy and Compensation Individual's absolute deviation from the target F0 (an index of their target note matching ability) is plotted against their magnitude of compensation to the FAF.

In addition, we investigated whether the variability of the participants' baseline periods could be indicative of the magnitude of their compensation to the FAF. The variability of the participants' baseline periods were found to differ significantly as a function of note, $F_{2,76} = 11.014$, $P < 0.001$, thus correlational analyses were calculated separately for each note. Significant correlations were found between the standard deviation of the participants' baseline periods, and their magnitude of compensation to the FAF for each note (C: $r = 0.611$, $P_{(2\text{-tailed})} < 0.001$, $n = 39$; C#: $r = 0.406$, $P_{(2\text{-tailed})} = 0.010$, $n = 39$; D: $r = 0.637$, $P_{(2\text{-tailed})} < 0.001$, $n = 39$).

The magnitude of compensation and baseline period standard deviations for each note are plotted against each other in Figure 13.

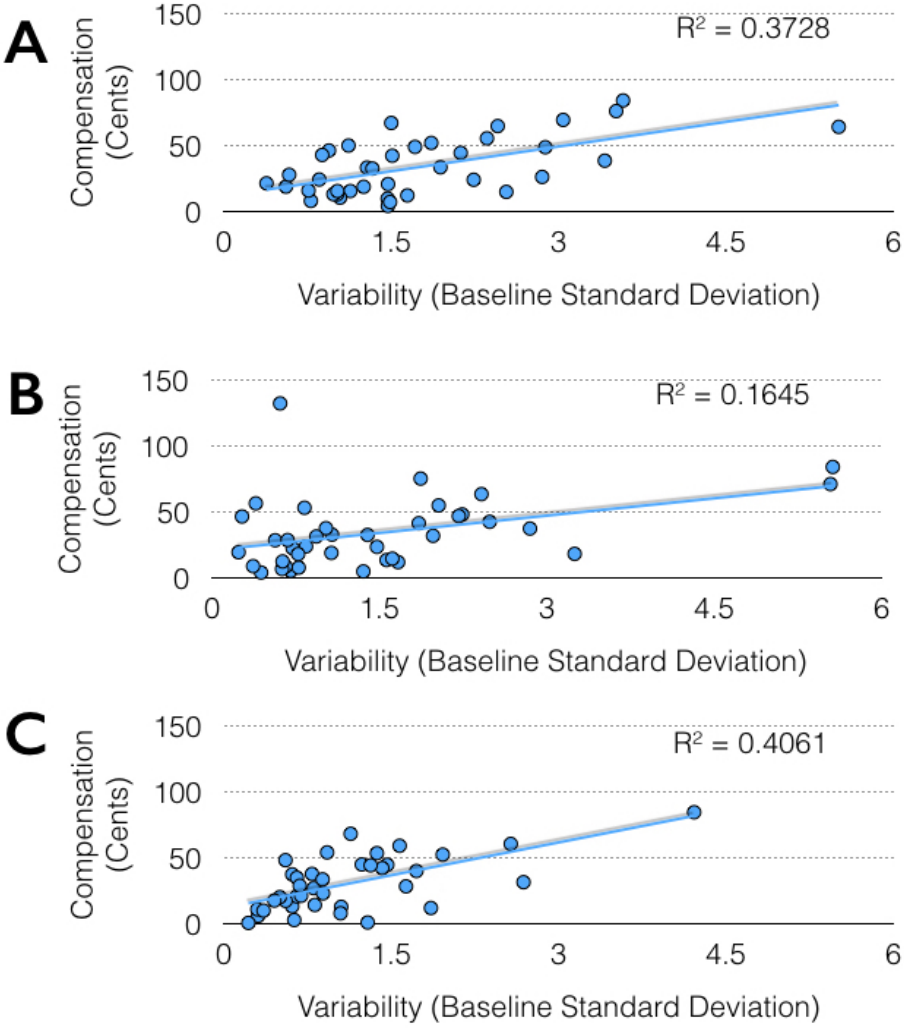

Figure 13: Relationship between Baseline Standard Deviation and Compensation Individual's baseline period standard deviation (an index of their F0 control) is plotted against their magnitude of compensation to the FAF for all three notes (a = C, b = C#, c=D).

Discussion

In this study, individuals' ability to match a target note, as well as their magnitude of compensation to online pitch shifts were examined in order to investigate if accuracy at matching a target note could predict the magnitude of compensation to a pitch shift. It was found that the ability to match a target note did not correlate with the magnitude of compensation to feedback pitch shifts, thus accuracy was not predictive of compensation magnitude. Although accuracy did not predict compensation, the variability of the participants' baseline periods were found to correlate significantly with their magnitude of compensation to the pitch shift. A highly variable baseline F0 likely reflects poor F0 control.

Speech motor commands are refined and F0 stability improves as F0 control transitions from feedback to feedforward control (Schmidt & Lee, 2005; Civier et al., 2010). The transition from feedback to feedforward control is driven by the stability of the sensorimotor representation, or the mapping between motor commands and the resultant sensory consequences. As the sensorimotor representation becomes more reliable and consistently allows for the execution of the correct motor command required to produce the intended sound, auditory feedback becomes redundant. This is because auditory feedback is no longer necessary to refine the mapping between motor commands and their resultant sensory consequences. As a result of this decreased reliance on auditory feedback, feedforward control becomes predominant. Despite this transition, individuals can never rely completely on a feedforward control system, as this would prevent detection of errors in auditory feedback. Civier and colleagues (2010) suggest that successful execution of motor commands for speech production involves a weighting of feedback and feedforward control systems. We suggest that when F0 control is poor, there is an increased reliance, or 'weighting,' of feedback control, relative to feedforward control. This suggests that individuals with more variability in their baseline periods have less consistent vocal output, thus they weight their auditory feedback more heavily to ensure production errors are detected, and their sensorimotor representations are updated. On the other hand, participants relying more on sensorimotor representations, or 'weighting' the feedforward system more heavily, have more consistent vocal output. The higher weighting these individuals place on their sensorimotor representation causes them to 'ignore' sounds that are in violation of their sensorimotor representation.

This weighting of feedback control vs. feedforward control has direct implications for the amount of compensation to FAF. When there is an increased weighting of auditory feedback, or feedback control, individuals are more susceptible to FAF, because they constantly monitor and adjust their vocal output in order to keep their auditory feedback in line with their perception of the target note. Thus when a pitch shift occurs, it results in a large compensatory response as the participant uses this feedback to guide their vocal output back in line with the target note. On the other hand, when there is an increased weighting of sensorimotor representations, or feedforward control, deviant auditory feedback is more readily ignored, and thus participants are less susceptible to changes in their auditory feedback. For this reason, when their vocal output is pitch shifted they will show less of a compensatory response.

The fact that we did not find a correlation between the magnitude of compensation and the ability to match a target note is in contrast with the results found by Natke and colleagues (2003), who found that the degree of compensation to FAF correlated inversely with accuracy at matching a target note. Although the results of this current study are in contrast to those of Natke and colleagues (2003), it is likely due to differences in experimental design, as the current study differs from their study in three fundamental ways, including number of trials, type of pitch shift, and target stimuli.

In the study by Natke and colleagues (2003), participants were only exposed to six FAF trials, whereas in the current study participants performed vocalizations during 150 FAF trials. The small number of trials used to produce individual averages in the Natke and colleagues

(2003) study, relative to the current study, would suggest that this current data is a more reliable representation of the individuals' performances. Furthermore, to eliminate the possibility that the lack of effect found in this study is the result of habituation over the large number of trials, we also examined the data for the first 6 trials of this study. Consistent with the previously reported results over 150 trials, a non-significant correlation was found between compensation to the pitch shift and accuracy at matching the target note over the first 6 trials ($r = -0.279$, $P_{(2\text{-tailed})} = 0.052$, $n = 49$). The consistency of the results when examining either 6 trials or 150 trials, provides further support for the notion that there is no significant relationship between compensation to a pitch shift and accuracy at matching a target note.

Examining further differences between the current study and the study by Natke and colleagues (2003) it was found that Natke and colleagues utilized whole utterance pitch shifts, whereas in the current study participants experienced mid-utterance pitch shifts. In a study by Hawco and Jones (2009), they found that compensation responses to mid-utterance pitch shifts were consistently smaller than compensation responses to shifts occurring at utterance onset. Furthermore, they concluded that different control strategies are utilized prior to vocalization onset, compared to within an ongoing vocalization (Hawco & Jones, 2009). For this reason, it may not be surprising that these different FAF techniques elicited different responses when comparing the current study to that of Natke and colleagues (2003).

Lastly, in the current study participants were required to match three different notes that were pseudorandomly presented throughout the experiment. It was found that participants' baseline variability differed significantly as a function of the note being matched. This result suggests that F0 control is differentially affected by different target notes. In the Natke and colleagues (2003) study, the participants were matching the same note for all six trials. This difference in design may have further magnified the discrepancy in results found in these two studies. A study by Liu and Larson (2007) found that participants compensate more to difficult (high) target pitches, compared to easy target pitches. The results from Liu and Larson (2007) would suggest that the compensation magnitudes calculated in this current study, as well as the study by Natke and colleagues (2003) are modulated by the proximity of the target note to the participants' baseline F0. In the Natke and colleagues (2003) study, the fact that all participants were matching the same target note may indicate that the inverse correlation found is partially related to closeness of the target note to the participants' baseline F0s. Participants may have compensated more to the FAF simply because the target was farther away from their baseline F0, irrespective of their ability to match the target note. On the other hand, in this current study participants matched three target notes. Since these three notes were chosen in a range that corresponds to the average F0 for women, it is likely that one of the three notes was close to the baseline F0 of each participant, while the other two notes would be further away from their preferred F0. This combination of easy and difficult target notes for each participant, relative to their preferred baseline F0, should cancel out any advantage to a single participant due to the proximity of the target note to their own baseline F0.

It is important to note that in the current study only female participants were recruited. While gender related differences in vocal responses to FAF have not been previously identified (Scheerer et al., 2013b), future research should also investigate whether this pattern is maintained with male participants.

Conclusion

Previous research has shown that singers compensate less to pitch shifts in their feedback, relative to untrained singers during whole utterance shifts (Jones & Keough, 2008). For this reason, it was expected that accuracy at matching a target note would be related to smaller compensatory responses. Instead, the variability of the participants' baseline periods was more predictive of the degree of compensation to FAF. In order to see if this finding holds true, a future study looking at the adaptive behaviours of these two groups would help to validate this finding. If individuals with stable baseline periods also display robust aftereffects as seen in the group of singers participating in the Jones and Keough (2008) adaptation study, it would lend support to the idea that individuals with more stable F0s rely more on sensorimotor representations. Similarly, if individuals with more variable baseline periods show greater compensation and fewer aftereffects, then it is likely that they rely more on auditory feedback, as seen in the group of untrained singers in the Jones and Keough (2008) adaptation study.

In closing, the results of this study suggest that the ability to match a target note is not predictive of the magnitude of compensation to frequency altered feedback. Instead, F0 control, as indexed by the variability of participants' baseline periods, is a better indicator of the degree of compensation to FAF.

Study 5 - The Relationship between Vocal Variability and Sensorimotor Learning

Scheerer, N.E., & Jones, J.A. (In Prep). The relationship between vocal variability and sensorimotor learning.

Jones and Keough (2008) demonstrated that trained singers produced smaller compensatory responses, but larger after-effects following repeated exposure to predictable changes in their auditory feedback, relative to non-singers. In a follow-up study, Keough and Jones (2009) found that when exposed to predictable changes in their auditory feedback, singers started subsequent utterances at a similar F0 as the previous trial, suggesting they were updating their sensorimotor representations, while non-singers appeared to 'search' for the target F0. Together, these results suggest that superior vocal control influences the plasticity of one's sensorimotor representations. Furthermore it suggests that although vocally stable individuals may be less susceptible to deviant auditory feedback, the deviant feedback is still used to modify their sensorimotor representations. The aim of this experiment was to identify whether non-singers who demonstrate superior speech motor control would also demonstrate increased plasticity of their sensorimotor representations, relative to more variable non-singers.

To investigate this research question we had non-singers produce vocalizations as part of a FAF adaptation paradigm. It was expected that individuals who demonstrated increased vocal variability would also produce larger compensatory responses, but smaller after-effects. On the other hand, it was expected that individuals who demonstrated relatively less vocal variability would produce smaller compensatory responses, but larger after effects. If the expected pattern of results are obtained, it will suggest that vocal stability results in increased weighting of the feedforward control system.

Abstract

Speech motor control is regulated by auditory feedback driven closed-loop control, and sensorimotor representation driven open-loop control. However, the factors that dictate the relative weighting of these two control systems are unclear. This study aimed to investigate the relationship between vocal variability and the relative weighting of feedback versus feedforward control. In this study, participants produced vocalizations while they heard their vocal pitch persistently shifted downward a semi-tone. Participants' vocal pitch was measured across each vocalization, to index the extent to which ongoing vocalizations were modified by deviant auditory feedback. The size of participants' compensatory responses to the pitch manipulation correlated with the variability of participants' habitual fundamental frequency (F0), with more variable participants producing larger compensatory responses. This suggests that increased vocal variability leads to increased weighting of the feedback control system. Participants' vocal pitch was also measured at the beginning of each vocalization, before auditory feedback was available, to assess the extent to which the deviant auditory feedback modified subsequent

speech motor commands. Changes in vocal pitch at vocalization onset were observed in response to the feedback manipulation. These changes were also found to persist once the manipulation was removed, and these after-effects demonstrate that sensorimotor learning occurred. Importantly, the size of these after-effects correlated with the variability of participants' habitual F0, with less variable participants producing larger after-effects. This suggests that decreased vocal variability leads to increased feedforward control. Together the results of this study suggest that vocal variability can predict the weighting of feedback versus feedforward control.

Introduction

Speech production is a highly complex motor behaviour that involves the precise coordination of a complex network of muscles and articulators. In order to produce fluent speech, speakers rely on sensory feedback, such as auditory feedback, to monitor and correct for speech production errors (Burnett et al., 1997, 1998; Liu et al., 2011; Scheerer & Jones, 2012; Scheerer et al., 2013a,b, 2014). The importance of auditory feedback for the acquisition and maintenance of fluent speech becomes evident when observing children who are born deaf and fail to acquire fluent speech (Oller & Eilers, 1988), and adults who are deafened postlingually and experience a progressive decline in the quality of their speech (Cowie & Douglas-Cowie, 1992), respectively. However, since individuals can also maintain fluent speech when auditory feedback is masked or unavailable (Lane & Tranel, 1971), for instance when speaking in a noisy room, speech fluency must not be regulated solely by auditory feedback.

Neurocomputational models of speech motor control, such as the directions into velocities of articulators (DIVA; Guenther, 2006; Tourville et al., 2008; Civier et al., 2010) and state-feedback control (SFC; Hickok et al., 2011) models, as well as experimental evidence (Jones & Munhall, 2005; Jones & Keough, 2008; Hawco & Jones, 2009; Keough & Jones, 2009, 2011; Scheerer & Jones, 2012, 2014; Chen et al., 2013), suggest that fluent speech production relies on the combined effort of a feedback and a feedforward control system. The feedback control system relies on sensory feedback, such as auditory and somatosensory feedback. In contrast, the feedforward control system relies on sensorimotor representations, which encode the relationship between the motor commands responsible for producing speech, and the sensory consequences of these motor commands (Guenther, 2006; Civier et al., 2010). Early in speech acquisition, the feedback control system is important for exploring auditory space, and learning associations between motor commands and their auditory consequences (Bailly, 1997; Callan et al., 2000; Jones & Munhall, 2000, 2002; Guenther, 2006; Civier et al., 2010). However, since feedback control operates in a closed-loop manner, relying solely on feedback control is implausible, as the delays associated with the processing of sensory information would result in disfluent speech (Houde et al., 2002; Guenther, 2006; Civier et al., 2010; Perkell, 2012). For this reason, speech motor control also relies on feedforward control. Since the feedforward control system operates in an open-loop manner, speech is executed by virtue of stored motor commands, which prevents the delays associated with the processing of peripheral sensory feedback. This being said, strict reliance on feedforward control is also unfeasible, as sensory information is important for maintaining the precision and accuracy of the mapping of the sensorimotor representations that drive feedforward control (Civier et al., 2010). While it is clear that both feedback and feedforward control contribute to fluent speech production, the factors that dictate the relative weighting of these systems remain unclear.

The importance of auditory feedback for speech motor control is often investigated using the frequency-altered feedback (FAF) paradigm. As part of this paradigm, researchers synthetically alter the fundamental frequency (F0; or pitch) of participants' vocalizations and instantaneously present the altered auditory feedback to the participants through headphones (Elman, 1981; Burnett et al., 1997). When exposed to this FAF, participants tend to compensate by shifting the F0 of their voice in the opposite direction of the manipulation. This compensatory response is often only a fraction of the size of the manipulation (Burnett et al., 1997, 1998; Liu et al., 2011; Korzyukov et al., 2012b; Scheerer et al., 2013a,b), which highlights the fact that auditory feedback is not the sole input to the speech motor control system. Importantly, by measuring responses to FAF, it becomes possible to isolate the factors that dictate the relative contribution of auditory feedback for speech motor control.

One factor that has been suggested to influence the contribution of auditory feedback for speech motor control is vocal variability. Scheerer and Jones (2013b) exposed children and adults aged 4-30 years old to brief FAF perturbations. The results of this study indicated that vocal variability, rather than age, predicted the size of the compensatory responses to the FAF perturbations. Similarly, Scheerer and Jones (2012) found that within a group of young adults, vocal variability also predicted the size of the compensatory responses elicited by mid-utterance FAF manipulations. Together, these results suggest that individuals who produce highly variable speech tend to rely more heavily on their auditory feedback. This increased reliance on auditory feedback ensures speech is produced fluently, and also provides information to calibrate one's sensorimotor representations. However, this increased reliance on feedback control also increases one's susceptibility to FAF manipulations and results in larger compensatory vocal responses. On the other hand, individuals with more stable vocal productions, likely have sensorimotor representations that are better at predicting the sensory consequences of their associated motor movements. When sensorimotor representations are precisely predicting the sensory consequences of upcoming movements, the information provided by auditory feedback becomes redundant. For this reason, individuals with increased vocal stability increase the weighting of their feedforward control system, which makes them less susceptible to FAF manipulations, and results in smaller compensatory vocal responses (Scheerer & Jones, 2012).

One factor that has been suggested to influence vocal stability is vocal training. Jones and Keough (2008) exposed trained singers and non-singers to predictable changes in their auditory feedback. The results of this study indicated that the non-singers were faster to compensate, and produced larger responses, to changes in their auditory feedback. On the other hand, singers responded more gradually to changes in their auditory feedback, but the compensatory response persisted after the FAF manipulation was removed, an effect that was not observed in the non-singers. Based on the results of this study, it was suggested that relative to singers, non-singers have an increased reliance on auditory feedback for regulating their speech motor control. On the other hand, as a result of extensive training, singers rely less on auditory feedback and more on their precisely tuned feedforward system. Interestingly, although singers produced smaller compensatory responses, the larger after-effects witnessed for singers, once their auditory feedback was returned to normal, suggest that an increased reliance on feedforward control is also related to increased sensorimotor learning. In other words, although the singers were not producing large compensatory responses to the FAF, it appears as though they were using information from the deviant auditory feedback to update the mapping of their sensorimotor representations. As a result of this remapping, their vocal productions were different than those produced prior to the exposure to the predictable FAF, even when the FAF manipulation was removed. Based on the results of this study, it appears that vocal proficiency, achieved through extensive vocal training, may increase the weighting of the feedforward control system.

Previous studies have demonstrated that increased vocal variability leads to larger compensatory responses to FAF (Scheerer & Jones, 2012; Scheerer et al., 2013b), which suggests that increased vocal variability results in increased weighting of the feedback control system. On the other hand, trained singers who have more precise speech motor commands have been shown to produce smaller compensatory responses to FAF, and have faster rates of sensorimotor learning (Jones & Keough, 2008), which suggests that vocal proficiency results in increased weighting of the feedforward control system. However, it is currently unclear if this finding can be observed in individuals who have not received formal vocal training. For this reason, we sought to determine whether vocal variability predicts the degree of sensorimotor learning that occurs when individuals are exposed to predictable FAF. In order to investigate this research question, we exposed non-singers to predictable changes in their auditory feedback and evaluated their compensatory responses, as well as any after-effects, or persistence of the compensatory response once the FAF manipulation was removed. Based on the results of previous studies, we predicted that individuals who have a high degree of vocal variability would produce large compensatory responses to the FAF, but these responses would not persist when the FAF manipulation was removed. On the other hand, we predicted that individuals who produce more stable vocal productions would produce smaller compensatory responses to the FAF, but these smaller responses would persist for a period of time after the FAF manipulation was removed. If, as expected, compensatory responses and after-effects vary as a function of vocal variability, these results would suggest that the weighting of the feedback and feedforward control systems is influenced by vocal variability.

Methods

i. Participants

Forty-eight participants between the ages of 17 and 30 years (*M* = 20.20, *SD* = 2.57; 28 females) participated in this study. All participants were Canadian-English speakers who did not speak a tonal language, and had no formal vocal training. Participants received course credit, or financial compensation for participating in this study. All procedures were approved by the Wilfrid Laurier Research Ethics Board and were in accordance with the World Medical Association 2013 Declaration of Helsinki.

ii. Procedure

Participants were seated in front of a computer monitor that displayed a Microsoft Powerpoint presentation composed of blank slides that were solid green and solid red in colour. Each time the display was red, participants were presented with a note (G3 for men, G4 for women) for 2000 ms. Following the presentation of the note, participant were told that the screen would turn green for 3000 ms, and this was their cue to start vocalizing an /a/ sound. Participants were instructed to vocalize at a loud, but comfortable amplitude. Also, participants were instructed to use the note as a reference to help them vocalize at a consistent pitch throughout the experiment, thus consistency was emphasized over accuracy. Participants produced 140 vocalizations in total, and these vocalizations were played back to the participants in real time via headphones. Participants were also told that during the experiment their auditory feedback might be altered, but they should ignore any alterations and continue producing their vocalizations at a consistent pitch.

The experiment consisted of one block of 140 trials, and lasted approximately 20 minutes. The block was divided into three phases: the pre-shift phase (trials 1-20), the shifted phase (trials 21-120) and the post-shift phase (trials 121-140). During the shifted phase, the pitch of the participants' auditory feedback was shifted downwards 100 cents (1 semitone) for the entire utterance. During the pre-shift and post-shift phases, participants' auditory feedback was unaltered.

iii. Apparatus

Participants were seated in a double-walled sound attenuated booth (Industrial Acoustics Company, Model 1601-01) and fitted with a headset-microphone (Sennheiser HMD 280-13 Pro, Sennheiser Electronics, Germany). Both the visual stimuli as well as the notes were presented using Microsoft Powerpoint. The two notes were generated by combining three sine wave tones to create a harmonic tone using the program Praat (Boersma, 2001). The two target notes G3 (196 Hz) and G4 (392 Hz) were selected as they are typically within the vocal range of healthy men and women, respectively (Fitch & Holbrook, 1970; Moore et al., 2007). During the experiment vocalizations were sent to a mixer (Mackie Oynx 1640, Loud Technologies, Woodinville, WA), followed by a digital signal processor (VoiceOne, T.C. Hellicon, Victoria, BC), which altered the pitch of the participant's voice. This process introduced an approximately 10 second delay to the feedback signal, which was then presented back to the participant as auditory feedback. The onset and offset of the pitch alteration was controlled manually by the experimenter between the experimental phases. Both the altered and the unaltered voice signals were digitally recorded (TASCAM HD-P2, Montebello, CA) at a sampling rate of 44.1 kHz for later analysis.

iv. Data Analysis

The digital recording of the vocalizations was segmented into separate utterances and F0 values were calculated for each utterance using the SWIPE algorithm (Camacho & Harris, 2008). F0 values were normalized to the participants habitual F0, by converting Hertz values to cents using the following formula:

$$\boxed{\text{Cents} = 100\ (12 \log2\ F/B)}$$

In this formula, F is the F0 value in Hertz and B is the participant's habitual F0. Participants' habitual F0 values were estimated by calculating the average median 1500 of the participants' vocalizations from trial 11-20, the last 10 trials of the pre-shift phase. Cents values were calculated for the first 1500 ms of each vocalization.

For each vocalization the median F0 of the first 80 ms (the ‘median 80’ value) and median F0 of the first 1500 ms (the ‘median 1500’ value) were calculated. Since neural processing delays prevent the auditory feedback resulting from a vocal motor command from being processed for at least 100 ms (Burnett et al., 1997, 1998; Hawco & Jones, 2009; Keough & Jones, 2009, 2013), the median 80 value provides an index of the F0 at which the vocalization was initiated, before auditory feedback was available. On the other hand, median 1500 values provide an index of the F0 of the vocalization once auditory feedback becomes available. In addition, in order to index the normal amount of variability in each participant’s habitual F0, the standard deviation of the median 1500 of trials 11-20, each of the last 10 trials of the pre-shift phase, was calculated and then averaged over the 10 trials.

v. Statistical Analysis

Four groups of experimental trials were used for statistical analysis: baseline (trials 11-20), shift-start (trials 21-30), shift-end (trials 111-120), and post-shift (trials 121 and 122). The size of each participant’s compensatory response at the end of the shift-phase was calculated by subtracting the average of the baseline trials from the average of the shift-end trials, for both median 80 and median 1500. In addition, to assess whether the compensatory behaviour persisted after the participant’s auditory feedback was returned to normal, an after-effect value was calculated by subtracting the average of the baseline trials from the average of the post-shift trials, for both median 80 and median 1500.

A repeated measures analysis of variance (RM-ANOVA) was performed to assess the influence of the 4 trial groups (baseline, shift-start, shift-end, and post-shift) on median 80 and median 1500 values. In instances where the sphericity assumption was violated, the Greenhouse-Geisser correction was applied (Greenhouse & Geisser, 1959). However, for ease of interpretation original degrees of freedom are reported. Pairwise comparisons were conducted to examine potential differences in the size of responses across the 4 trial groups. Pearson correlations were also calculated to investigate the relationship between habitual vocal variability and both the size of the compensatory responses and the after-effects.

Results

i. Median 80

The RM-ANOVA conducted to investigate the effect of trial group (baseline, shift-start, shift-end, and post-shift) on median 80 values, revealed a main effect of trial group, $F_{3,\,141} = 6.919$, $P < .001$ (see Figure 14). Pairwise comparisons indicated that median 80 values were significantly smaller in the baseline group, relative to the shift-start ($P < .001$), shift-end ($P = .001$), and the post-shift ($P = .002$) groups.

Correlational analyses were conducted using median 80 compensatory response and after-effect values, in order to determine whether habitual vocal variability was related to the size of a participant's compensatory response, and the size of their after-effects. The correlation between vocal variability and compensatory responses was not significant, $r_{48} = -.044$, $P = .767$. However, the correlation between vocal variability and after-effects was significant, $r_{48} = -.291$, $P = .045$ (see Figure 15).

ii. Median 1500

The RM-ANOVA conducted to investigate the effect of trial group (baseline, shift-start, shift-end, and post-shift) on median 1500 values, revealed a main effect of trial group, $F_{3,\,141} = 33.639$, $P < .001$ (see Figure 14). Pairwise comparisons indicated that median 1500 values were significantly smaller in the baseline group, relative to the shift-start ($P < .001$), shift-end ($P < .001$), and the post-shift ($P = .038$) groups. Median values were also significantly smaller in the shift-start group, relative to the shift-end group (*P* = .008), and in the post-shift group, relative to the shift-start (*P* = .002) and shift-end (*P* < .001) groups.

Correlational analyses were conducted using median 1500 compensatory response and after-effect values, in order to determine whether habitual vocal variability was related to the size of a participant's compensatory response, and the size of their after-effects. The correlation between vocal variability and compensatory responses was significant, r_{48} = .338, *P* = .019. However, the correlation between vocal variability and after-effects was not significant, r_{48} = .230, *P* = .117 (see Figure 15).

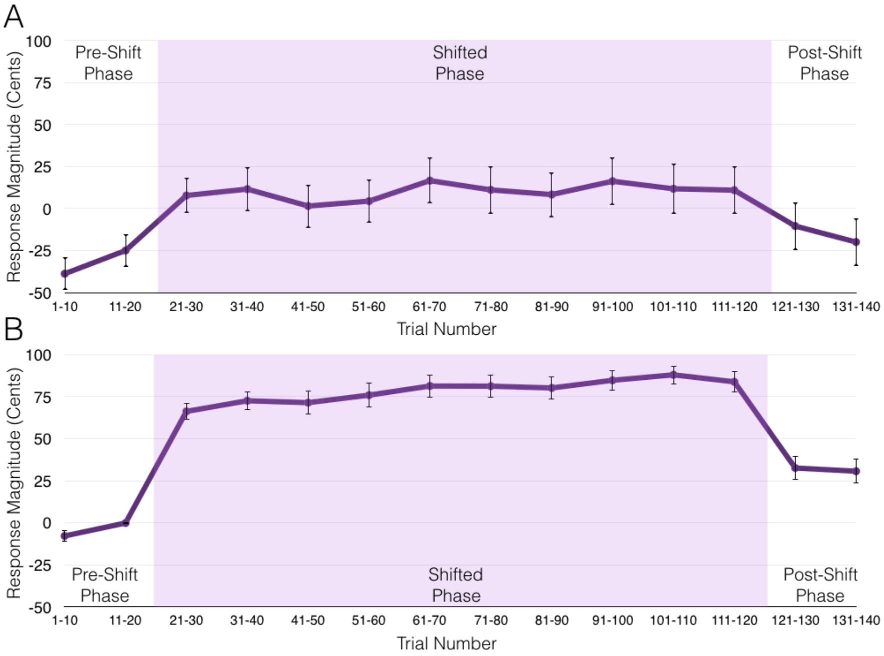

Figure 14: Median 80 (top) and Median 1500 (bottom) Values

An average of the participants' median fundamental frequencies during the first 80 ms (a) and 1500 ms (b) of each utterance, for trials 1-140. Each data point represents an average of 10 trials across all participants. Error bars indicate the standard error of the mean. During trials 1-20 (pre-shift phase) and 121-140 (post-shift trials), participants auditory feedback was unaltered. During trials 21-120 (the shifted phase), the pitch of participants' auditory feedback was shifted downward 100 cents (one semitone).

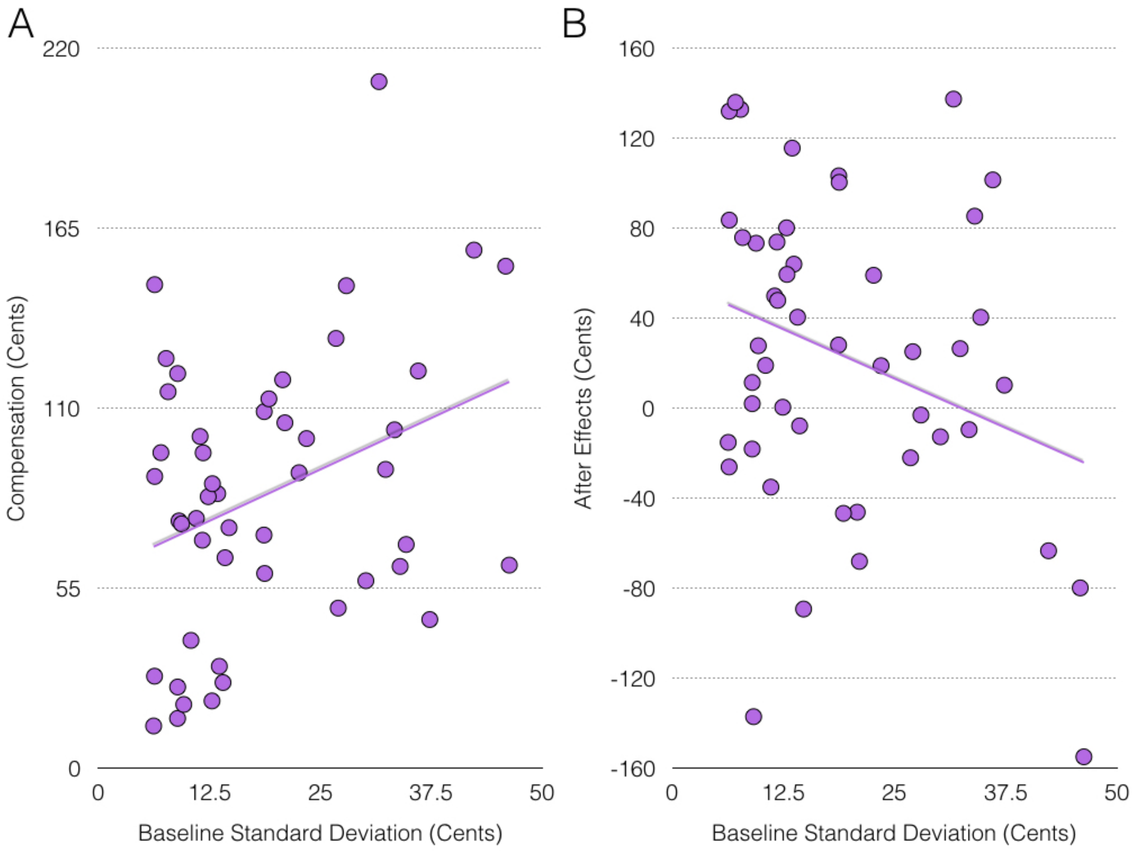

Figure 15: Vocal Variability Correlated with Median 1500 Compensation (Left) and Median 80 After-Effects (Right)

(a) Participant's baseline standard deviation is plotted against their compensation magnitude, calculated as the difference between the shift-end phase and the baseline phase. (b) Participants baseline standard deviation is plotted against the size of their after-effects, calculated as the difference between the test-phase and the baseline phase.

Discussion

The aim of the current study was to investigate whether vocal variability could predict the degree of sensorimotor learning that occurs when participants are exposed to FAF. During this study participants produced vocalizations while exposed to predictable changes in their auditory feedback. Sensorimotor learning was assessed by evaluating the F0 values for the early portion of participants' vocalizations. Since auditory feedback is delayed by cortical processing, the F0 at the beginning of participants' vocalizations can be used to index feedforward control, or the extent to which deviant auditory feedback was used for the planning of subsequent speech motor commands (Keough & Jones, 2009, 2011, 2013; Hawco et al., 2009). On the other hand, by looking at participants' F0 across each vocalization, it was possible to index the extent to which the deviant auditory feedback modulated the ongoing vocalization.

As expected, median 80 values increased after persistent exposure to the FAF, suggesting that sensorimotor learning occurred. More specifically, median 80 values were larger in the shift-start, shift-end, and post-shift phases, relative to the baseline phase. The fact that the median 80 values changed when the FAF manipulation was introduced, demonstrates that the way in which speech motor commands were executed by the feedforward controller changed as a result of exposure to FAF. Furthermore, the fact that median 80 values in the post-shift phase were larger than those recorded in the baseline phase, suggests that the effects of sensorimotor learning persisted even when the FAF manipulation was removed. Importantly, the variability of participants' habitual F0 correlated significantly with the size of the after-effects. That is, participants who had a higher degree of vocal variability displayed smaller after-effects, relative to participants who had a more stable habitual F0. Together, these results suggest that persistent exposure to FAF drove sensorimotor learning, but this effect was more evident in individuals who possessed a higher degree of vocal stability.

The fact that individuals with a more stable habitual F0 demonstrated more sensorimotor learning is very similar to the findings of Jones and Keough (2008) who found that trained singers showed more sensorimotor learning than non-singers when exposed to a persistent FAF manipulation. Together these results suggest that vocal stability, which may be intrinsic to the speaker, or may be achieved through extensive vocal training, results in an increased weighting of the feedforward control system. As a result of this increased weighting of the feedforward control system, information from auditory feedback is more readily used to update the mapping of one’s sensorimotor representations. As a result of this remapping, responses to the FAF persist, even after the FAF manipulation is removed.

While it is clear that vocal variability modulated the weighting of the feedforward control system, by calculating the size of participants’ compensatory responses using their F0 across the vocalization, it was possible to investigate the influence of vocal variability on the weighting of the feedback control system. As anticipated, median 1500 values increased following the initiation of the FAF manipulation, demonstrating that the FAF elicited compensatory responses. These compensatory responses were larger during the shift-end phase, relative to the shift-start phase, suggesting that the size of the compensatory response increased over the duration of the exposure to the FAF. Interestingly, the F0 of participants’ vocalizations during the test phase was larger than their F0 during the baseline phase, but smaller than their F0 during the shift-start and shift-end phases. This finding suggests that when participants’ auditory feedback was returned to normal their F0 decreased, however, there were still residual effects from the previous exposure to the FAF. This result demonstrates the interaction between the feedback and feedforward controllers. Remapping of the sensory-motor representations as a result of exposure to the FAF caused the feedforward controller to produce a vocalization at a higher F0, however, once the no-longer-deviant auditory feedback was processed, the feedback controller reduced the F0 of the vocalization. In addition to the significant experimental phase effects, the variability of participants’ habitual F0 correlated significantly with the size of their compensatory responses. This finding is consistent with previous FAF studies that have shown that participants who have a higher degree of vocal variability produce larger compensatory responses to FAF perturbations (Scheerer & Jones, 2013b) and FAF mid-utterance shifts (Scheerer & Jones, 2012). Together, these studies suggest that vocal variability results in increased weighting of the feedback control system for monitoring and maintaining speech motor commands.

Speech motor control involves the combined effort of a feedback and a feedforward control system. Computational models (Guenther, 2006; Tourville et al., 2008; Civier et al., 2010, Hickok et al., 2011) and experimental evidence (Jones & Munhall, 2005; Jones & Keough, 2008; Hawco & Jones, 2009; Keough & Jones, 2009, 2011; Scheerer & Jones, 2012, 2014; Chen et al., 2013) suggest that throughout development speech motor control transitions from a feedback dominant control system, to feedforward dominant control system. However, the factors that dictate the relative weighting of these control systems are unclear. The results of this study suggest that vocal variability influences the relative weighting of feedback and feedforward control. When individuals are highly variable, they increase the weighting of their feedback control system in order to monitor and maintain the fluency of their speech. On the other hand, when individuals have more stable speech motor control, they increase the weighting of their feedforward control system, as the information provided by auditory feedback is less imperative for the production of fluent speech.

Study 6: The Predictability of Frequency-Altered Auditory Feedback Changes the Weighting of Feedback and Feedforward Input for Speech Motor Control.

Published as:

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

Sensory feedback plays a crucial role in the regulation and planning of our movements, and allows us to produce fluid movements despite our constantly changing environment (Bays & Wolpert, 2006). As sensory feedback plays a crucial role in the regulation of voluntary movements, an understanding of how the information from sensory feedback is used to modify ongoing and future movements is paramount to the understanding of these sensorimotor interactions. Previous studies have demonstrated that vocal and ERP responses to FAF are modulated both by the prediction created by one's sensorimotor representations (Houde et al., 2002; Heinks-Maldonado et al., 2005, 2006; Scheerer et al., 2013a) and by the predictability of experimentally induced manipulations of the direction (Korzyukov et al., 2012) and latency (Burnett, McCurdy, & Bright, 2008; Chen et al., 2012a) of feedback perturbations. Together, these studies suggest that being able to predict the properties of upcoming auditory feedback results in a decreased reliance on feedback control, and smaller responses. However, it is currently unknown whether being able to predict the magnitude of experimentally induced feedback errors influences both behavioural and neural responses to these errors. In this experiment individuals produced vocalizations while being exposed to blocks of FAF perturbations that were either predictable in magnitude (consistently either 50 or 100 cents) or unpredictable in magnitude (50 and 100 cent perturbations varying randomly within each vocalization). If, as expected, vocal and neural responses differ as a function of stimulus predictability, these results would suggest that the weighting of the feedback and feedforward speech motor control systems is influenced by the predictability of the magnitude of auditory feedback errors.

Abstract

Speech production requires the combined effort of a feedback control system driven by sensory feedback, and a feedforward control system driven by internal models. However, the factors that dictate the relative weighting of these feedback and feedforward control systems are unclear. In this event-related potential (ERP) study, participants produced vocalizations while being exposed to blocks of frequency-altered feedback (FAF) perturbations that were either predictable in magnitude (consistently either 50 or 100 cents) or unpredictable in magnitude (50 and 100 cent perturbations varying randomly within each vocalization). Vocal and P1–N1–P2 ERP responses revealed decreases in the magnitude and trial-to-trial variability of vocal responses, smaller N1 amplitudes, and shorter vocal, P1 and N1 response latencies following predictable FAF perturbation magnitudes. In addition, vocal response magnitudes correlated with N1 amplitudes, vocal response latencies, and P2 latencies. This pattern of results suggests that after repeated exposure to predictable FAF perturbations, the contribution of the feedforward control system increases. Examination of the presentation order of the FAF perturbations revealed smaller compensatory responses, smaller P1 and P2 amplitudes, and shorter N1 latencies when the block of predictable 100-cent perturbations occurred prior to the block of predictable 50-cent perturbations. These results suggest that exposure to large perturbations modulates responses to subsequent perturbations of equal or smaller size. Similarly, exposure to a 100-cent perturbation prior to a 50-cent perturbation within a vocalization decreased the magnitude of vocal and N1 responses, but increased P1 and P2 latencies. Thus, exposure to a single perturbation can affect responses to subsequent perturbations.

Introduction

In everyday life we face the challenge of having to function in dynamic and unpredictable surroundings. Despite this challenge, with the aid of sensory feedback we are able to regulate and successfully plan the execution of our movements, leading to fluid interactions within our environment (Bays & Wolpert, 2006). As sensory feedback plays a crucial role in the regulation of voluntary movements, an understanding of how the information from sensory feedback is used to modify ongoing and future movements is paramount to the understanding of these sensorimotor interactions. Specifically, in the study of speech motor control, it is important to understand how auditory feedback is used to regulate future and ongoing speech motor commands in order to maintain fluent speech.

Early in speech development, infants rely heavily on auditory feedback to explore the relationship between articulator movements and the resultant sounds (Bailly, 1997; Callan et al., 2000; Jones & Munhall, 2000, 2002; Guenther, 2006; Civier et al., 2010). As development progresses, associations between the environment, musculature, and motor commands, as well as the sensory consequences of the motor commands required to perform speech-related movements are learned, establishing sensorimotor representations (Civier et al., 2010). These sensorimotor representations support the speech motor control system by predicting the outcome of speech-related movements and providing internal feedback to the planning and control subsystems (Miall & Wolpert, 1996; Kawato, 1999; Jones & Munhall, 2000, 2002, 2005; Jones & Keough, 2008). This internal feedback helps the motor control system to plan and regulate future movements, without relying on peripheral sensory feedback, which is generally not available for at least 80 ms after the speech movements are performed (Beal et al., 2011). In order to accommodate the growth of the articulators, vocal folds, and musculature, as well as changes in lung capacity, these sensorimotor representations must remain plastic throughout development (Guenther, 2006). However, empirical evidence suggests that as development progresses individuals rely less on auditory feedback, and more on sensorimotor representations (Liu et al., 2010a; Scheerer et al., 2013b).

Accordingly, computational models of speech motor control (e.g. the directions into velocities of the articulators model; Guenther, 2006; Tourville et al., 2008; Civier et al., 2010) and experimental evidence (e.g. Jones & Munhall, 2005; Jones & Keough, 2008; Hawco &

Jones, 2009; Keough & Jones, 2009, 2011; Scheerer & Jones, 2012) suggest that fluent speech production relies on the combination of a feedback control system, driven by sensory feedback, and a feedforward control system, driven by sensorimotor representations. The feedforward control system is necessary for maintaining fluent speech, as strict reliance on the feedback control system would induce disfluencies during speech production (Houde et al., 2002; Guenther, 2006; Civier et al., 2010; Perkell, 2012). On the other hand, the feedback control system is crucial for detecting and correcting speech production errors, and for providing feedback to update the mapping of the sensorimotor representations (Civier et al., 2010). However, it is currently unclear what factors dictate the relative weighting of feedback and feedforward input for speech motor control.

The importance of auditory feedback for speech motor control is often investigated with the frequency-altered feedback (FAF) paradigm. Typical FAF experiments involve manipulating the fundamental frequency (F0) of an individual's vocalization, and instantaneously presenting the altered auditory feedback back to the individual through headphones (Elman, 1981; Burnett et al., 1997). FAF is perceived by speakers as an error in their vocal pitch production, thus upon exposure to FAF, speakers tend to compensate, or shift their vocal pitch in the opposite direction of the manipulation. However, this compensatory response is often only a fraction of the size of the manipulation. For example, brief 100–500 ms auditory feedback perturbations that are 100 cents (one semitone) in magnitude have consistently been shown to elicit responses that are < 50 cents in magnitude (Burnett et al., 1997, 1998; Liu et al., 2011; Korzyukov et al., 2012b; Scheerer et al., 2013a,b), with responses as small as 8 cents having been reported (Scheerer et al., 2013a). As the compensatory responses are often only a fraction of the size of the manipulation, it has been suggested that these compensatory responses play a role in stabilizing voice F0 around a desired target (Hain et al., 2000; Natke et al., 2003; Hawco & Jones, 2009). These partial compensatory responses also highlight the fact that auditory feedback is not the sole input to the speech motor control system; somatosensory input also facilitates fluent speech motor control (Tremblay, Shiller, & Ostry, 2003; Lametti, Nasir, & Ostry, 2012). Importantly, by measuring compensatory responses to FAF, it is possible to index the relative contribution of information from auditory feedback to ongoing speech production.

One factor that has been shown to influence the size of compensatory responses to FAF, and has been implicated in the weighting of feedback relative to feedforward input, is vocal variability (Scheerer & Jones, 2012; Scheerer et al., 2013b). Scheerer & Jones (2012) investigated the relationship between vocal variability and compensatory responses to FAF. Scheerer and Jones (2012) found that individuals with more variable vocal pitch, produced larger compensatory responses to FAF. These results suggested that less consistent vocal output led to an increased reliance on feedback input for monitoring and maintaining speech motor control. As a result of this increased reliance on auditory feedback, these individuals became more susceptible to deviant auditory feedback. On the other hand, the results also suggested that individuals with more consistent vocal output were less susceptible to deviant auditory feedback. For this reason, the authors concluded that vocal consistency (typically achieved as individuals develop, or through intense vocal training) contributes to the internal model's ability to reliably predict the sensory consequences of upcoming motor movements. When individuals reliably predict the sensory consequences of their motor movements, the weighting of their feedforward control system is increased. For this reason, individuals with increased vocal consistency are less susceptible to deviant auditory feedback, resulting in smaller vocal responses when auditory feedback is altered.

As the ability to predict future events increases as events become more frequent and consistent, it is reasonable to expect that varying any aspect of the sensory or motor system in a predictable manner may result in increased weighting of feedforward input, and habituated responses to deviant auditory feedback. For example, in a study by Burnett and colleagues

(2008), the effects of temporal and direction predictability on vocal F0 responses was investigated. During this experiment, participants performed a button press to elicit a FAF perturbation as they vocalized a vowel sound. This button press resulted in either a FAF perturbation that occurred immediately, and was thus temporally predictable, or a FAF perturbation that occurred after a random delay. In addition, the FAF perturbation was either always upwards, or varied randomly between upwards and downwards, depending on the condition. Direction predictability was not found to influence vocal responses, however, increased temporal predictability resulted in faster vocal responses. These results indicate that predictable changes to some aspects of auditory feedback may result in modified vocal responses, suggesting that learning has occurred. However, the extent to which the predictability of FAF perturbation magnitudes can guide learning in dynamic environments, and thus alter responses to brief perturbations of auditory feedback, is currently unclear.

The influence of learning on the speech motor control system is often evaluated by monitoring auditory–cortical processing using electroencephalography (EEG) or magnetoencephalography responses. In previous studies, it has been shown that the N1 event related potential (ERP), and its magnetic equivalent the M1, are maximally attenuated during the perception of one’s own unaltered speech, relative to FAF (Heinks-Maldonado et al., 2005, 2006), alien speech (Heinks-Maldonado et al., 2005, 2006), and the playback of self-produced speech (Houde et al., 2002; Beal et al., 2011). In addition, single-cell recordings in the auditory cortex of neurosurgery patients (Greenlee et al., 2011; Chang et al., 2013) and marmoset monkeys (Eliades & Wang, 2005, 2008) have revealed suppressed neural activity during self-produced speech, relative to the playback of self-produced speech. It has been suggested that this neural modulation observed in humans and marmoset monkeys reflects suppression in the auditory cortex during the perception of one’s own unaltered speech, as a result of a match between perceived and predicted sensory feedback. On the other hand, increases in activation when sensory feedback is incongruent with the predicted feedback is thought to reflect a violation of the prediction created by the sensorimotor representation (Heinks-Maldonado et al., 2006; Beal et al., 2011). In line with this notion, Sitek et al. (2013) found that, relative to instances where the variability of vocalizations were high from trial to trial, stable vocalizations resulted in suppressed N1 amplitudes. These studies suggest that when auditory feedback is predictable, fewer neural resources are dedicated to processing auditory feedback, reducing the size of the N1 component. On the other hand, when auditory feedback is novel or unpredictable, more neural resources are dedicated to the processing of auditory feedback.

Similarly to the N1 component, previous FAF studies have shown that the amplitude of the P2 ERP component increases linearly with increasing feedback perturbation magnitudes (Behroozmand et al., 2009; Scheerer et al., 2013a). As the size of the P2 response varies with the size of the feedback manipulation, it has been suggested that the P2 component indexes the size of the mismatch between the perceived and predicted auditory feedback (Scheerer et al., 2013a). Together, these results suggest that the N1 and P2 neural responses are modulated by the prediction created by the sensorimotor representation. Furthermore, these results suggest that any predictable aspects of auditory feedback may have the potential to habituate ERP responses, meaning that auditory learning may be assessed by monitoring changes in these ERP responses.

Korzyukov et al. (2012b) exposed participants to blocks of FAF perturbations that were either predictable in direction or varied randomly in direction. They found that predictable perturbations led to a reduction of the amplitude of the N1 ERP component, relative to the unpredictable perturbations. Similarly, Chen et al. (2012a) found that varying the temporal predictability of FAF perturbations modified ERP responses. The results of their study indicated that temporally predictable auditory feedback perturbations led to suppressed P2 ERP responses, relative to temporally unpredictable auditory feedback perturbations. Together, these results suggest that foreknowledge of the properties of an upcoming auditory perturbation results in smaller ERP responses to that perturbation, relative to unexpected auditory perturbations.

Previous studies have demonstrated that vocal and ERP responses to FAF are modulated both by the prediction created by the sensorimotor representation (Houde et al., 2002; Heinks-Maldonado et al., 2005, 2006; Scheerer et al., 2013a) and by the predictability of experimentally induced manipulations of the direction (Korzyukov et al., 2012b) and the latency (Burnett et al., 2008; Chen et al., 2012a) of feedback perturbations. Together, these studies suggest that being able to predict the properties of upcoming auditory feedback results in decreased reliance on feedback control, and smaller responses. However, it is currently unknown whether being able to predict the magnitude of experimentally induced feedback errors influences both behavioural and neural responses to these errors. For this reason, in this experiment, participants produced vocalizations while being exposed to blocks of FAF perturbations that were either predictable in magnitude (consistently either 50 or 100 cents) or unpredictable in magnitude (50 and 100 cent perturbations varying randomly within each vocalization). As both behavioural and neural responses to FAF have shown signs of habituation following the presentation of repetitive predictable stimuli, we expected that when the magnitude of the feedback perturbations was predictable, vocal and ERP responses would be smaller, as the predictable nature of the stimuli would result in increased weighting of the feedforward control system. On the other hand, when feedback perturbation magnitudes varied randomly, the unpredictable nature of the sensory feedback would drive an increase in the weighting of the feedback control system, allowing for rapid correction of auditory feedback errors. Accordingly, we expected that predictable feedback perturbation magnitudes would elicit smaller vocal and ERP responses than randomly varying feedback perturbation magnitudes, reflecting habituation as a result of the repetitive and predictable nature of the stimuli. If, as expected, vocal and neural responses differed as a function of stimulus predictability, these results would suggest that the weighting of the feedback and feedforward speech motor control systems is influenced by the predictability of the magnitude of auditory feedback errors.

Methods

i. Participants

Thirty-six participants between the ages of 18 and 25 years [mean, 19.33 years; SD, 1.85 years; 27 females and nine males] participated in this study. Participants were right-handed native Canadian English speakers, and did not speak a tonal language. Prior to testing, participants' hearing was assessed with a diagnostic audiometer (Midimate 602; Madsen Electronics, Minneapolis, MN, USA) to ensure hearing thresholds of ≤20 dB HL at 250, 500, 1000, 2000, 4000 and 8000 Hz. Participants also completed a music experience questionnaire, and two participants reported having received formal vocal training. Participants gave informed consent and received course credit or financial compensation for their participation. All procedures used in this study were approved by the Wilfrid Laurier University Research Ethics Board, and were in accordance with the World Medical Association 2013 Declaration of Helsinki.

ii. Procedure

Participants vocalized the vowel sound /a/, 120 times over four blocks, while being exposed to unaltered feedback and FAF. Participants were cued to start vocalizing by a green box on a computer screen, while a red box indicated that they should stop vocalizing. Each experimental block contained 30 trials, in which the participants' auditory feedback was perturbed either 50 or 100 cents downwards, four times per vocalization, or left unaltered. Each perturbation had a fixed duration of 200 ms and occurred with an inter-stimulus interval of 700–900 ms, resulting in vocalizations that were ~5 s in length. The four experimental blocks were divided into two conditions: the predictable condition, in which perturbation magnitudes were consistently either 50 or 100 cents within each vocalization, and the unpredictable condition, in which perturbation magnitudes were different combinations of 50 and 100 cents within each vocalization (see Figure 16). The 50-cent predictable block contained 25 trials in which the participant's voice was perturbed 50 cents downwards four times per vocalization, and five trials in which the participant's voice was left unaltered, but sampled four times per vocalization. This resulted in 100 50-cent predictable trials, and 20 unaltered predictable trials. Similarly, the 100-cent predictable block contained 25 trials in which the participant's voice was perturbed 100 cents downwards four times per vocalization, and five trials in which the participant's voice was left unaltered, but sampled four times per vocalization. This resulted in 100 100-cent predictable trials, and 20 unaltered predictable trials. The two unpredictable blocks each contained 24 trials in which the participant's voice was perturbed 50 and 100 cents downwards, in different combinations, resulting in a total of four perturbations per vocalization, while the other six vocalizations were left unaltered. A latin-square was used to determine all possible combinations of four 50-cent or 100-cent perturbations for the unpredictable condition. This resulted in 16 different combinations of 50- and 100-cent perturbations. Each combination of 50- and 100-cent perturbations occurred three times per experimental session, and was pseudorandomly presented throughout the two unpredictable experimental blocks. This resulted in 96 50-cent unpredictable trials, 96 100-cent unpredictable trials, and 48 unaltered unpredictable trials, split between the two blocks of unpredictable trials. FAF and unaltered trials were pseudorandomly presented within each block. The block order for all participants was as follows: unpredictable, predictable, unpredictable, and predictable. However, the presentation order of the predictable blocks was counterbalanced across participants.

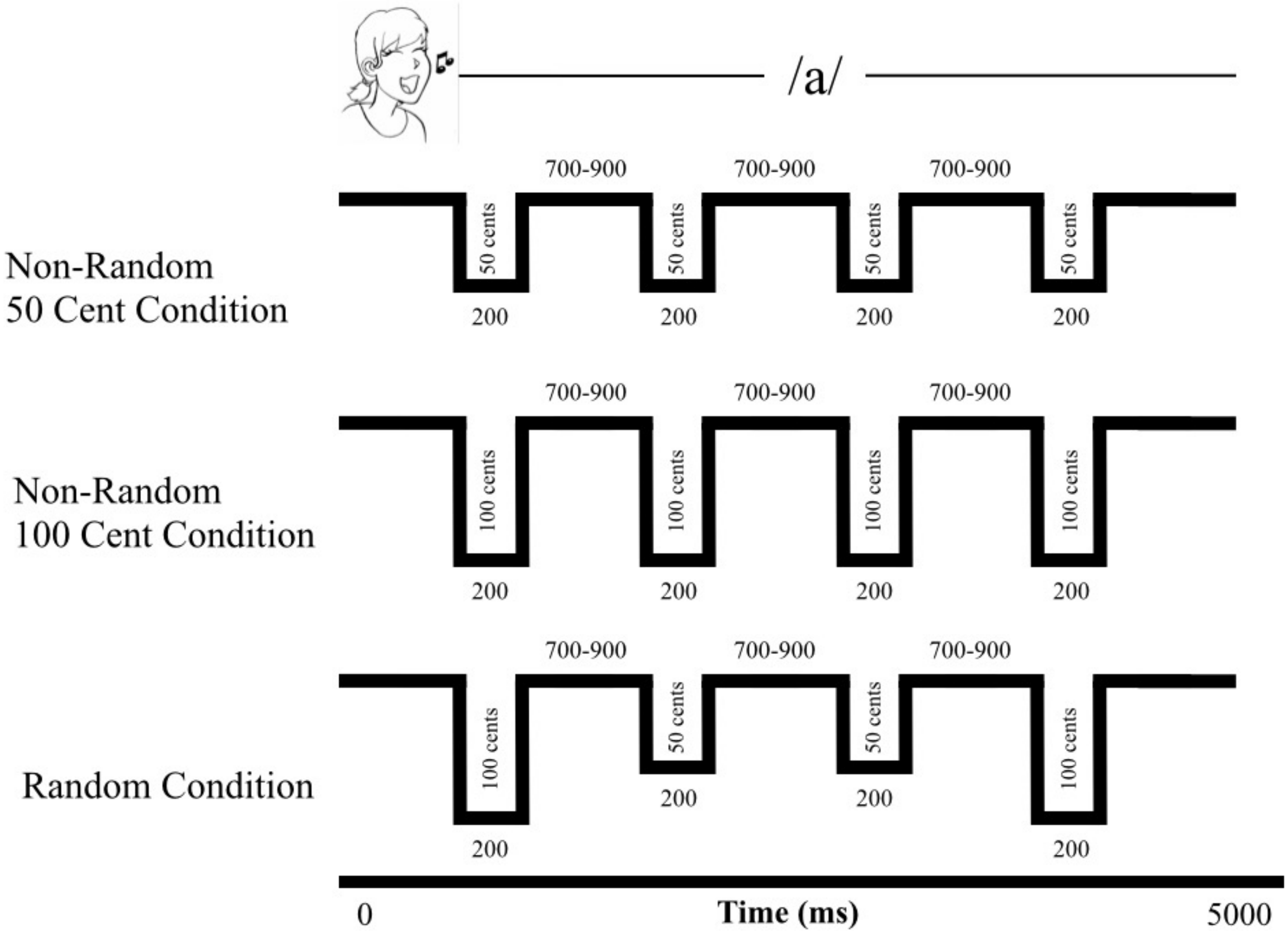

Figure 16: Schematic of the experimental design. Participants were exposed to three different FAF manipulations. During the predictable 50- and 100-cent conditions, participants were exposed to four 200-ms auditory feedback perturbations within each vocalization that were predictable in magnitude, either 50 cents downwards (top), or 100 cents downwards (middle). During the unpredictable condition, participants were exposed to four 200-ms auditory feedback perturbations within each vocalization that were one of 16 combinations of 50- and 100-cent perturbations.

iii. Apparatus

Participants wore a HydroCel GSN 64 1.0 electrode cap (Electrical Geodesics, Eugene, OR, USA), Etymotic ER-3 insert headphones (Etymotic Research, Elk Grove Village, IL, USA), and a headset microphone (Countryman Isomax E6 Omnidirectional Microphone), and were tested in an electrically shielded booth (Raymond EMC, Ottawa, ON, Canada). The onset and offset of each pitch perturbation was controlled by MAX/MSP 4 (Cycling ‘74, San Francisco, CA, USA). Voice signals were sent to a mixer (Mackie Onyx 1220; Loud Technologies, Woodinville, WA, USA), followed by a digital signal processor (VoiceOne; T.C. Helicon, Victoria, BC, Canada), which altered the fundamental frequency of the voice signal. The altered voice signal was then presented back to the participant through headphones as FAF. The unaltered voice signal was digitally recorded (HD-P2; TASCAM, Montebello, CA, USA) at a sampling rate of 44.1 kHz for later analysis.

iv. Behavioural Recording and Analysis

The unaltered voice signal was segmented into separate vocalizations, and F0 values were calculated for each vocalization with the SWIPE algorithm (Camacho & Harris, 2008). Each vocalization was then segmented on the basis of the onset of the four perturbations. F0 values for each of the four perturbed segments were normalized to the baseline period, which was the portion of the segment 200 ms prior to the onset of the perturbation, by converting hertz values to cents with the formula:

$$\boxed{\text{Cents} = 100\ (12 \log_2 F/B)}$$

where F is the F0 value in hertz, and B is the mean frequency of the baseline period.

Cent values were calculated for the 200 ms prior to the perturbation (the baseline period) and the 1000 ms after the perturbation. An averaged F0 trace was constructed for each perturbation magnitude, that is, 0 (unaltered), 50, and 100 cents, in each condition (unpredictable and predictable) for each participant. Additional F0 traces were also constructed for the unpredictable perturbation condition to determine whether the magnitude of the preceding perturbation modulated the size of the response to the subsequent perturbation. This resulted in four conditions: 50 cents preceded by 50 cents, 50 cents preceded by 100 cents, 100 cents preceded by 50 cents, and 100 cents preceded by 100 cents.

Vocal responses were quantified by examining the response magnitude and latency. The magnitude of the compensatory response was determined by finding the point at which the participant's averaged F0 trace deviated maximally from the baseline mean, and the latency was calculated as the time at which this maximal deviation occurred. In addition, to index the variability in response magnitudes, the SD of the response magnitude was calculated for each experimental condition.

v. ERP Recording and Analysis

EEG signals were recorded from 64 scalp electrodes and referenced online to the vertex (Cz) electrode. Data were bandpass-filtered (1–30 Hz) and digitized (12-bit precision) at 1000 Hz. Electrode impedances were maintained below 50 kΩ throughout the duration of the experiment (Ferree et al., 2001). After data acquisition, EEG voltage values were re-referenced to the average voltage across all electrode sites. The data were then epoched into segments from 100 ms before the perturbation onset to 500 ms after the perturbation onset. Data were analyzed offline for movement artifacts, and any segment with voltage values exceeding 55 µV of the moving average over an 80-ms span were rejected. In addition, a visual inspection of all data was completed to ensure that artifacts were being adequately detected. Eight subjects were eliminated from further analyses, as they had <50% of their trials retained across all perturbation magnitudes. Across all other participants, on average, 84–92% of trials were retained for each perturbation magnitude.

Nine electrodes were included in the analysis: FC1, FCz, FC2, C1, Cz, C2, CP1, Pz, and CP2. These electrodes were then grouped into three electrode sites: left (CP1, C1, and FC1), medial (FCz, Cz, and Pz), and right (FC2, C2, and CP2). These electrodes were chosen on the basis of visual inspection of the regions showing the most robust P1–N1–P2 components, as well as previous research suggesting that fronto-medial and centro-frontal regions are optimal for recording P1– N1–P2 responses to FAF (Behroozmand et al., 2009; Chen et al., 2012a; Korzyukov et al., 2012a; Scheerer et al., 2013a,b).

For each participant, averaged waveforms were created for the unaltered and the FAF conditions for each electrode. Grand-averaged waveforms were created for all conditions by averaging the data from all participants for each electrode, and this was followed by baseline correction. For all average files for each participant, the peak amplitude and the latency of the peak amplitude were calculated for the ERP components of the P1–N1–P2 complex. On the basis of visual inspection of the latency of the most prominent ERP peaks, these components were extracted at time windows from 50 to 100 ms, from 100 to 200 ms, and from 200 to 300 ms, respectively.

vi. Statistical analyses

Preliminary statistical analyses were conducted to investigate the behavioural and ERP responses in the 0-cent (unaltered) conditions. Responses were not found to vary as a function of experimental condition (unpredictable vs. predictable), so all control responses were collapsed into a single category.

In order to investigate the effect of the predictability of the perturbation magnitude within a vocalization, five experimental categories were examined: 0 cents (unaltered), 50 cents (unpredictable), 50 cents (predictable), 100 cents (unpredictable), and 100 cents (predictable). In addition, to investigate whether the magnitude of the preceding perturbation modulated the size of the response to the subsequent perturbation, four experimental categories were examined: 50-cent perturbations preceded by a 50-cent perturbation, 50-cent perturbations preceded by a 100-cent perturbation, 100-cent perturbations preceded by a 50-cent perturbation, and 100-cent perturbations preceded by a 100-cent perturbation.

To examine the influence of the predictability of the magnitude of perturbations within a vocalization, SPSS (v. 19.0) was used to conduct a 5 (experimental condition) x 2 (block order) repeated-measures ANOVA (RM-ANOVA) on vocal response magnitudes and the variability of vocal response magnitudes. A 5 (experimental condition) x 3 (electrode site) x 2 (block order) RM-ANOVA was conducted on the amplitudes of the P1–N1–P2 complex. A 4 (experimental condition) x 2 (block order) RM-ANOVA was also conducted on the latencies of the vocal responses, and a 4 (experimental condition) x 3 (electrode site) x 2 (block order) RM-ANOVA was conducted on the latencies of the P1–N1–P2 ERP components. The unaltered (0-cent) condition was not analyzed with regards to ERP and vocal latencies, as stimuli were not presented during the unaltered trials, thus data were randomly sampled with no true reference, rendering latency information meaningless.

Similarly, to assess whether responses to 50- and 100-cent perturbations were modulated by the magnitude of the preceding perturbation, a 4 (experimental category) x 2 (block order) RM-ANOVA was conducted on the amplitude of the vocal response magnitudes. A 4 (experimental category) x 3 (electrode site) x 2 (block order) RM-ANOVA was conducted on P1–N1–P2 ERP amplitudes and latencies.

Correlational analyses were also performed on both the data from the perturbation magnitude predictability manipulation and the data from the preceding perturbation magnitude manipulation, in order to investigate the relationship between the behavioural and ERP responses.

For all RM-ANOVAs, post hoc least significant difference tests were conducted to examine differences in the recorded responses as a function of the experimental condition. The Greenhouse–Geisser (Greenhouse & Geisser, 1959) correction was used in instances where Mauchley's assumption of sphericity was violated. However, original degrees of freedom are reported for ease of interpretation.

Results

i. Vocal Responses

Perturbation Magnitude Predictability Effects

The RM-ANOVA conducted to investigate the effect of varying the predictability of perturbation magnitudes within a vocalization on vocal response magnitude showed a significant main effect of experimental condition, $F_{4,104} = 13.036$, $P < 0.001$ (see Figure 17). There was also a main effect of block order, $F_{1,26} = 13.113$, $P = 0.001$, because, overall, participants had larger vocal responses when the 50-cent predictable block occurred prior to the 100-cent predictable block. However, the interaction between experimental condition and block order was non-significant, $F_{4,104} = 1.618$, $P > 0.05$. Pairwise comparisons indicated that all FAF conditions elicited significantly larger vocal responses than the unaltered condition ($P < 0.05$). In addition, the 50-cent predictable condition elicited significantly smaller vocal responses than all other FAF conditions ($P < 0.05$), and the 50-cent unpredictable condition elicited smaller vocal responses than the 100-cent unpredictable condition ($P < 0.05$), but not the 100-cent predictable condition ($P > 0.05$). Finally, the 100-cent unpredictable condition elicited significantly larger vocal responses than all other conditions ($P < 0.01$).

A RM-ANOVA was also conducted to investigate the effect of varying the predictability of perturbation magnitudes within a vocalization, as well as block order, on the variability of response magnitudes. A main effect of experimental block, $F_{4,104} = 6.426$, $P < 0.001$, was found (see Figure 18); however, the effect of block order, $F_{1,26} = 1.931$, $P > 0.05$, and the interaction between experimental block and block order were non-significant, $F_{4,104} = 1.636$, $P > 0.05$. Pairwise comparisons indicated that the magnitudes of the vocal responses in the 100-cent unpredictable block were more variable than in all other experimental blocks ($P < 0.05$), and that the magnitudes of the vocal responses in the 50-cent unpredictable block were also more variable than in the 100-cent predictable block ($P < 0.05$).

The RM-ANOVA investigating the influence of experimental condition on vocal response latencies showed a main effect of experimental condition, $F_{3,78} = 2.987$, $P < 0.05$. However, the main effect of block order, $F_{1,26} = 3.245$, $P > 0.05$, and the interaction between experimental condition and block order failed to reach significance, $F_{3,78} = 0.585$, $P > 0.05$. Pairwise comparisons indicated that response latencies were significantly shorter following predictable 50-cent perturbations relative to those following 100-cent perturbations in both conditions ($P < 0.05$).

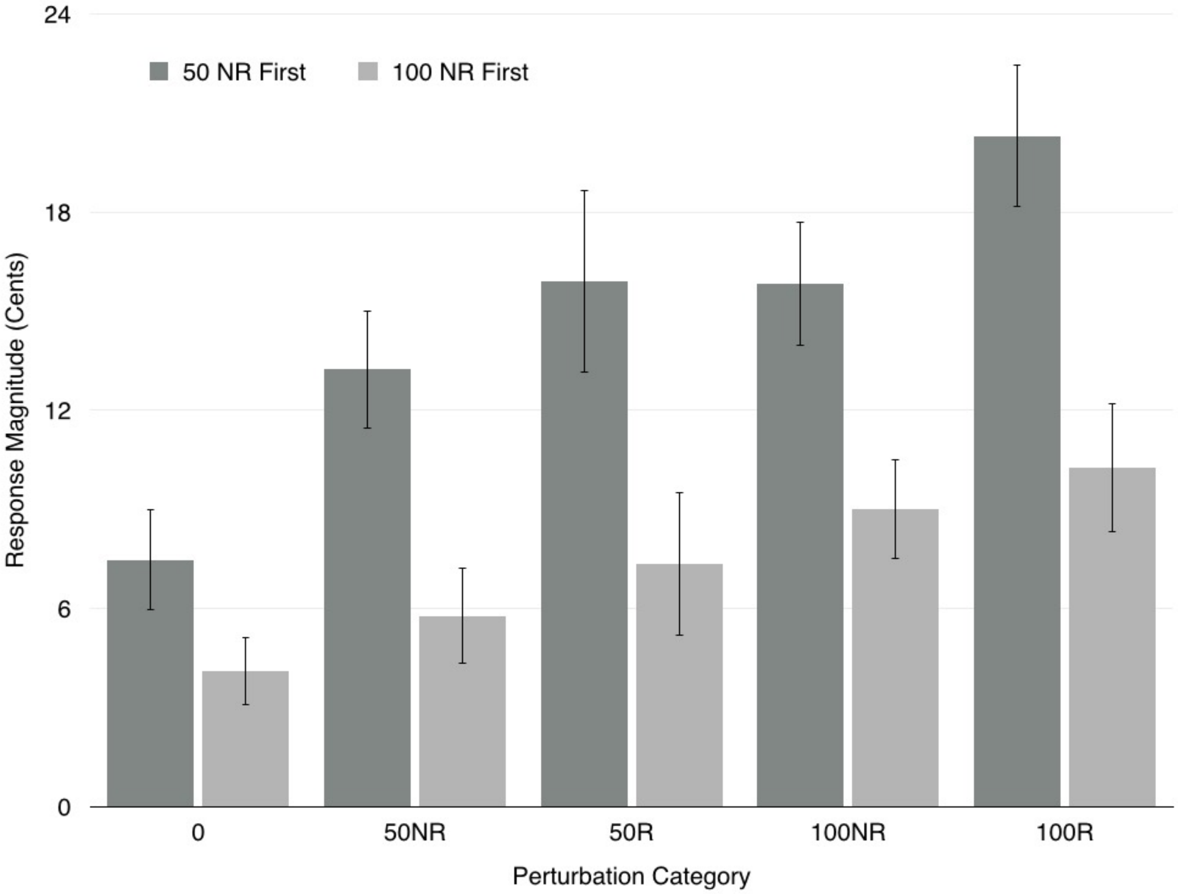

Figure 17: Averaged vocal response magnitudes in cents for all participants as a function of perturbation category (0, unaltered; 50NR, 50-cent predictable perturbations; 50R, 50-cent unpredictable perturbations; 100NR, 100-cent predictable perturbations; and 100R, 100-cent unpredictable perturbations) and block order (50 NR first, the block of 50-cent predictable perturbations prior to the block of 100-cent perturbations; and 100 NR first, 100-cent predictable perturbations prior to the block of 50-cent predictable perturbations).

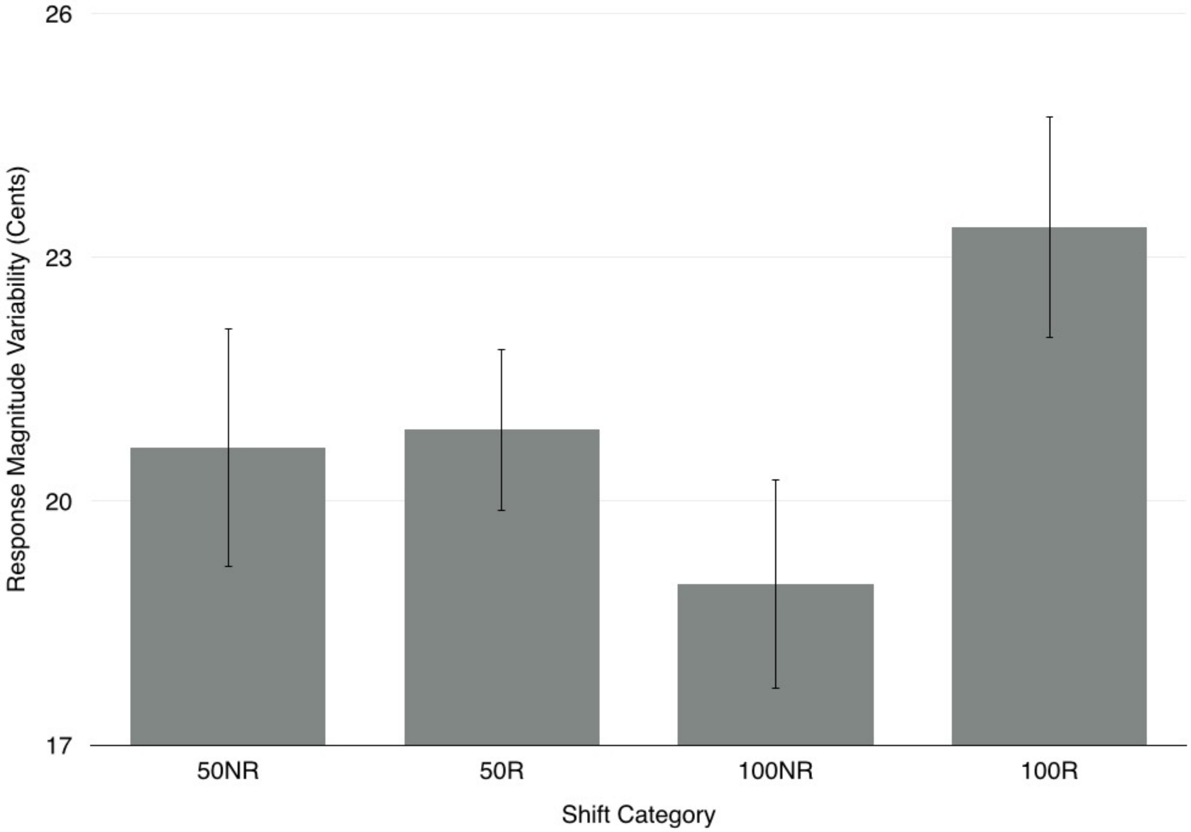

Figure 18: Averaged response magnitude variability in cents for all participants as a function of perturbation category (50NR, 50-cent predictable perturbations; 50R, 50-cent unpredictable perturbations; 100NR, 100-cent predictable perturbations; and 100R, 100-cent unpredictable perturbations).

Preceding Perturbation Magnitude Effects.

An RM-ANOVA investigating whether vocal response magnitudes were modulated by the size of the preceding perturbation showed a main effect of preceding perturbation magnitude, $F_{3,78} = 13.524$, $P < 0.001$ (see Figure 19) and a main effect of block order, $F_{1,26} = 8.992$, $P < 0.01$. However, the interaction between preceding perturbation magnitude and block order was non-significant $F_{3,78} = 0.869$, $P > 0.05$. Pairwise comparisons indicated that 100-cent perturbations always elicited larger vocal responses than 50-cent perturbations, regardless of the size of the previous perturbation ($P < 0.05$). In addition, 50-cent perturbations following 50-cent perturbations elicited larger responses than 50-cent perturbations following 100-cent perturbations ($P < 0.05$), and 100-cent perturbations following 50-cent perturbations elicited larger responses than 100-cent perturbations following 100-cent perturbations, although this result was only marginally significant ($P = 0.078$). Consistent with the perturbation magnitude predictability effects, pairwise comparisons indicated that the block order effect was driven by larger vocal responses overall when the 50-cent predictable block occurred prior to the 100-cent predictable block.

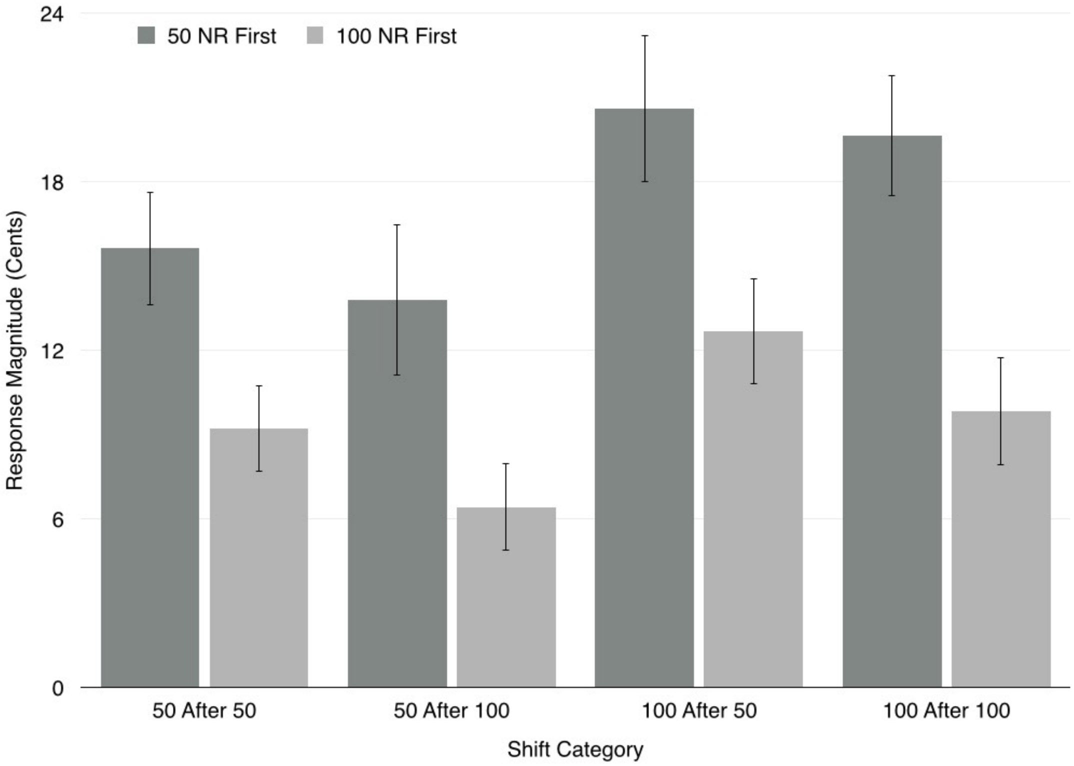

Figure 19: Averaged vocal response magnitudes in cents for all participants as a function of within-vocalization perturbation exposure (50 after 50, 50-cent unpredictable perturbations following 50-cent unpredictable perturbations; 50 after 100, 50-cent unpredictable perturbations following 100-cent unpredictable perturbations; 100 after 50, 100-cent unpredictable perturbations following 50-cent unpredictable perturbations; and 100 after 100, 100-cent unpredictable perturbations following 100-cent unpredictable perturbations) and block order (50 NR first, the block of 50-cent predictable perturbations prior to the block of 100-cent perturbations; and 100 NR first, 100-cent predictable perturbations prior to the block of 50-cent predictable perturbations).

ii. ERP Results

P1 Amplitudes and Latencies

Perturbation Magnitude Predictability Effects

A three-way RM-ANOVA was conducted to investigate the influence of experimental condition, electrode site, and block order on P1 amplitudes. The main effects of experimental condition, $F_{4,104} = 0.684$, $P > 0.05$, electrode site, $F_{2,52} = 0.683$, $P > 0.05$, and block order, $F_{1,26} = 0.020$, $P > 0.05$, all failed to reach significance. The interaction between experimental condition and block order was significant, $F_{4,104} = 3.556$, $P < 0.01$; however, all other interactions failed to reach significance ($P > 0.05$). Follow-up one-way RM-ANOVAs were conducted to investigate the influence of experimental condition on P1 amplitudes for each block order separately. Experimental condition significantly modulated P1 amplitudes when the 50-cent predictable block occurred prior to the 100-cent predictable block, $F_{4,52} = 2.733$, $P < 0.05$, but not when the 100-cent predictable block occurred prior to the 50-cent predictable block, $F_{4,52} = 1.330$, $P > 0.05$. Furthermore, pairwise comparisons indicated that for the 50-cent predictable first block order, the 100-cent unpredictable perturbations resulted in larger P1 amplitudes than the 0- and 50-cent predictable perturbations, and that the 100-cent predictable perturbations resulted in larger P1 amplitudes than the 50-cent predictable perturbations ($P < 0.05$; see Figure 20).

The three-way RM-ANOVA investigating the influence of experimental condition, electrode site and block order on P1 latency showed a main effect of experimental condition, $F_{3,78} = 5.239$, $P < 0.01$. However, all other main effects and interactions failed to reach significance ($P > 0.05$). Pairwise comparisons investigating latency differences across the experimental conditions indicated that the 50-cent unpredictable perturbations resulted in later P1s than the other FAF conditions ($P < 0.05$).

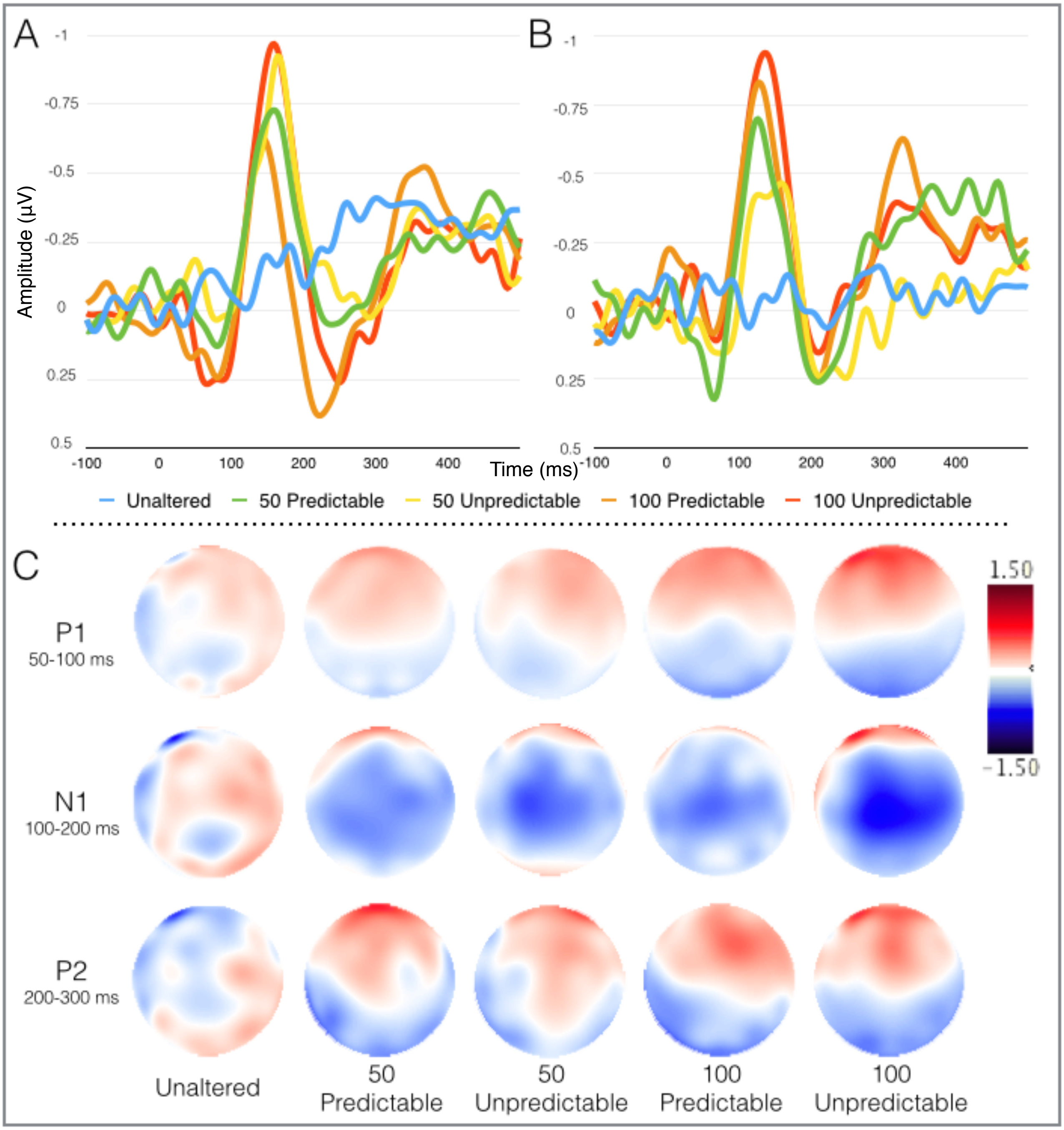

Figure 20: ERP waveforms averaged across all participants and all electrodes [left (CP1, C1, and FC1), medial (FCz, Cz, and Pz), and right (FC2, C2, and CP2)]. Separate lines represent each of the five perturbation categories (unaltered, 50-cent predictable perturbations, 50-cent

unpredictable perturbations, 100-cent predictable perturbations, and 100-cent unpredictable perturbations). (A) ERP responses when the block of 50-cent predictable perturbations was presented prior to the block of 100-cent perturbations. (B) ERP responses when the block of 100-cent predictable perturbations was presented prior to the block of 50-cent predictable perturbations. (C) The topographical scalp distribution of ERPs in all five experimental conditions, collapsed across the two block orders. The top row (P1) represents cortical activity between 50 and 100 ms, the middle row (N1) represents cortical activity between 100 and 200 ms and the bottom row (P2) represents cortical activity between 200 and 300 ms following the feedback perturbation.

Preceding Perturbation Magnitude Effects

A three-way RM-ANOVA investigating the influence of the magnitude of the preceding perturbation, electrode site, and block order on P1 amplitudes following a subsequent perturbation failed to show any main effects or interactions ($P > 0.05$).

A three-way RM-ANOVA investigating the influence of the magnitude of the preceding perturbation, electrode site, and block order on the latency of the P1 peak following a subsequent perturbation showed a main effect of preceding perturbation magnitude, $F_{3,78} = 4.573$, $P < 0.01$. All other main effects and interactions failed to reach significance ($P > 0.05$). Pairwise comparisons investigating latency differences across the preceding magnitude conditions indicated that 100-cent perturbations after 100-cent perturbations elicited faster responses than all other preceding magnitude conditions ($P < 0.05$).

N1 Amplitudes and Latencies

Perturbation Magnitude Predictability Effects

A three-way RM-ANOVA was conducted to investigate the influence of experimental condition, electrode site and block order on N1 amplitudes. Significant main effects of experimental condition, $F_{4,104} = 6.313$, $P = 0.001$, and electrode site, $F_{2,52} = 2.063$, $P < 0.05$, were found (see Figure 20). However, the main effect of block order and all interactions failed to reach significance ($P > 0.05$). Pairwise comparisons indicated that all FAF conditions elicited significantly larger N1 amplitudes (absolute value) than the unaltered condition. In addition, the 100-cent unpredictable condition elicited significantly larger N1 amplitudes (absolute value) than all other conditions ($P < 0.05$). With regard to electrode site, smaller N1 amplitudes were recorded from right lateralized electrode sites relative to central and left electrode sites ($P < 0.05$).

The three-way RM-ANOVA investigating the influence of experimental condition, electrode site and block order on N1 latency revealed main effects of experimental condition, $F_{3,78} = 4.698$, $P < 0.01$ (see Figure 20), and block order, $F_{1,26} = 8.690$, $P = 0.01$. However, the main effect of electrode site and all interactions failed to reach significance ($P > 0.05$). Pairwise comparisons indicated that N1 latencies were shorter following 50-cent random perturbations relative to all other perturbations ($P < 0.05$). Post hoc comparisons investigating block order indicated that participants who were exposed to the 50-cent predictable condition prior to the 100-cent predictable condition had later N1 latencies in all experimental conditions relative to participants who were exposed to the 100-cent predictable condition first.

Preceding perturbation magnitude effects.

A three-way RM-ANOVA investigating the influence of the magnitude of the preceding perturbation, electrode site and block order on N1 amplitudes following a subsequent perturbation showed main effects of preceding perturbation magnitude, $F_{3,78} = 3.136$, $P < 0.05$ (see Figure 21), and electrode site, $F_{2,52} = 3.737$, $P < 0.05$. However, all other main effects and interactions failed to reach significance ($P > 0.05$). Pairwise comparisons indicated that 100-cent perturbations after 50-cent perturbations elicited larger responses than 50-cent perturbations, regardless of the magnitude of the preceding perturbation ($P < 0.05$). Additionally, central electrode sites elicited larger N1 amplitudes than right lateralized sites ($P = 0.01$).

A three-way RM-ANOVA investigating the influence of the magnitude of the preceding perturbation, electrode site and block order on the latency of the N1 peak following a subsequent perturbation showed a main effect of block order, $F_{1,26} = 5.219$, $P < 0.05$. However, all other main effects and interactions failed to reach significance ($P > 0.05$). Pairwise comparisons investigating latency differences across the preceding magnitude conditions indicated that overall latencies were longer when the 50-cent predictable condition occurred first ($P < 0.05$).

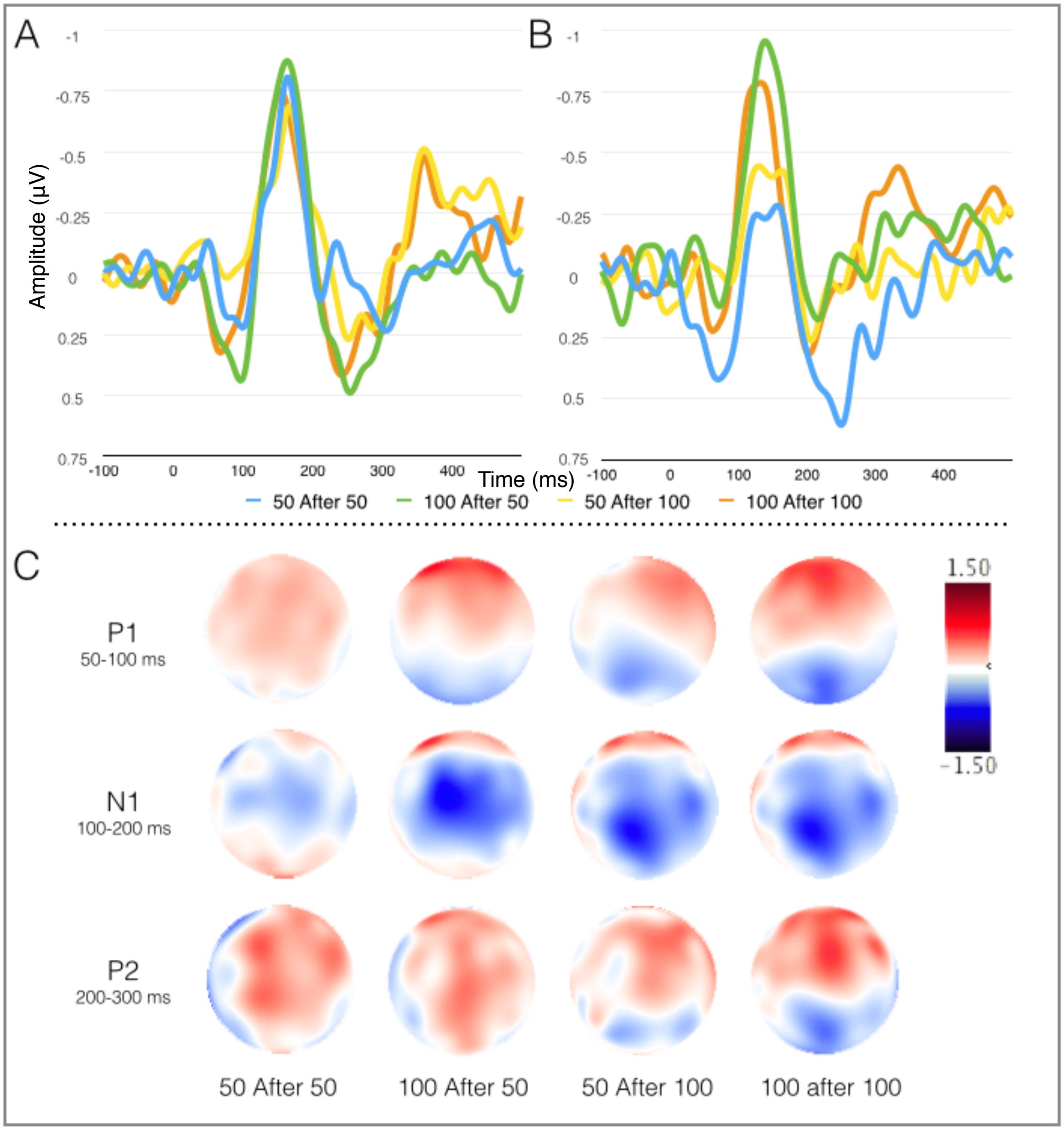

Figure 21: ERP waveforms averaged across all participants and all electrodes [left (CP1, C1, and FC1), medial (FCz, Cz, and Pz), and right (FC2, C2, and CP2)]. Separate lines represent each of the four possible within-vocalization perturbation presentation orders (50 after 50, 50-cent unpredictable perturbations following 50-cent unpredictable perturbations; 50 after 100, 50-cent unpredictable perturbations following 100-cent unpredictable perturbations; 100 after 50, 100-cent unpredictable perturbations following 50-cent unpredictable perturbations; and 100

after 100, 100-cent unpredictable perturbations following 100-cent unpredictable perturbations). (A) ERP responses when the block of 50-cent predictable perturbations was presented prior to the block of 100-cent perturbations. (B) ERP responses when the block of 100-cent predictable perturbations was presented prior to the block of 50-cent predictable perturbations. (C) The topographical scalp distribution of ERPs in all four experimental conditions, collapsed across the two block orders. The top row (P1) represents cortical activity between 50 and 100 ms, the middle row (N1) represents cortical activity between 100 and 200 ms and the bottom row (P2) represents cortical activity between 200 and 300 ms following the feedback perturbation.

P2 Amplitude and Latencies

Perturbation Magnitude Predictability Effects

A three-way RM-ANOVA was conducted to investigate the influence of experimental condition, electrode site, and block order on P2 amplitudes. The interaction between experimental condition and block order was significant, $F_{4,104} = 3.745$, $P < 0.05$ (see Figure 20). However, all other main effects and interactions failed to reach significance ($P > 0.05$). Follow-up one-way RM-ANOVAs were conducted to investigate the influence of experimental condition on P2 amplitudes for each block order separately. Experimental condition significantly modulated P2 amplitudes when the 50-cent predictable block order occurred prior to the 100-cent predictable block, $F_{4,52} = 3.640$, $P < 0.05$, but not when the 100-cent predictable block occurred prior to the 50-cent predictable block, $F_{4,52} = 0.515$, $P > 0.05$. Furthermore, pairwise comparisons indicated that both predictable and unpredictable perturbations elicited larger P2 amplitudes than the unaltered condition, and that predictable 100-cent perturbations also elicited larger P2 amplitudes than both predictable and unpredictable 50-cent perturbations.

The three-way RM-ANOVA investigating the influence of experimental condition, electrode site, and block order on P2 latency showed a main effect of electrode site, $F_{2,52} =$

7.760, $P = 0.004$. However, all other main effects and interactions failed to reach significance ($P > 0.05$). Pairwise comparison indicated that P2 responses were later at the right electrode sites relative to the central and left sites ($P < 0.01$).

Preceding perturbation magnitude effects.

A three-way RM-ANOVA investigating the influence of the magnitude of the preceding perturbation, electrode site and block order on P2 amplitudes following the subsequent perturbation failed to show any main effects or interactions ($P > 0.05$).

A three-way RM-ANOVA investigating the influence of the magnitude of the preceding perturbation, electrode site, and block order on the latency of the P2 peak following the subsequent perturbation found main effects of preceding perturbation magnitude, $F_{3,78} = 3.769$, $P < 0.05$, and electrode site, $F_{2,52} = 7.286$, $P < 0.01$. However, the main effect of block order and all interactions failed to reach significance ($P > 0.05$). Pairwise comparisons investigating latency differences across the preceding perturbation magnitude conditions indicated that 100-cent perturbations after 100-cent perturbations elicited faster responses than all other preceding perturbation magnitude conditions ($P < 0.05$). In addition, P2 responses were later at the right lateralized electrodes than at the central and left sites ($P < 0.05$).

iii. Correlational Analyses

Correlational analyses were conducted in order to determine whether P1, N1 or P2 amplitude and latency modulation was related to changes in vocal response magnitudes and latencies. The results indicated that, across the perturbation magnitude predictability conditions, vocal response magnitudes and N1 amplitudes, $r = 0.166$, $P_{\text{(two-tailed)}} < 0.05$, $n = 140$, and vocal response latencies and P2 latencies, $r = 0.213$, $P_{\text{(two-tailed)}} < 0.05$, $n = 112$, correlated significantly. However, all other correlations failed to reach significance ($P > 0.05$).

Discussion

In this study, speakers were exposed to FAF perturbations that were either predictable in magnitude or unpredictable in magnitude. Behavioural and neurological responses to these FAF perturbations were examined in order to investigate whether being able to predict the magnitude of brief (temporally unpredictable) FAF perturbations altered responses to these perceived speech production errors.

Overall, the behavioural results indicated that vocal responses to the FAF perturbations were influenced by the perturbation magnitude predictability manipulation, as unpredictable 100-cent perturbations elicited larger vocal responses than all other perturbations. In addition, unpredictable 50-cent perturbations elicited larger vocal responses than predictable 50-cent perturbations. Although this is the first study to directly investigate how manipulating the predictability of FAF perturbation magnitudes within a vocalization influences compensatory responses to these perturbations, previous studies have suggested that the magnitude of the compensatory response to FAF is related to the prediction created by a sensorimotor representation (Heinks-Maldonado et al., 2005, 2006; Beal et al., 2011; Korzyukov et al., 2012b; Scheerer et al., 2013a). Our results indicate that, similarly to the modulation by the prediction created by the sensorimotor representation, compensatory responses to FAF can be modulated by experimentally induced predictability. We hypothesize that when the magnitude of FAF perturbations is predictable, the perturbations are more readily distinguished from self-produced variability. Following repeated exposure to predictable FAF perturbations, the weighting of feedforward input is increased, as the information provided by auditory feedback becomes unreliable. As a result of the increased weighting of the feedforward input, individuals are less susceptible to deviant auditory feedback, as auditory perturbations are regarded as externally induced noise, rather than violations of the prediction created by the internal model. For this reason, compensatory responses to predictable FAF perturbations are smaller. Conversely, when FAF perturbation magnitudes are unpredictable, the unpredictable nature of these perturbations makes the deviant auditory feedback resemble normal variability in the voice, and it becomes difficult to distinguish self-produced variability from the experimentally induced perturbations. This increased vocal variability, experimentally induced or otherwise, leads to increased weighting of the feedback input. As a result, deviant auditory feedback is processed as a violation of the prediction created by the internal model, and compensatory responses are initiated in an attempt to minimize the deviation between the perceived auditory feedback and the auditory feedback predicted by the internal model. For this reason, compensatory responses to unpredictable FAF perturbations are larger. Accordingly, the results of this study indicate that both the predictable 50- and 100-cent perturbations elicited smaller vocal responses than the unpredictable 100-cent perturbations, and that the predictable 50-cent perturbations also resulted in smaller vocal response magnitudes than the unpredictable 50-cent perturbations. These results suggest that presenting brief predictable perturbations of auditory feedback may result in increased weighting of feedforward input, relative to unpredictable perturbations of auditory feedback, and smaller vocal responses.

In addition to modulating vocal response magnitudes, vocal response magnitude variability, or the size of compensatory responses from trial to trial, were also found to differ as a function of predictability. The variability of the magnitude of the compensatory responses elicited by unpredictable 100-cent perturbations was much larger than for all other perturbations. Similarly, the variability of vocal response magnitudes elicited by unpredictable 50-cent perturbations was larger than the variability of response magnitudes elicited by predictable 100-cent perturbations. Studies investigating reaching behaviours in variable environments have suggested that after exposure to perturbations of unpredictable magnitudes, future sensory predictions are based on a moving average of the magnitude of the perturbations experienced over the previous few reaches (Thoroughman & Shadmehr, 2000; Scheidt & Dingwell, 2001;

Takahashi & Scheidt, 2001). In the case of the unpredictable perturbation magnitudes in this study, utilizing information from the previous few perturbations would have resulted in highly variable sensory predictions, as the pattern of previously experienced perturbations was constantly changing. As a result of these dynamic sensory predictions, the difference between the predicted and perceived auditory feedback was constantly changing, resulting in variable trial-to-trial compensatory response magnitudes following perturbations that varied randomly in magnitude. On the other hand, when individuals were exposed to predictable FAF perturbations, increased weighting of the feedforward system would have reduced the influence of the FAF perturbations on vocal response magnitudes. Furthermore, as the predictable FAF perturbations were not regarded as self-produced, future sensory predictions would not have been influenced by the predictable FAF perturbations. As a result, smaller variability in trial-to-trial compensatory response magnitudes occurred in the predictable perturbation magnitude conditions.

Vocal response latency was also found to differ as a function of predictability in this study. Vocal responses were faster following predictable 50-cent perturbations relative to 100-cent perturbations, regardless of their predictability. The fact that responses were faster following 50-cent predictable perturbations compared to 100-cent unpredictable perturbations is consistent with the results of Burnett et al. (2008), who found that temporal predictability resulted in faster vocal responses. However, if increased predictability does in fact result in faster vocal responses, it is unclear why the 100-cent predictable perturbations did not also result in faster vocal responses relative to the unpredictable perturbations.

In addition to modulating behavioural responses, predictability was found to influence neural responses indexed by the P1–N1–P2 ERP components. It has previously been shown that these ERP components are sensitive to FAF manipulations (Heinks-Maldonado et al., 2005; Behroozmand et al., 2009; Scheerer et al., 2013a,b). However, as each component represents a different stage of auditory processing, P1–N1–P2 responses to auditory feedback manipulations are not uniform. Accordingly, the results of this study indicate that the patterns of ERP responses following predictable and unpredictable FAF perturbations were not uniform.

The P1 ERP component plays a role in detecting early changes in auditory feedback (Chait, Simon, & Poeppel, 2004; Nakagawa, Otsuru, Inui, & Kakigi, 2014). Accordingly, Scheerer et al. (2013a) found that P1 amplitudes were elicited in an all-or-nothing manner by FAF perturbations. For this reason, they suggested that the P1 ERP component is sensitive to deviant auditory feedback, but not the specific properties of the deviant auditory feedback. In this study, P1 amplitudes were not found to differ as a function of FAF perturbation magnitude predictability, however, this is not surprising in light of previous research. On the other hand, P1 latencies were influenced by the predictability of the perturbation magnitudes. P1 latencies were longest following unpredictable 50-cent perturbations, suggesting that unpredictable FAF perturbation magnitudes may result in later neural responses than predictable FAF perturbation magnitudes. However, as longer latencies were only observed for unpredictable 50-cent perturbations, and not unpredictable 100-cent perturbations, further investigation is required to elucidate the relationship between FAF magnitude predictability and P1 latencies.

Modulation of the N1 ERP component by FAF perturbation magnitude predictability was observed in this study. N1 amplitudes were largest following unpredictable 100-cent FAF perturbations. These findings are congruent with the results reported by Korzyukov et al. (2012b), who observed larger N1 amplitudes in conditions where participants were unable to predict the direction of FAF perturbations, relative to conditions where the direction of the FAF perturbations was predictable. Together, these results suggest that both the predictability of the magnitude, and the direction, of FAF perturbations can modulate N1 responses.

Previous research has indicated that the auditory N1 response can be modulated by the efference copy issued by the motor system (Houde et al., 2002; Heinks-Maldonado et al., 2005; Beal et al., 2011), neural adaptation (Jaaskelainen et al., 2004; Jaaskelainen, Ahveninen, Belliveau, Raij, & Sams, 2007), habituation (Butler, 1968), and refractory period effects (Budd, Barry, Gordon, & Rennie, 1998; Brattico, Tervaniemi, & Picton, 2003; Coch, Skendzel, & Neville, 2005), among other things. These studies have suggested that the amplitude of the N1 component reflects the amount of neural resources allocated to processing a stimulus (Brattico et al., 2003; Jaaskelainen et al., 2004, 2007; Sitek et al., 2013). Without an auditory-only listening condition, the precise mechanism driving the reduction in N1 amplitudes following predictable FAF perturbations in this study is unclear. However, as the N1 responses recorded in this study were elicited by auditory perturbations occurring within ongoing vocalizations, auditory–motor interactions are quite plausible. Congruent with this notion, a correlation between N1 amplitudes and vocal response magnitudes was found. We suggest that when the magnitude of the FAF perturbations was unpredictable, the FAF perturbations were not readily distinguished from normal variability in the voice. As a result, large N1 responses were elicited, reflecting increased neural resources being allocated to the processing of auditory feedback. This increased allocation of neural resources to the processing of auditory feedback resulted in increased feedback input for ongoing speech motor control, and larger compensatory responses. On the other hand, when the magnitude of the FAF perturbations was predictable, the perturbations were easily identified as externally induced. As a result, small N1 responses were elicited, reflecting a decrease in the neural resources being allocated to the processing of auditory feedback. This also resulted in increased weighting of feedforward input, which reduced the impact of the deviant auditory feedback, and resulted in smaller compensatory responses.

In addition to being modulated by experimental condition, N1 amplitudes were found to vary as a function of electrode site. N1 amplitudes were larger at the left and central recording sites than at the right recording sites. This result is not surprising, as previous studies have reported that N1 responses to FAF are maximal at central sites relative to all other recording sites (Behroozmand et al., 2009; Scheerer et al., 2013a).

In addition to N1 amplitude modulation, N1 latencies were also influenced by FAF perturbation magnitude predictability. N1 latencies were the longest following unpredictable 50-cent perturbations. The lengthy N1 responses following unpredictable 50-cent perturbations mirror the P1 latency results found in this study. Although, to our knowledge, the effect of predictability on neural response timing has not been previously reported, Burnett et al. (2008) reported faster vocal responses to temporally predictable stimuli. We suggest that when FAF perturbation magnitudes were unpredictable, additional processing was required in order to calculate the mismatch between the perceived auditory feedback and the prediction created by the sensorimotor representation, resulting in later responses. On the other hand, when FAF perturbation magnitudes were predictable, the FAF perturbations were deemed to be externally induced. As externally induced FAF perturbations are not compared with the prediction created by the sensorimotor representation, less processing was required, which resulted in faster responses. Accordingly, these results suggest that the predictable nature of FAF perturbations decreases neural processing time, resulting in shorter P1 and N1 response latencies.

Although the main effect of experimental condition was not significant for P2 latencies, P2 latencies were found to correlate with vocal response latencies. Previous studies have reported an association between P2 amplitudes and vocal responses to FAF, suggesting that the P2 component may reflect the computation of the mismatch between perceived and expected feedback, which dictates the size of the compensatory vocal responses to FAF (Scheerer et al., 2013a). On the basis of these results, it is not surprising that P2 latencies and vocal latencies correlate, as it would be expected that later processing of the mismatch between perceived and expected auditory feedback would also result in later compensatory vocal responses. P2 latencies were also found to vary as a function of electrode site in this study. P2 latencies were later at the right recording sites relative to the central and left recording sites. This result is consistent with previous studies that have reported later P2 latencies at right lateralized electrode sites compared to medial (Scheerer et al., 2013a) and left (Scheerer et al., 2013b) lateralized sites.

Decreases in the magnitude and trial-to-trial variability of vocal responses, smaller N1 amplitudes and shorter vocal, P1 and N1 response latencies following predictable FAF perturbation magnitudes support the notion that experimentally induced predictability can modulate responses to FAF. Furthermore, correlations between vocal response magnitudes and N1 amplitudes, and between vocal response latencies and P2 latencies, suggest that the P1–N1–P2 ERP complex may reflect changes in auditory processing that are related to speech motor control. As previously discussed, we suggest that increasing the predictability of the magnitude of FAF perturbations makes these perturbations easier to distinguish from internally induced vocal variability. After repeated exposure to these predictable FAF perturbations, the weighting of the feedforward system increases. As a result of the increased weighting of the feedforward control system, deviant auditory feedback is less salient, resulting in smaller responses, both behaviourally and neurologically. Modifying the weighting of the feedback and feedforward control systems in different contexts is physiologically advantageous. Increasing the weighting of the feedback system is advantageous in situations where the information from auditory feedback is reliable and can be used to update the mapping of sensorimotor representations. For example, throughout development, auditory feedback is required to maintain the mapping of the sensorimotor representations as growth-related changes to the articulators, vocal folds, musculature, and lung capacity occur (Bailly, 1997; Callan et al., 2000; Jones & Munhall, 2000, 2002; Guenther, 2006; Civier et al., 2010). Even in adulthood, increased weighting of the feedback control system can help to rapidly update the mapping of the sensorimotor representations following the acquisition of dental appliances, or as aging-related changes to the articulators, vocal folds, musculatures and lung capacity occur (Jones & Munhall, 2003; Perkell, 2012). On the other hand, increased weighting of the feedforward system can also be physiologically advantageous. As development halts, the mapping of the sensorimotor representations should remain relatively stable. For this reason, the additional information provided by auditory feedback becomes redundant. Increasing the weighting of the feedforward control system not only increases the fluidity of speech, but also reduces susceptibility to externally generated noise, and frees attentional resources for the processing of potentially important stimuli, rather than predictable auditory errors (Heinks-Maldonado et al., 2005; Wang et al., 2014).

Although the primary aim of this study was to investigate the influence of being able to predict FAF perturbation magnitudes on vocal and neural responses to these perturbations, relative to unpredictable FAF perturbation magnitudes, the experimental design also permitted the investigation of order effects. More specifically, by presenting blocks of predictable and unpredictable FAF perturbation magnitudes, the effects of being exposed to consistent and variable stimulus magnitudes on vocal and neural responses to subsequent FAF perturbations could be investigated. In addition, within the unpredictable FAF perturbation magnitude blocks, perturbation magnitudes were varied within each utterance. This manipulation made it possible to investigate the influence of exposure to smaller (50-cent) vs. larger (100-cent) perturbations on subsequent perturbations.

The results of these investigations indicated that vocal response magnitudes varied as a function of the block presentation order. When the block of predictable 100-cent perturbations occurred prior to the block of predictable 50-cent perturbations, overall vocal response magnitudes were smaller. We suggest that repeated exposure to predictable 100-cent perturbations increased the weighting of the feedforward system, resulting in smaller compensatory responses to these perturbations. As perturbations of up to 100 cents in magnitude became expected, weighting of the feedforward system increased, decreasing the saliency of perturbations of 100 cents and smaller. As a result, exposure to 50-cent perturbations following predictable exposure to 100-cent perturbations resulted in smaller vocal response magnitudes, as the 50-cent perturbations were less salient. On the other hand, when the block of predictable 50-cent perturbations occurred prior to the block of predictable 100-cent perturbations, overall vocal response magnitudes were larger. As previously discussed, we suggest that repeated exposure to predictable 50-cent perturbations increased the weighting of the feedforward system, resulting in smaller compensatory responses to these perturbations. In this case, perturbations of up to 50 cents in magnitude became expected, decreasing the saliency of perturbations of 50 cents and smaller. However, with exposure to 100-cent perturbations following predictable exposure to 50-cent perturbations, larger compensatory responses were elicited, as the 100-cent perturbations were more salient. Together, these block order effects suggest that exposure to 50-cent perturbations prior to 100-cent perturbations increases the saliency of the 100-cent perturbations, whereas exposure to 100-cent perturbations prior to 50-cent perturbations only reduces the saliency of 50-cent perturbations.

As the presentation order of 50- and 100-cent perturbations modulated vocal responses across blocks, we also investigated the influence of the presentation order of 50- and 100-cent perturbations within single vocalizations in the unpredictable FAF perturbation magnitude blocks. Similarly to the previously reported block effects, vocal response magnitudes were found to vary as a function of the magnitude of the preceding perturbation. Larger compensatory responses were recorded to 50- and 100-cent perturbations that occurred following a 50 cent perturbation than to those following a 100-cent perturbation. This result provides further support for the notion that exposure to large perturbations reduces the saliency of subsequent perturbations of equal or smaller size, whereas exposure to smaller perturbations does not reduce the saliency of subsequent perturbations that are larger in size. More importantly, these results suggest that exposure to a single perturbation can affect responses to subsequent perturbations.

Although P1 amplitudes were not modulated by FAF perturbation magnitude predictability in this study, they were modulated by the interaction between experimental condition and block order. More specifically, when the block of predictable 50-cent perturbations occurred prior to the block of predictable 100-cent perturbations, P1 amplitudes were larger for 100-cent perturbations. On the other hand, when the block of predictable 100-cent perturbations occurred prior to the block of predictable 50-cent perturbations, P1 amplitudes did not differ across the experimental conditions. As previous studies have suggested that the P1 component is not sensitive to the magnitude of feedback errors, it is currently unclear why these differences were observed in this study (Scheerer et al., 2013a).

The influence of the magnitude of the preceding perturbation on the subsequent perturbation was also analyzed with regard to P1 latencies. The results indicated that P1 responses to 100-cent perturbations occurring after 100-cent perturbations were the fastest overall. Previous studies have suggested that larger amplitude perturbations result in faster neural responses, which is consistent with 100-cent perturbations following 100-cent perturbations resulting in faster responses overall (Liu et al., 2011; Scheerer et al., 2013a,b).

Investigation of the influence of block order on N1 amplitudes revealed no significant findings, however, N1 latencies were modulated by the block presentation order. N1 latencies were later overall when participants were exposed to the block of predictable 50-cent perturbations prior to the block of predictable 100-cent perturbations. These latency effects mirror the block order effects observed for vocal response magnitudes. As previously discussed, repeated exposure to predictable 100-cent perturbations, prior to predictable exposure to 50-cent perturbations, resulted in smaller vocal responses overall. We suggest that this is because predictable exposure to the 100-cent perturbations increased the weighting of the feedforward system, and also rendered feedback errors of 100 cents and smaller less salient. On the other hand, when exposure to the predictable 50-cent perturbations occurred first, increased weighting of the feedforward system also occurred, however, in this situation only the 50-cent perturbations became less salient. For this reason, we suggest that when the block of 50-cent predictable perturbations occurred prior to the block of 100-cent predictable perturbations, more processing was required, resulting in later responses than when the block of 100-cent predictable perturbations occurred prior to the block of 50-cent predictable perturbations. Interestingly, the same effect was observed within vocalizations in the blocks of unpredictable perturbation magnitudes. The fact that the block order effect was still significant when only the unpredictable perturbation magnitude blocks were analyzed suggests that the effect was not solely driven by the predictable perturbation blocks, rather, the effects of the predictable blocks carried over to the unpredictable blocks.

Although N1 amplitudes were not influenced by the block presentation order, within vocalizations in the blocks of unpredictable perturbation magnitudes, N1 amplitudes were influenced by the magnitude of the previously experienced perturbation. Within the blocks of unpredictable perturbation magnitudes, 100-cent feedback perturbations following 50-cent feedback perturbations elicited larger responses than 50-cent perturbations, regardless of the size of the feedback perturbation that they followed. The fact that 100-cent perturbations elicited larger N1 responses than 50-cent perturbations is not surprising, given the previously discussed N1 amplitude results and previous research (Liu et al., 2011; Scheerer et al., 2013a). However, the fact that 100-cent perturbations following 50-cent perturbations, but not 100-cent perturbations following 100-cent perturbations elicited larger N1 amplitudes than 50-cent perturbations, further supports the notion that exposure to large perturbations reduces the saliency of subsequent perturbations of equal or smaller size, resulting in smaller responses. In addition to modulation by the size of the previous perturbation magnitude, N1 amplitudes were also larger at central recording sites, relative to right recording sites. This result is congruent with the N1 laterality effects previously discussed.

Much like the results obtained for P1 amplitudes, P2 amplitudes were modulated by the block presentation order. When the block of predictable 50-cent perturbations was presented prior to the block of predictable 100-cent perturbations, P2 amplitudes elicited by 100-cent perturbations were larger overall. In addition, when the block of predictable 50-cent perturbations was presented prior to the block of predictable 100-cent perturbations, predictable 100-cent perturbations elicited larger P2 responses than both 50-cent conditions, as well as the unaltered condition. The unpredictable 100-cent perturbations also elicited larger P2 responses than the unaltered condition. However, when the block of 100-cent predictable perturbations occurred prior to the block of 50-cent predictable perturbations, there were no P2 differences across the experimental conditions. Previous studies have shown that P2 amplitudes increase in a linear fashion as the magnitude of feedback errors increases (Behroozmand et al., 2009; Scheerer et al., 2013a). For this reason, it has been suggested that P2 amplitudes reflect the size of the mismatch between perceived and predicted auditory feedback. As previously discussed, when the block of predictable 50-cent perturbations was presented prior to the block of predictable 100-cent perturbations, subsequent 100-cent perturbations were perceived as substantial deviations from the prediction created by the internal model, resulting in relatively large P2 amplitudes. On the other hand, when the block of predictable 100-cent perturbations occurred prior to the block of predictable 50-cent perturbations, the predictable nature of the 100-cent perturbations resulted in increased weighting of the feedforward system, and attenuated responses to deviant auditory feedback. As a result, subsequent exposure to both 50- and 100-cent perturbations resulted in smaller P2 amplitudes, as the perturbations were processed as externally induced errors, rather than as violations of the expectation created by the internal model.

In addition to P2 amplitude effects, investigation of the influence of the preceding perturbation magnitudes in the blocks of unpredictable perturbations revealed that P2 responses to 100-cent perturbations occurring after 100-cent perturbations were the fastest overall. This result mirrors the findings for P1 latency, and is congruent with previous findings that larger magnitude feedback perturbations result in faster neural responses (Liu et al., 2011; Scheerer et al., 2013a,b). In addition, similarly to the laterality effects previously discussed, P2 latencies were found to be later at right recording sites, relative to left and central recording sites.

Although the primary goal of this study was to investigate the effects of predictable FAF perturbation magnitudes on vocal and neural responses to FAF, the experimental design also permitted the investigation of FAF perturbation magnitude order effects. Smaller compensatory responses, smaller P1 and P2 amplitudes, and shorter N1 latencies when the block of predictable 100-cent perturbations occurred prior to the predictable block of 50-cent perturbations suggest that exposure to large perturbations may habituate responses to subsequent perturbations of equal or smaller size, resulting in smaller and faster responses. This exposure effect was also evident within vocalizations in the blocks of unpredictable perturbation magnitudes, where compensatory responses and N1 latencies were longer when the block of predictable 100-cent perturbations occurred prior to the predictable block of 50-cent perturbations. In addition to replicating the block order effects, these within-vocalization effects also demonstrate that exposure to a single perturbation can affect responses to subsequent perturbations.

Despite the fact that in everyday life we function in dynamic and unpredictable environments, with the aid of sensory feedback, we are able to maintain fluent speech. The aim of the current study was to investigate whether predictable sensory feedback, specifically auditory feedback, is processed differently than unpredictable auditory feedback at both a behavioural and a neural level. We found that exposing participants to predictable auditory feedback errors resulted in faster, smaller, and less variable behavioural and neural responses than exposing them to unpredictable auditory feedback errors. Without a listening-only condition, we are unable to isolate the specific contribution of the motor system to the pattern of ERP responses reported in this study. However, the parity between the vocal and ERP responses in this study supports the notion that experimentally induced predictability can modulate the relative contribution of auditory feedback to ongoing speech motor control. Increasing the weighting of feedback input in situations where it is providing reliable information can allow for rapid error correction, and, by providing information for updating the sensorimotor representations, can increase the accuracy of subsequent vocal productions. On the other hand, increasing the weighting of feedforward input in situations where auditory feedback is unreliable or redundant can minimize unnecessary vocal adjustments, and can free attentional resources for the processing of potentially important stimuli, respectively (Heinks-Maldonado et al., 2005; Wang et al., 2014).

General Discussion

Neurocomputational models of speech motor control (Guenther, 1994, 1995, 2006; Guenther et al., 1998, 2006; Tourville et al., 2011; Hickok et al., 2011; Guenther & Vladusich, 2012; Hickok, 2012; Houde & Chang, 2015) and experimental evidence (Jones & Munhall, 2005; Jones & Keough, 2008; Hawco & Jones, 2009; Keough & Jones, 2009, 2011; Scheerer & Jones, 2012) both support the notion that fluent speech production relies on the combined efforts of a feedback and a feedforward control system. Children born deaf often fail to acquire fluent speech (Smith, 1975; Svirsky et al., 2004), yet once speech is acquired adults who lose their hearing are able to maintain relatively fluent speech (Cowie et al., 1982; Goehl & Kaufman, 1984; Cowie & Douglas-Cowie, 1992). For this reason, it appears that the relative contribution of feedback and feedforward control change throughout the lifespan. The aim of this thesis was to examine the role of auditory feedback for speech motor control throughout development, as well as the factors that may influence the role of auditory feedback for speech motor control in adulthood.

Significant Findings

i. Study 1

The first study presented in this thesis examined the role of auditory feedback for speech motor control in toddlers. Despite behavioural, clinical, and neurophysiological evidence demonstrating that auditory feedback aids in the acquisition and lifelong maintenance of the sensorimotor mechanisms that support fluent speech production in both humans and songbirds (Guenther, 1994, Perkell et al., 1997, Doupe & Kuhl, 1999, Callan et al., 2000, Sakata & Brainard, 2008, Civier et al., 2010; Kelly & Sober, 2014), a study by MacDonald and colleagues (2012) reported that when exposed to altered auditory feedback regarding their formant frequencies, toddlers did not modify their vocal output to compensate for the deviant auditory feedback. Based on these results the authors of that study suggested that the speech motor control system of toddlers may “suppress” the influence of auditory feedback, especially when speech is not being produced in a social context (MacDonald et al., 2012). While this explanation is certainly plausible, as social reinforcement has been shown to influence the rate of speech acquisition (Benoit et al., 1996; Kuhl & Metzoff, 1996, 1997; Locke & Snow, 1997; Goldstein et al., 2003; Kuhl, 2004), it is hard to imagine how toddlers can acquire the sensorimotor associations required for feedforward control of speech without auditory feedback. For this reason, we exposed toddlers to brief perturbations of the F0 of their auditory feedback. The toddlers compensated for these perturbations by changing the F0 of their vocalization in the opposite direction of the perturbation, producing significantly different responses to upwards and downwards perturbations. The results of this study provide empirical support for the notion that toddlers use auditory feedback to regulate their speech motor commands.

In addition to demonstrating that toddlers use auditory feedback to regulate their speech motor commands, the results of this study also suggest that vocal variability may decrease with age. In order to allow for a meaningful comparison with the results of MacDonald and colleagues (2012), the toddlers tested in this study were divided into two age groups: younger toddlers who were between 24 and 35 months, and older toddlers who were between 40 and 46 months. While both groups of toddlers were found to produce similar sized compensatory responses, the older toddlers were found to be less variable than the younger toddlers. Based on these results we suggest that a decrease in vocal variability may reflect a gradual maturation of the speech motor control system with development, and thus may potentially serve as an index of speech proficiency.

Lastly, the fact that toddlers produced compensatory responses to changes to the F0 of their vocalizations, but not manipulations of the formant frequencies of their vocalizations (MacDonald et al., 2012), suggests that the postural and phonemic settings of speech may be controlled independently. This result is in accordance with a previous study that demonstrated that formant frequency, a property of auditory feedback related to phonemic control, is less sensitive to the loss of auditory feedback, relative to F0, a property related to postural control (Perkell et al., 1997). Since toddlers appear to be able to modify their auditory feedback in response to changes in F0, but not formant frequencies, these results also suggest that the ability to regulate F0 may develop prior to the ability to regulate formant frequencies. We propose that toddler’s increased proficiency at regulating the postural properties of speech, specifically F0, may be the result of exposure to IDS. Early in development, caregivers use IDS to communicate with prelinguistic children (Fernald, 1993, Saint-Georges et al., 2013). Since this IDS provides an effective way to convey affect and intentions with these prelinguistic children (Saint-Georges et al., 2013), we believe IDS may increase the salience of prosodic features of speech, such as F0. As a result of this increase salience, toddlers may prioritize the monitoring and correcting of prosodic changes in their speech over non-prosodic features of the speech signal, such as formant frequencies.

In conclusion, the results of this first study not only provide empirical support for the notion that toddlers use auditory feedback for the regulation of their speech motor commands, but these results also suggest that vocal variability may serve as an index of speech proficiency, and that the control of postural aspects of the speech signal, such as F0, may develop prior to control of the phonemic aspects, such as formant frequencies. Since risk factors for many speech disorders often appear early in childhood (Schneider et al., 2010), it is important to understand the normal development of the speech motor control system. Thus understanding the normal development of auditory feedback control of speech may have future applications in the diagnosis of disorders with known disturbances in vocal control, such as autism spectrum disorders, which are lacking viable diagnostic markers (Filipek et al., 2000; Russo, Larson, & Kraus, 2008).

ii. Study 2

The first study presented in this thesis provides support for the notion that auditory feedback is important during the development of speech motor control. However, it failed to provide information regarding the time course of the development of auditory feedback control of speech motor control. For this reason, the second study presented in this thesis examined the time course of auditory feedback control of speech using both behavioural and EEG measures. Behaviourally, vocal responses to FAF were modulated as a function of age, as both response latency and vocal variability decreased with age. There was also a correlation between vocal variability and vocal response magnitude, further highlighting the relationship between speech motor control and responses to FAF. Neurological changes in response to the FAF were also observed, as the amplitude and latency of the P1-N1-P2 ERP components were found to vary as a function of age. Changes to the P1 and P2 components were consistent with previous AEP research, and did not correlate with the behavioural measures, suggesting the modulation of these components reflects age related changes in auditory processing. On the other hand, N1 amplitude was found to correlate with vocal variability, suggesting the changes in N1 amplitudes across age groups was not purely the result of developmental changes in auditory processing.

Based on the results of previous studies (Liu et al., 2010a), it was expected that the magnitude of the vocal responses to FAF would decrease with age, reflecting increased speech motor control, and a transition from feedback to feedforward speech motor control. However, this was not the case. Vocal variability, rather than vocal response magnitudes, was found to decrease with age. In addition, vocal variability was found to predict both vocal response magnitudes and N1 amplitudes. So although vocal variability was also found to decrease with age, the results of this study suggest that vocal variability, rather than age, is best for predicting responses to FAF.

In addition to vocal variability, vocal response latencies were also found to decrease with age. Changes in response latencies have been suggested to reflect neurophysiological maturation (Rojas et al., 1998; Kotecha et al., 2009). Specifically, latencies changes are thought to reflect changes in synaptic density and efficacy in the auditory cortex (Eggermont, 1989). For this reason, the modulation of the vocal response latencies as a function of age found in this study suggests that vocal response latency may be useful for indexing the maturation of the speech motor control system.

One of the more interesting findings of this study is that vocal variability predicted N1 amplitudes. Previous studies have suggested that the N1 component is sensitive to whether auditory feedback is deviant, but still considered self-produced, or deviant, but large enough to be classified as externally produced (Scheerer et al., 2013a). Based on the current findings, we suggest that the N1 component is also modulated by the proficiency of speech motor control. As previously discussed, increased stability of one's sensorimotor representations drives a transition from feedback to feedforward control (Schmidt & Lee, 2005; Civier et al., 2010). We suggest that individuals who are more variable, and thus rely more on feedback control, are likely to accept a wider range of deviant auditory feedback as internally produced. On the other hand, individuals who are less variable, and thus rely more on feedforward control, are likely to accept a smaller range of deviant auditory feedback as internally produced as they have more robust sensorimotor representations. As a result, individuals who are more variable will be more likely to accept a feedback deviation as internally produced, relative to individuals who are less variable. As a result, the same sized feedback perturbation may elicit only a small N1 response in individuals who are highly variable, but a large response in individuals who are less variable. This relationship between N1 amplitude and vocal variability highlights the complex interaction between vocal and neurological responses to FAF.

In conclusion, the results of the second study demonstrate age-related differences in both behavioural and neurological responses to FAF. These results indicate that the auditory feedback system undergoes robust changes with age and physiological development. Although many of the results appear to be attributable to developmentally related changes in auditory processing, the relationship between vocal variability and N1 amplitudes suggest that a complex interaction between vocal and neural responses to FAF exists. Furthermore, the fact that vocal variability, rather than age, predicts the magnitude of vocal responses to FAF suggests that the maturation of the speech motor control system is not strictly dependent on age.

iii. Study 3

The first two studies reported in this thesis investigated children's responses to brief FAF perturbations. While examining responses to FAF perturbations can provide valuable information about the way in which the feedback component of the speech motor control system operates, these responses provide little information about feedforward control. For this reason, study 3 utilized the FAF adaptation paradigm to investigate children and adults responses to predictable changes in their auditory feedback. During this study participants produced vocalizations while they heard the pitch of their auditory feedback predictably or unpredictably shifted. Participants' vocal pitch was measured at the beginning of each vocalization, before auditory feedback was available, which allowed the extent to which the deviant auditory feedback modified subsequent speech motor commands to be assessed. Sensorimotor learning was observed for both the children and the adults, as all participants were found to increase their vocal pitch following trials where they were exposed to the predictable FAF manipulation, but not when they were exposed to the unpredictable FAF manipulation. When a speaker's F0 values change at the beginning of their vocalizations following exposure to a predictable FAF manipulation, these changes indicate that the way in which the speaker's speech motor commands were executed changed as a result of exposure to the deviant auditory feedback. In other words, these results suggest that the speaker used their auditory feedback for motor planning. Participants' vocal pitch was also measured across each vocalization, to index the extent to which deviant auditory feedback was used to modify ongoing vocalizations. Both children and adults were found to increase their vocal pitch following predictable and unpredictable changes to their auditory feedback, however, adults were found to produce larger compensatory responses. Lastly, consistent with study 1 and 2, vocal variability was found to decrease with age, as adults were less variable than children.

The fact that adults were found to produce larger compensatory responses in this study was unexpected. However, although previous studies have shown that children and adults produce similar sized responses to FAF, these studies utilized brief FAF perturbations (Liu et al., 2013; Scheerer et al., 2013b). These perturbation studies modified the speaker's auditory feedback mid-utterance, while the current study modified the speaker's auditory feedback at vocalization onset. A study by Hawco and Jones (2009) demonstrated that responses to feedback manipulations at vocalization onset and those induced mid-utterance are different. They stated that manipulations at voice onset produce larger responses because auditory feedback at voice onset is compared to the expectation dictated by the speaker's sensorimotor representation (Hawco & Jones, 2009). On the other hand, manipulations that occur mid-utterance produce smaller responses because mid-utterance auditory feedback is compared to the F0 of the vocalization just prior to the manipulation in attempt to stabilize the vocalization (Hawco & Jones, 2009). Based on these findings we suggest that the current results reflect the fact that adults are more proficient at comparing incoming auditory feedback with the feedback predicted by their sensorimotor representations, as a result of possessing more precisely mapped sensorimotor representations.

In conclusion, the results of study 3 demonstrate that both children and adults can rapidly integrate information derived from their auditory feedback to modify subsequent speech motor commands. Based on these results it appears as though once speech is acquired, speakers are proficient at using sensory information to modify the planning of future motor commands. Although this result was not predicted, since auditory feedback has been suggested to function as an adaptive signal that guides the movements of the articulators during development (Callan et al., 2000), it would make sense that the ability to use auditory feedback for sensorimotor learning would develop at a young age. The results of this study also provide support for the notion that adults have more finely tuned sensorimotor representations, which makes it easier for adult speakers to resolve the discrepancy between deviant auditory feedback and the feedback predicted by their sensorimotor representation.

iv. Study 4

A relationship between vocal variability and responses to FAF was highlighted in each of the first three studies presented in this thesis. The fourth study presented in this thesis specifically investigated the relationship between vocal accuracy, vocal variability, and compensatory responses to FAF. The results of this study indicated that vocal variability, but not vocal accuracy, predicted the size of compensatory responses to FAF. We believe this is an important finding, as it suggests that individuals with poor vocal control, demonstrated by increased vocal variability, respond differently to FAF. Moreover, this result suggests that these individuals with poor vocal control rely more on auditory feedback to ensure that production errors are detected and corrected for. As a result of this increased reliance on auditory feedback, individuals with poor vocal control are more susceptible to deviant auditory feedback. On the other hand, individuals with better vocal control, reflected by less vocal variability, rely less on auditory feedback and more on their well tuned sensorimotor representations. As a result of this increased reliance on sensorimotor representations, or feedforward control, it is easier for these individuals to ignore deviant auditory feedback.

In conclusion, the results of study 4 indicate that vocal variability may be a useful index for predicting the size of compensatory responses to FAF. Since the size of a speaker's compensatory response may be an indication of their reliance on feedback control, vocal variability may also provide an index of a speaker's reliance on auditory feedback for speech motor control.

v. Study 5

The results of the fourth study reported in this thesis indicated that vocal variability may be a good indication of a speaker's reliance on auditory feedback for speech motor control. In order to validate this finding, we investigated the relationship between vocal variability and responses to FAF using an adaptation paradigm. Previous research has suggested that individuals who rely more on feedforward control produce smaller compensatory responses, but show larger after-effects when exposed to persistent changes to the pitch of their auditory feedback (Jones & Keough, 2008). On the other hand, individuals who rely more on feedback control produce larger compensatory responses and smaller after-effects when exposed to persistent changes to the pitch of their auditory feedback (Jones & Keough, 2008). If vocal variability can in fact be used to index a speaker's reliance on auditory feedback, we expected that highly variable speakers would produce larger compensatory responses and small after-effects, while less variable speakers would produce smaller compensatory responses and large after-effects when participating in this FAF adaptation study. The results of this study indicated that speakers who were more variable did in fact produce larger compensatory responses to the FAF, providing support for the notion that increased vocal variability is related to an increased reliance on auditory feedback. In addition, as expected individuals who were less variable also produced larger after-effects when their auditory feedback was returned to baseline following persistent exposure to the FAF. This result suggests that speakers who are less variable rely more on feedforward control during speech production.

Together, the results of study 4 and study 5 provide a strong argument for the notion that vocal variability can be used to index a speaker's reliance on auditory feedback for speech motor control. Based on the results of these studies, we can conclude that vocal variability is one of the factors that dictates the relative weighting of feedback and feedforward control.

vi. Study 6

The first five studies presented in this thesis all identified a relationship between vocal variability and the use of auditory feedback for speech motor control. Overall, it appears as though speakers who demonstrate less vocal variability have a decreased reliance on auditory feedback for fluent speech production. Based on these results we have suggested that vocal consistency, whether achieved throughout development or as a result of intense vocal training, results in more stable sensorimotor representations that are better at predicting the sensory consequences of upcoming motor movements. As the ability to predict future events increases as events become more frequent and consistent, we believe it is reasonable to expect that varying any aspect of the sensory or motor system in a predictable manner should increase the weighting of feedforward control, and habituate responses to deviant auditory feedback. To test this theory, in study 6, we exposed speakers to FAF perturbations that were either predictable or unpredictable in magnitude, and measured the speaker's vocal and ERP responses to these perturbations. The results revealed decreases in the magnitude and trial-to-trial variability of vocal responses, smaller N1 amplitudes, and shorter vocal, P1 and N1 response latencies following predictable FAF perturbation magnitudes. As expected, this pattern of results suggests that after repeated exposure to predictable FAF perturbations, the contribution of the feedback control system decreases.

Previous research has demonstrated that the magnitude of the compensatory response to FAF is related to the prediction created by a sensorimotor representation (Heinks-Maldonado et al., 2005, 2006; Beal et al., 2011; Korzyukov et al., 2012b; Scheerer et al., 2013a). The results of the current study suggest that similarly to the modulation by the prediction created by the sensorimotor representation, compensatory responses to FAF can be modulated by experimentally induced predictability. Based on the pattern of results observed, we suggest that when FAF perturbations are predictable, these perturbations are more readily distinguished from self-produced variability. Following repeated exposure to predictable FAF perturbations, it is clear that the information provided by auditory feedback is unreliable, and thus the weighting of feedforward input is increased. When the weighting of the feedforward input is increased, auditory feedback is less salient, thus speaker's are less susceptible to the deviant auditory feedback. Since auditory perturbations are now regarded as externally induced noise, rather than violations of the prediction created by the sensorimotor representation, compensatory responses to predictable FAF perturbations are smaller. In addition, fewer neural resources are dedicated to processing the deviant auditory feedback, resulting in smaller ERP amplitudes. Conversely, when FAF perturbation magnitudes are unpredictable, the deviant auditory feedback resembles normal variability in the voice, and it becomes difficult to distinguish self-produced variability from the experimentally induced perturbations. Much like the increased variability reported in the first 5 studies, the experimentally induced variability in this study resulted in increased weighting of the feedback control system. Consequently, deviant auditory feedback was processed as a violation of the prediction created by the sensorimotor representation, and a compensatory responses was initiated in an attempt to minimize the deviation between the perceived auditory feedback and the auditory feedback predicted by the sensorimotor representation. As a result of the increased weighting of the feedback system, vocal and ERP responses to these unpredictable perturbations were larger.

Interestingly, in addition to the predictability effects, strong block order and stimulus presentation order effects were also found. Specifically, smaller compensatory responses, smaller P1 and P2 amplitudes, and shorter N1 latencies were found when the block of predictable 100-cent perturbations occurred prior to the block of predictable 50-cent perturbations. These results suggest that exposure to large perturbations habituates responses to subsequent perturbations of equal or smaller size. Similarly, exposure to a 100-cent perturbation prior to a 50-cent perturbation within a vocalization decreased the magnitude of vocal and N1 responses, but increased P1 and P2 latencies. Thus, exposure to a single perturbation can affect responses to subsequent perturbations.

In conclusion, the results of study 6 indicate that being able to predict the size of an error in one's auditory feedback modulates both the vocal and ERP responses to the error. Furthermore, the results of this study suggest that experimentally induced predictability and variability can be used to induce increases in feedforward and feedback control, respectively.

Major Themes

i. Vocal Variability

Across all 6 studies reported in this thesis, a robust relationship between vocal variability and responses to FAF was identified. Based on these findings it appears as though vocal variability, whether naturally occurring or experimentally induced, can serve as an indication of the relative importance of auditory feedback control during speech production. Interestingly, a similar pattern of results has also been observed in songbirds. Kelly and Sober (2014) demonstrated that the size of songbirds' compensatory responses to changes in their auditory feedback were best predicted by the animals' baseline pitch distributions. When the songbirds' baseline pitch distributions overlapped with the experienced auditory feedback errors, larger compensatory responses were observed. Based on these results, Kelly and Sober (2014) suggested that when baseline pitch variability is high, the probability that a feedback manipulation will fall within the normal distributions of produced pitches increases, making it more plausible that a compensatory response will be elicited. This explanation is consistent with our suggestion that individuals who are more variable are more likely to accept deviant auditory feedback as self-produced, thus resulting in the production of a compensatory response. Together with the results reported in this thesis, these findings further support the notion that increased variability promotes the use of auditory feedback for speech motor control.

Both the results reported in this thesis, and those reported by Kelly and Sober (2014) suggest that individuals who are more variable produce larger compensatory responses to deviant auditory feedback. However, it is important to note that with the exception of study 6, these studies describe variability as a global trait, making comparisons across speakers who differ in their degree of variability; these studies are not indicative of how a single instance of increased variability may influence speech motor control. Since study 6 examined within-subject differences in processing predictable and unpredictable feedback perturbations, it was possible to derive some information about how a variable (unpredictable) vocalization is processed relative to a predictable vocalization. The results of study 6 demonstrated that increased variability resulted in increased vocal responses, trial-to-trial variability, and increased N1 amplitudes. Based on these results, we concluded that larger N1 and vocal responses were indicative of a mechanism that detects and corrects for mismatches between the perceived auditory feedback and the prediction created by the sensorimotor representation, respectively. However, there is research to suggest that mid-utterance perturbations are processed differently than perturbations that occur at voice onset (Hawco & Jones, 2009). Hawco and Jones suggest that at utterance onset auditory feedback is compared to the feedback predicted by the sensorimotor representation in order to ensure the intended F0 is produced. On the other hand, after utterance onset auditory feedback is compared to the F0 of the auditory feedback received prior to the manipulation, in attempt to stabilize the F0 of the vocalization (Hawco and Jones, 2009). Based on these findings, it is possible that the responses recorded in study 6 reflect a comparison between the perceived auditory feedback and the previously experienced feedback, rather than the feedback predicted by the sensorimotor representation. That being said, since the unpredictable feedback manipulation in study 6 also increased trial-to-trial variability, it is possible that a more general increase in variability across the vocalizations resulted in the larger vocal and N1 responses, rather than deviations from the prediction created by the sensorimotor representation. This alternate explanation comes in light of recent findings from the speech induced suppression literature.

Although there appears to be a clear link between vocal variability and the way in which deviant auditory feedback is processed, there is some debate with regards to the type of variability that drives these changes in auditory feedback processing. Two recent studies investigated the relationship between vocal variability and speech induced suppression of the N1 ERP component (Niziolek et al., 2013; Sitek et al., 2013). Niziolek and colleagues (2013) reported that speech induced suppression was greater, or N1 amplitudes were smaller, when vowel productions were closer to the speaker's median production of that vowel. For this reason, Niziolek and colleagues (2013) concluded that "sensory predictions" are actually sensory goals that predict the desired outcome of the motor plan, rather than the outcome of the ongoing motor commands. Thus, productions that deviate farther from this sensory goal result in less suppression by the motor system. On the other hand, Sitek and colleagues investigated the relationship between pairwise variability and speech induced suppression of the N1 ERP component. Sitek and colleagues (2013) reported that it was the trial-to-trial change in formant frequency that predicted the degree of N1 suppression, with less change from trial-to-trial resulting in more suppression, rather than the absolute deviation from the median of all utterances. Sitek and colleagues attribute the differences across these studies to the fact that Niziolek and colleagues had speakers produce 3 different vowels, while in their own study speakers were repeating the same vowel. It was argued that since producing a different vowel across each utterance does not afford a situation under which the current vocalization can be compared with the previous production, a different control strategy might be utilized (Sitek et al., 2013). More specifically, similar to the explanation provided by Hawco and Jones (2009) it was suggested that when the vowel was varied from trial-to-trial, the current production was compared to a stored representation of that vowel. However, when the vowel was held constant across vocalizations, formants were tracked relative to the most recently produced vocalization (Sitek et al., 2013). In light of these findings, it is clear that more research is required to fully understand the relationship between vocal variability and the processing of auditory feedback, particularly on a trial-to-trial basis. That being said, regardless of whether incoming auditory feedback is being compared with a prediction created by a sensorimotor representation, or previously experienced auditory feedback, the results reported in studies 1 through 5 provide ample evidence to support the notion that individuals who are more variable rely more on their feedback control system.

ii. Individuals Differences

Across the 6 studies reported in this thesis, a large number of individual differences in both vocal and ERP responses to FAF were reported. For example, in many of the studies reported in this thesis, as well as in previously conducted studies (Hain et al., 2000; Bauer & Larson, 2003; Liu et al., 2010a; Korzyukov et al., 2012b; Behroozmand et al., 2013; Patel et al., 2014), speakers' data were excluded from the analyses as their vocal responses followed the direction of the manipulation rather than opposing it. Also, from a purely observational standpoint, data derived from individual participants was often qualitatively different from the averaged waveform. In order to truly understand, and be able to predict how the speech motor control system will function in a given circumstance, it is important to understand why these large individual differences exist (Purcell & Munhall, 2006a). To date, efforts to understand the influence of somatosensory feedback, singing experience, tonal language speaking experience, and attentional load on responses to FAF, have furthered our understanding of feedback control of speech.

Somatosensory Feedback Control of Speech

Throughout this thesis when discussing the role of the feedback control system in speech production, the discussion has focused primarily on auditory feedback. While it is difficult to acquire fluent speech without auditory feedback (Oller & Eilers, 1988; Hickok et al., 2011), the fact that deaf speakers can produce intelligible speech suggests that auditory feedback is not the sole input into the feedback control system (Lane & Wozniak-Webster, 1991; Tremblay et al., 2003; Nasir and Ostry, 2008). From the moment a child begins making speech like sounds, the auditory feedback they receive is correlated with the movements responsible for producing the sounds (Gracco & Löfqvist, 1994; Lametti et al., 2012). For this reason, it is likely that somatosensory feedback also plays a role in monitoring the accuracy of speech production (Tremblay et al., 2003; Lametti et al., 2012). This notion is supported by previous research that has shown that speakers compensate for perturbations to their lips and jaw (Abbs & Gracco, 1984; Gracco & Abbs, 1985) as well as their larynx (Lofqvist & Gracco, 1991; Munhall, Lofqvist & Kelso, 1994). In addition, speakers have been shown to adapt to jaw movements in the absence of changes to their auditory feedback, suggesting that somatosensory information is also closely regulated during speech production (Tremblay et al., 2003). With this in mind, could it be an increased use of somatosensory feedback, relative to auditory feedback, that accounts for the individual differences in responses to FAF?

Lametti and colleagues (2012) investigated the relative influence of auditory and somatosensory feedback during speech production. To do this, auditory and somatosensory feedback were simultaneously altered as speakers vocalized. The results of this study revealed an inverse relationship between reliance on auditory and somatosensory feedback. Specifically, speakers who adapted to the somatosensory perturbation showed a smaller response to the auditory perturbation, relative to speakers who failed to adapt to the somatosensory perturbation, and vice versa (Lametti et al., 2012). Based on these results, the authors suggested that there is a ‘sensory preference,’ with speakers having a preferred type of sensory feedback. While this sensory preference could certainly account for some of the individual differences observed in response to FAF, further research is required to determine just how much of the variance in individuals responses to FAF can be accounted for by this sensory preference.

Singing Experience

Speaking and singing, while similar in many respects, differ in many ways. For example, regulating vocal pitch it important for both singing and speaking (Zatorre, Belin, & Penhune, 2002; Zarate & Zatorre, 2008). However, when speaking vocal pitch is regulated in order to achieve the function goal of transferring information (Burnett et al., 1997; Perkell, 2012). On the other hand, when singing the regulation of pitch, or the production of specific pitch targets, is the functional goal (Liu, Jiang, Wang, Xu, & Patel, 2015). For this reason, singing requires more precise control over the speech motor control system (Zarate et al., 2008; Zarate, 2013). In accordance with this notion, trained singers have been shown to have superior speech motor control (Zarate et al., 2008; Zarate, 2013). For example, previous research has shown that singers are better at ignoring deviant auditory feedback (Zarate et al., 2008). Differences in neural activity have also been observed when singers and non-singers are asked to ignore deviant auditory feedback (Zarate et al., 2008). Based on these results we now know that vocal experience not only improves voluntary speech motor control, but also changes the functional network involved in processing auditory feedback (Zarate et al., 2008). For this reason, singing experience represents another individual difference that may contribute to the differences in responses observed across individuals participating in FAF studies.

Tonal Language Experience

Similar to singing experience, the ability to speak a tonal language may also contribute to individual differences in the processing of FAF. In English, fundamental frequency is used to control postural settings of speech (Jones & Munhall, 2002), such as those concerning stress or sentence structure (Giuliano, Pfordresher, Stanley, Narayana, & Wicha, 2011). However, in tonal languages, such as Mandarin, Cantonese, Thai, and Vietnamese, words take on different lexical meanings depending on the fundamental frequency in which they are enunciated (Jones & Munhall, 2002; Deutsch, Henthorn, & Dolson, 2004). Thus unlike English speakers, in order to convey the intended meaning of a word, tonal language speakers have an explicit pitch target they must produce (Jones & Munhall, 2002). Previous research has demonstrated that Mandarin speakers were faster and more accurate at detecting small pitch changes relative to non-tonal language speakers (Giuliano et al., 2011). In another study, Mandarin speakers were found to produce smaller compensatory responses to FAF relative to non-tonal language speakers (Ning, Loucks, & Shih, 2015). Together, these results suggest that tonal language speakers not only have better pitch discrimination abilities (Giuliano et al., 2011; Ning et al., 2015), but tonal language speakers also have better vocal control, making them less susceptible to deviant auditory feedback (Ning et al., 2015).

Attentional Load

The last factor that may have contributed to individual differences in the studies reported in this thesis is attentional load. A discussion of attentional load is particularly relevant to the results of the studies reported in this thesis, as many of the studies utilized speakers of different age groups, and attentional resources have been shown to vary with age (Kok, 2000). Moreover, attention deficits are becoming more prevalent in the general population, with a 2010

United States consensus reporting the prevalence of attention deficit/ hyperactivity disorder at ~10% (Bloom, Cohen, & Freeman, 2010).

Since attention is a limited resource it must be divided amongst all incoming sensory inputs (Wickens, 2002). If the processing demands imposed by the sensory inputs exhaust the available attentional resources, then performance on ongoing tasks may suffer (Wickens, 2002). In everyday life, as well as in the laboratory setting, speakers process auditory feedback while simultaneously processing input from other modalities (Tumber et al., 2014). For this reason, it is important to understand how divided attention may influence the processing of auditory feedback. To date, the results of studies investigating the influence of attention on responses to FAF have been mixed. Alsius and colleagues (2013) reported that dividing a speaker's attention did not modulate responses to FAF. Similarly, an ERP study investigating responses to FAF under divided attention also revealed no change in the size of speaker's compensatory responses under divided attention, however, they did find that P2 responses decreased when attention was divided (Hu et al., 2015). Contrary to these findings, Tumber and colleagues (2014) reported that speakers produced smaller compensatory responses to FAF perturbations under divided attention. Similarly, Scheerer and colleagues (2015) reported smaller compensatory responses and less sensorimotor adaptation when speaker's attention was divided while participating in a FAF adaptation task. The disparity amongst the studies conducted to date leaves the role of attention during the processing of auditory feedback unclear. However, since in some cases attention has been shown to modulate responses to FAF, attentional load must be taken into account when considering potential sources of individual differences in FAF responses.

Previous efforts to understand the influence of somatosensory feedback, singing experience, tonal language speaking experience, and attentional load on responses to FAF, have furthered our understanding of feedback control of speech. From these studies it is clear that experience controlling fundamental frequency in different contexts and with different functional objectives can have long term effects on how auditory feedback regarding one's fundamental frequency is processed. Based on these studies efforts have been made to control for singing experience, tonal language experience, and attentional load across participants. Despite these controls, individual differences persist. Based on the results of the studies reported in this thesis it appears as though vocal variability is an additional factor that must be controlled for when designing FAF studies, however, future research is required to fully understand the individual differences in responses observed following exposure to FAF.

Future Directions

i. Development of Formant vs. Fundamental Frequency Control

The results of study 1, together with the results of MacDonald and colleagues (2012), suggest that the control of formant frequencies and fundamental frequency may develop at different rates. However, direct comparison of formant and fundamental frequency control throughout development would help to validate this suggestion. If in fact children are faster to develop fundamental frequency control, then it is also possible that children are also better at perceiving perceptual differences in fundamental frequency, relative to formant frequency. Future research on these differences may help to highlight developmental differences in speech processing that may in turn facilitate techniques for communicating with young infants, and for promoting speech and language acquisition.

ii. Vocal Training

The results of study 6 suggest that exposing speakers to brief perturbations to the pitch of their auditory feedback can induce changes in the weighting of their feedback and feedforward control systems. If this brief exposure to auditory feedback perturbations can already change the relative input from each of these control systems, then more extensive exposure to predictable auditory feedback may have potential utility for vocal training. For example, implementing a training program that focuses on decreasing a speaker's vocal variability, and thus making their auditory feedback more predictable, may promote the use of feedforward control. Since singers often perform in environments with extraneous ambient noise, this increased weighting of feedforward control would help to ensure that the singer is not affected by the extraneous background noise.

iii. Speech Motor Control in Individuals with Autism Spectrum Disorders (ASDs)

Autism Spectrum Disorder (ASD) is a heterogeneous disorder, which has led to difficulty clearly classifying manifestations of this disorder. However, biological and physiological deviations can be readily uncovered by investigating each core symptom of ASD separately. One of the more commonly reported symptoms of ASD is speech irregularity, including difficulties with speech motor control (Russo et al., 2008).

Based on the results of study 3, we now have an idea of how the P1, N1, and P2 ERPs are modulated by FAF in typically developing children and adults. However, these neural responses to FAF have yet to be documented in children and adults with ASD. Although there is a lack of research investigating neural responses to FAF, more generalized auditory cortical responses to auditory stimuli have been studied in individuals with ASD. These studies suggest that P1, N1, and P2 amplitudes are smaller in individuals with ASD relative to typically developing individuals (Novick et al., 1980). While these studies highlight auditory processing differences in individuals with ASD relative to typically developing individuals, a specific investigation of auditory cortical responses to FAF would help to identify potential neural dysfunctions related to speech motor control in individuals with ASD.

Language development is significantly disrupted in adults with ASD, yet the underlying cause of this disruption is currently unknown (Russo et al., 2008). Determining whether adults with ASD deviate from typically developing adults in the development of speech motor control may help uncover functional abnormalities underlying these language and speech disturbances.

iv. Persistent Developmental Stuttering

Persistent developmental stuttering is characterized by disturbances in the coordination and movements of the respiratory, phonatory, and articulatory systems (Max et al., 2004; Beal et al., 2010). Delayed, frequency altered, and masked auditory feedback have been shown to influence stuttering severity (Van Riper, 1982; Howell, El-Yaniv, & Powell, 1987; Kalinowski, Armson, Stuart, & Gracco, 1993; Kalinowski, Armson, Stuart, & Gracco, 1996; Stuart, Kalinowski, Armson, Stenstrom, & Jones, 1996; Stager, Denman, & Ludlow, 1997; Natke, Grosser, & Kalveram, 2001; Stuart & Kalinowski, 2004). However, the mechanisms by which these auditory feedback manipulations affect speech motor control in individuals who stutter are currently unclear (Loucks, Chon, & Han, 2012).

Neurocomputational models of speech production posit that the speech disfluencies demonstrated by individuals who stutter are the result of an over reliance on auditory feedback during speech production (Max et al., 2004; Civier et al., 2010). However, it has also been suggested that individuals who stutter have difficulty extracting information from their auditory feedback to modifying their ongoing speech motor commands, as well as to plan future speech motor commands (Max et al., 2004; Cai et al., 2012). By exposing individuals who stutter and healthy controls to FAF, it may be possible to identify whether individuals who stutter do in fact rely too much on auditory feedback, or whether they are unable to utilize their auditory feedback for sensorimotor planning.

As demonstrated throughout this thesis, individuals who have an increased reliance on auditory feedback tend to produce larger compensatory responses to FAF. Thus, if individuals who stutter do in fact rely too much on auditory feedback, we would expect that they would produce larger responses to FAF perturbations. On the other hand, if individuals who stutter are unable to extract meaningful information from their auditory feedback for sensorimotor planning, we would expect that these individuals would show slower rates of sensorimotor adaptation when exposed to persistent changes to their auditory feedback. Thus comparing responses to FAF across individuals who stutter and healthy controls may reveal differences in the way in which speech motor control is regulated by auditory feedback in these populations. Identifying potential differences in the way in which individuals who stutter and fluent speakers process auditory feedback is the first step in developing more effective techniques for alleviating stuttering severity.

Conclusion

The aim of this thesis was to examine changes in the role of auditory feedback for speech motor control throughout development, as well as factors that may influence the role of auditory feedback for speech motor control in adulthood. The first study in this thesis demonstrated that toddlers use auditory feedback to regulate their speech motor commands, supporting the long held notion that auditory feedback is important during the acquisition of speech. While mapping out the developmental trajectory of vocal and ERP responses to FAF, the second study also demonstrated that vocal variability, rather than age, is best for predicting responses to FAF. Importantly, this suggests that maturation of the speech motor control system is not strictly dependent on age. The third study in this thesis demonstrated that children and adults show similar rates of sensorimotor adaptation, suggesting that once speech is acquired, speakers are proficient at using sensory information to modify the planning of future motor commands. However, since adults produced larger compensatory responses, these results also suggested that adults are more proficient at comparing incoming auditory feedback with the feedback predicted by their sensorimotor representations, as a result of possessing more precisely mapped sensorimotor representations. The results of studies four and five demonstrated that vocal variability can be used to predict the size of compensatory responses and sensorimotor adaptation to changes in one’s auditory feedback, respectively. Furthermore, these studies demonstrated that increased variability was related to increased auditory feedback control of speech. Finally, the sixth study in this thesis demonstrated that experimentally induced predictability and variability can be used to induce increases in feedforward and feedback control, respectively. In conclusion, the results of this thesis demonstrate that age and vocal variability, both naturally occurring and experimentally induced, are important determinants of the role of auditory feedback in speech motor control.

References

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: orofacial muscle responses to load perturbations of lip during speech. Journal of Neurophysiology, 51(4), 705–723. http://doi.org/10.1152/jn.00147.2013

Albrecht, R., Suchodoletz, W.V., & Uwer, R. (2000). The development of auditory evoked dipole source activity from childhood to adulthood. Clinical Neurophysiology, 111, 2268–2276.

Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. Journal of Cognitive Neuroscience, 21(4), 791–802. http://doi.org/10.1162/jocn.2009.21055

Allison, T., Hume, A.L., Wood, C.C., & Goff, W.R. (1984). Developmental and aging changes in somatosensory, auditory, and visual evoked potentials. Electroencephalography and Clinical Neurophysiology, 58, 14–24.

Alsius, A., Mitsuya, T., & Munhall, K. (2013). Does compensation in auditory feedback require attention? (pp. 060098–060098). Presented at the ICA 2013 Montreal, ASA. http://doi.org/10.1121/1.4799040

Bailly, G. (1997). Learning to speak. Sensorimotor control of speech movements. Speech Communication, 22(2-3), 251–267. http://doi.org/10.1016/S0167-6393(97)00025-3

Banaschewski, T., & Brandeis, D. (2007). Annotation: What electrical brain activity tells us about brain function that other techniques cannot tell us – a child psychiatric perspective. Journal of Child Psychology and Psychiatry, 48(5), 415–435. http://doi.org/10.1111/j.1469-7610.2006.01681.x/pdf

Bauer, J. J., & Larson, C. R. (2003). Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique. The Journal of the Acoustical Society of America, 114(2), 1048–1054. http://doi.org/10.1121/1.1592161

Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. The Journal of the Acoustical Society of America, 119(4), 2363–2371. http://doi.org/10.1121/1.2173513

Bays, P. M., & Wolpert, D. M. (2006). Computational principles of sensorimotor control that minimize uncertainty and variability. The Journal of Physiology, 578(2), 387–396. http://doi.org/10.1113/jphysiol.2006.120121

Beal, D. S., Cheyne, D. O., Gracco, V. L., Quraan, M. A., Taylor, M. J., & De Nil, L. F. (2010). Auditory evoked fields to vocalization during passive listening and active generation in adults who stutter. NeuroImage, 52(4), 1645–1653. http://doi.org/10.1016/j.neuroimage.2010.04.277

Beal, D. S., Quraan, M. A., Cheyne, D. O., Taylor, M. J., Gracco, V. L., & De Nil, L. F. (2011). Speech-induced suppression of evoked auditory fields in children who stutter. NeuroImage, 54(4), 2994–3003. http://doi.org/10.1016/j.neuroimage.2010.11.026

Behroozmand, R., Karvelis, L., Liu, H., & Larson, C. R. (2009). Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. Clinical Neurophysiology, 120(7), 1303–1312. http://doi.org/10.1016/j.clinph.2009.04.022

Behroozmand, R., Korzyukov, O., & Larson, C. R. (2011a). Effects of voice harmonic complexity on ERP responses to pitch-shifted auditory feedback. Clinical Neurophysiology, 122(12), 2408–2417. http://doi.org/10.1016/j.clinph.2011.04.019

Behroozmand, R., Korzyukov, O., & Larson, C. R. (2012). ERP correlates of pitch error detection in complex Tone and Voice auditory feedback with missing fundamental. Brain Research, 1448(C), 89–100. http://doi.org/10.1016/j.brainres.2012.02.012

Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard, M. A., III, & Greenlee, J. D. W. (2015). Sensory–motor networks involved in speech production and motor control: An fMRI study. NeuroImage, 109(C), 418–428. http://doi.org/10.1016/j.neuroimage.2015.01.040

Benoit, T. C., Jocelyn, L. J., Moddemann, D. M., & Embree, J. E. (1996). Romanian adoption: the Manitoba experience. Archives of Pediatrics and Adolescent Medicine, 150, 1278–82.

Blakemore, S. J., Rees, G., & Frith, C. D. (1998). How do we predict the consequences of our actions? a functional imaging study. Neuropsychologia, 36(6), 521–529. http://doi.org/10.1016/S0028-3932(97)00145-0

Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? Neuroreport, 11(11), R11-16.

Bloom, B., Cohen, R. A., & Freeman, G. (2010). Summary health statistics for U.S. children: National Health Interview Survey, 2009. National Center for Health Statistics. Vital Health Stat 10(247).

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. Glot International, 5(9/10), 341–345.

Brattico, E., Tervaniemi, M., & Picton, T. W. (2003). Effects of brief discrimination-training on the auditory N1 wave. Neuroreport, 14(18), 2489–2492. http://doi.org/10.1097/01.wnr.0000098748.87269.a1

Braun, A. R., Varga, M., Stager, S., Schulz, G., & Selbie, S. (1997). Altered patterns of cerebral activity during speech and language production in developmental stuttering. An H2 (15) O positron emission tomography study. Brain. 120, 761-784.

Bruneau, N., & Gomot, M. (1998). Auditory evoked potentials (N1 wave) as indices of cortical development. In Neuroimaging in child neuropsychiatric disorders (pp. 113–123). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-95848-9_10

Bruneau, N., Roux, S., Guerin, P., Barthelemy, C., & Lelord, G. (1997) Temporal prominence of auditory evoked potentials (N1 wave) in 4–8 year old children. Psychophysiology, 34, 32–38.

Budd, T. W., Barry, R. J., Gordon, E., & Rennie, C. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. International Journal of Psychophysiology, 31(1), 51–68. http://doi.org/10.1016/S0167-8760(98)00040-3

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

Burnett, T. A., & Larson, C. R. (2002). Early pitch-shift response is active in both steady and dynamic voice pitch control. The Journal of the Acoustical Society of America, 112(3), 1058–1063. http://doi.org/10.1121/1.1487844

Burnett, T. A., McCurdy, K. E., & Bright, J. C. (2008). Reflexive and volitional voice fundamental frequency responses to an anticipated feedback pitch error. Experimental Brain Research, 191(3), 341–351. http://doi.org/10.1007/s00221-008-1529-z

Burnett, T. A., Senner, J. E., & Larson, C. R. (1997). Voice Fo Responses to Pitch-Shifted Auditory Feedback: A Preliminary Study. Journal of Voice, 11(2), 202–211.

Butler, R. A. (1968). Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. The Journal of the Acoustical Society of America, 44(4), 945–950. http://doi.org/10.1121/1.1911233

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012a). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. Nature Reviews Neuroscience, 13, 407–420. http://doi.org/10.1038/nrn3241

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak Responses to Auditory Feedback Perturbation during Articulation in Persons Who Stutter: Evidence for Abnormal Auditory-Motor Transformation. PLoS ONE, 7(7), e41830. http://doi.org/10.1371/journal.pone.0041830

Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An Auditory-Feedback-Based Neural Network Model of Speech Production That is Robust to Developmental Changes in the Size and Shape of the Articulatory System. Journal of Speech, Language, and Hearing Research, 43, 721–736.

Camacho, A., & Harris, J. G. (2008). Computing pitch of speech and music using a sawtooth waveform inspired pitch estimator. The Journal of the Acoustical Society of America, 122(5), 2960–2961. http://doi.org/10.1121/1.2942550

Ceponiene, R., Rinne, T., & Nantanen, R. (2002). Maturation of cortical sound processing as indexed by event-related potentials. Clinical Neurophysiology, 113, 870-882.

Chait, M., Simon, J. Z., & Poeppel, D. (2004). Auditory M50 and M100 responses to broadband noise: functional implications. Neuroreport, 15(16), 2455–2458.

Chase, R. A., Sutton, S., First, D., & Zubin, J. (1961). A Developmental Study of Changes in Behavior under Delayed Auditory Feedback. The Journal of Genetic Psychology, 99(1), 101–112. http://doi.org/10.1080/00221325.1961.10534396

Chen, Z., Chen, X., Liu, P., Huang, D., & Liu, H. (2012a). Effect of temporal predictability on the neural processing of self-triggered auditory stimulation during vocalization. BMC Neuroscience, 13(1), 1–10. http://doi.org/10.1186/1471-2202-13-55

Chen, Z., Jones, J. A., Liu, P., Li, W., Huang, D., & Liu, H. (2013). Dynamics of Vocalization-Induced Modulation of Auditory Cortical Activity at Mid-utterance. PLoS ONE, 8(3), e60039. http://doi.org/10.1371/journal.pone.0060039

Chen, Z., Liu, P., Jones, J.A., Dongfeng, H., & Liu, H. (2010). Sex-related differences in vocal responses to pitch feedback perturbations during sustained vocalization. The Journal of the Acoustical Society of America, 128, 355–360.

Chen, Z., Liu, P., Wang, E. Q., Larson, C. R., Huang, D., & Liu, H. (2012b). ERP correlates of language-specific processing of auditory pitch feedback during self-vocalization. Brain and Language, 121(1), 25–34. http://doi.org/10.1016/j.bandl.2012.02.004

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. The Journal of the Acoustical Society of America, 121(2), 1157–1163. http://doi.org/10.1121/1.2404624

Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. Human Brain Mapping, 28(9), 868–879. http://doi.org/10.1002/hbm.20315

Civier, O., Tasko, S. M., & Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: simulations of stuttering and fluency-inducing conditions with a neural model of speech production. Journal of Fluency Disorders, 35(3), 246–279. http://doi.org/10.1016/j.jfludis.2010.05.002

Coch, D., Skendzel, W., & Neville, H. J. (2005). Auditory and visual refractory period effects in children and adults: An ERP study. Clinical Neurophysiology, 116(9), 2184–2203. http://doi.org/10.1016/j.clinph.2005.06.005

Cooper, R. P., & Aslin, R. N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. Child Development, 65(6), 1663–1677.

Cowie, R., Douglas-Cowie, E., & Kerr, A. G. (1982). A study of speech deterioration in post-lingually deafened adults. The Journal of Laryngology and Otology, 96, 101–112.

Cowie, R., & Douglas-Cowie, E. (1992). Postlingually acquired deafness: Speech deterioration and the wider consequences. Berlin: Mouton de Gruyter, pp. ix + 304.

Cunningham, J., Nicol, T., Zecker, S., & Kraus, N. (2000). Speech-evoked neurophysiological responses in children with learning problems: development and behavioural correlates of perception. Ear and Hearing, 21, 554–568.

de Boisson-Bardies, B. (1999). How language comes to children. MIT Press, Cambridge.

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute Pitch, Speech, and Tone Language: Some Experiments and a Proposed Framework. Music Perception: an Interdisciplinary Journal, 21(3), 339–356. http://doi.org/10.1525/mp.2004.21.3.339?ref=search-gateway:19c72d3a5e8a6b4c1588c61eee11f88e

Devous, M. D., Altuna, D., Ngai, W. T., Chiu, S., & Harris, T. S. (2006). Maturation of speech and language functional neuroanatomy in pediatric normal controls. Journal of Speech Language and Hearing Research, 49, 856–866.

Donath, T. M., Natke, U., & Kalveram, K. T. (2002). Effects of frequency-shifted auditory feedback on voice F0 contours in syllables. The Journal of the Acoustical Society of America, 111(1), 357–366. http://doi.org/10.1121/1.1424870

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. Annual Review of Neuroscience, 22(1), 567–631. http://doi.org/10.1146/annurev.neuro.22.1.567

Eggermont, J.J. (1989). The onset and development of auditory function: contributions of evoked potential studies. Journal of Speech and Language Pathology and Audiology, 13, 5–16.

Eliades, S. J. (2002). Sensory-Motor Interaction in the Primate Auditory Cortex During Self-Initiated Vocalizations. Journal of Neurophysiology, 89(4), 2194–2207. http://doi.org/10.1152/jn.00627.2002

Eliades, S. J., & Wang, X. (2005). Dynamics of Auditory–Vocal Interaction in Monkey Auditory Cortex. Cerebral Cortex, 15(10), 1510–1523. http://doi.org/10.1093/cercor/bhi030

Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature, 453(7198), 1102–1106. http://doi.org/10.1038/nature06910

Elliott, L., & Niemoeller, A. (1970). The role of hearing in controlling voice fundamental frequency, International Audiology IX, 47–52.

Elman, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. The Journal of the Acoustical Society of America, 70(1), 45–50. http://doi.org/10.1121/1.386580

Fernald, A. (1993). Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages. Child Development, 64(3), 657–674.

Ferree, T. C., Luu, P., Russell, G. S., & Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. Clinical Neurophysiology, 112(3), 536–544. http://doi.org/10.1016/S1388-2457(00)00533-2

Filipek, P.A., Accardo, P.J., Ashwal, S., Baranek, G.T., Cook, E.G., Dawson, G., Gordon, B., Gravel, J.S., Johnson, C.P., Kallan, R.J., Levy, S.E., Minshew, N.J., Ozonoff, S., Prizant, B.M., Rapin, I., Rogers, S.J., Stone, W.L., Teplin, S.W., Tuchman, R.F., Volkmar, F.R. (2000). Practice parameter: screening and diagnosis of autism: report of the quality standards subcommittee of the American Academy of Neurology and the Child Neurology Society. Neurology, 55, 468-479.

Ford, J. M., & Mathalon, D. H. (2005). Corollary discharge dysfunction in schizophrenia: Can it explain auditory hallucinations? International Journal of Psychophysiology, 58(2-3), 179–189. http://doi.org/10.1016/j.ijpsycho.2005.01.014

Fox, P. T., Ingham, R. J., Ingham, J. C., Hirsch, T. B., Downs, J. H., Martin, C., et al. (1996). A PET study of the neural systems of stuttering. Nature, 382(6587), 158–162. http://doi.org/10.1038/382158a0

Fox, P. T., Ingham, R. J., Ingham, J. C., Zamarripa, F., Xiong, J.-H., & Lancaster, J. L. (2000). Brain correlates of stuttering and syllable production. Brain, 123(10), 1985–2004. http://doi.org/10.1093/brain/123.10.1985

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. Neuroscience & Biobehavioral Reviews, 25(4), 355–373.

Gage, N.M., Siegel, B., Callan, M., & Roberts, T.P. (2003). Cortical sound processing in children with autism disorder: an MEG investigation. NeuroReport, 14, 2047–2051.

Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., & Wicha, N. Y. Y. (2011). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. Frontiers in Psychology, 2, 146. http://doi.org/10.3389/fpsyg.2011.00146

Goehl, H., & Kaufman, D. K. (1984). Do the Effects of Adventitious Deafness Include Disordered Speech? Journal of Speech and Hearing Disorders, 49(1), 58–64.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: testing parallels between birdsong and speech. Proceedings of the National Academy of Sciences, 100(13), 8030–8035. http://doi.org/10.1073/pnas.1332441100

Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. NeuroImage, 55(3), 1324–1338. http://doi.org/10.1016/j.neuroimage.2010.12.065

Goodin, D.S., Squires, K.C., Henderson, B.H., & Starr, A. (1978). Age-related variations in evoked potentials to auditory stimuli in normal human subjects. Electroencephalography and Clinical Neurophysiology, 44, 447–458.

Grabski, K., Tremblay, P., Gracco, V. L., Girin, L., & Sato, M. (2013). A mediating role of the auditory dorsal pathway in selective adaptation to speech: A state-dependent transcranial magnetic stimulation study. Brain Research, 1515, 55–65. http://doi.org/10.1016/j.brainres.2013.03.024

Gracco, V. L., & Abbs, J. H. (1985). Dynamic control of the perioral system during speech: kinematic analyses of autogenic and nonautogenic sensorimotor processes. Journal of Neurophysiology.

Gracco, V. L., & Lofqvist, A. (1994). Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. Journal of Neuroscience, 14(11), 6585–6597.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24(2), 95–112. http://doi.org/10.1007/BF02289823

Greenlee, J. D. W., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Kawasaki, H., et al. (2011). Human Auditory Cortical Activation during Self-Vocalization. PLoS ONE, 6(3), e14744. http://doi.org/10.1371/journal.pone.0014744

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

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. Psychological Review, 102(3), 594–621.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. Journal of Communication Disorders, 39(5), 350–365.

Guenther, F. H., & Bohland, J. W. (2002). Learning sound categories: A neural model and supporting experiments. Acoustical Science and Technology, 23(4), 213–221.

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

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. Psychological Review, 105(4), 611–633.

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

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., & Kenney, M. K. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. Experimental Brain Research, 130(2), 133–141. http://doi.org/10.1007/s002219900237

Hawco, C. S., & Jones, J. A. (2009). Control of vocalization at utterance onset and mid-utterance: Different mechanisms for different goals. Brain Research, 1276, 131–139. http://doi.org/10.1016/j.brainres.2009.04.033

Hawco, C. S., & Jones, J. A. (2010). Multiple instances of vocal sensorimotor adaptation to frequency-altered feedback within a single experimental session. The Journal of the Acoustical Society of America, 127(1), EL13. http://doi.org/10.1121/1.3272633

Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. Psychophysiology, 42(2), 180–190. http://doi.org/10.1111/j.1469-8986.2005.00272.x

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. Neuroreport, 17(13), 1375–1379. http://doi.org/10.1097/01.wnr.0000233102.43526.e9

Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. Journal of Communication Disorders, 45(6), 393–402.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. Neuron, 69(3), 407–422.

Houde, J. F., & Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. Current Opinion in Neurobiology, 33, 174–181.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor Adaptation in Speech Production. Science, 279(5354), 1213–1216. http://doi.org/10.2307/2894820?ref=no-x-route:8d1e2c40125a44080fa3c1858f698ae9

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor Adaptation of Speech: Compensation and Adaptation. Journal of Speech, Language, and Hearing Research, 45(2), 295–310. http://doi.org/10.1044/1092-4388(2002/023)

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. Frontiers in Human Neuroscience, 5, 1–14.

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(8), 1125–1138. http://doi.org/10.1162/089892902760807140

Howell, P., El-Yaniv, N., & Powell, D. J. (1987). factors affecting fluency in stutterers. Speech Motor Dynamics in Stuttering. HF MPeters, WHulstijn. Springer, New York, NY, 361-369.

Hu, H., Liu, Y., Guo, Z., Li, W., Liu, P., Chen, S., & Liu, H. (2015). Attention Modulates Cortical Processing of Pitch Feedback Errors in Voice Control. Scientific Reports, 5, 7812. http://doi.org/10.1038/srep07812

Ingram, D. (1989). First Language Acquisition. Cambridge University Press.

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levanen, S., Lin, F., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., & Belliveau, J. W. (2004). Human Posterior Auditory Cortex Gates Novel Sounds to Consciousness. Proceedings of the National Academy of Sciences of the United States of America, 101(17), 6809–6814. http://doi.org/10.2307/3372165?ref=search-gateway:4bb0923f315428f13da1be3bf7faeb47

Jääskeläinen, I. P., Ahveninen, J., Belliveau, J. W., Raij, T., & Sams, M. (2007). Short-term plasticity in auditory cognition. Trends in Neurosciences, 30(12), 653–661. http://doi.org/10.1016/j.tins.2007.09.003

Jeste, S. S., & Nelson, C. A., III. (2008). Event Related Potentials in the Understanding of Autism Spectrum Disorders: An Analytical Review. Journal of Autism and Developmental Disorders, 39(3), 495–510. http://doi.org/10.1007/s10803-008-0652-9

Johnstone, S.J., Barry, R.J., Anderson, J.W. & Coyle, S.F. (1996) Age- related changes in child and adolescent event-related potential component morphology, amplitude and latency to standard and target stimuli in an auditory oddball task. International Journal of Psychophysiology, 24, 223–238.

Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. Experimental Brain Research, 190(3), 279–287. http://doi.org/10.1007/s00221-008-1473-y

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: studies of Mandarin tone production. Journal of Phonetics, 30(3), 303–320. http://doi.org/10.1006/jpho.2001.0160

Jones, J. A., & Munhall, K. G. (2005). Remapping Auditory-Motor Representations in Voice Production. Current Biology, 15(19), 1768–1772. http://doi.org/10.1016/j.cub.2005.08.063

Jürgens, U. (2002). Neural pathways underlying vocal control. Neuroscience & Biobehavioral Reviews, 26(2), 235–258.

Jürgens, U. (2009). The Neural Control of Vocalization in Mammals: A Review. Journal of Voice, 23(1), 1–10.

Kalinowski, J., Armson, J., Stuart, A., & Gracco, V. L. (1993). Effects of Alterations in Auditory Feedback and Speech Rate on Stuttering Frequency. Language and Speech, 36(1), 1–16. http://doi.org/10.1177/002383099303600101

Kalinowski, J., Stuart, A., Sark, S., & Armson, J. (1996). Stuttering amelioration at various auditory feedback delays and speech rates. International Journal of Language & Communication Disorders, 31(3), 259–269. http://doi.org/10.3109/13682829609033157

Kawato, M. (1999). Internal models for motor control and trajectory planning. Current Opinion in Neurobiology, 9(6), 718–727. http://doi.org/10.1016/S0959-4388(99)00028-8

Kelly, C. W., & Sober, S. J. (2014). A simple computational principle predicts vocal adaptation dynamics across age and error size. Frontiers in Integrative Neuroscience, 8, 75. http://doi.org/10.3389/fnint.2014.00075

Kent, R. D. (1999). Motor control: Neurophysiology and functional development, In A. J. Caruso & E. A. Strand (Eds.), Clinical management of motor speech disorders in children (pp. 29-71). New York: Thieme Medical and Scientific Publishers.

Kent, R. D., & Vorperian, H. K. (1995). Anatomic development of the craniofacial-oral-laryngeal systems: A review. Journal of Medical Speech-Language Pathology, 3, 145-190.

Keough, D., Hawco, C., & Jones, J. A. (2013). Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. BMC Neuroscience, 14(1), 1–11. http://doi.org/10.1186/1471-2202-14-25

Keough, D., & Jones, J. A. (2009). The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing. The Journal of the Acoustical Society of America, 126(2), 837–846. http://doi.org/10.1121/1.3158600

Keough, D., & Jones, J. A. (2011). Contextual cuing contributes to the independent modification of multiple internal models for vocal control. Journal of Neurophysiology, 105(5), 2448–2456. http://doi.org/10.1152/jn.00291.2010

Kim, K. S., & Max, L. (2014). Estimating feedforward vs. feedback control of speech production through kinematic analyses of unperturbed articulatory movements. Frontiers in Human Neuroscience, 8, 1–15. http://doi.org/10.3389/fnhum.2014.00911

Kingyon, J., Behroozmand, R., Kelley, R., Oya, H., Kawasaki, H., Narayanan, N. S., & Greenlee, J. D. W. (2015). High-gamma band fronto-temporal coherence as a measure of functional connectivity in speech motor control. Neuroscience, 305(C), 15–25. http://doi.org/10.1016/j.neuroscience.2015.07.069

Kok, A. (2000). Age-related changes in involuntary and voluntary attention as reflected in components of the event-related potential (ERP). Biological Psychology, 54(1-3), 107–143.

Korzyukov, O., Karvelis, L., Behroozmand, R., & Larson, C. R. (2012a). ERP correlates of auditory processing during automatic correction of unexpected perturbations in voice auditory feedback. International Journal of Psychophysiology, 83(1), 71–78. http://doi.org/10.1016/j.ijpsycho.2011.10.006

Korzyukov, O., Sattler, L., Behroozmand, R., & Larson, C. R. (2012b). Neuronal Mechanisms of Voice Control Are Affected by Implicit Expectancy of Externally Triggered Perturbations in Auditory Feedback. PLoS ONE, 7(7), e41216. http://doi.org/10.1371/journal.pone.0041216

Kotecha, R., Pardos, M., Wang, Y., Wu, T., Horn, P., Brown, D., et al. (2009). Modeling the Developmental Patterns of Auditory Evoked Magnetic Fields in Children. PLoS ONE, 4(3), e4811. http://doi.org/10.1371/journal.pone.0004811

Kraus, N., Micco, A., Koch, D., McGee, T., Carrell, T., Sharma, A., Wiet, R., & Weingarten, C. (1993). The mismatch negativity cortical evoked potential elicited by speech in cochlear-implant users. Hearing Research, 65, 118–124.

Kuebrich, B. D., & Sober, S. J. (2015). Variations on a theme: Songbirds, variability, and sensorimotor error correction. Neuroscience, 296, 48–54. http://doi.org/10.1016/j.neuroscience.2014.09.068

Kuhl, P. K. (1994). Learning and representation in speech and language. Current Opinion in Neurobiology, 4(6), 812–822.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. Nature Reviews Neuroscience, 5(11), 831–843. http://doi.org/10.1038/nrn1533

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. Journal of the Acoustic Society of America, 100, 2425–38.

Kuhl, P. K., & Meltzoff, A. N. (1997). Evolution, nativism, and learning in the development of language and speech. In The Inheritance and Innateness of Grammars, ed. M Gopnik, pp. 7–44. New York: Oxford Univ. Press

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory Preference in Speech Production Revealed by Simultaneous Alteration of Auditory and Somatosensory Feedback. Journal of Neuroscience, 32(27), 9351–9358. http://doi.org/10.1523/JNEUROSCI.0404-12.2012

Lane, H., & Tranel, B. (1971). The Lombard Sign and the Role of Hearing in Speech. Journal of Speech, Language, and Hearing Research, 14(4), 677–709. http://doi.org/10.1044/jshr.1404.677

Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. The Journal of the Acoustical Society of America, 89(2), 859–866. http://doi.org/10.1121/1.1894647

Lapenta, O. M., & Boggio, P. S. (2014). Motor network activation during human action observation and imagery: Mu rhythm EEG evidence on typical and atypical neurodevelopment. Research in Autism Spectrum Disorders, 8(7), 759–766. http://doi.org/10.1016/j.rasd.2014.03.019

Larson, C. R., Carrell, T. D., Senner, J. E., Burnett, T. A., & Nichols, L. L. (1995). A proposal for the study of voice F0 control using the pitch shifting technique. Vocal fold physiology: Voice quality control, 321-31.

Larson, C. R., Altman, K. W., Liu, H., & Hain, T. C. (2008). Interactions between auditory and somatosensory feedback for voice F 0 control. Experimental Brain Research, 187(4), 613–621. http://doi.org/10.1007/s00221-008-1330-z

Liu, P., Chen, Z., Jones, J. A., Wang, E. Q., Chen, S., Huang, D., & Liu, H. (2013). Developmental sex-specific change in auditory vocal integration: ERP evidence in children. Clinical Neurophysiology, 124(3), 503–513. http://doi.org/10.1016/j.clinph.2012.08.024

Liu, P., Chen, Z., Larson, C. R., Huang, D., & Liu, H. (2010a). Auditory feedback control of voice fundamental frequency in school children. The Journal of the Acoustical Society of America, 128(3), 1306–1312. http://doi.org/10.1121/1.3467773

Liu, F., Jiang, C., Wang, B., Xu, Y., & Patel, A. D. (2015). A music perception disorder (congenital amusia) influences speech comprehension. Neuropsychologia, 66(C), 111–118. http://doi.org/10.1016/j.neuropsychologia.2014.11.001

Liu, H., & Larson, C. R. (2007). Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. The Journal of the Acoustical Society of America, 122(6), 3671. http://doi.org/10.1121/1.2800254

Liu, H., Meshman, M., Behroozmand, R., & Larson, C. R. (2011). Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. Clinical Neurophysiology, 122(5), 951–957. http://doi.org/10.1016/j.clinph.2010.08.010

Liu, H., Russo, N. M., & Larson, C. R. (2010b). Age-related differences in vocal responses to pitch feedback perturbations: A preliminary study. The Journal of the Acoustical Society of America, 127(2), 1042–1046. http://doi.org/10.1121/1.3273880

Locke, J. L., & Snow, C. (1997). Social influences on vocal learning in human and nonhuman primates. See Snowdon & Hausberger (1997), pp. 274–92

Loucks, T., Chon, H., & Han, W. (2012). Audiovocal integration in adults who stutter. International Journal of Language & Communication Disorders, 47(4), 451–456. http://doi.org/10.1111/j.1460-6984.2011.00111.x

Lofqvist, A., & Gracco, V. L. (1991). Discrete and continuous modes in speech motor control. PERILUS, XIV, 27-34.

Lombard, E. (1911). Le signe de l'evelation de la voix (The sign of a rising voice). Annales Des Maladies de l'oreille et du Larynx, 37(1), 101-119.

MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., & Munhall, K. G. (2012). Children's development of self-regulation in speech production. Current Biology, 22, 113–117. http://doi.org/10.1016/j.cub.2011.11.052

MacKay, D. G. (1968). Metamorphosis of a Critical Interval: Age-Linked Changes in the Delay in Auditory Feedback that Produces Maximal Disruption of Speech. The Journal of the Acoustical Society of America, 43(4), 811. http://doi.org/10.1121/1.1910900

Max, L., Guenther, F. H., Gracco, V. L., & Ghosh, S. S. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. Contemporary Issues in Communication Science and Disorders, 31, 105–122.

McAllister, A., Sederholm, E., & Sundberg, J. (1993). Quarterly Progress and Status Report: Acoustic and perceptual analysis of vocal registers in children. Department for Speech, Music, and Hearing, 34(4), 029–034.

Miall, R. C., & Wolpert, D. M. (1996). Forward Models for Physiological Motor Control. Neural Networks, 9(8), 1265–1279. http://doi.org/10.1016/S0893-6080(96)00035-4

Moore, D. R. (2002). Auditory development and the role of experience. British Medical Bulletin, 63(1), 171-181.

Moore, R. E., Keaton, C., & Watts, C. (2007). The Role of Pitch Memory in Pitch Discrimination and Pitch Matching. Journal of Voice, 21(5), 560–567. http://doi.org/10.1016/j.jvoice.2006.04.004

Möttönen, R., van de Ven, G. M., & Watkins, K. E. (2014). Attention Fine-Tunes Auditory–Motor Processing of Speech Sounds. Journal of Neuroscience, 34(11), 4064–4069. http://doi.org/10.1523/JNEUROSCI.2214-13.2014

Munhall, K. G., Löfqvist, A., & Kelso, J. A. S. (1994). Lip–larynx coordination in speech: Effects of mechanical perturbations to the lower lip. The Journal of the Acoustical Society of America, 95(6), 3605–3616. http://doi.org/10.1121/1.409929

Nakagawa, K., Otsuru, N., Inui, K., & Kakigi, R. (2014). Change-related auditory P50: A MEG study. NeuroImage, 86(C), 131–137. http://doi.org/10.1016/j.neuroimage.2013.07.082

Nanova, P., Lyamova, L., Hadjigeorgieva, M., Kolev, V., & Yordanova, J. (2008). Gender-specific development of auditory information processing in children: An ERP study. Clinical Neurophysiology, 119(9), 1992–2003. http://doi.org/10.1016/j.clinph.2008.05.002

Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. Nature Neuroscience, 11(10), 1217–1222. http://doi.org/10.1038/nn.2193

Natke, U., Donath, T. M., & Kalveram, K. T. (2003). Control of voice fundamental frequency in speaking versus singing. The Journal of the Acoustical Society of America, 113(3), 1587–1593. http://doi.org/10.1121/1.1543928

Natke, U., Grosser, J., & Kalveram, K. T. (2001). Fluency, fundamental frequency, and speech rate under frequency-shifted auditory feedback in stuttering and nonstuttering persons. Journal of Fluency Disorders, 26(3), 227–241. http://doi.org/10.1016/S0094-730X(01)00099-7

Nelson, M.D., Hall, J.W. 3rd & Jacobson, G.P. (1997). Factors affecting the recordability of auditory evoked response component Pb (P1). Journal of the American Academy of Audiology, 8, 89–99.

Ning, L.-H., Loucks, T. M., & Shih, C. (2015). The effects of language learning and vocal training on sensorimotor control of lexical tone. Journal of Phonetics, 1–20. http://doi.org/10.1016/j.wocn.2014.12.003

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What Does Motor Efference Copy Represent? Evidence from Speech Production. Journal of Neuroscience, 33(41), 16110–16116. http://doi.org/10.1523/JNEUROSCI.2137-13.2013

Novick, B., Vaughan, H. G., Kurtzberg, D., & Simson, R. (1980). An electrophysiologic indication of auditory processing defects in autism. Psychiatry Research, 3(1), 107–114.

Oades, R.D., Dittman-Balear, A., & Zerbin, D. (1997). Development and topography of auditory event-related potentials (ERPs): mismatch and processing negativity in individuals 8–22 years of age. Psychophysiology, 34, 677–693.

Oller, D. K., & Eilers, R. E. (1988). The Role of Audition in Infant Babbling. Child Development, 59(2), 441–449.

Oram Cardy, J. E., Ferrari, P., Flagg, E. J., Roberts, W., & Roberts, T. P. L. (2004). Prominence of M50 auditory evoked response over M100 in childhood and autism. Neuroreport, 15(12), 1867-1870.

Paetau, R., Ahonen, A., Salonen, O., & Sams, M. (1995). Auditory evoked magnetic fields to tones and pseudowords in healthy children and adults. Journal of Clinical Neurophysiology, 12, 177–185.

Patel, S., Nishimura, C., Lodhavia, A., Korzyukov, O., Parkinson, A., Robin, D. A., & Larson, C. R. (2014). Understanding the mechanisms underlying voluntary responses to pitch-shifted auditory feedback. The Journal of the Acoustical Society of America, 135(5), 3036–3044. http://doi.org/10.1121/1.4870490

Peeva, M. G., Tourville, J. A., Agam, Y., Holland, B., Manoach, D. S., & Guenther, F. H. (2013). White matter impairment in the speech network of individuals with autism spectrum disorder. NeuroImage: Clinical, 3, 234–241. http://doi.org/10.1016/j.nicl.2013.08.011

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

Perkell, J., Matthies, M., Lane, H., & Guenther, F. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. Speech Communication, 22, 227–250.

Picton, T. W., & Hillyard, S. A. (1988). Endogenous event-related potentials. Handbook of electroencephalography and clinical neurophysiology, 3, 361-426.

Picton, T. W., Hillyard, S. A., & Krausz, H. I. (1974). Human auditory evoked potentials. I: Evaluation of components. Electroencephalography and Clinical Neurophysiology, 36, 179–190.

Polich, J., Ladish, C. & Burns, T. (1990) Normal variation of P300 in children: age, memory span, and head size. International Journal of Psychophysiology, 9, 237–248.

Ponton, C.W., & Eggermont, J.J. (2001). Of kittens and kids: altered cortical maturation following profound deafness and cochlear implant use. Audiology and Neurotology, 6, 363–380.

Ponton, C.W., & Eggermont, J.J. (2007). Auditory Evoked Potentials: Basic Principles and Clinical Application. Lippincott Williams & Wilkins, Philadelphia, pp. 385-402.

Ponton, C., Eggermont, J.J., Khosla, D., Kwong, B., & Don, M. (2002). Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. Clinical Neurophysiology, 113, 407–420.

Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2000). Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. Clinical Neurophysiology, 111, 220–236.

Proteau, L., Marteniuk, R. G., Girouard, Y., & Dugas, C. (1987). On the type of information used to control and learn an aiming movement after moderate and extensive training. Human Movement Science, 6, 181–199.

Proteau, L., Marteniuk, R. G., & Levesque, L. (1992). A sensorimotor basis for motor learning: evidence indicating specificity of practice. Quarterly Journal of Experimental Psychology, 44A, 557–575.

Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. The Journal of the Acoustical Society of America, 120(2), 966-977. http://doi.org/10.1121/1.2217714

Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. The Journal of the Acoustical Society of America, 119(4), 2288–2297. http://doi.org/10.1121/1.2173514

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nature Neuroscience, 12(6), 718–724.

Roach, B. J., & Mathalon, D. H. (2008). Event-Related EEG Time-Frequency Analysis: An Overview of Measures and An Analysis of Early Gamma Band Phase Locking in Schizophrenia. Schizophrenia Bulletin, 34(5), 907–926. http://doi.org/10.1093/schbul/sbn093

Rojas, D.C., Walker, J.R., Sheeder, J.L., Teale, P.D., & Reite, M.L. (1998). Developmental changes in refractoriness of the neuromagnetic M100 in children. NeuroReport, 9, 1543–1547.

Russo, N., Larson, C., & Kraus, N. (2008). Audio–vocal system regulation in children with autism spectrum disorders. Experimental Brain Research, 188(1), 111–124. http://doi.org/10.1007/s00221-008-1348-2

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Muratori, F., Laznik, M., & Cohen, D. (2013). Motherese in Interaction: At the Cross-Road of Emotion and Cognition? (A Systematic Review). PLoS ONE, 8(10), e78103–17. http://doi.org/10.1371/journal.pone.0078103

Sakata, J. T., & Brainard, M. S. (2008). Online Contributions of Auditory Feedback to Neural Activity in Avian Song Control Circuitry. The Journal of Neuroscience, 28(44), 11378–11390. http://doi.org/10.1523/JNEUROSCI.3254-08.2008

Scheerer, N. E., Behich, J., Liu, H., & Jones, J. A. (2013a). ERP correlates of the magnitude of pitch errors detected in the human voice. Neuroscience, 240, 176–185. http://doi.org/10.1016/j.neuroscience.2013.02.054

Scheerer, N. E., Liu, H., & Jones, J. A. (2013b). The developmental trajectory of vocal and event-related potential responses to frequency-altered auditory feedback. European Journal of Neuroscience, 38(8), 3189–3200. http://doi.org/10.1111/ejn.12301

Scheerer, N. E., & Jones, J. A. (2014). The predictability of frequency-altered auditory feedback changes the weighting of feedback and feedforward input for speech motor control. European Journal of Neuroscience, 40(12), 3793–3806. http://doi.org/10.1111/ejn.12734

Scheerer, N. E., & Jones, J. A. (2012). The relationship between vocal accuracy and variability to the level of compensation to altered auditory feedback. Neuroscience Letters, 529, 128–132. http://doi.org/10.1016/j.neulet.2012.09.012

Scheerer, N.E., & Jones, J.A. (2015). The relationship between vocal variability and sensorimotor learning. Neuroscience. Accepted: November 18, 2015.

Scheerer, N.E., Tumber, A.K., & Jones, J.A. (2015). Attentional Demands Modulate Sensorimotor Learning Induced by Persistent Exposure to Changes in Auditory Feedback. Journal of Neurophysiology. Accepted: November 19, 2015.

Scheidt, R. A., & Dingwell, J. B. (2001). Learning to move amid uncertainty. Journal of Neurophysiology, 86, 971–985.

Schmidt, R.A., & Lee, T.D. (2005). Motor Control and Learning: A Behavioral Emphasis. Human Kinetics Publishers, Champaign, IL, USA.

Schneider, B., Zumtobel, M., Prettenhofer, W., Aichstill, B., & Jocher, W. (2010). Normative Voice Range Profiles in Vocally Trained and Untrained Children Aged Between 7 and 10 Years. Journal of Voice, 24(2), 153–160. http://doi.org/10.1016/j.jvoice.2008.07.007

Shadmehr, R., Smith, M.A., & Krakauer, J.W. (2010). Error correction, sensory prediction, and adaptation in motor control. Annual Review of Neuroscience, 33, 89–108.

Sharma, A., Kraus, N., McGee, T.J., & Nicol, T.G. (1997). Developmental changes in P1 and N1 central auditory responses elicited by consonant– vowel syllables. Electroencephalography and Clinical Neurophysiology, 104, 540–545.

Siegel, G. M., Pick, H. L., Olsen, M. G., & Sawin, L. (1976). Auditory feedback on the regulation of vocal intensity of preschool children. Developmental Psychology, 12(3), 255–261. http://doi.org/10.1037/0012-1649.12.3.255

Singleton, D. M., & Ryan, L. (2004). Language Acquisition. Multilingual Matters.

Sitek, K. R., Mathalon, D. H., Roach, B. J., Houde, J. F., Niziolek, C. A., & Ford, J. M. (2013). Auditory Cortex Processes Variation in Our Own Speech. PLoS ONE, 8(12), e82925. http://doi.org/10.1371/journal.pone.0082925

Smith, C.R. (1975). Residual Hearing and Speech Production in Deaf Children. Journal of Speech, Language, and Hearing Research, 18, 795-811.

Smotherman, M. S. (2007). Sensory feedback control of mammalian vocalizations. Behavioural Brain Research, 182(2), 315–326. http://doi.org/10.1016/j.bbr.2007.03.008

Sober, S. J., & Brainard, M. S. (2012). Vocal learning is constrained by the statistics of sensorimotor experience. Proceedings of the National Academy of Sciences, 109(51), 21099–21103. http://doi.org/10.1073/pnas.1213622109

Sommer, M., Koch, M. A., Paulus, W., Weiller, C., & Büchel, C. (2002). Disconnection of speech-relevant brain areas in persistent developmental stuttering. The Lancet, 360(9330), 380–383. http://doi.org/10.1016/S0140-6736(02)09610-1

Stager, S. V., Denman, D. W., & Ludlow, C. L. (1997). Modifications in Aerodynamic Variables by Persons Who Stutter Under Fluency-Evoking Conditions. Journal of Speech, Language, and Hearing Research, 40(4), 832–847. http://doi.org/10.1044/jslhr.4004.832

Stuart, A., Kalinowski, J., Armson, J., Stenstrom, R., & Jones, K. (1996). Fluency Effect of Frequency Alterations of Plus/Minus One-Half and One-Quarter Octave Shifts in Auditory Feedback of People Who Stutter. Journal of Speech, Language, and Hearing Research, 39(2), 396–401. http://doi.org/10.1044/jshr.3902.396

Stuart, A., & Kalinowski, J. (2004). The Perception of Speech Naturalness of Post-Therapeutic and Altered Auditory Feedback Speech of Adults with Mild and Severe Stuttering. Folia Phoniatrica Et Logopaedica, 56(6), 347–357. http://doi.org/10.1159/000081082

Sussman, E., Steinschneider, M., Gumenyuk, V., Grushko, J., & Lawson, K. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. Hearing Research, 236, 61–79.

Svirsky, M. A., Teoh, S.W., & Neuburger, H. (2004). Development of language and speech perception in congenitally, profoundly deaf children as a function of age at cochlear implantation. Audiology & Neuro-Otology, 9(4), 224–233.

Takahashi, C. D., & Scheidt, R. A. (2001). Impedance control and internal model formation when reaching in a randomly varying dynamical environment. Journal of Neurophysiology, 86, 1047–1051.

Thoroughman, K. A., & Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. Nature, 407(6805), 742–747. http://doi.org/10.1038/35037588

Tian, X., & Poeppel, D. (2014). Dynamics of Self-monitoring and Error Detection in Speech Production: Evidence from Mental Imagery and MEG. Journal of Cognitive Neuroscience, 1–13. http://doi.org/10.1162/jocn_a_00692

Todorov, E. (2004). Optimality principles in sensorimotor control. Nature Neuroscience, 7(9), 907–915.

Tonnquist-Uhlen, I. (1996). Topography of auditory evoked cortical potentials in children with severe language impairment. Scandinavian Audiology, 44, 1–40.

Tonnquist-Uhlen, I., Borg, E. & Spens, K.E. (1995) Topography of auditory evoked long-latency potentials in normal children, with particular reference to the N1 component. Electroencephalography and Clinical Neurophysiology, 95, 34–41.

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. Language and Cognitive Processes, 26(7), 952–981.

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

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. Nature, 423(6942), 866–869. http://doi.org/10.1038/nature01710

Tumber, A. K., Scheerer, N. E., & Jones, J. A. (2014). Attentional Demands Influence Vocal Compensations to Pitch Errors Heard in Auditory Feedback. PLoS ONE, 9(10), e109968. http://doi.org/10.1371/journal.pone.0109968

Van Riper, C. (1982). The nature of stuttering. Prentice Hall.

Ventura, M. I., Nagarajan, S. S., & Houde, J. F. (2009). Speech target modulates speaking induced suppression in auditory cortex. BMC Neuroscience, 10(1), 58. http://doi.org/10.1186/1471-2202-10-58

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. The Journal of the Acoustical Society of America, 122(4), 2306–2319.

Wang, J., Mathalon, D. H., Roach, B. J., Reilly, J., & Keedy, S. (2014). Action planning and predictive coding when speaking. NeuroImage, 91, 91–98. http://doi.org/10.1016/j.neuroimage.2014.01.003

Watts, C., Murphy, J., & Barnes-Burroughs, K. (2003). Pitch Matching Accuracy of Trained Singers, Untrained Subjects with Talented Singing Voices, and Untrained Subjects with Nontalented Singing Voices in Conditions of Varying Feedback. Journal of Voice, 17(2), 185–194. http://doi.org/10.1016/S0892-1997(03)00023-7

Weiskrantz, L., Elliott, J., & Darlington, C. (1971). Preliminary observations on tickling oneself. Nature, 230(5296), 598–599. http://doi.org/10.1038/230598a0

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science, 3(2), 159–177. http://doi.org/10.1080/14639220210123806

Yates, A. J. (1963). Delayed auditory feedback. Psychological Bulletin, 60, 213–232.

Zarate, J. M. (2013). The neural control of singing. Frontiers in Human Neuroscience, 7, 1–12.

Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. NeuroImage, 40(4), 1871–1887. http://doi.org/10.1016/j.neuroimage.2008.01.026

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. Trends in Cognitive Sciences, 6(1), 37-46.