

Electronic Thesis and Dissertation Repository

---

8-14-2013 12:00 AM

## Investigation of Auditory Encoding and the Use of Auditory Feedback During Speech Production

Laura E. Beamish  
*The University of Western Ontario*

Supervisor  
Dr. David Purcell  
*The University of Western Ontario*

Graduate Program in Neuroscience  
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science  
© Laura E. Beamish 2013

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Neuroscience and Neurobiology Commons](#), and the [Speech and Hearing Science Commons](#)

---

### Recommended Citation

Beamish, Laura E., "Investigation of Auditory Encoding and the Use of Auditory Feedback During Speech Production" (2013). *Electronic Thesis and Dissertation Repository*. 1591.  
<https://ir.lib.uwo.ca/etd/1591>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

INVESTIGATION OF AUDITORY ENCODING AND THE USE OF AUDITORY  
FEEDBACK DURING SPEECH PRODUCTION

(Thesis format: Monograph)

by

Laura Beamish

Graduate Program in Neuroscience

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Master of Science

The School of Graduate and Postdoctoral Studies  
Western University  
London, Ontario, Canada

© Laura Beamish 2013

## Abstract

Responses to altered auditory feedback during speech production are highly variable. The extent to which auditory encoding influences this varied use is not well understood. Thirty-nine normal hearing adults completed a first formant (F1) manipulation paradigm where F1 of the vowel /ε/ was shifted upwards in frequency towards an /æ/-like vowel in real-time. Frequency following responses (FFRs) and envelope following responses (EFRs) were used to measure neuronal activity to the same vowels produced by the participant and a prototypical talker. Cochlear tuning, measured by SFOAEs and a psychophysical method, was also recorded. Results showed that average F1 production changed to oppose the manipulation. Three metrics of EFR and FFR encoding were evaluated. No reliable relationship was found between speech compensation and evoked response measures or measures of cochlear tuning. Differences in brainstem encoding of vowels and sharpness of cochlear tuning do not appear to explain the variability observed in speech production.

## Keywords

Auditory feedback, human frequency following response, human envelope following response, speech encoding, speech compensation, speech production, speech perception, vowel formants, stimulus frequency otoacoustic emissions, psychoacoustic tuning curves

## Acknowledgments

First and foremost I would like to thank my supervisor Dr. David Purcell for his unwavering support and guidance over the past two years. I am extremely grateful for all your encouragement, ideas and kindness over the past two years.

I would like to thank my advisory committee, Dr. Jessica Grahn, Dr. Lisa Archibald and Dr. Derek Mitchell for their knowledge and expertise throughout this research project.

To my SAFER lab mates and friends, thank-you for your support and advice over the past two years.

Finally, I am eternally grateful for my loving and supportive family, without whom this would not be possible.

# Table of Contents

Abstract .....	ii
Acknowledgments .....	iii
Table of Contents .....	iv
List of Tables .....	vii
List of Figures .....	viii
List of Appendices .....	x
List of Abbreviations .....	xi
Chapter 1 .....	1
1 Introduction .....	1
1.1 Introduction to Canadian English Vowels .....	2
1.2 Auditory Feedback .....	4
1.2.1 Pitch Shifts .....	6
1.2.2 Formant Perturbations .....	6
1.3 Psychoacoustic and Physiological measures of cochlear tuning .....	9
1.4 Auditory Evoked potentials .....	12
1.5 Frequency Following Response (FFR) .....	14
1.6 Envelope Following Response (EFR) .....	16
1.7 Rationale .....	17
1.8 Hypotheses .....	17
Chapter 2 .....	19
2 Methods .....	19
2.1 Participants .....	19
2.2 Summary of Procedures .....	19
2.3 Perceptual Measures .....	20
2.3.1 Vowel goodness .....	20
2.3.2 F1 discrimination threshold .....	22
2.3.3 Psychoacoustic tuning curves .....	22
2.4 Altered auditory feedback .....	23

2.4.1	Equipment .....	23
2.4.2	Formant Estimation .....	24
2.4.3	Procedure and experimental conditions .....	24
2.4.4	Online voice detection and formant shifting .....	25
2.4.5	Offline formant analysis .....	26
2.5	Otoacoustic Emissions .....	26
2.5.1	Stimulus generation and recording .....	26
2.6	Evoked Potentials .....	27
2.6.1	Stimuli .....	27
2.6.2	Polarity asymmetry in the EFR .....	28
2.6.3	Stimulus presentation and response recording .....	29
2.6.4	Offline response analysis .....	30
2.6.5	Envelope and frequency following response estimation .....	30
2.6.6	Response detection .....	30
Chapter 3	.....	32
3	Results .....	32
3.1	Speech .....	32
3.1.1	F1 Discrimination threshold .....	32
3.1.2	Vowel goodness ratings .....	32
3.1.3	Speech compensation for English /ε/ in “head” .....	32
3.2	Relationships between perception and production .....	33
3.3	Auditory Filter Bandwidth .....	33
3.3.1	Fast psychoacoustic tuning curves .....	33
3.3.2	Stimulus frequency otoacoustic emissions .....	37
3.3.3	Comparison between SWPTC and SFOAE .....	37
3.4	Electrophysiological measures .....	37
3.4.1	Envelope following response and frequency following response .....	37
3.4.2	Relationships between the EFR and FFR and speech compensation .....	51
Chapter 4	.....	68
4	Discussion .....	68
4.1	Compensation .....	68
4.2	Polarity Asymmetry .....	71

4.3	Envelope Following Response .....	71
4.4	Frequency Following Response .....	72
4.5	Compensation versus EFR and FFR .....	73
4.6	Tuning and Compensation.....	77
4.7	Perception & Compensation.....	79
4.8	Closing remarks, limitations, and future work .....	79
	Curriculum Vitae .....	100

## List of Tables

Table 1. Significant individual correlations ( $p < 0.05$ ) between vowel goodness ratings and F1 compensation in vowel production. ....	42
Table 2. EFR Responses for Polarity A. ....	53
Table 3. EFR Responses for Polarity B. ....	54
Table 4. EFR Response magnitude for Polarity A and B. ....	55
Table 5. FFR Responses for Polarity A. ....	57
Table 6. FFR Responses for Polarity B. ....	58
Table 7. EFR and speech compensation linear correlations for polarity A. ....	62
Table 8. EFR and speech compensation linear correlations for polarity B. ....	63
Table 10. FFR and speech compensation linear correlations for polarity B. ....	67
Table 9. FFR and speech compensation linear correlations for polarity A. ....	66



## List of Figures

Figure 1. Canadian-English vowel space.....	3
Figure 2. Representation of the phases in formant-shift paradigms. ....	8
Figure 3. Example of psychoacoustic tuning curves. ....	11
Figure 4. Overview of study methodology and measurements.....	21
Figure 5. Plot of group and individual F1 discrimination thresholds. ....	34
Figure 6. Mean vowel goodness ratings. ....	35
Figure 7. Average normalized F1 compensation during altered auditory feedback.....	36
Figure 8. Average normalized F1 results for English /ε/ as in “head” across the Ramp phase, the Hold phase and the End phase. ....	38
Figure 9. Individual variation in F1 production during altered auditory feedback.....	39
Figure 10. Correlation between goodness ratings and F1 compensation for /ε/ in “head” on a continuum towards /æ/ in had. ....	40
Figure 11. Plot of individual vowel goodness ratings compared to F1 compensation values for each Ramp step. ....	41
Figure 12. Example of individual trial from the SWPTC program. ....	43
Figure 13. Estimated individual and group average auditory filter bandwidth from the SWPTC program.....	44
Figure 14. Correlation between SWPTC filter bandwidth and compensation.....	45
Figure 15. Example of individual SFOAE analysis.....	46
Figure 16. SFOAE filter bandwidth group and individual results. ....	47

Figure 17. Correlation between SFOAE filter bandwidth and compensation. ....	48
Figure 18. Correlation between SFOAE bandwidth and SWPTC bandwidth. ....	49
Figure 19. Average EFR Amplitude for significant responses. ....	52
Figure 20. Average FFR Amplitude for significant responses. ....	56
Figure 21. Absolute amplitude (nV) of EFR to “head” correlated with compensation magnitude (Hz). ....	59
Figure 22. Change in EFR amplitude (nV) from “head” to “had” (no phase) correlated with compensation magnitude (Hz). ....	60
Figure 23. EFR Change in magnitude (nV) from “head” to “had” (including response phase) correlated with compensation magnitude (Hz). ....	61
Figure 24. Absolute amplitude (nV) of FFR to “head” correlated with compensation magnitude (Hz). ....	64
Figure 25. Change in FFR amplitude (nV) from “head” to “had” (no phase) correlated with compensation magnitude (Hz). ....	65
Figure 26. The DIVA Model of speech-production. Adapted from Figure 1 in Guenther (2006). ....	70

## List of Appendices

Appendix A: Ethics approval notice .....	92
Appendix B: Forms and questionnaires for participant .....	93

## List of Abbreviations

2AFC	Two-alternative forced choice
$\mu$ s	Microseconds
AEP	Auditory evoked potentials
ASSR	Auditory steady-state response
BMO	Best model order
BM	Basilar membrane
CN	Cochlear nucleus
CVC	Consonant vowel consonant word
DAF	Delayed auditory feedback
dB	Decibels
DPOAE	Distortion product otoacoustic emissions
DIVA	Directions in an orosensory space Into Velocities of Articulators
DNLL	Dorsal nucleus of the lateral lemniscus
EEG	Electroencephalography
EFR	Envelope following response
ERB	Equivalent rectangular bandwidth
$f_0$	Fundamental frequency
F1	First formant
F2	Second formant
F3	Third formant
FFR	Frequency following response
HL	Hearing level
Hz	Hertz
kHz	Kilohertz

LPC	Linear predictive coding
ms	Milliseconds
OAE	Otoacoustic emissions
OHC	Outer hair cells
nV	Nano volts
PTC	Psychoacoustic tuning curve
SD	Standard deviation
SFOAE	Stimulus frequency otoacoustic emissions
SNR	Signal to noise ratio
SOAE	Spontaneous otoacoustic emissions
SPL	Sound pressure level
SWPTC	Fast psychoacoustic tuning curve program
TEOAE	Transient evoked otoacoustic emissions
VNLL	Ventral nucleus of the later lemniscus
VOT	Voice onset time
yr	Years of age

## Chapter 1

### 1 Introduction

Auditory information from one's own voice during speech production plays a role in maintaining its accuracy and fluency. Auditory feedback provides talkers with information regarding different elements of their ongoing speech (i.e. intensity, spectral, and temporal information) and allows individuals to monitor and adjust their production when required. Speech production can be greatly affected if the feedback received is disrupted while talking.

Studies have examined the consequences of perturbations to auditory feedback through pitch-shifted auditory feedback (Burnett, Freedland, Larson, & Hain, 1998), temporal disruptions to running speech (Yates, 1963), loudness changes (Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988) and spectral changes (Garber, Seigel, & Pick, 1981). Results from these studies demonstrate that individuals vary substantially in their use of auditory feedback, however, in general, all perturbations are met with a speech production response that opposes the manipulation.

The influence of auditory feedback during speech production is well documented, although the mechanisms underlying the processing of acoustic information and how this in turn influences production are not completely understood. The role the auditory system plays in encoding the acoustic signal into useful information to guide production requires further investigation.

Incoming auditory information is processed by both the peripheral and central auditory systems. The peripheral auditory system is obviously essential for using auditory feedback during speech production, but further investigation into what aspects of peripheral function are influencing how individuals use auditory feedback is required. One way to assess peripheral auditory function is with the measurement of otoacoustic emissions (OAEs). OAEs can provide a physiological measure of cochlear tuning (Souter, 1995), which may influence how frequency changes in auditory feedback are

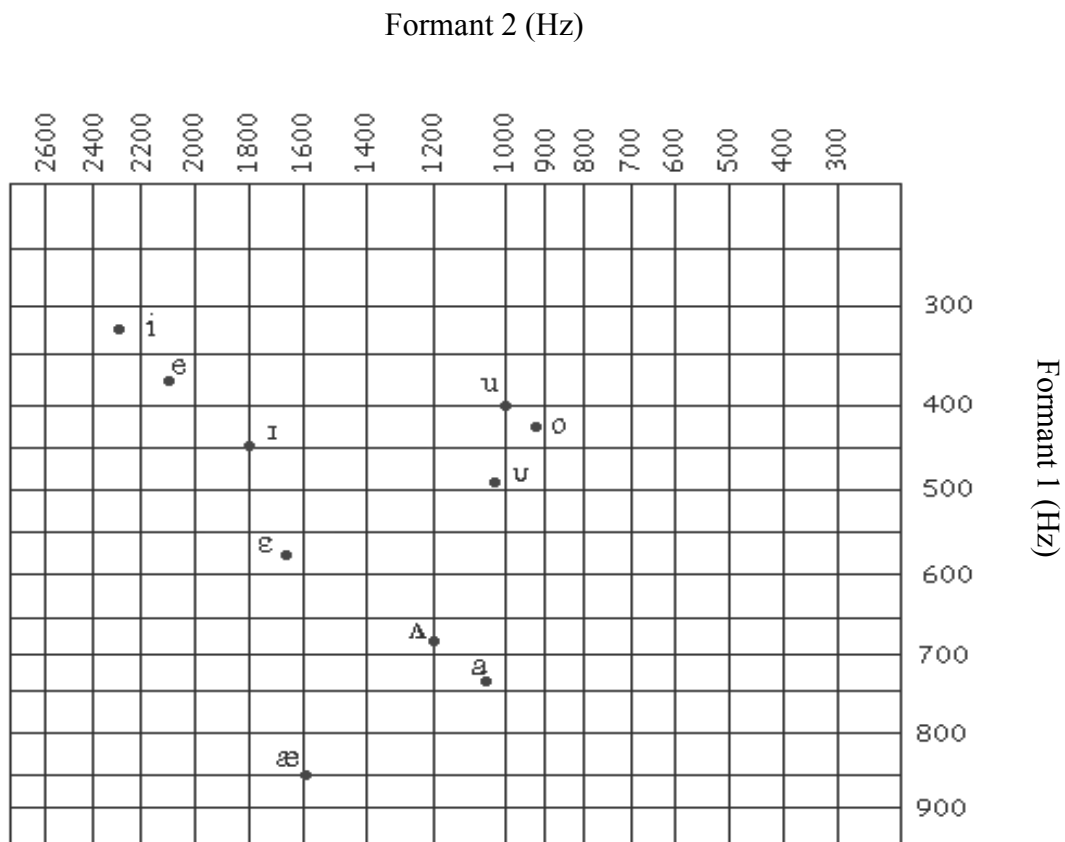
encoded. No current research has explored if there is a relationship between cochlear tuning (measured by OAEs) and the use of auditory feedback in real-time.

Electroencephalography (EEG) techniques can show that complex spectral and temporal aspects of speech are encoded in the central auditory system at the level of the brainstem with the synchronous firing of neurons (Worden & Marsh, 1968; Greenberg, Marsh, Brown, & Smith, 1987). Neurophysiological approaches to speech production and perception may allow a more comprehensive understanding of how both the cochlea and brainstem neurons process complex acoustic stimuli such as auditory feedback and how this affects speech.

This research project investigated the peripheral and neural encoding of vowels at the level of the brainstem and how this might influence the use of auditory feedback in real-time. In the following introduction, a review of Canadian English vowels, auditory feedback, and speech perception will be presented, followed by a detailed summary of research on cochlear tuning and auditory evoked potentials from the brainstem.

## 1.1 Introduction to Canadian English Vowels

The Canadian English language has ten different vowels /i, e, ɪ, ε, æ, ɑ, ʌ, ʊ, o, u/ (Hagiwara, 2006). Each vowel has its own unique spectral characteristics, which can be represented most simply by the frequencies of the first and second formants, F1 and F2 respectively (see Figure 1). Pioneering research demonstrated that only F1 and F2 are required for accurate vowel recognition although higher formants also contribute (Peterson & Barney, 1952). Unlike other English dialects, the vowels /a/ and /ɔ/ in Canadian English overlap substantially, a phenomenon called the Canadian Shift (Clarke, Elms, & Youssef, 1995). This shift has been documented in both Ontario (Clarke et al., 1995) and Manitoba (Hagiwara, 2006). Although this shift is well documented, it does not occur in all regions within Canada, such as the Maritime provinces (Boberg, 2000). Due to the variability of Canadian English vowels, only those who were raised in Ontario or Western Canada were included in the study.



**Figure 1. Canadian-English vowel space.**

Vowels are represented by their first formant on the vertical axis and their second formant on the horizontal axis. This chart was adapted from <http://www.ic.arizona.edu/~lsp/Canadian/canphon2.html> (Mendoza-Denton, Hendricks, & Kennedy, 2001).



## 1.2 Auditory Feedback

The role of auditory feedback during language learning and speech production is well established (Callan, Kent, Guenther & Varperian, 2000; Guenther, Ghosh & Tourville, 2006). Dynamic acoustic environments require moment-to-moment adaptations to maintain accurate and fluent speech. These changes are similar to motor changes in adaptation studies examining perturbations of the arm and hand. When an individual grasps an object, the force of their grasp changes with the load force of the object, such that an increase in load force results in an increase in grasp (Flanagan & Wing, 1993). In bimanual reaching, compensation to moment-to-moment changes in force applied by a robotic arm suggested that participants made pre-planned adjustments to the perturbation and could correct for it rapidly (Jackson & Miall, 2007). This idea of motor adaptation is relevant to speech production as many of the same principles apply. It is important for the talker to monitor and adjust ongoing speech to ensure accurate and appropriate production in changing environments. Speech production relies on two types of feedback: somatosensory and auditory. Somatosensory feedback guides production based on the position of the articulators such as the jaw, lips and tongue (Lindblom & Sundberg, 1971) and works with auditory feedback to control speech production.

Theoretical models of the speech-motor system create a framework from which to interpret experimental results. A number of models have been proposed, establishing a relationship between somatosensory and auditory feedback mechanisms and internal feed-forward models involved in speech production (Perkell et al., 1997; Guenther, Hampson, & Johnson, 1998; Callan et al., 2000; Guenther, 2006; Guenther et al., 2006). Perkell and colleagues (1997) theorized that segmental speech production (e.g. vowels and consonants) involves auditory perceptual goals, which are based on a harmony between articulation and sound. Due to the latency of auditory processing, they hypothesized that solely relying on auditory feedback to guide production is unlikely, thus the system must rely on a sophisticated feed-forward internal model. The internal model is proposed to arise during development, and maps on to different anatomical areas in the brain. The model is made up of a series of auditory perceptual goals that act as a set of targets for the speaker in different environmental conditions. Auditory and

somatosensory feedback play the role of training and maintaining the internal model. Once established, the internal model contributes information in a feed-forward manner alongside the feedback. Guenther and colleagues (1998) postulated that the auditory perceptual goals that make up this internal model are acquired during development and create a network of acoustic and somatosensory information within the auditory system. Although both senses play a role in speech production, our focus is on auditory feedback.

The importance of auditory feedback and its role in speech production was first investigated over a century ago. It was recognized that when talking in noise, individuals raise the intensity of their voice (Lombard, 1911), a result that is now called the Lombard effect. To better characterize these changes in intensity under more controlled conditions, researchers recorded subjects' speech while talking in different noise levels ranging from 0 to 100 dB sound pressure level (SPL). In addition to an increase in amplitude, increases in duration and vocal pitch were noted as well as changes to vowel formant frequencies compared to speech in a quiet environment (Summers et al., 1988; Siegel & Pick, 1974).

Auditory feedback also plays a role in the temporal accuracy of speech production. This was first identified with delayed auditory feedback (DAF). Lee (1950) demonstrated that when speech is played back to an individual with a slight delay during production, the speaker becomes disfluent. A similar study using DAF revealed that the temporal information in auditory feedback influences not only the timing of production, but also other characteristics such as duration and accuracy (Yates, 1963).

Changes to other properties of speech production have been observed when the feedback received is altered in some way. Young cochlear implant users asked to produce the vowel / $\epsilon$ / in "head" had significant changes in formant frequencies when their implants were off, thus receiving no auditory feedback, compared to when their implant was on (Tobey & Murchison, 1989). Studies examining speech production in post-lingually deafened adults reveal changes in voice-onset timing for voiceless stops, more restricted vowel spaces, increased vowel duration, and longer sentence duration compared to normal hearing individuals (Waldstein, 1990). The literature cited above shows that

alterations to auditory feedback result in clear and significant changes in speech production.

### 1.2.1 Pitch Shifts

Pitch is an important characteristic of the voice and carries perceptual information such as emotion and talker identity. Voice pitch is strongly influenced by the fundamental frequency ( $f_0$ ) of the voice, which, in turn is determined by the mass, tension, and length of the vocal folds. Studies show that upward shifts to the  $f_0$  of speech feedback resulted in compensation of voice  $f_0$  in the downward direction and vice versa (Elman, 1981). Opposition is the most typical response to a given manipulation (Larson, Burnett, Kiran, & Hain, 2000; Jones & Munhall, 2000) and suggests the use of internal pitch representation. In some instances however, individuals will follow the manipulation suggesting the feedback is used as an external cue (Burnett et al., 1998). Similar results have been found in cross cultural studies using tonal languages such as Mandarin (Yi Xu, Larson, Bauer, & Hain, 2004; Jones & Munhall, 2002).

Compensation to pitch-shifts is generally only a fraction of the manipulation introduced. Burnett et al. (1998) noted that responses to shift magnitude were not proportional, suggesting that vocal motor control does not rely entirely on auditory feedback. This result agrees with other similar studies (Larson, 1998; Larson et al., 2000; Chen, Liu, Xu, & Larson, 2007).

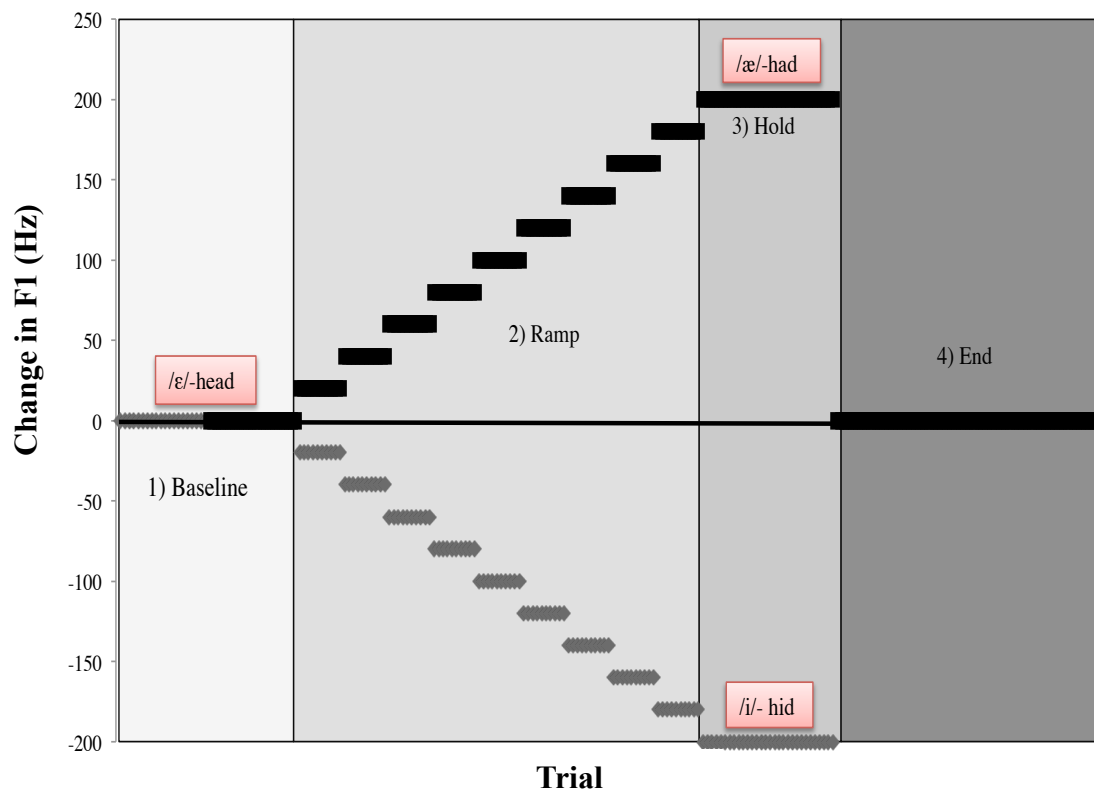
### 1.2.2 Formant Perturbations

Various other laboratory studies have investigated speech compensation during vowel formant manipulation. Positive and negative frequency shifts in the F1 of an isolated English vowel resulted in a compensatory response in F1 production in the opposing direction (Purcell & Munhall, 2006a). Similar results have been found using normally voiced words (Purcell & Munhall, 2006b; Villacorta, Perkell & Guenther, 2007) and whispered speech (Houde & Jordan, 1998). The formant manipulation paradigm is often organized into four distinct phases: the Baseline phase, the Ramp phase, the Hold phase and the End phase (see Figure 2). The F1 shift takes place during the Ramp phase once a baseline of production is achieved. In a study by Purcell and Munhall (2006), the vowel

/ɛ/ in “head” was shifted in 4 Hz steps across 50 trials (+200 Hz total) to produce the vowel /æ/ in “had” while production was recorded (Purcell et al., 2006a). In the Ramp phase, F1 is gradually filtered and increased or decreased in frequency so the change goes undetected by the talker. However, compensation to the manipulation appears to be an unconscious process, occurring automatically (Munhall et al., 2009; Elman, 1981).

A challenge in formant shifting paradigms is to maintain a natural sounding vowel throughout the manipulation. One way to accomplish this is by shifting both F1 and F2 at the same time. However, estimates of F2 can be quite variable, which can result in undesirable feedback during real-time processing. Studies have demonstrated that the speech motor control system can independently adjust for changes in F1 and F2 (MacDonald, R. Goldberg, & Munhall, 2010; Munhall, MacDonald, Byrne, & Johnsrude, 2009). Further, results showed that changes to F1 did not affect the whole vowel spectrum, just the energy around the manipulated formant (MacDonald, Purcell, & Munhall, 2011). Therefore, the independence of formant control does not necessitate manipulations of both formants together, allowing researchers to manipulate F1 with its more stable estimates.

Compensation to formant manipulations is not complete and generally is only a fraction the manipulation. Studies have found that on average, subjects compensate around 25% to 50% of the manipulation (Houde & Jordan, 1998; Purcell & Munhall, 2006b; Villacorta, Perkell, & Guenther, 2007; MacDonald et al., 2010; MacDonald et al., 2011). Studies examining the effects of post-lingual deafness indicate the importance of auditory feedback for accurate production (Waldstein, 1991). However, the incomplete compensation observed in this paradigm indicates that other types of feedback are contributing to the control of speech. Studies such as Tremblay, Shiller, & Ostry (2003) and Dhanjal, Handunnetthi, Patel, & Wise (2008) outline the role of the somatosensory system in the control of speech. One explanation for this partial compensation is an integration of the two signals into a speech-motor control system, which then weighs the importance of each signal based on the feedback received (MacDonald et al., 2010). At a point during the manipulation, there may be such a discrepancy between the different



**Figure 2. Representation of the phases in formant-shift paradigms.**

The formant-shift paradigm consists of four phases. 1) Baseline, 2) Ramp, 3) Hold, 4) End. This figure was adapted from Mitsuya et al. (2011). The black rectangles represent the formant manipulation and the grey points represent hypothetical production that perfectly opposes the manipulation.

inputs that the system relies more heavily on the somatosensory feedback to guide production. Based on the altered auditory feedback literature, incomplete compensation is expected.

The perception of auditory feedback is critical for speech-motor control. Perceptual organization of the vowel space, vowel categories, and vowel goodness all influence formant control (Mitsuya et al., 2011). Vowel goodness is defined as the ability of an exemplar to fit into a specific category (Kuhl, 1991). Goodness ratings are established by having participants rate a vowel prototype on its apparent “goodness”. Individuals tend to give high ratings to the prototype and lower ratings to exemplars that move farther away from the prototype (Iverson & Kuhl, 1996). A robust correlation has been found between individuals’ vowel goodness ratings and compensation measures (Nguyen, 2012).

Auditory feedback plays an important role in guiding speech production while working concurrently with learned, internal models of speech. Having established an introductory understanding of how alterations to auditory feedback manifest at the behavioural level, the next step is to determine how alterations are represented in the peripheral and central auditory systems. The following sections will review past research investigating physiological and psychoacoustic measures of cochlear tuning as well as brainstem auditory evoked potentials.

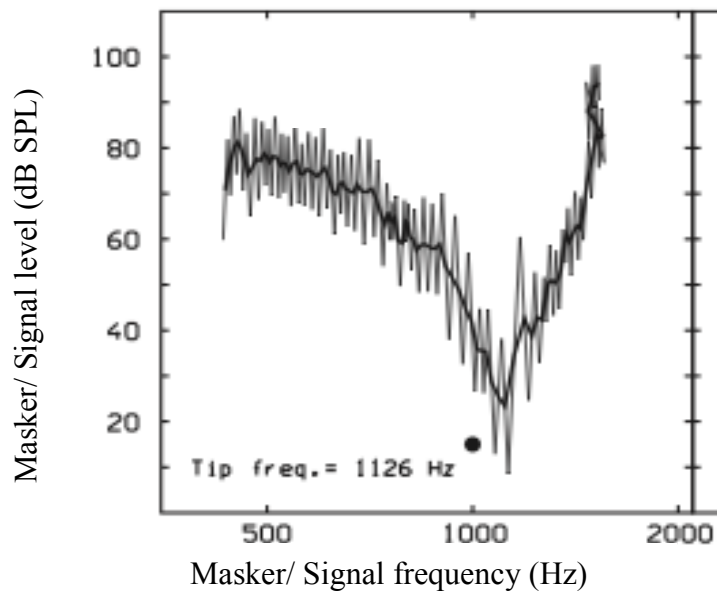
### 1.3 Psychoacoustic and Physiological measures of cochlear tuning

The peripheral auditory system (i.e. the human ear) is a highly complex sensory organ that is not completely understood. Sound travels to the cochlea via the tympanic membrane, setting into motion the ossicles, which in turn set the fluid of the cochlea in motion. The mechanical sound wave energy in the cochlea is then transformed into electrical signals via the hair cells. These signals travel to the central auditory system via the auditory nerve (Seikel, King & Drumright, 2010). The aim of this research project was to investigate the role of acoustic output from the cochlea (via the hair cells) and behavioural measures of cochlear function in the use of auditory feedback during speech production.

The tonotopic organization of the auditory system begins with the basilar membrane (BM). The frequency selectivity of the BM can be represented as a series of auditory filters, with the centre of each filter corresponding to a specific frequency location (Sek, Alcantara, Moore, Kluk, & Whicer, 2005). Cochlear tuning, the frequency selectivity of the cochlea, can be investigated through the use of psychoacoustic tuning curves (PTC). One way to measure PTCs is with a sinusoidal signal presented at a low level and a set frequency. A narrow-band noise masker is added and the level required to just mask the signal is determined (Small, 1959). PTC results are graphed on a logarithmic scale with masker/signal level on the vertical axis and frequency on the horizontal axis. In normal hearing individuals, the low frequency part of the curve is negative sloping followed by a steep positive slope above the signal frequency (Sek et al., 2005; see Figure 3).

Traditional PTC measures require testing times of approximately 1 hour (Small, 1959), however, new attempts to reduce this time requirement have been made. Sek et al (2005) have begun work on a program that aims to measure PTCs in less than 5 minutes. Their results demonstrated that the PTCs obtained in both normal and hearing-impaired listeners were highly correlated with PTCs measures using traditional methods. A similar method has also been developed for testing in children (Malicka, Munro, & Baker, 2009). A faster PTC measurement would be beneficial in both research and clinical settings.

When the cochlea is stimulated with acoustic input, the outer hair cells (OHCs) depolarize and hyperpolarize causing them to move. This motility results in the OHCs acting as an amplifier to the acoustic signal as it travels up the auditory system. The additional energy can also reverse and travel back out to the ear canal (Kemp; Robinette & Glatke, 2007; pg. 56). This additional energy from the OHCs results in acoustic outputs from the cochlea called otoacoustic emissions (OAEs), which are detectable in the ear canal using a sensitive microphone (Kemp, 1978; Kemp, 1979). A number of different types of OAEs can be recorded. Click or transient evoked OAEs (TEOAE) and distortion product OAEs (DPOAEs) are commonly used in a clinical setting as a measure of cochlear health (i.e. hearing loss). Two other types of OAEs commonly encountered are spontaneous OAEs (SOAEs) and stimulus frequency OAEs (SFOAEs). Each type



**Figure 3. Example of psychoacoustic tuning curves.**

Example of PTC data. The vertical axis represents the noise masker and sinusoidal signal level in dB SPL and the vertical axis represents the noise masker and sinusoidal signal frequency in Hz. Figure was adapted from Sek et al., (2005)



of OAE is evoked with different stimuli and thus has a unique response pattern (Kemp; Robinette & Glattke, 2007; pg. 28). TEOAEs are evoked by a brief stimulus and reflect all the frequency components of the evoking stimulus in a complex sound waveform. SOAEs are natural pure tones produced by the cochlea at certain frequencies unique to the individual. DPOAEs are evoked by a pair of pure tone stimuli, which stimulate the OHCs at the same time and produce a distortion emission at their place of interaction on the basilar BM. Finally, SFOAEs are evoked by pure tone stimuli and can be used to estimate cochlear tuning (Kemp; Robinette & Glattke, 2007; pg. 28). Cochlear tuning plays a role in our ability to distinguish between acoustic stimuli (Shera & Guinan, 2003), which is important for speech perception. Tuning therefore was selected as an appropriate measure of peripheral auditory function for the current study.

PTCs and SFOAEs measures can provide behavioural and physiological information about the frequency selectivity of the cochlea, respectively. Knowledge of cochlear frequency selectivity may reveal that more narrow auditory filters might better detect formant changes, which may influence how speech errors are remedied in real-time.

## 1.4 Auditory Evoked potentials

Currently, an understanding of how acoustic elements of speech are encoded and how this neural representation may influence the changes observed in speech production is missing from the literature. To establish a thorough understanding of the neural processes involved in the human auditory system, the following section will review the methodology and findings from a number of auditory evoked potential investigations. Due to the scope of this project, we will only be considering brainstem auditory evoked potentials. Although the cortex evidently plays an integral role in speech perception, our focus will remain on the brainstem (which precedes cortical processing) and more specifically, the frequency following response (FFR) and the envelope following response (EFR).

The electrical signals produced by the brain can be measured using a technique called electroencephalography (EEG). During EEG measurements, surface electrodes placed on the scalp pick up changes in the ionic current flow of neurons and record voltage changes

in response to brain activity. EEG can be employed to record both cortical and brainstem electrical activity to specific sensory events in time. Evoked potentials (EPs) are the summed, time-locked activity from a large number of neurons produced by the presentation of a sensory stimulus (Kandel, Schwartz & Jessel, 2000) and can provide us with information (spectral and temporal) with regards to how the brain processes different sensory stimuli (i.e. somatosensory, visual, and auditory).

When the human auditory system is presented with a sound, the EEG signal undergoes specific changes that are related to the spectral and temporal properties of that stimulus (Burkard, Don & Eggermont, 2007). Auditory evoked potential (AEPs) are measured from the surface of the scalp and reflect neural activity in response to acoustic stimuli. AEPs can provide information about the function and integrity of the auditory pathway and can reveal pathology that may not be detectable at the level of the cochlea (Berger & Blum, 2006; p. 475) or through traditional behavioural methods.

The AEP signal is composed of contributions from different neural generators in the central auditory system. AEPs can be characterized into near-field and far-field potentials, depending on the location of the electrode placement. Near-field recordings are those collected from electrodes placed directly on structures of the auditory nervous system (e.g. cochlear nucleus), whereas far-field potentials are recorded from electrodes more removed (i.e. scalp) from their source (Moller et al., 2006; p. 152). Far-field potentials are less specific, because they receive inputs from a number of different neural and anatomical sources as well as from muscle activity (e.g. eye blinks and swallowing). The sensitivity of the EEG to muscle movements makes it necessary to collect a large number of samples in order to reduce the influence of unwanted artifacts through averaging. The technique of averaging allows the response, which is time-locked to an acoustic stimulus, to be emphasized while all the random background noise is reduced.

Input from multiple anatomical generators along the auditory pathway results in the neural response containing multiple components with different latencies, reflecting the different origins (Moller, 2006; p. 164). The different components with varied latencies reveal important information about the spectral and temporal characteristics of the

acoustic stimulus. A number of different AEPs have been identified and characterized based on their response characteristics.

In 1974, Picton and colleagues identified 15 discrete components of AEPs in humans using vertex-mastoid electrode placements (Picton, Hillyard, Krausz, & Galambos, 1974). These components were reliably evoked using tone bursts and clicks at 60 dB SPL in a number of participants. The 15 identified components were divided into early, middle and late responses depending on their latency, with each representing different locations along the auditory pathway from the cochlea to the cortex. It was determined that the early components (occurring within 8 ms of stimulus presentation) represent activity at the cochlea and the brainstem auditory nuclei (Gerken, Moushegian, Stillman, & Rupert, 1975). The later components are from generators located higher up in the auditory pathway. Studies revealed that no significant changes in the peak latency of any brainstem EP component are observed when an individual is asked to attend versus ignore the stimuli (Picton & Hillyard, 1974). The same study showed that AEP measurements taken while the participant was sleeping were less noisy. Three common AEPs encountered in the literature include the auditory brainstem response (ABR), the FFR and the EFR. Our focus will be on the FFR and EFR because they readily reflect the frequency characteristics of vowels.

## 1.5 Frequency Following Response (FFR)

The human FFR was first described in the early 1970s. The FFR represents synchronous neural activity in upper brainstem structures, with response spectrum peaks corresponding to the periodicity of the stimulus frequency (Moushegian, Rupert, & Stillman, 1973). Band-pass filtered recordings from implanted electrodes in a cat brain demonstrated a persistent electrical response that recreated the sine wave of the auditory stimulus (Worden & Marsh, 1968). The response and the stimulus spectral profiles were very similar and suggested that the central auditory system was capable of closely representing acoustic stimuli.

Experiments using multi-electrode recordings in the cat revealed the FFR was made up of phase-locked synchronous inputs from structures such as the cochlear nucleus (CN), the

ventral nucleus of the lateral lemniscus (VNLL), the dorsal nucleus of the lateral lemniscus (DNLL), and the inferior colliculus with some contribution from the superior olivary complex (Marsh, Brown, & Smith, 1974; Smith, Marsh, & Brown, 1975). There is agreement that the synchronous phase-locked activity of upper brainstem nuclei are involved in the generation of the FFR (Greenberg et al., 1987), however there was some dispute regarding the degree to which each neural site contributes to the response (see Gardi, Merzenich, & McKean, 1979). Animal studies revealed that electrode placement influences the degree to which different structures contribute to the FFR signal (Davis & Britt, 1984).

Worden et al (1968) ruled out electrical inputs solely from peripheral auditory structures (e.g. cochlear microphonic) as the source of the FFR due to the long onset latency (approximately 6 ms) and a reduction in the amplitude of the FFR in the presence of masking noise (Glaser, Suter, Dasheiff, & Goldberg, 1976). The long onset latency suggests neural origins within the classical auditory pathway, more specifically, within nuclei in the upper brainstem region (Batra, Kuwada, & Maher, 1986). The discovery of the FFR and its role in brainstem level encoding of basic sound stimulus properties in humans led researchers to investigate how this response might play a role in representing more complex sounds such as frequency modulated tones, synthetic speech and natural speech.

Phase-locking in the FFR may play a role in representing speech and processing information at the level of the brainstem that is critical for perception. Perceptual elements of speech such as pitch, intonation, prosody and loudness all carry information that influences the speech signals' intelligibility. In normal hearing individuals, the FFR was recorded using spectrally complex tones and responses were found to contain energy concentrated at the  $f_0$ , with pitch-relevant information being encoded by phase locked activity in upper brainstem nuclei (Greenberg et al., 1987). This demonstrates that the brainstem is robustly encoding important elements of the speech signal such as pitch. In a more recent study, the FFR waveform showed clear spectral peaks at the two formant frequencies of three English 'vowel-like' sounds (Krishnan, 1999). These results suggest that the FFR reflects brainstem activity that is phase-locked to the individual frequency

components of the stimuli, whereby the first and second formant frequency are robustly represented. Similar results were shown for more complex synthetic English vowel sounds (Krishnan, 2002), natural English vowels (Aiken & Picton, 2008a), and four different Mandarin tones (Krishnan, Yisheng Xu, Gandour, & Cariani, 2004). Other laboratory studies have measured FFRs in response to the  $f_0$  of a synthetic speech sound [da], with more robust FFRs occurring after auditory training (Hayes, Warrier, Nicol, Zecker, & Kraus, 2003; Russo, Nicol, Zecker, Hayes, & Kraus, 2005; Russo, Nicol, Musacchia, & Kraus, 2004).

## 1.6 Envelope Following Response (EFR)

The EFR is an AEP where the neural activity follows the periodicity of the stimulus envelope (Hall, 1979). The EFR is recorded from surface electrodes placed on the scalp and can be used to objectively assess the hearing of individuals that cannot participate in traditional behavioural tests of hearing. The EFR is elicited by complex auditory stimuli such as a modulated sinusoidal or noise carriers and speech (Levi et al., 1995; Levi, Folsom, & Dobie, 1993). It has been demonstrated that when adult participants are presented with amplitude modulated tones ranging from 150 to 450 Hz, the neural response closely follows the amplitude modulated envelope of the stimulus (Kuwada, Batra, & Maher, 1986). Similar results have been found when presenting young infants with 80 Hz amplitude modulated tones (Levi et al., 1995).

Prosodic features of sound such as rhythm and intonation carry a lot of communicative information in the envelope of speech. The ability to follow and perceive changes in the speech envelope is important for accurate speech perception. In studies such as Purcell, John, Schneider, & Picton (2004), behavioural measures of temporal acuity (e.g. gap and modulation detection tasks) were closely related to the frequency at which the EFR was no longer detected. A similar study by Dajani, Purcell, Wong, Kunov, & Picton (2005) noted that the human EFR accurately tracks the pitch contour of a natural vowel, and reflects small changes in the periodicity of speech which can be detected behaviourally.

In another study, EFRs were found to follow the speech envelope of three different natural English vowels (/a/, /i/, /u/) in normal hearing individuals (Aiken & Picton, 2006).

Similarly, significant EFR peaks were detected at the fundamental frequency of two different vowels (/a/ and /i/) in all normal hearing participants included (Aiken & Picton, 2008a). Most recently Choi et al. (2013) recorded EFRs to five English vowels present in three different sentences or as a steady-state string of vowels. Both the steady-state vowels and the vowels embedded within sentences elicited significant responses.

## 1.7 Rationale

The objective of the proposed research study is to better understand how the human auditory system encodes the information in auditory feedback at the peripheral and central levels and how this information influences production. More specifically, this project sought to investigate how the auditory brainstem encodes changes in vowel formants during speech production. From this information, it may be possible to determine if individual differences in peripheral and neural encoding are related to the varied use of auditory feedback across different individuals. No study has investigated if SFOAE measures and AEP measures (EFRs and FFRs) influence the use of auditory feedback in real-time. In order to investigate this relationship further, measures of central and peripheral function were paired with a real-time auditory feedback perturbation task.

## 1.8 Hypotheses

For the present study, it is hypothesized that 1) individuals who produce a greater compensation response (i.e. reduction in vowel F1 frequency) to real-time perturbations in vowel F1 feedback, will have greater amplitude AEPs (EFR and FFR), 2) individuals who produce a greater compensation response to real time perturbations in vowel F1 feedback will have a greater difference between AEP response amplitude (excluding phase) to two different vowels (in the present case /ε/ and /æ/) and 3) that individuals who produce a greater compensation response to real time perturbations in vowel F1 feedback will have a greater vector difference between AEP response magnitude (including phase) to two different vowels (in the present case /ε/ and /æ/). The predictions were generated under the assumption that those who have higher amplitude responses and who show more of a change from “head” to “had” are receiving better auditory information and therefore will produce a larger behavioural response to remedy the perceived error during

the formant manipulation. Further, it is hypothesized that 4) individuals who produce a greater compensation response to real-time perturbations in vowel F1 feedback will have narrower auditory filters when measured both physiologically (i.e. using SFOAEs) and behaviourally (i.e. using PTCs) than those who compensate to a lesser degree. More narrow auditory filters may allow improved detection of frequency changes in the vowel formant harmonics, resulting in greater compensation.

## Chapter 2

### 2 Methods

#### 2.1 Participants

Thirty-nine participants were recruited from the Western University community and the city of London. All participants were English talkers (25 females, 15 males; ages 17-29 yr, mean: 22, SD: 3.35). Participants had learned English as their first language in Canada, predominantly in Ontario. Hearing thresholds were measured for each ear at octave intervals between 250 Hz and 4 kHz. Individuals were included if their thresholds were in the normal range ( $\leq 20$  dB HL). There was one participant with a slightly elevated threshold at 2000 Hz in one ear. This was not expected to influence the results using supra threshold speech so the participant was retained. Each participant attended two testing sessions. No participants had known neurological, language, hearing, or speech impairments as determined by questionnaires.

#### 2.2 Summary of Procedures

The following paragraph provides a brief summary of all the experimental procedures carried out for this research project. Detailed descriptions of each procedure follow in subsequent sections (see Figure 4 for a brief overview). This study was completed over two separate testing sessions lasting for approximately 1 hour and 1.5 hours, respectively. The experiment was first explained to the participant and she/he was asked for informed consent and to complete some short questionnaires concerning demographic information, language experience, and music history. Audiometric thresholds were then determined to ensure that the participant's hearing fell within normal limits. An altered auditory feedback task was then performed with the participant seated comfortably in an Eckoustic C-26 sound booth. After completion of the altered feedback task, participants completed two perceptual tasks in a quiet laboratory environment to determine vowel goodness ratings and F1 discrimination thresholds. In the second testing session, middle ear function was evaluated using a tympanometer, to ensure typical middle ear function. Participants then completed a perceptual task in the sound booth to obtain a behavioural

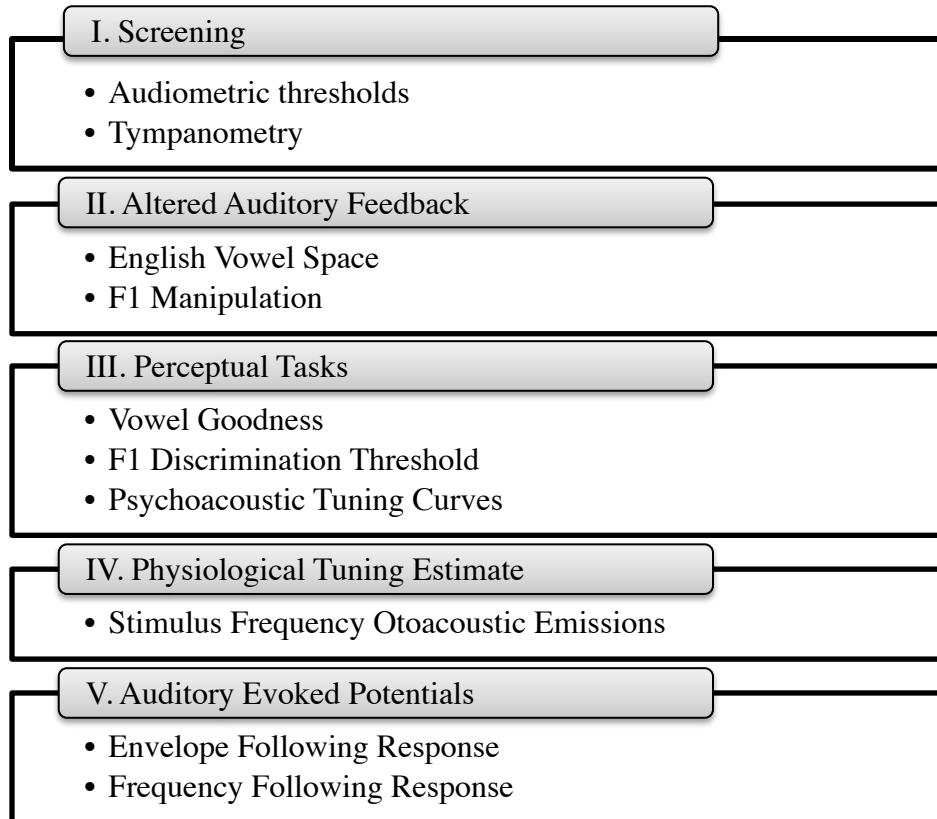


measure of cochlear tuning (Sek, *et al.*, 2005). Once the behavioural measure was complete, a physiological estimate of tuning was obtained with SFOAEs. Participants were then fitted with surface electrodes on the scalp, and brainstem FFRs and EFRs were recorded. Once all tasks were complete, participants were provided with a summary of the experiment and compensation for their time and effort. The Western University Health Sciences Research Ethics Board approved all questionnaires and experimental procedures (see Appendix A).

## 2.3 Perceptual Measures

### 2.3.1 Vowel goodness

Vowel goodness is a perceptual measure of vowel quality that is highly correlated with speech compensation (Nguyen, 2012). The term goodness is defined as the ability of an exemplar of a specific sound to fit into its respective category (Kuhl, 1991). This measure allows for identification of individuals who do not perceive vowel goodness in a typical fashion. To determine the goodness of different exemplars of the vowel / $\epsilon$ /, 11 different versions were created on an F1 continuum from “head” to “had” using filtering similar to during online formant shifting (see 2.4.4). The F1 of the unaltered “head” was shifted upwards in 20 Hz steps to +200 Hz (i.e. +20, +40, +60, +80 ... +200 Hz) towards / $\text{æ}$ /. The 11 utterances were randomly ordered and presented nine different times for a total of 99 trials. The first four repetitions of the set of sounds were not used in order to allow the participant to know the full range of / $\epsilon$ -like sounds. Each participant was asked to rate the versions of “head” on a scale of 1 to 7, where 1 is a very poor version of the word “head” and 7 is an excellent version. These vowel goodness ratings were taken to represent the participant’s perceptual organization of the vowel / $\epsilon$ /.



**Figure 4. Overview of study methodology and measurements**

### 2.3.2 F1 discrimination threshold

F1 Discrimination threshold is the smallest change in F1 that the listener can detect perceptually. A two-alternative forced choice procedure (2AFC) was used to determine the F1 discrimination threshold for /ɛ/ in “head”, with shifts of F1 in the positive direction towards /æ/ in “had”. A continuum of “head” was produced by shifting F1 upwards in 5 Hz steps using a method similar to that done in the goodness task and online (see 2.3.1 and 2.4.4). An adult male whose first language was English produced the unaltered version of “head”. Dinosaur, an AXB 2AFC program developed by Dorothy Bishop (Oxford University), was used to complete this measure. During the program, participants were asked to make a judgment about which sound, the first or the last, was most like the unaltered, middle presentation of “head”. As the Dinosaur program continued, the F1 difference between the two sounds became smaller and more difficult to detect. When the participant made two correct selections consecutively, the task was made harder, by having the participant hear a smaller shift, and when the participant made an incorrect selection, the task was made easier by having the participant hear a larger shift: this was considered one reversal. After eight reversals, the program ended and the participant’s F1 discrimination threshold was found by averaging the shift magnitude for the final four reversals.

### 2.3.3 Psychoacoustic tuning curves

Psychoacoustic tuning curves (PTCs) can be used to measure the frequency selectivity of the auditory system. This measurement was performed using a fast PTC measurement program (SWPTC) developed by Aleksander Sek at Mickiewicz University (see Sek, *et al.*, 2005 for more information).

The SWPTC program (Sek, *et al.*, 2005) was run on a laptop computer while participants were seated comfortably in a sound booth. The stimuli were presented over Sennheiser HD 280 pro over-the-ear headphones to the left and then right ear. Trials were approximately 3 minutes in duration.

Following methods used by Sek and colleagues (2005), participants were instructed to attend to the 1000 Hz pure-tone beep throughout the experiment in either their left or

right ear. The signal beep was 200 ms in duration, with a 200 ms gap between each beep. The task was initiated by the participant and began with repetitions of the pure-tone beep in isolation presented at 40 dB SPL. Following this introduction to the pure-tone beeps, a noise masker was added with a centre frequency of 500 Hz at 40 dB SPL and was swept upwards to 1500 Hz. Participants were instructed to hold down a button until they could no longer hear the pure-tone beep. Holding down the button increased the level of the noise at a rate of 2 dB per second. To avoid any discomfort the maximum output level of the masker was 80 dB SPL. Once the beep was inaudible, participants were instructed to let go of the button: this was considered one reversal. The frequency of the masker changed only after the first four reversals. As soon as the beep was heard again, participants once again pressed the button. Regression lines were fit to each side of the PTC and the width of the PTC was measured 10 dB above where the lines intersected (Malicka, Munro, & Baker, 2009).  $Q_{10\text{dB}}$  was calculated to measure the sharpness of the PTC and therefore cochlear tuning.

## 2.4 Altered auditory feedback

### 2.4.1 Equipment

Participants were prompted on a computer screen to speak the target word at a rate of approximately one word every two seconds. Participants wore a Shure WH20 headset microphone. The microphone signal was amplified using a microphone amplifier (Tucker-Davis Technologies MA3) with a +20 dB gain switch active and adjustable gain set individually as described below. The signal was low pass filtered with a cut-off frequency of 4500 Hz (Frequency Devices type 901). The analogue signal was then digitized at a 10 kHz sampling rate with 18-bit precision (National Instruments PXI-6289M input/output board). During altered auditory feedback, the signal was analyzed and filtered in real time to create the formants shifts (National Instruments PXI-8106). The digital signal was converted back to analogue sound at 10 kHz with 16-bit precision by the National Instruments PXI-6289M and routed to a Madsen Itera audiometer for amplification. During practice trials, the microphone MA3 amplifier gain was adjusted between 20 and 40 dB gain for each participant. The setting chosen for each talker ensured the vocal sounds reaching the Madsen Itera were approximately 0 dB on its input

VU meter. This VU meter reading corresponded to 80 dBA SPL at the listeners' ears using Sennheiser "HD 265 linear" headphones. The Madsen Itera audiometer also added background speech shaped noise of 50 dBA SPL to hide small imperfections that may have occurred during filtering. All equipment reported was similar to Purcell and Munhall (2006b).

## 2.4.2 Formant Estimation

Estimating formants in speech signals is commonly approached through LPC, linear predictive coding (O'Shaugessy, 1988). Linear filter coefficients are determined by the LPC method, which can predict the current speech sample from a weighted combination of previous samples. When the coefficients' filtering characteristic is represented in the frequency domain as a spectrum, it resembles a spectral envelope fitted over the actual speech harmonics. Formant estimates are given by the peaks in this LPC envelope, where the number of formants is set by the model order. An optimization procedure was carried out to determine the best model order (BMO) for producing stable formant estimates before the altered auditory feedback was completed. Tokens of /ε/ in "head", recorded with the English vowel space, were used to calculate formant estimates using various models from 8 to 12. The model order that produced the least variable F1 and F2 estimates was considered the best.

## 2.4.3 Procedure and experimental conditions

After participants arrived, informed consent was obtained and three short questionnaires were completed (medical background, language background and music history; see Appendix B). Screening questionnaires were completed to ensure that participants were in good health, had normal hearing and were native English speakers. Participants were asked about their music history because of the potential influence musical training could have on the results. Participants' hearing thresholds were then tested using a pure-tone audiogram. Normal thresholds were  $\leq 20$  dB HL at octave intervals between 250 Hz and 4 kHz using TDH-296 headphones and a Madsen Itera Audiometer.

Participants were seated comfortably in a chair in the sound booth. The task was explained to participants and they were asked to produce all of the consonant-vowel-

consonant (CVC) words used in the study. Individuals were asked to speak normally and to keep the loudness and pitch of their voice relatively consistent as they uttered each of the prompted words on the monitor. Microphone amplification adjustments took place at this time and then participants were prompted with the following words: head, had, heed, hid, hayed, hawed, and who'd to collect their English vowel space. Talkers then went through five phases of an F1 positive shift for the English vowel / $\epsilon$ /, where they repeated the word "head" 220 times. The five phases were Acclimatization, Baseline, Ramp, Hold and End. In the Acclimatization phase (first 40 utterances) and the Baseline phase (utterances 41 to 60) individuals received normal, unaltered feedback. In the Ramp phase, (utterances 61 to 140) auditory feedback was shifted upwards by 20 Hz every 10 utterances to a maximum shift of +200 Hz. In the Hold phase (utterances 141 to 160) participants received the maximum +200 Hz F1 shift. Finally, in the End phase (utterances 161 to 220), the manipulation was removed and participants received unaltered feedback.

#### 2.4.4 Online voice detection and formant shifting

Auditory feedback was altered in real-time by filtering the utterance during the voiced part of speech. A statistical amplitude threshold technique was used to detect the onset of voicing in each trial. This was accomplished by determining the mean and standard deviation of the microphone input level during a quiet period prior to the prompt. When the microphone input level exceeded this mean input level by six standard deviations, voice onset was assumed to have occurred. From this point onwards, the voice was filtered using coefficients determined from real-time LPC formant estimates, which were updated every 900  $\mu$ s. The formant manipulations were achieved through two filters that simultaneously processed the speech signal. One filter deemphasized harmonics near the current F1 and the second emphasized harmonics near the desired F1, thereby shifting the formant.

### 2.4.5 Offline formant analysis

Prior to analysis, all trials containing overt pronunciation errors were removed from the data set. Subsequently, each vowel was cropped from its utterance by a semi-automated program. The experimenter then verified vowel boundaries.

Offline estimates were calculated for the first three formants (F1, F2 and F3) for each utterance. A single steady-state value for each formant was calculated by averaging the estimates from the middle 60% of the vowel for that formant. The analysis only includes the middle 60% of the vowel because the first and last 20% of the vowel have formants that may be in transition or estimates that are unstable. A graph of all the F1, F2 and F3 values for each participant was inspected for any incorrect categorization of formants (i.e. F1 being characterized as F2, etc.) by the offline LPC algorithm. If categorization errors were present, the experimenter corrected them. Formant values were graphed in the order that they were produced during the experiment.

## 2.5 Otoacoustic Emissions

### 2.5.1 Stimulus generation and recording

Pure tone stimuli used to elicit the SFOAE ranged in frequency from 960 Hz to 1920 Hz with a resolution of 48 Hz and were digitally generated using Matlab (Mathworks Inc, MA, USA). This frequency range was selected under the assumption that filter bandwidths will be similar near the F1 of the vowel / $\epsilon$ / (approximately 530 Hz and 610 Hz in men and women, respectively; Baken, 1987) to filter bandwidths near 1 kHz.

Practically, it is challenging to measure SFOAE below about 750 Hz due to background noise. A custom LabView program (National Instruments, TX, USA) was used to record responses. The total measurement duration was approximately 15 minutes.

The digital stimulus was converted to analog signals in the digital-to-analog (and analog-to-digital) converter at a sampling rate of 32000 Hz (National Instruments, TX, USA, type 6289M series acquisition card). The levels of all output signals were controlled using PA5 attenuators (TDT Tucker-Davis Technologies, FL, USA). Following attenuation, the signals were power amplified using TDT SA1 amplifiers driving

Etymotic ER2 transducers connected to an ER-10B+ otoacoustic emission probe that delivered the signals in the ear-canal. The system was calibrated using a Bruel and Kjaer sound level meter and ear simulator. An online in-the-canal calibration was also performed at the beginning of every frequency to adjust the level of the stimulus to produce the desired SPL at the probe tip regardless of the size and acoustic impedance of the individual ear canal. The minimum acceptable signal to noise ratio (SNR) to consider a response an OAE was set at 12 dB. In the SFOAE recording, it is common to see poor SNRs at some frequencies due to the interaction of the forward and reverse traveling waves called microstructure (Goodman et al., 2003). Participants whose responses did not meet the SNR criteria (except for microstructure) were not included for further analysis in SFOAE.

Participants sat in a comfortable chair in a sound attenuated booth and were encouraged to relax and try their best to swallow as little as possible and sleep if possible. SFOAEs were recorded from only the left ear. To extract SFOAEs, the “suppression method” (Brass and Kemp, 1993; Kalluri & Shera, 2007) was used. In this method the stimulus tone was presented at 40 dB SPL continuously. When the stimulus tone is presented in isolation, the recording contains the stimulus and OAE. Periodically, a suppressor tone of 60 dB SPL and frequency +16 Hz relative to the stimulus is introduced to eliminate the OAE. This results in just the stimulus tone being recorded. A vector subtraction between the two conditions is then done to eliminate the stimulus tone and obtain an estimate of the SFOAE. Tuning is determined from the SFOAE measure through SFOAE group delay. Group delay is determined by calculating the slope of the SFOAE phase across frequency. From group delay, we can calculate the equivalent rectangular bandwidth (ERB), a simplified estimate of the filter bandwidth, which can be used to estimate  $Q_{\text{ERB}}$  ( $Q_{\text{ERB}} = 1 \text{ kHz/ERB}$ ).

## 2.6 Evoked Potentials

### 2.6.1 Stimuli

The stimuli for the EEG recording were developed from two separate sources. A standard version of the English vowel /ε/ in “head” was produced by a 28 year old prototypical



male talker with most of his schooling in Western Canada and Ontario. A version of the vowel /æ/ in “had” was created by shifting the F1 of this standard “head” upwards by 200 Hz. In addition to the standard talker, a token of the subject’s own version of the English vowel /ε/ was selected from the 20 baseline trials in the formant manipulation paradigm mentioned above. Each baseline utterance of the word “head” was analyzed to determine its duration. Of the five longest vowel productions, the  $f_0$ , and quality were determined using Praat (Boersma & Weenink, University of Amsterdam). The quality of tokens was assessed based on the overall perceptual quality of the vowel, the stability of the pitch, the duration, and the absence of any glottal fry (i.e. creaky voice). Based on these criteria, the best exemplar was selected and it was filtered using MATLAB (Math Works, Natick MA) to produce an exemplar of the English vowel /æ/ in “had”. Again, this was accomplished by shifting the first formant of the vowel /ε/ upwards by 200 Hz. These tokens were then combined into a single stimulus consisting of the standard talker’s versions of the words “head” and “had” and the subject’s versions of the words “head” and “had”. The stimulus was presented in its original polarity, then inverted and presented in the opposite polarity. Together these were considered one full stimulus sweep. The duration of each polarity presentation varied between subjects because each participant’s vowels were different durations. Vowel duration ranged from 0.13 s to 0.25 s. The utterances were presented repeatedly at an overall level of 80 dB SPL for 500 sweeps or a total duration of approximately 55 minutes.

## 2.6.2 Polarity asymmetry in the EFR

Early in analysis, it was noted that responses elicited by the speech stimuli in polarity A differed in amplitude from the responses elicited by the speech stimuli in polarity B (a stimulus flip of 180° relative to polarity A). The typical procedure for EFR analysis is to average the two individual polarities, however this could result in a significant reduction in the overall response because sometimes the response to one polarity was very small. This phenomenon was not observed in every individual and was not consistent across a specific polarity. This interesting observation was independently verified using different EFR data recorded at the laboratory of our collaborator Dr. Steve Aiken (Dalhousie

University). Moving forward, responses from polarity A were treated separately from responses to polarity B.

### 2.6.3 Stimulus presentation and response recording

Participants were fitted with three disposable MEDI-TRACE Ag/AgCl electrodes placed at the vertex, just below the hairline at the posterior midline of the neck, and on the collarbone (as a ground). Electrode impedances were measured using an F-EZM5 GRASS impedance meter to ensure impedances were <5000 Ohm with inter-electrode differences <2000 Ohm. The stimulus was presented to the left ear of each subject using an Etymotic ER2 earphone, sealed in the ear-canal with a disposable foam insert. The experiment was controlled by software developed using LabVIEW (Version 8.5, National Instruments, Austin TX). Digital-to-analog conversion of the stimuli and analog-to-digital capture of the EEG were performed by a National Instruments PCI-6289 M-series acquisition card. Stimuli were output at 32000 S/s with 16-bit resolution and responses were recorded at 8000 S/s with 18-bit resolution. A Tucker-Davis Technologies PA5 attenuator and SA1 power amplifier controlled stimulus levels at 80 dB SPL through the Etymotic ER2 earphone acoustic transducer.

Participants were seated comfortably in a reclined chair in a sound insulated and electromagnetically shielded sound attenuated booth. A rolled towel was placed under their neck to help support their head and a blanket was provided for comfort. The booth lights were turned off and the participants were encouraged to sleep for the 55-minute duration of the measurement.

The stimulus transducer leads and the recording leads were positioned as far apart as possible to reduce the possibility of stimulus artifacts during the recording. An artifact check was also performed. The system was set up as usual with an individual fitted with electrodes; however, the acoustic tube from the ER2 was sealed in a Zwislocki coupler while the EEG was measured from the individual. In this set up, the transducer experiences a typical acoustic load however no true response is present, as the stimulus is not delivered to the ear. The recording showed typical EEG and myogenic noise without any response detection beyond the expected false positive rate.

## 2.6.4 Offline response analysis

While the measurement was running, the EEG time waveform and spectrum were displayed; however analysis was completed offline using noise rejection and a Fourier analyzer developed in MATLAB (Math Works, Natick MA). Noise metrics for each subject's EEG data were calculated from a frequency band of 80 to 120 Hz. Certain 1.024 s data blocks whose noise metric exceeded the mean noise metric plus two standard deviations were rejected (see Choi et al., 2013). Remaining data were analyzed independently for Polarity A and Polarity B. To isolate the brain's response to vowels, the time segments of the average EEG that corresponded with vowel boundaries were selected. This was performed manually, such that the central part of each vowel was selected to exclude the brief ramp-in and ramp-out sections at the beginning and end of each vowel segment.

## 2.6.5 Envelope and frequency following response estimation

The EFR to each vowel condition (i.e. standard head, standard had, subject head and subject had) was estimated from the averaged EEG for each polarity (A and B) using a Fourier analyzer (Choi et al., 2013). Using the instantaneous frequencies in the stimulus  $f_0$  track, reference cosine and sine sinusoids were created. The average EEG data were corrected back 10 ms to account for brainstem processing delays for the EFR (Aiken & Picton, 2006; Purcell et al., 2004). The data were then multiplied with the reference sinusoids to obtain real and imaginary components of the EFR. An identical procedure was used for the FFR, except the  $f_0$  track was multiplied by a positive integer to mimic the frequency track followed by the harmonic closest to F1.

## 2.6.6 Response detection

EEG amplitudes in ten frequency tracks adjacent to the  $f_0$  track for both vowels were estimated (five above  $f_0$  and five below). The distance in Hz between each track was determined by the reciprocal of vowel duration (i.e. the duration submitted to analysis), which is the bandwidth of the Fourier analyzer. These ten frequency tracks were averaged to obtain an estimate of the noise at frequencies neighbouring  $f_0$ . An F-ratio (John & Picton, 2000; p 143-144) was used to determine whether the observed response estimate

was likely to be from the distribution of the observed noise. This statistical approach determines if a significant EFR was present at  $p < 0.05$ . As above, the same method was used for the FFR, but with the harmonic closest to F1 and its neighbouring frequencies.

## Chapter 3

### 3 Results

#### 3.1 Speech

##### 3.1.1 F1 Discrimination threshold

The average F1 discrimination threshold for participants was 28.1 Hz (SD: 6.3 Hz; see Figure 5). This threshold is in agreement with previous data from 21 English monolinguals from our laboratory and did not statistically differ using an independent samples t-test ( $t = 0.89$ ,  $p = 0.38$ ). This indicates that participants were typical in their F1 discrimination thresholds and were capable of detecting the 200 Hz manipulation.

##### 3.1.2 Vowel goodness ratings

Participants' vowel goodness ratings (Figure 6) indicate that higher vowel goodness ratings were given to sounds with small F1 changes and lower ratings to sounds with large F1 changes. A Pearson's correlation was conducted and showed a robust correlation between goodness ratings and change in change in F1 ( $r = 0.978$ ,  $p < 0.001$ ). Error bars are large for the F1 shifts around 100 Hz due to the highly categorical changes in goodness observed in some participants compared to others.

##### 3.1.3 Speech compensation for English /ε/ in "head"

There is variation in the average F1 production across participants for a given vowel. To account for this variability, F1 was normalized to allow for comparisons across individuals. This was accomplished by subtracting each individual's average F1 of the Baseline phase from all trials. Normalized average group results are plotted in Figure 7. The average speech compensation threshold for all subjects, defined as the trial where the average change in F1 production was two standard deviations from average Baseline, was found at 40 Hz. This value is only slightly higher than the behavioural F1 discrimination threshold obtained. Although the group average showed a consistent and near monotonic growth of compensation with shift magnitude (see Figure 8), there was significant

individual variability in F1 production between participants (see Figure 9 for examples of individual responses).

## 3.2 Relationships between perception and production

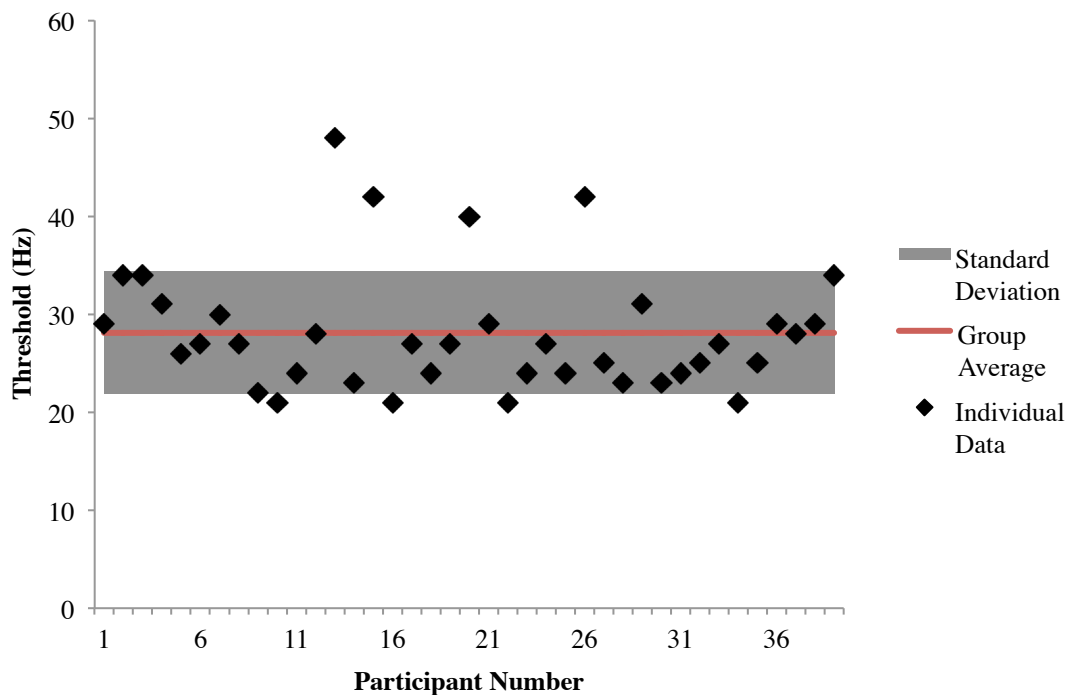
From previous laboratory data (Nguyen, 2012), it was expected that vowel goodness and speech compensation would be related. Speech compensation values for each Ramp step were correlated with the corresponding vowel goodness ratings (Figure 10). A Pearson correlation found a robust relationship between average speech compensation and vowel goodness ratings [ $r^2 = 0.962$   $p < 0.001$ ]. In general, greater compensations in speech production corresponded with lower goodness ratings (which themselves had been associated with larger F1 shifts) and low compensations in speech production corresponded with higher goodness ratings. In the average group data the relationship was robust, however, there was a great deal of individual variability (see Figure 11). Of the 39 participants, 22 had statistically significant linear correlations between goodness ratings and compensation for each Ramp step (see Table 1).

## 3.3 Auditory Filter Bandwidth

Auditory filter bandwidth was determined for each participant using both a behavioural and a physiological approach. Other related measures of cochlear tuning, group delay and  $Q_{\text{ERB}}$ , were also calculated and compared. Further, a correlation analysis was conducted to determine if there was a relationship between auditory filter bandwidth and speech compensation during the hold phase of altered auditory feedback.

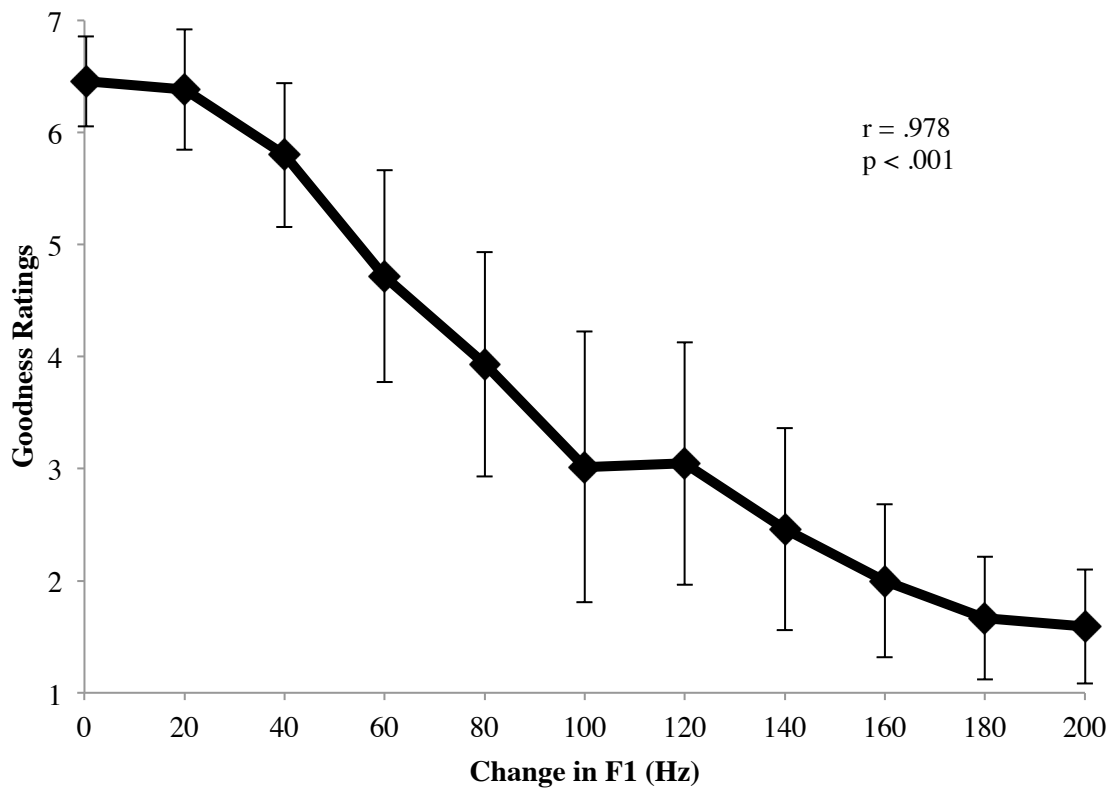
### 3.3.1 Fast psychoacoustic tuning curves

The psychoacoustic tuning curve program (SWPTC; Sek et al., 2005) was used to collect behavioural measures of cochlear tuning (see figure 14 for an individual example). The program's double regression value of  $Q$  (mean  $Q = 4.18$ ,  $SD = 0.85$ ) was used to calculate cochlear filter bandwidth (mean = 255.27 Hz,  $SD = 78.48$  Hz; see figure 13) by dividing the centre frequency (1 kHz) by the  $Q$  value for each individual. A Pearson's correlation was conducted to determine if a relationship existed between speech



**Figure 5. Plot of group and individual F1 discrimination thresholds.**

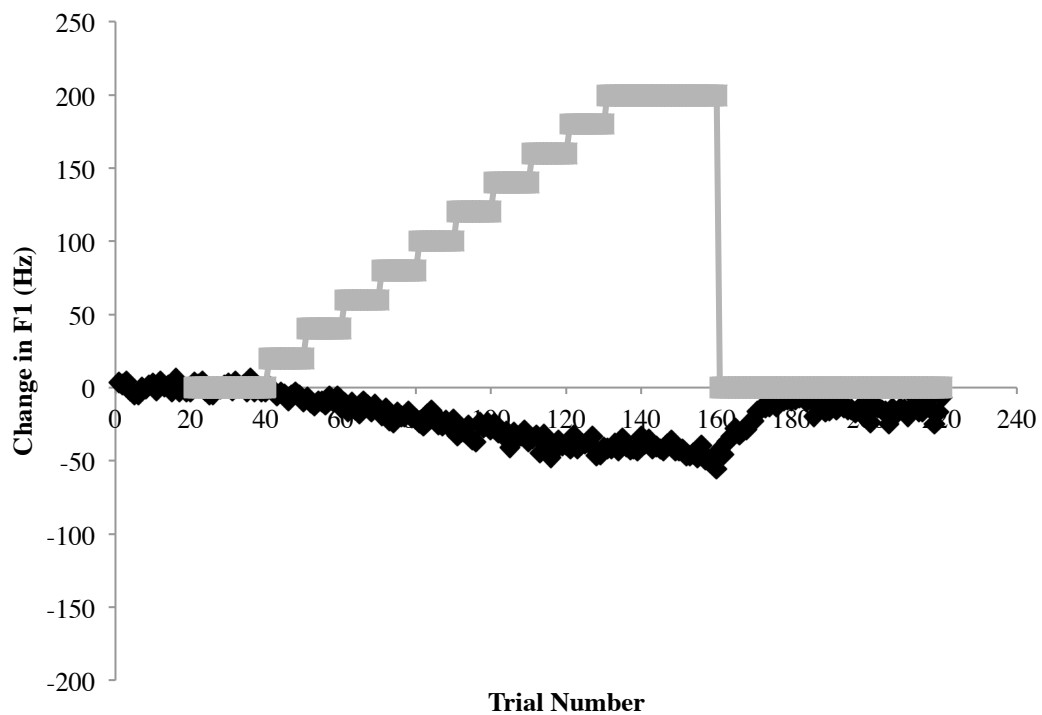
The average F1 discrimination threshold is 28.1 Hz (SD: 6.3 Hz). Axes are the participant's number and the threshold value. The largest shift during the speech manipulation was 200 Hz. The red line represents group average and the grey area represents  $\pm$  one standard deviation. The black diamonds represent individual threshold measures.



**Figure 6. Mean vowel goodness ratings.**

Error bars represent one standard deviation.





**Figure 7. Average normalized F1 compensation during altered auditory feedback.**

Light grey lines represent the F1 manipulation in Hz. Black points represents average normalized F1 production.

compensation magnitude and filter bandwidth (see Figure 14). No linear relationship was found [ $r(37) = 0.061$ ,  $p = 0.71$ ;  $N=39$ ].

### 3.3.2 Stimulus frequency otoacoustic emissions

A physiological measure of cochlear tuning was recorded (see figure 15) using SFOAEs. Group delay (mean = 8.51 ms, SD = 1.93 ms), filter bandwidth (mean = 108.7 Hz, SD = 31.78 Hz; see figure 16) and  $Q_{\text{ERB}}$  (mean = 9.78, SD = 2.22) were calculated. A Pearson's correlation was performed to investigate a relationship between filter bandwidth and compensation (see figure 17). No linear correlation was found [ $r(36) = 0.001$ ,  $p = 0.99$ ;  $N=38$ ].

### 3.3.3 Comparison between SWPTC and SFOAE

Linear correlations between the behavioural and physiological measures of cochlear tuning were evaluated. Unexpectedly, no relationship was found between the Q values [ $r(36) = -0.23$ ,  $p = 0.16$ ;  $N = 38$ ] or between the measures of filter bandwidth [ $r(36) = 0.2$ ,  $p = 0.23$ ;  $N = 38$ ; see figure 18].

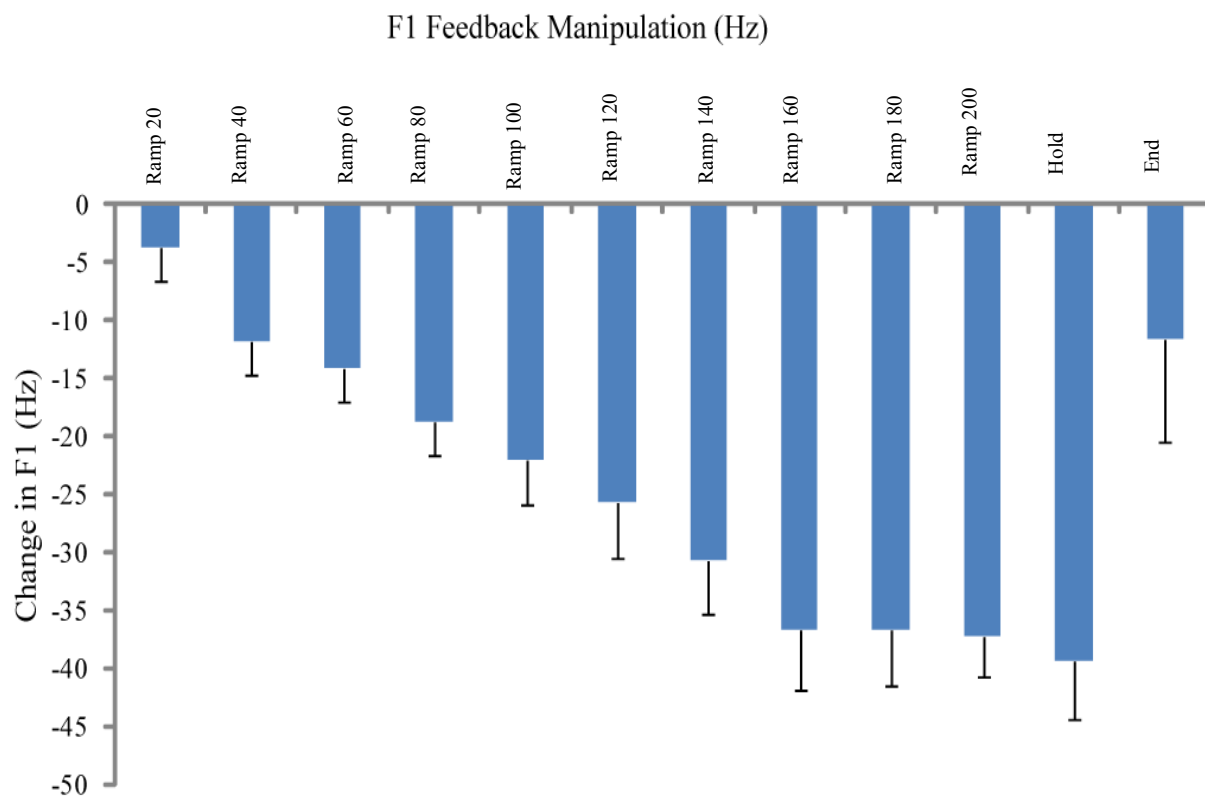
## 3.4 Electrophysiological measures

Metrics of brainstem speech encoding were determined for each participant using the EFR and FFR. An explanation of how each metric was determined will follow.

Correlations were conducted between these metrics of encoding and speech compensation during the hold phase of altered auditory feedback.

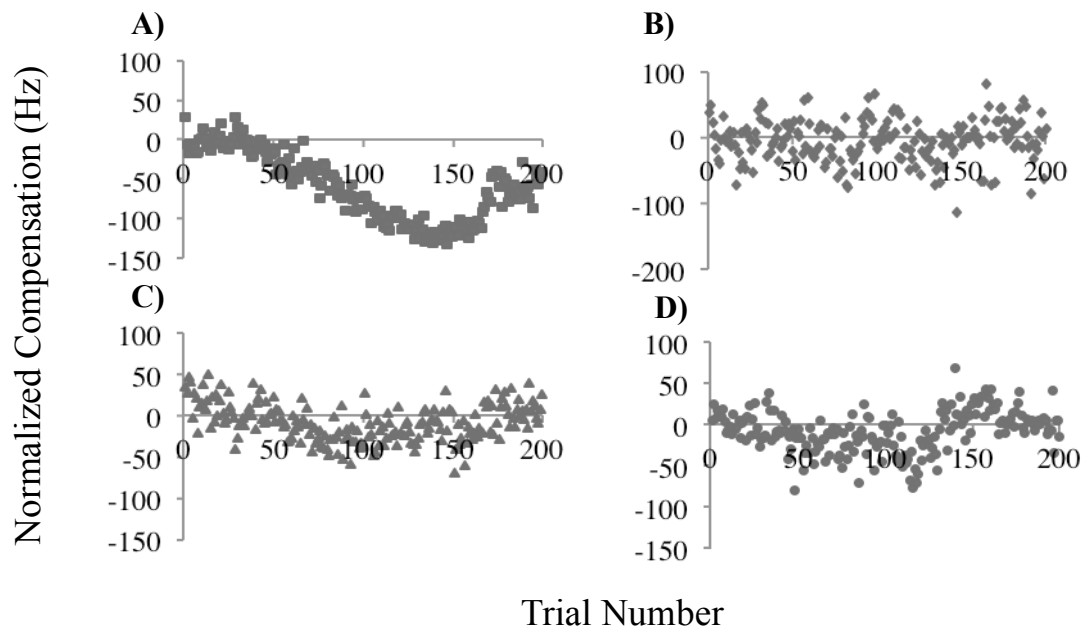
### 3.4.1 Envelope following response and frequency following response

EFR amplitude was estimated for each vowel (/ε/ and /æ/) and each talker (standard talker and the subject's own voice) in both polarity A and polarity B (see Figure 19). A 2X2X2 repeated measures ANOVA was conducted to determine if there were any amplitude differences between polarity (A and B), talker (standard and subject), and vowel (/ε/ and



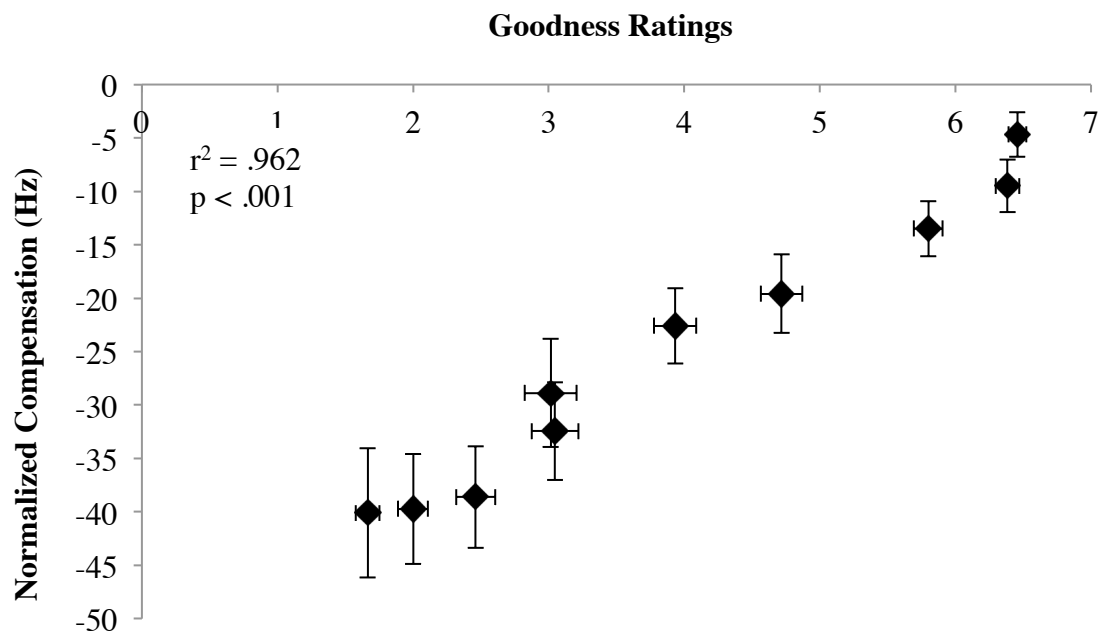
**Figure 8. Average normalized F1 results for English / $\epsilon$ / as in “head” across the Ramp phase, the Hold phase and the End phase.**

Ramp values (20, 40, ... 200 Hz), Hold phase, and End phase indicate average change in F1 for each phase of the manipulation. Error bars indicate standard error.



**Figure 9. Individual variation in F1 production during altered auditory feedback.**

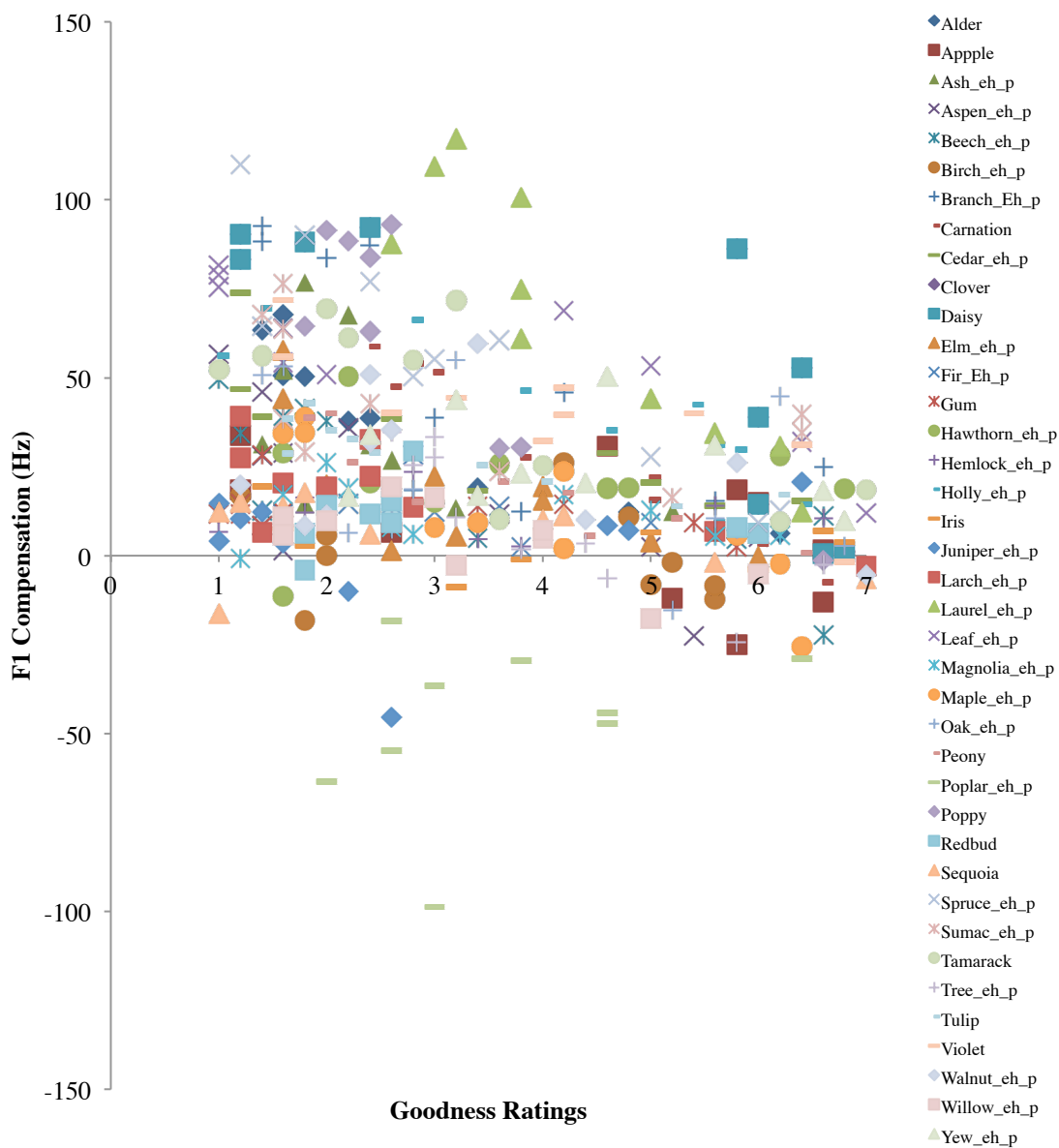
A) Subject with a large compensation response to the manipulation. B) Subject with almost no compensatory response. C) Subject with a small compensatory response. D) Subject who followed the manipulation.



**Figure 10. Correlation between goodness ratings and F1 compensation for /ɛ/ in “head” on a continuum towards /æ/ in had.**

Seven indicates an excellent version of the word head, and one indicates a poor version.

Error bars represent  $\pm$  one standard deviation.

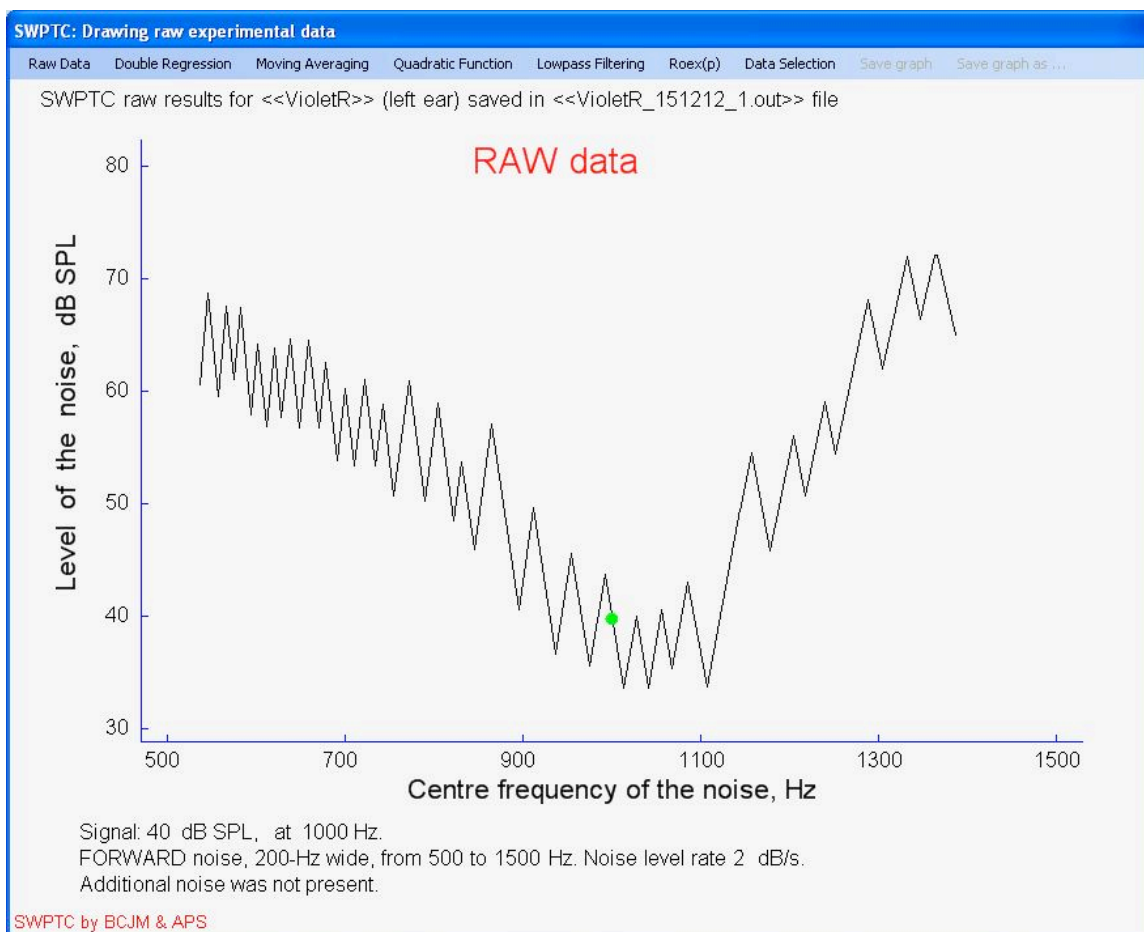


**Figure 11. Plot of correlation between individual goodness ratings and F1 compensation / $\varepsilon$ / in “head” on a continuum towards / $\varepsilon$ / in had.**

Seven indicates an excellent version of the word “head” and one indicates a poor version (N = 39).

<b>Participant Number</b>	<b>r</b>	<b>p</b>
1	0.87	0.001205
4	0.78	0.008267
5	0.81	0.004741
7	0.75	0.011777
8	0.98	1.01E-06
9	0.82	0.003603
10	0.79	0.00627
12	0.71	0.022024
14	0.93	0.000108
17	0.86	0.001302
20	0.72	0.019369
21	0.91	0.000253
22	0.84	0.002564
24	0.86	0.001231
26	0.79	0.006541
28	0.90	0.000328
31	0.93	0.000121
33	0.74	0.014143
34	0.84	0.002424
35	0.88	0.00077
36	0.82	0.003327
38	0.66	0.03759

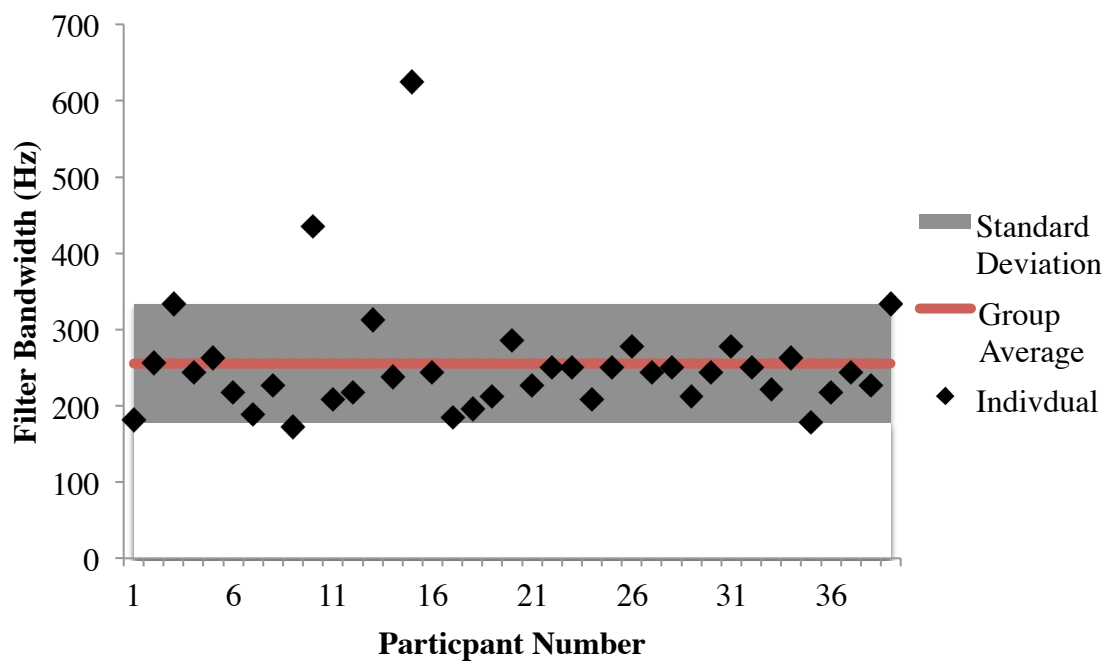
**Table 1. Significant individual correlations (p<0.05) between vowel goodness ratings and F1 compensation in vowel production.**



**Figure 12. Example of individual trial from the SWPTC program.**

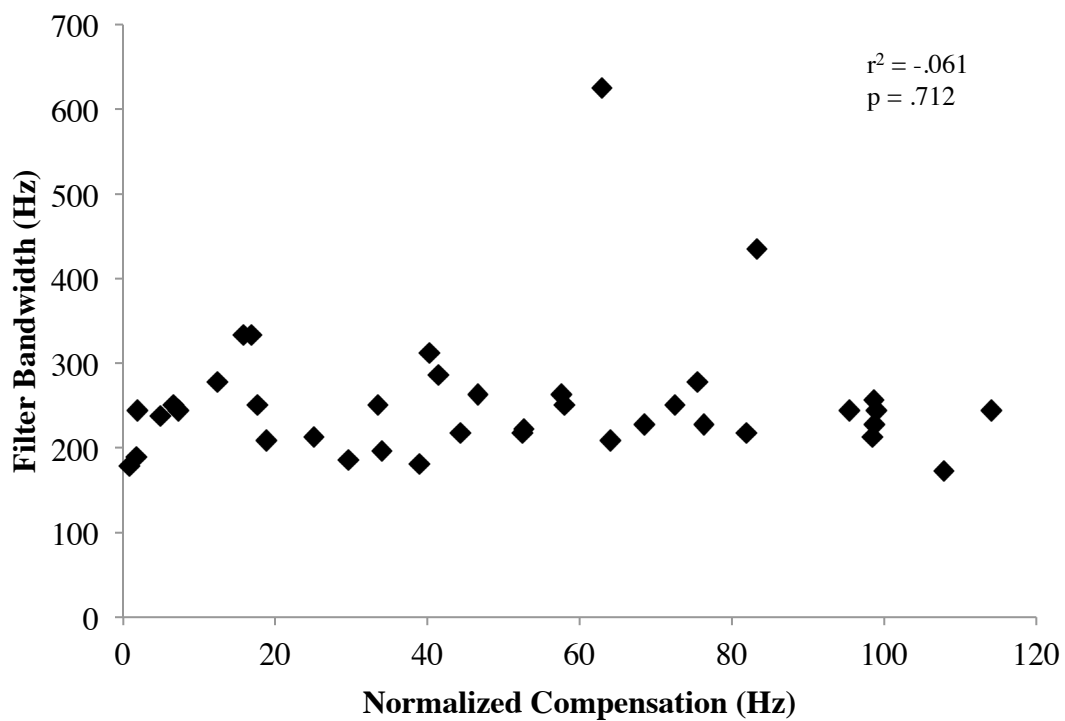
Screen capture from the SWPTC program (Sek et al., 2005). The green dot represents the centre frequency (1000 Hz). The jagged line represents the level of the noise in dB SPL across frequency. The program outputs a Q value that is calculated by dividing the centre frequency by the measured bandwidth (the width of the tuning curve 10 dB above its tip).



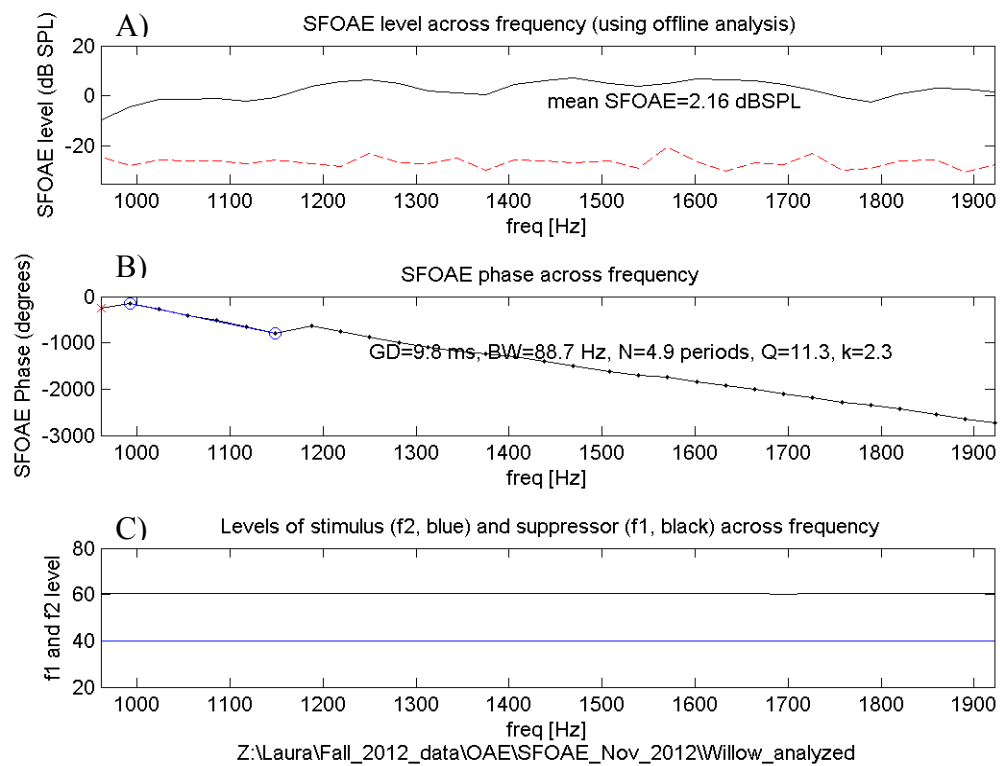


**Figure 13. Estimated individual and group average auditory filter bandwidth from the SWPTC program.**

The red line represents group average and the grey area represents  $\pm$  one standard deviation. The black diamonds represent individual bandwidth measures.

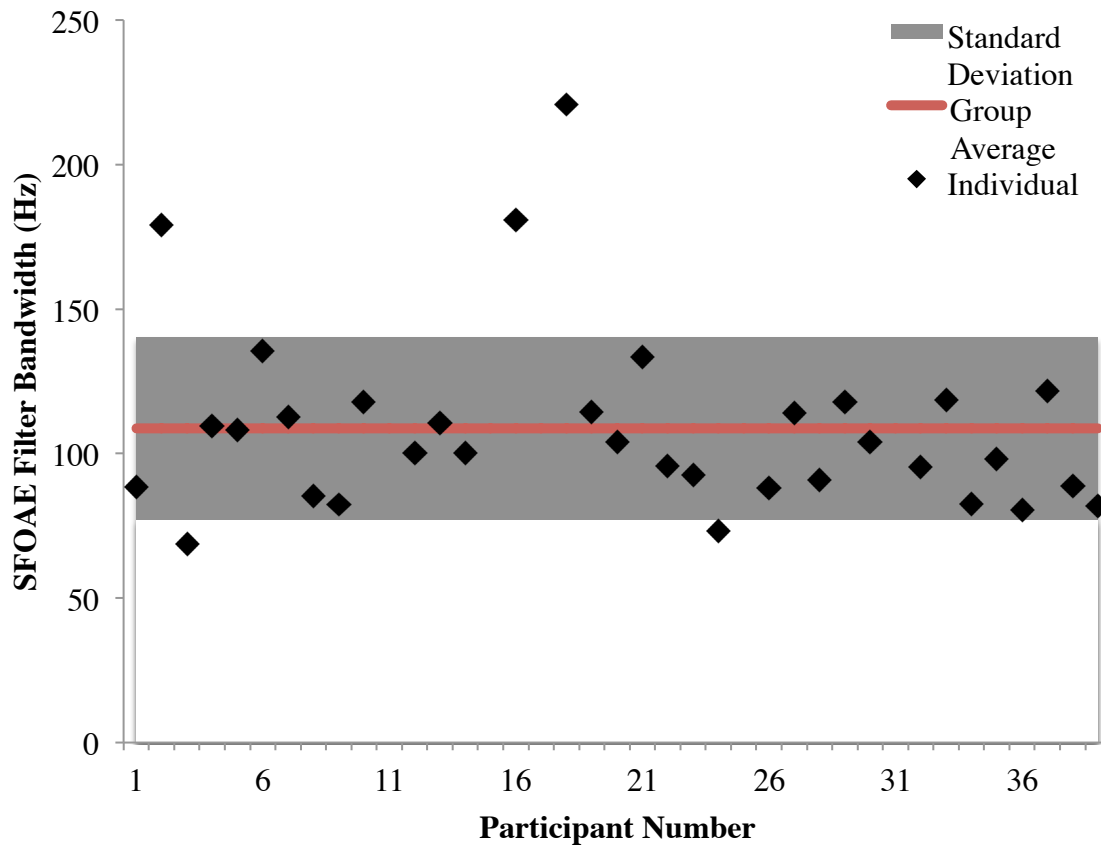


**Figure 14. Correlation between SWPTC filter bandwidth and compensation.** Auditory filter bandwidth estimated by the SWPTC program correlated with normalized F1 compensation.



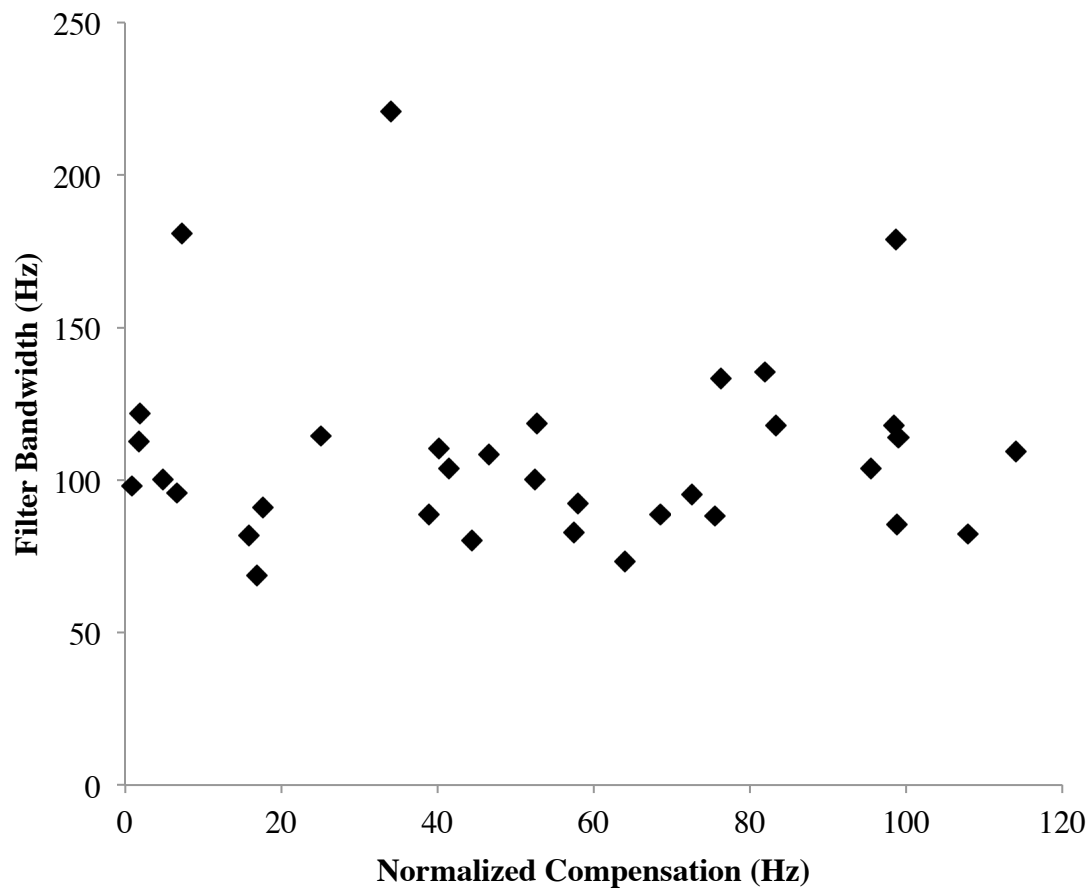
**Figure 15. Example of individual SFOAE analysis.**

A) The solid black line represents the SFOAE level in dB SPL across frequency; the hashed red line represents the noise level across frequency in dB SPL. B) The black line represents SFOAE phase in degrees across frequency. Group delay is determined by calculating the slope of the phase/frequency line (the blue line represents the measurement bandwidth manually selected for analysis). The  $Q_{ERB}$  is calculated by dividing the centre frequency (1 kHz) by the equivalent rectangular bandwidth, which is estimated from the group delay. C) The black line represents the level of the suppressor tone across frequency and the blue line represents the level of the stimulus across frequency.



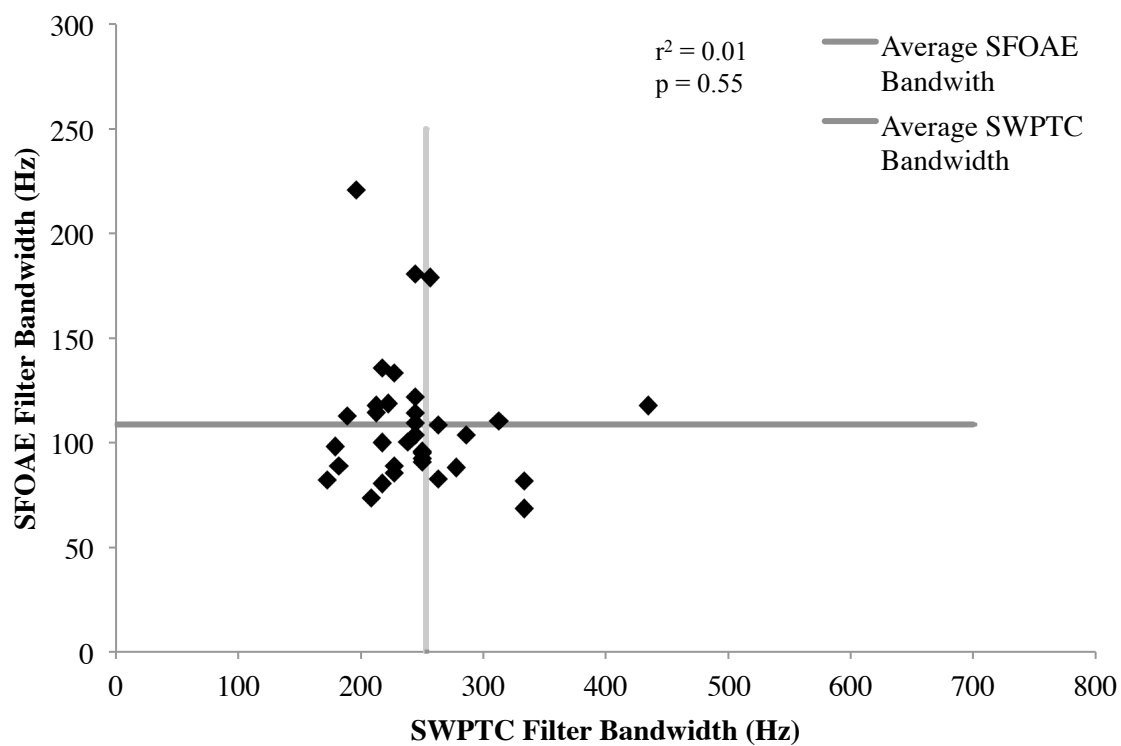
**Figure 16. SFOAE filter bandwidth group and individual results.**

The red line represents group average and the grey area represents  $\pm$  one standard deviation. The black diamonds represent individual bandwidth values.



**Figure 17. Correlation between SFOAE filter bandwidth and compensation.**

Auditory filter bandwidth estimated by the SFOAE program correlated with normalized F1 compensation.



**Figure 18. Correlation between SFOAE bandwidth and SWPTC bandwidth.**

/æ/). There were no significant differences between polarity A and polarity B [ $F(2, 39) = 2.97, p = 0.093$ ] or between talkers [ $F(2, 39) = 3.19, p = 0.082$ ], however, there was a significant difference between responses to the vowel /ε/ in “head” and /æ/ “had” [ $F(2,39) = 11.23, p = 0.002$ ]. Vowel /ε/ elicited slightly larger amplitudes. Similar amplitude estimates and an ANOVA analysis were completed for the FFR (see Figure 20). There was a significant difference between polarity A and polarity B [ $F(2, 39) = 0.136, p = 0.004$ ] where polarity A amplitudes were slightly higher. There was also a significant difference between talkers [ $F(2, 39) = 0.954, p = 0.024$ ] where the subject’s own voice elicited slightly larger amplitudes. There was however no significant difference between responses to the vowel /ε/ in “head” and /æ/ in “had” [ $F(2, 39) = 4.39, p = 0.104$ ].

To investigate potential relationships between the EFRs and speech compensation, the EFR results were considered three ways to serve as metrics of vowel encoding. The first measure was the absolute amplitude of the EFR (in nV) to the vowel /ε/ in “head”. This measure was used as an overall metric of encoding quality to determine if the stimuli evoked significant responses and if the amplitude was related to the compensation observed. The second measure was the change in EFR amplitude (in nV) from the vowel /ε/ in “head” to /æ/ in “had” (i.e. phase not included). This measure was used under the assumption that the differences observed between the amplitude of the EFR to /ε/ and /æ/ might reveal how brainstem encoding of these different vowels is involved in the changes observed in F1 production. See Tables 2 and 3 for the number of significant responses and the average amplitude and noise values for these two metrics of encoding. The final measure was the magnitude change from the vowel /ε/ in “head” to /æ/ in “had” determined with vector subtraction, which uses response amplitude and phase (see Table 3).

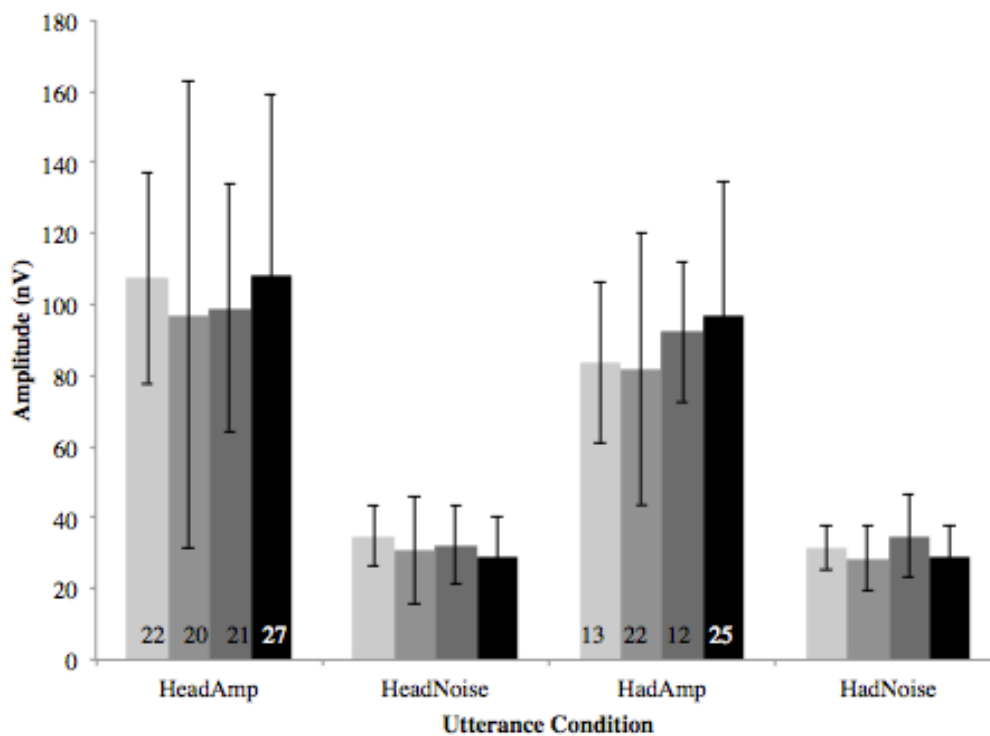
Absolute amplitude and change in amplitude were evaluated for the FFR (see Table 5 and 6) as well as the change in magnitude. A table for change in FFR magnitude is not included for brevity due to the small number of individuals where the FFR was detected for both vowels (N = 3 and 4 for polarities A and B, respectively). Detection of responses to both vowels is required for vector subtraction since a meaningful phase value is necessary.

### 3.4.2 Relationships between the EFR and FFR and speech compensation

A linear correlation analysis was conducted to determine if the absolute EFR amplitude to the vowel /ε/ in “head” was related to F1 compensation values (see Figure 21). A linear correlation was also done to determine if the change in EFR amplitude from “head” to “had” was related to F1 compensation values. Figure 22 shows this for an analysis where at least one word elicited a detectable EFR. The analysis with both words significant had a similar appearance and is not included for brevity. We chose to include the “at least one word significant” case under the interpretation that if one word was significantly detected and the other was not, that this indicated a change in encoding between vowels. The undetected response was not significantly different from noise, but served as the best estimate available for that vowel. A final correlation was completed between the change in EFR magnitude (including phase) between “head” and “had” and F1 compensation values (see Figure 23) for individuals where both words elicited significant responses. There were no significant correlations and correlation statistics are given in Tables 7 and 8.

The same correlations were completed examining the absolute FFR amplitude to the vowel /ε/ in “head (Figure 24), and the change in amplitude from “head to “had” (Figure 25). A figure for the change in FFR magnitude (including phase) is omitted because there were too few cases where both words were detected, as mentioned above. See Tables 9 and 10 for correlation statistics. None of the FFR comparisons revealed significant relationships ( $p > 0.05$ ).





**Figure 19. Average EFR Amplitude for significant responses.**

Numbers given within columns represent the number of significant responses for each condition. Legend letter (A) represents responses from polarity A and (B) represents responses from polarity B. Error bars are one standard deviation.

EFR Responses Polarity A								
Condition	Subject's Own Voice				Standard Talker			
	Absolute Head Amp (nV)	Absolute Had Amp (nV)	Delta Amp (nV) Both words	Delta Amp (nV) One word	Absolute Head Amp (nV)	Absolute Had Amp (nV)	Delta Amp (nV) Both words	Delta Amp (nV) One word
Significant responses	N = 20	N = 22	N = 14	N = 28	N = 22	N = 13	N = 8	N = 27
Response	97.12 (65.91)	81.89 (38.35)	24.71 (31.60)	30.38 (26.19)	107.51 (29.86)	83.68 (22.66)	38.81 (35.83)	47.00 (28.47)
Noise	30.90 (15.12)	28.53 (9.00)	-	-	34.87 (8.52)	31.54 (6.18)	-	-

**Table 2. EFR Responses for Polarity A.**

Absolute EFR amplitudes of “head” and “had” for both the subject’s own voice (columns 1 and 2) and the standard talker (columns 5 and 6). Delta amplitude is the change in amplitude between the EFR from “head” to “had”. “Both words” indicates that EFR responses to both “head” and “had” were significant (columns 3 and 7). “One word” indicates that the EFR response was significant to at least one word either “head” or “had” (columns 4 and 8). We chose to include this as a metric under the assumption that if one word was significantly detected and the other was not, that this indicated a change in encoding between vowels. The undetected response is not significantly different from noise, but serves as the best estimate available for that vowel. Brackets indicate standard deviation.

EFR Responses Polarity B								
Condition	Subject's Own Voice				Standard Talker			
	Absolute Head Amp (nV)	Absolute Had Amp (nV)	Delta Amp (nV) Both words	Delta Amp (nV) One word	Absolute Head Amp (nV)	Absolute Had Amp (nV)	Delta Amp (nV) Both words	Delta Amp (nV) One word
Significant responses	N = 27	N = 25	N = 21	N = 31	N = 21	N = 12	N = 8	N = 24
Response	108.31 (50.63)	97.01 (37.81)	31.76 (19.80)	38.63 (23.17)	98.92 (34.90)	92.26 (19.78)	24.22 (16.50)	44.43 (32.08)
Noise	29.13 (10.97)	29.24 (8.39)	-	-	32.17 (11.00)	34.85 (11.67)	-	-

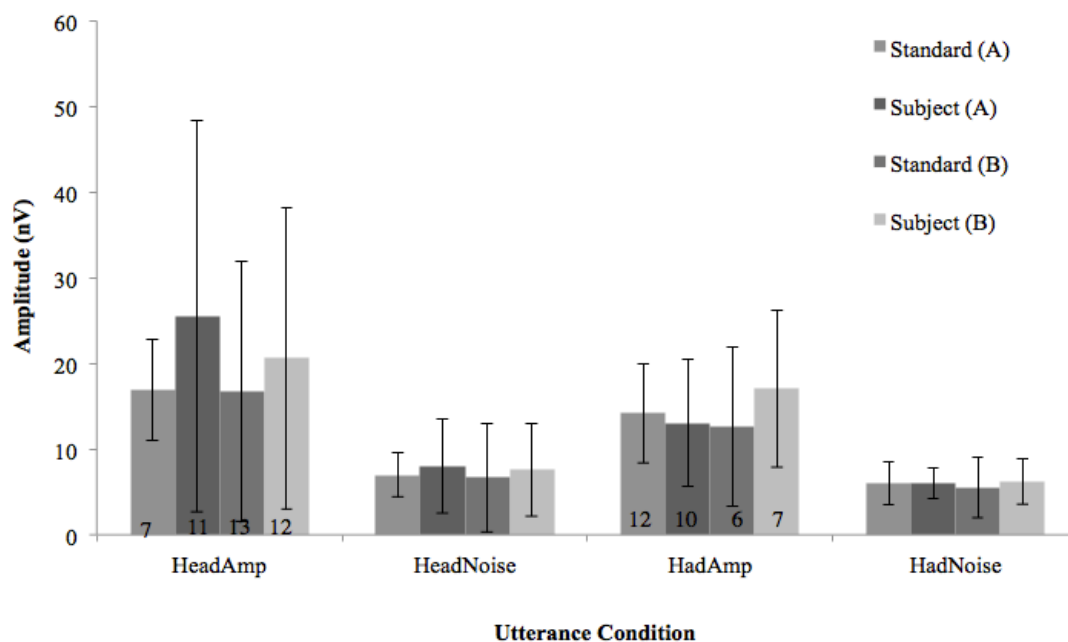
**Table 3. EFR Responses for Polarity B.**

Absolute EFR amplitudes of “head” and “had” for both the subject’s own voice (columns 1 and 2) and the standard talker (columns 5 and 6). Delta amplitude is the change in amplitude between the EFR from “head” to “had”. “Both words” indicates that EFR responses to both “head” and “had” were significant (columns 3 and 7). “One word” indicates that the EFR response was significant to at least one word either “head” or “had” (columns 4 and 8). We chose to include this as a metric under the assumption that if one word was significantly detected and the other was not, that this indicated a change in encoding between vowels. The undetected response is not significantly different from noise, but serves as the best estimate available for that vowel. Brackets indicate standard deviation.

EFR Magnitude Polarity A			EFR Magnitude Polarity B		
Talker	Subject's Own Voice	Standard Talker	Talker	Subject's Own Voice	Standard Talker
Condition	Delta Mag (nV) Both words	Delta Mag (nV) Both words	Condition	Delta Mag (nV) Both words	Delta Mag (nV) Both words
Significant responses	N = 14	N = 8	Significant responses	N = 21	N = 8
Response	77 (79)	92 (62)	Response	71 (52)	56 (45)

**Table 4. EFR Response magnitude for Polarity A and B.**

Change in EFR magnitude (including phase) in polarity A from “head” to “had” for both the subject’s own voice (column 1) and the standard talker (column 2). Same is presented for polarity B in columns 3 and 4, respectively. “Both words” indicates that FFR responses to both “head” and “had” were significant. Responses to only one word were not included because vector subtraction requires a valid phase value. Brackets indicate standard deviation.



**Figure 20. Average FFR Amplitude for significant responses.**

Numbers given within columns represent the number of significant responses for each condition. Legend letter (A) represents responses from polarity A and (B) represents responses from polarity B. Error bars are one standard deviation.

FFR Responses Polarity A								
Condition	Subject's Own Voice				Standard Talker			
	Absolute Head Amp (nV)	Absolute Had Amp (nV)	Delta Amp (nV) Both words	Delta Amp (nV) One word	Absolute Head Amp (nV)	Absolute Had Amp (nV)	Delta Amp (nV) Both words	Delta Amp (nV) One word
N	N = 11	N = 10	N = 4	N = 17	N = 7	N = 12	N = 4	N = 15
Response	25.57 (22.89)	13.06 (7.45)	26.43 (21.24)	11.59 (13.73)	16.87 (5.87)	14.22 (5.82)	4.31 (4.39)	6.07 (4.98)
Noise	8.33 (5.47)	5.61 (1.77)	-	-	6.75 (2.56)	5.97 (2.48)	-	-

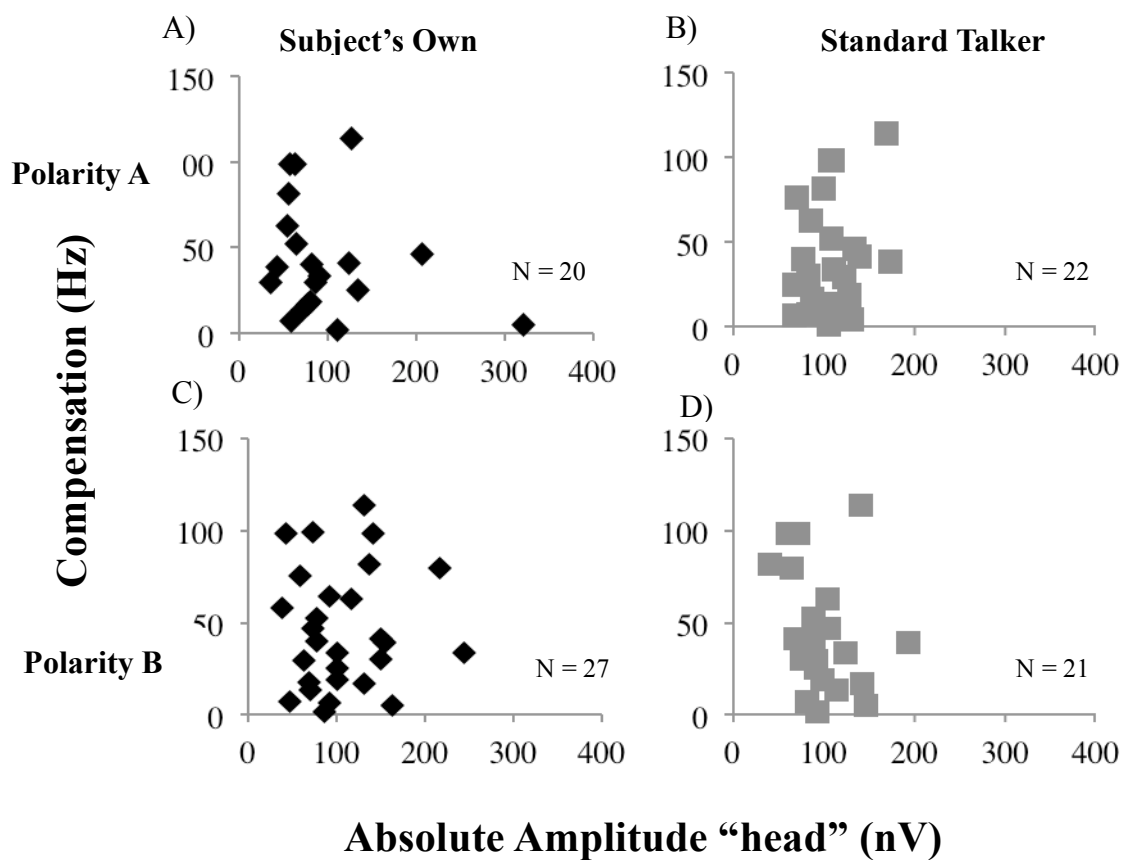
**Table 5. FFR Responses for Polarity A.**

Absolute FFR amplitudes of “head” and “had” for both the subject’s own voice (columns 1 and 2) and the standard talker (columns 5 and 6). Delta amplitude is the change in amplitude between the FFR from “head” to “had”. “Both words” indicates that EFR responses to both “head” and “had” were significant (columns 3 and 7). “One word” indicates that the FFR response was significant to at least one word either “head” or “had” (columns 4 and 8). We chose to include this as a metric under the assumption that if one word was significantly detected and the other was not, that this indicated a change in encoding between vowels. The undetected response is not significantly different from noise, but serves as the best estimate available for that vowel. Brackets indicate standard deviation.

FFR Responses Polarity B								
Condition	Subject's Own Voice				Standard Talker			
	Absolute Head Amp (nV)	Absolute Had Amp (nV)	Delta Amp (nV) Both words	Delta Amp (nV) One word	Absolute Head Amp (nV)	Absolute Had Amp (nV)	Delta Amp (nV) Both words	Delta Amp (nV) One word
N	N = 12	N = 7	N = 3	N = 16	N = 13	N = 6	N = 3	N = 16
Response	20.64 (17.63)	17.01 (9.12)	18.31 (20.89)	10.36 (9.54)	16.78 (15.21)	12.59 (9.28)	9.80 (15.90)	8.23 (9.78)
Noise	7.58 (5.42)	6.22 (2.65)	-	-	6.68 (6.36)	5.54 (3.54)	-	-

**Table 6. FFR Responses for Polarity B.**

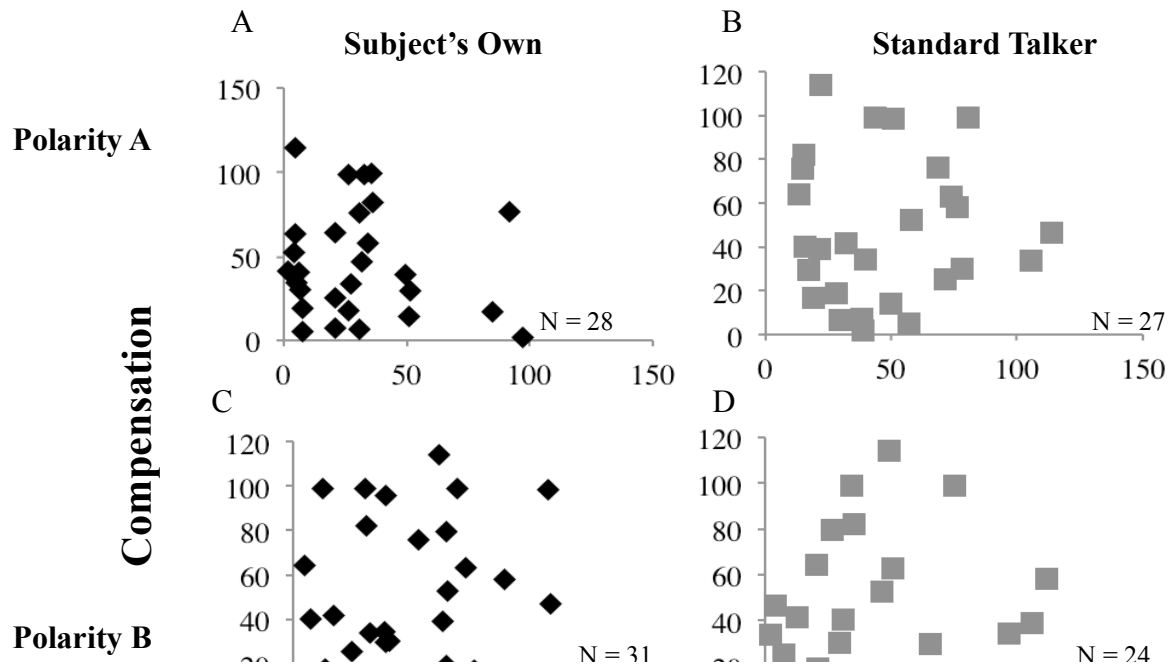
Absolute FFR amplitudes of “head” and “had” for both the subject’s own voice (columns 1 and 2) and the standard talker (columns 5 and 6). Delta amplitude is the change in amplitude between the FFR from “head” to “had”. “Both words” indicates that FFR responses to both “head” and “had” were significant (columns 3 and 7). “One word” indicates that the FFR response was significant to at least one word either “head” or “had” (columns 4 and 8). We chose to include this as a metric under the assumption that if one word was significantly detected there was and the other was not, that this indicated a change in encoding between vowels. The undetected response is not significantly different from noise, but serves as the best estimate available for that vowel. Brackets indicate standard deviation.



**Figure 21. Absolute amplitude (nV) of EFR to “head” correlated with compensation magnitude (Hz).**

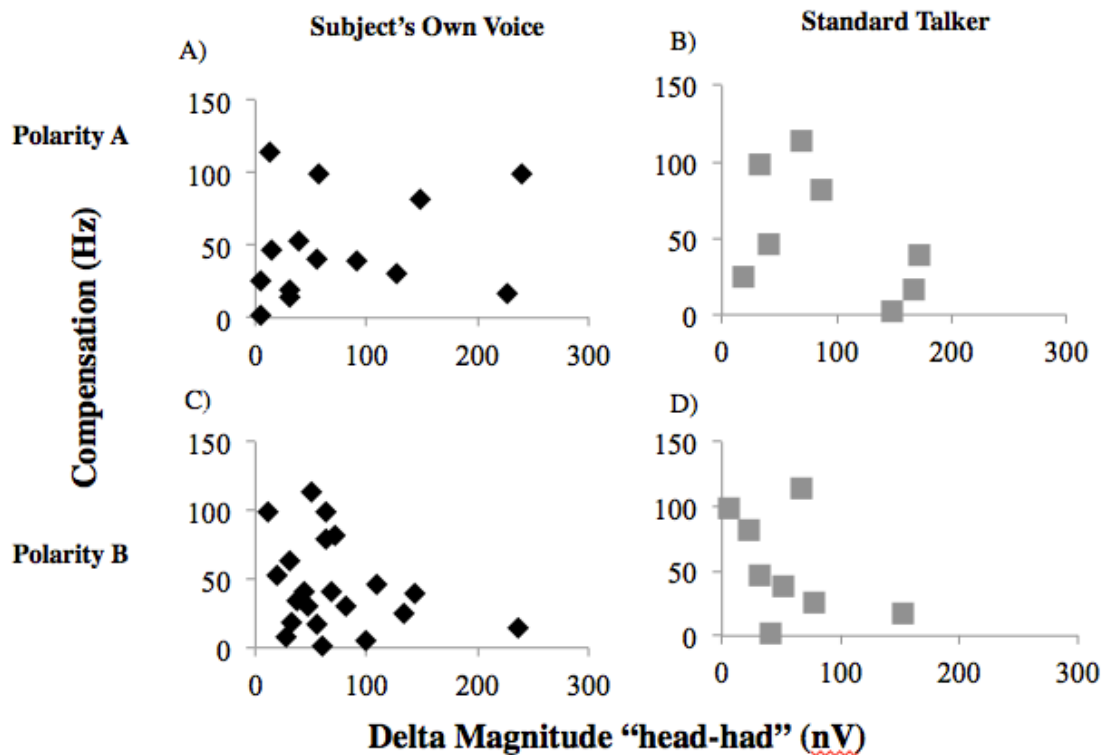
A) The absolute EFR amplitude to the subject’s own production of the word “head” in Polarity A. B) The absolute EFR amplitude to the standard talker’s production of the word “head” in Polarity A. Panels C) and D) are the same as A) and B), respectively, but for Polarity B.





**Figure 22. Change in EFR amplitude (nV) from “head” to “had” (no phase) correlated with compensation magnitude (Hz).**

A) The change in EFR amplitude to the subject’s own production of the words “head” to “had” in Polarity A. B) The change in EFR amplitude to the standard talker’s production of the words “head” to “had” in Polarity A. Panels C) and D) are the same as A) and B), respectively, but for Polarity B.



**Figure 23. EFR Change in magnitude (nV) from “head” to “had” (including response phase) correlated with compensation magnitude (Hz).**

A) The change in EFR amplitude to the subject’s own production of the words “head” to “had” in Polarity A. B) The change in EFR amplitude to the standard talker’s production of the words “head” to “had” in Polarity A. Panels C) and D) are the same as A) and B), respectively, but for Polarity B.

<b>EFR and Speech Compensation Linear Correlations Polarity A</b>				
<b>Subject's own Voice</b>				
<b>Condition</b>	<b>Absolute Head Amp (nV)</b>	<b>Delta Amp (nV) Both words</b>	<b>Delta Amp (nV) One word</b>	<b>Delta Mag (nV) Both words</b>
<b>Significant responses</b>	N = 20	N = 14	N = 28	N = 8
<b>r</b>	0.28	0.14	0.30	-0.28
<b>dof</b>	18	12	26	6
<b>t</b>	1.24	0.48	1.62	-0.72
<b>p</b>	0.23	0.64	0.11	0.50
<b>Standard Talker</b>				
	<b>Absolute Head Amp (nV)</b>	<b>Delta Amp (nV) Both words</b>	<b>Delta Amp (nV) One word</b>	<b>Delta Magnitude (nV) Both words</b>
<b>Significant responses</b>	N = 22	N = 8	N = 27	N = 14
<b>r</b>	-0.33	-0.65	-0.19	-0.49
<b>dof</b>	20	6	25	12
<b>t</b>	-1.55	-2.11	-0.97	-1.43
<b>p</b>	0.13	0.08	0.34	0.18

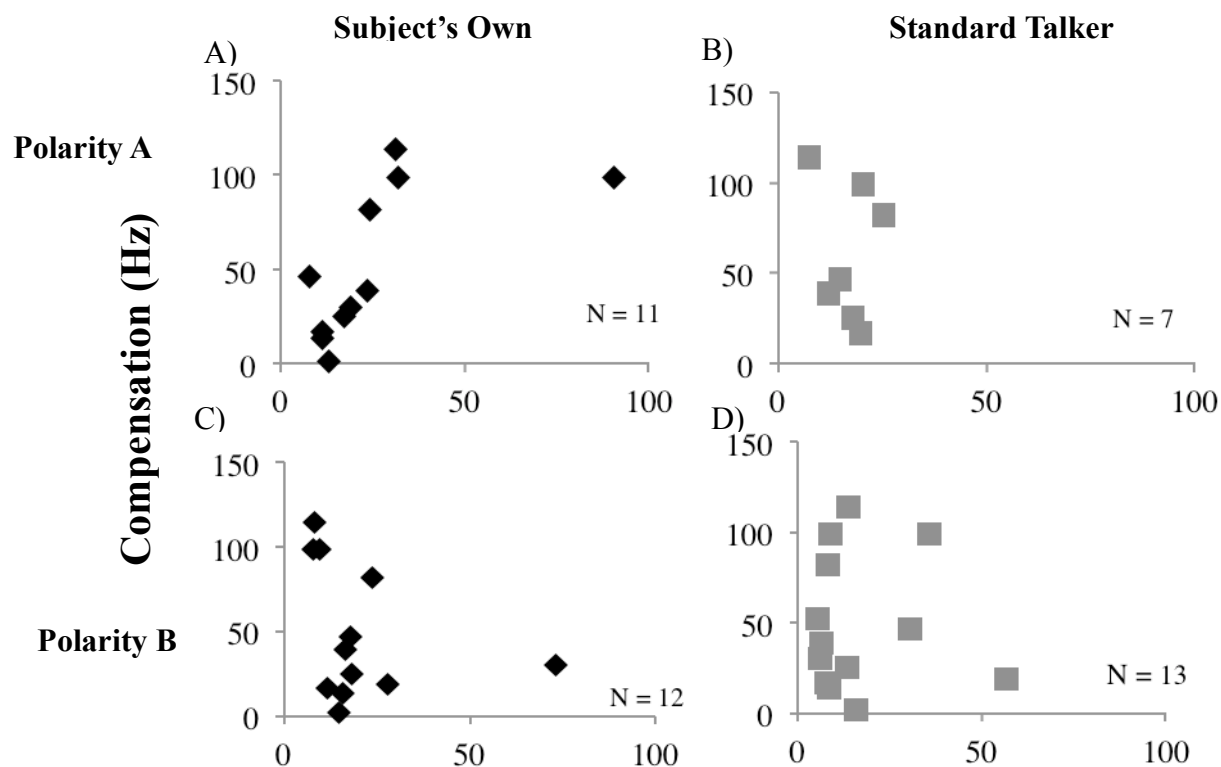
**Table 7. EFR and speech compensation linear correlations for polarity A.**

Each column represents a metric of brainstem encoding that was correlated with speech compensation. “Both words” indicates that EFR responses to both “head” and “had” were significant. “One word” indicates that the EFR response was significant to at least one word either “head” or “had”. Brackets indicate standard deviation. N denotes the number of subjects included in analysis.

<b>EFR and Speech Compensation Linear Correlations Polarity B</b>				
<b>Subject's Own Voice</b>				
<b>Condition</b>	<b>Absolute Head Amp (nV)</b>	<b>Delta Amp (nV) Both words</b>	<b>Delta Amp (nV) One word</b>	<b>Delta Mag (nV) Both words</b>
<b>Significant responses</b>	N = 27	N = 21	N = 31	N = 21
<b>r</b>	0.24	0.4	0.07	-0.01
<b>dof</b>	25	19	29	19
<b>t</b>	1.24	1.89	0.39	-0.44
<b>p</b>	0.22	0.07	0.70	0.66
<b>Standard Talker</b>				
<b>Condition</b>	<b>Absolute Head Amp (nV)</b>	<b>Delta Amp (nV) Both words</b>	<b>Delta Amp (nV) One word</b>	<b>Delta Magnitude (nV) Both words</b>
<b>Significant responses</b>	N = 21	N = 8	N = 24	N = 8
<b>r</b>	0.07	0.07	0.62	0.57
<b>dof</b>	19	6	22	6
<b>t</b>	0.33	1.94	0.34	1.69
<b>p</b>	0.75	0.10	0.73	0.14

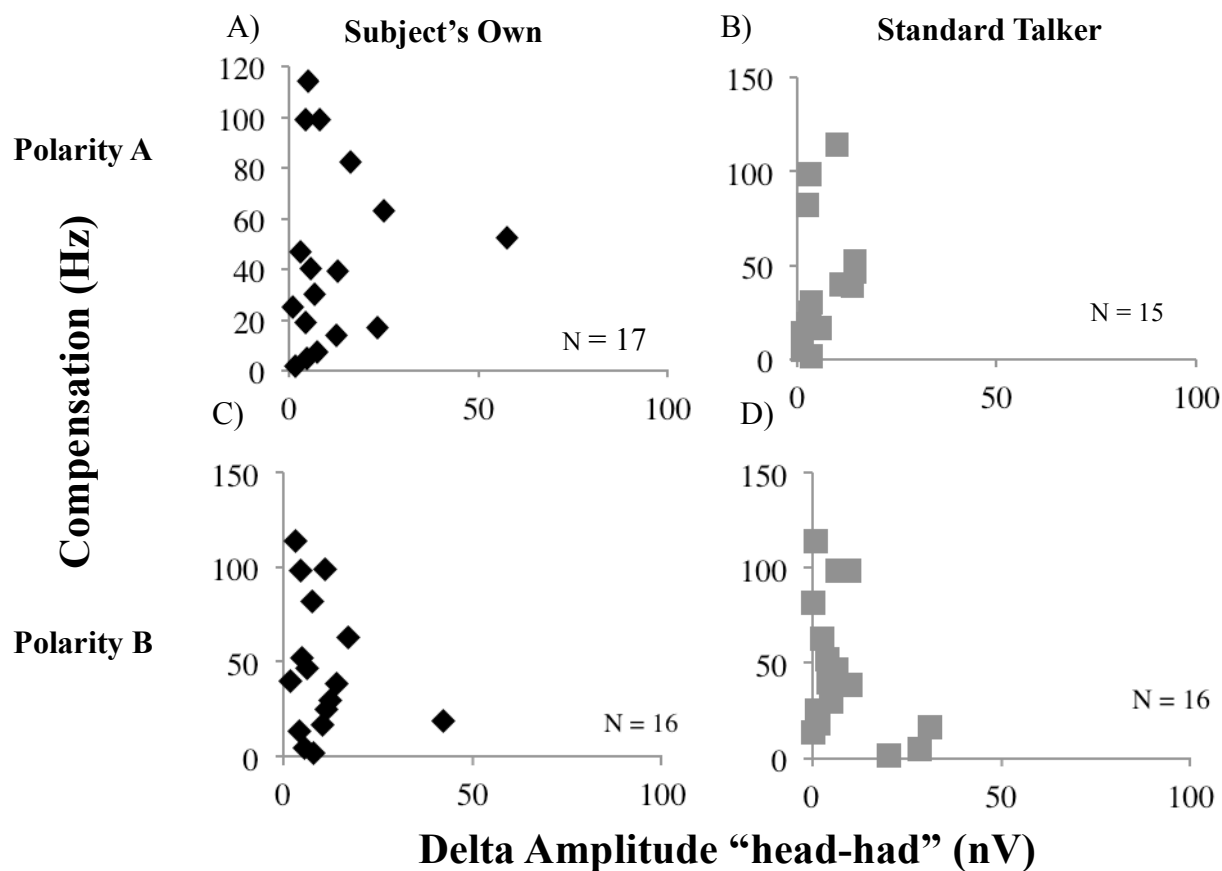
**Table 8. EFR and speech compensation linear correlations for polarity B.**

Each column represents a metric of brainstem encoding that was correlated with speech compensation. “Both words” indicates that EFR responses to both “head” and “had” were significant. “One word” indicates that the EFR response was significant to at least one word either “head” or “had”. Brackets indicate standard deviation. N denotes the number of subjects included in analysis.



**Figure 24. Absolute amplitude (nV) of FFR to “head” correlated with compensation magnitude (Hz).**

A) The absolute FFR amplitude to the subject’s own production of the word “head” in Polarity A. B) The absolute FFR amplitude to the standard talker’s production of the word “head” in Polarity A. Panels C) and D) are the same as A) and B), respectively, but for Polarity B.



**Figure 25. Change in FFR amplitude (nV) from “head” to “had” (no phase) correlated with compensation magnitude (Hz).**

A) The change in FFR amplitude to the subject’s own production of the words “head” to “had” in Polarity A. B) The change in FFR amplitude to the standard talker’s production of the words “head” to “had” in Polarity A. Panels C) and D) are the same as A) and B), respectively, but for Polarity B.

<b>FFR and Speech Compensation Linear Correlations Polarity A</b>			
<b>Subject's Own Voice</b>			
<b>Condition</b>	<b>Absolute Head Amp (nV)</b>	<b>Delta Amp (nV) Both words</b>	<b>Delta Amp (nV) One word</b>
<b>Significant responses</b>	N = 12	N = 3	N = 16
<b>r</b>	0.11	0.27	0.24
<b>dof</b>	10	1	14
<b>t</b>	0.34	0.28	0.92
<b>p</b>	0.74	0.83	0.37
<b>Standard Talker</b>			
<b>Condition</b>	<b>Absolute Head Amp (nV)</b>	<b>Delta Amp (nV) Both words</b>	<b>Delta Amp (nV) One word</b>
<b>Significant responses</b>	N = 13	N = 3	N = 16
<b>r</b>	0.01	0.11	-0.18
<b>dof</b>	11	1	14
<b>t</b>	0.02	0.11	-0.68
<b>p</b>	0.98	0.93	0.51

**Table 9. FFR and speech compensation linear correlations for polarity A.**

Each column represents a metric of brainstem encoding that was correlated with speech compensation. “Both words” indicates that FFR responses to both “head” and “had” were significant. “One word” indicates that the FFR response was significant to at least one word either “head” or “had”. Brackets indicate standard deviation.

<b>FFR and Speech Compensation Linear Correlations Polarity B</b>			
<b>Subject's Own Voice</b>			
<b>Condition</b>	<b>Absolute Head Amp (nV)</b>	<b>Delta Amp (nV) Both words</b>	<b>Delta Amp (nV) One word</b>
<b>Significant responses</b>	N = 11	N = 4	N = 17
<b>r</b>	0.36	0.29	0.00
<b>dof</b>	9	2	15
<b>t</b>	1.15	0.42	0.01
<b>p</b>	0.28	0.71	0.99
<b>Standard Talker</b>			
<b>Condition</b>	<b>Absolute Head Amp (nV)</b>	<b>Delta Amp (nV) Both words</b>	<b>Delta Amp (nV) One word</b>
<b>Significant responses</b>	N = 7	N = 4	N = 15
<b>r</b>	-0.13	0.88	-0.21
<b>dof</b>	5	2	13
<b>t</b>	-0.30	2.60	-0.78
<b>p</b>	0.70	0.12	0.45

**Table 10. FFR and speech compensation linear correlations for polarity B.**

Each column represents a metric of brainstem encoding that was correlated with speech compensation. “Both words” indicates that FFR responses to both “head” and “had” were significant. “One word” indicates that the FFR response was significant to at least one word either “head” or “had”. Brackets indicate standard deviation.



## Chapter 4

### 4 Discussion

The current study aimed to better understand both the peripheral and central auditory mechanisms underlying the use of auditory feedback to guide speech production. A formant manipulation paradigm was used to elicit a compensatory speech production response in real-time. Participants produced the vowel / $\epsilon$ / in “head” while their F1 feedback was shifted without their awareness to approximate the vowel / $\text{æ}$ / in “had”. Compensatory responses to the perturbation were highly variable. In an attempt to explain this variability, measures of brainstem auditory encoding (EFRs and FFRs) to the participant’s own productions of those vowels, as well as a standard talker producing the same vowels, were recorded. Measures of cochlear tuning within the frequency range of 960 Hz to 1920 Hz using SFOAEs were also collected. The following sections aim to elucidate the mechanisms underlying the use of auditory feedback during speech production and their relation to the current results through the discussion of past literature and theoretical models.

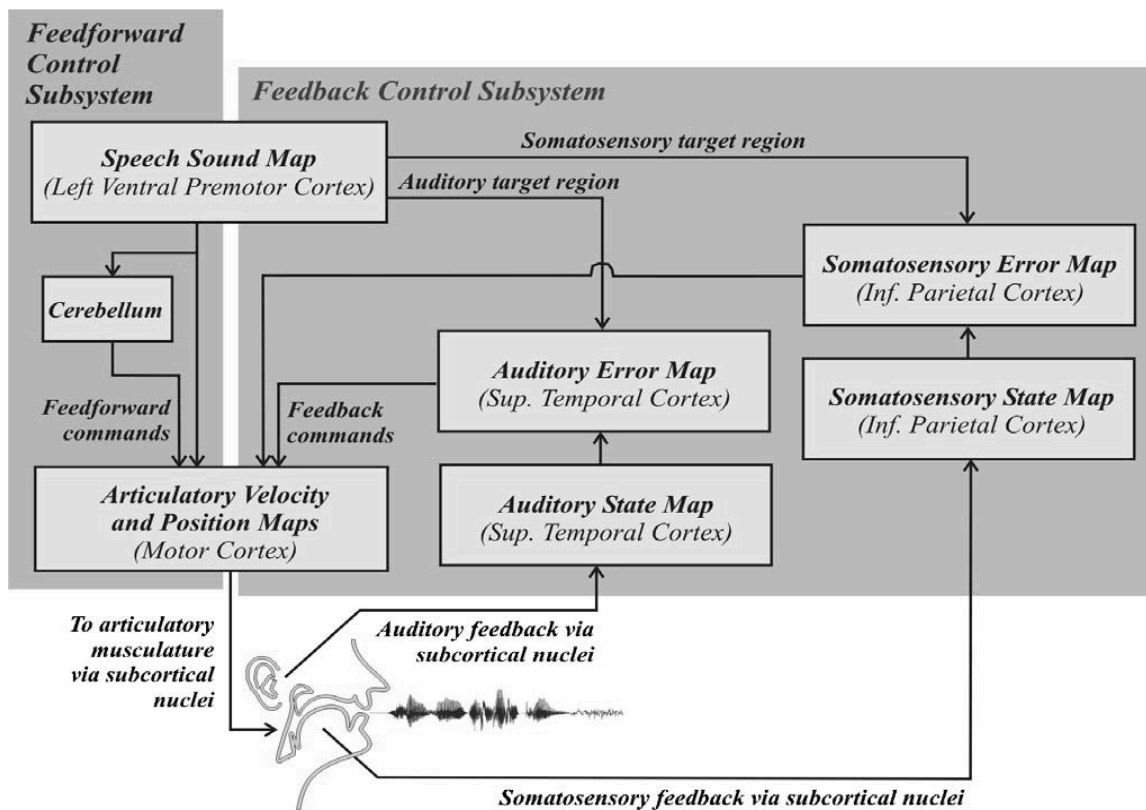
#### 4.1 Compensation

As expected, F1 perturbations in real-time resulted in a varied amount of compensation across participants that opposed the manipulation. Some showed no apparent change in production, however most opposed the manipulation. In line with previous research, average results showed that participants compensated by approximately 25% of the total manipulation (Houde & Jordan, 1998; Purcell & Munhall, 2006a; Villacorta, Perkell, & Guenther, 2007; MacDonald, Goldberg, & Munhall, 2010; MacDonald, Purcell, & Munhall, 2011). Varied response to feedback manipulation is reflected in past literature. Burnett, Freedland, Larson, & Hain (1998) found that in response to pitch-shifted auditory feedback, some individuals followed the manipulation, while most opposed. The same was found in Purcell & Munhall (2006b) where a small number of individuals’ F1 productions followed the manipulation, while the majority opposed. Our sample

consisted of heterogeneous compensators, suggesting that subjects are using the auditory feedback they receive differently.

Approaching these findings using theoretical models can help interpret the relationship between speech production and perception. A number of computational models have explored the idea of speech production as a collection of different auditory-perceptual goals that develop through motor and acoustic input from the articulators (Callan, Kent, Guenther, & Vorperian, 2000; Guenther, 1994; 2006; Perkell et al., 1997; Guenther, Hampson, & Johnson, 1998). In 1994, Guenther first introduced the computational DIVA (Directions in an orosensory space Into Velocities of Articulators) model of speech production. Since then, the model has continued to be adapted and refined.

The DIVA model is a computational network made up of both feed-forward and feedback control subsystems (see Figure 26). It proposes different neural processing steps involved in speech production and acquisition. This model provides a useful framework to interpret our results. The feedback control subsystems in the DIVA model are made up of both auditory target regions and somatosensory target regions that both influence how we perceive and produce speech. The relationship between the two types of feedback might help explain the incomplete nature of compensation. In this model, both the auditory and somatosensory target areas have error maps associated with them. These maps use information from the current auditory or somatosensory state and the target regions to remedy speech errors. If both subsystems are attempting to remedy the error and there is no resulting change in the auditory feedback, the somatosensory inputs may override the auditory inputs resulting in a cessation of compensation. MacDonald and colleagues (2010) suggested that incomplete compensation might be a result of too large a discrepancy between the auditory feedback and the expected production, causing the system to associate the auditory feedback with a source divorced from the subject's own voice. Measures of brainstem encoding that will be discussed in the following sections, were recorded to better understand the role of the auditory brainstem neurons in the process of speech production.



**Figure 26. The DIVA Model of speech-production.**

Adapted from Figure 1 in Guenther (2006).

## 4.2 Polarity Asymmetry

An initial discussion of the findings regarding stimulus presentation polarity and response amplitude is important. During piloting, a previously unknown trend was discovered in our EFR results. It was noted that responses elicited with stimulus polarity A differed from responses elicited by the same stimulus flipped 180° for polarity B. It was observed that one polarity could produce a response double or triple the response compared to the opposite polarity. This phenomenon was independently verified using different EFR data recorded at the laboratory of our collaborator Dr. Steve Aiken (Dalhousie University). As previously mentioned, we continued by treating each polarity as a separate stimulus and response, which is contrary to typical EFR methodology (Aiken & Picton, 2008). Typical EFR methodology has been based upon auditory steady-state response (ASSR) methods where the stimulus envelope is symmetrical, unlike a speech envelope. Treating the responses separately was done to reduce the likelihood of an unwanted extinguishing of the mean response during averaging of the two polarities. Additionally, results showed that one, both, or neither polarities might produce significant responses. There was no readily discernible pattern behind which polarity might elicit a larger response in a given individual. This finding is complex and has important implications for EFR methodology and analysis and requires additional investigation to determine what is contributing to the observed differences.

## 4.3 Envelope Following Response

EFR amplitudes were found to be significantly greater than the background noise amplitudes in response to the subject's own vowel productions (Polarity A: /ε/, N = 20 and /æ/, N = 22; Polarity B: /ε/, N = 27 and /æ/, N = 25) and the standard talker (Polarity A: /ε/, N = 22 and /æ/, N = 13; Polarity B: /ε/, N = 21 and /æ/, N = 12). These detection rates validated the use of the individual's own vowel productions as well as a standard voice because significant responses were elicited in an acceptable proportion of the subjects. This result is similar to previous work that found significant EFRs to the envelope of natural English vowels (/a/ and /i/) produced by a standard talker (Aiken & Picton, 2008). As with the present data, their EFR showed significant response peaks

corresponding to the fundamental frequency of each vowel. In another study Aiken and Picton (2006) found significant EFRs to natural English vowels /a/, /i/ and /u/ in all participants. Difference in detection rate could be explained by differences in vowel duration. The current study used tokens that were between 0.1 and 0.3 s with detection rates at approximately 50 to 70%, whereas the study by Aiken and Picton (2006) used stimuli that were 1.5 s in duration. Stimulus duration is proportional to the SNR when using the Fourier analyzer.

Similar to the speech compensation results, there was a great deal of variation between and within subjects, as well as across polarities A and B in the EFR response. In the current study, the EFR amplitudes in both polarities to a standard talker and the subject's own voice were similar to those found by Aiken and colleagues (2006): in the range of 60 to 110 nV. In the current study, EFR standard deviation to a standard talker in both polarities was in the range of 22 to 38 nV, similar to those found by Aiken and colleagues (2006). The variability to the subject's own voice was greater: approximately 66 nV.

#### 4.4 Frequency Following Response

Similar to the EFR, significant FFRs were found in response to the subject's own vowel productions (Polarity A: /ε/, N = 11 and /æ/, N = 10; Polarity B: /ε/, N = 7 and /æ/, N = 12) and the standard talker (Polarity A: /ε/, N = 7 and /æ/, N = 12; Polarity B: /ε/, N = 13 and /æ/, N = 6). The low detection rates for the FFR reflect the relatively small amplitudes produced by the FFR compared to the EFR. We chose the harmonic closest to F1 of the vowel /ε/ to determine the FFR because harmonics near F1 are presumably most relevant to detection of changes in F1. This result is similar to previous work that found response peaks in the FFR that corresponded to the first and second formants in two-tone approximations of steady-state vowels (Krishnan, 1999). More specifically, FFR amplitudes in the current study to the harmonic closest to F1 (ranging from h3 to h6) were comparable to FFR amplitudes in a study by Krishnan and colleagues (2002) to higher harmonics (h7 and h8) at moderate intensity levels that were thought to be harmonics representing F2: in the range of 15 to 35 nV. Amplitudes observed at lower harmonics (h2 and h3) at moderate intensity levels that were thought to be harmonics

representing F1 were greater in amplitude (90 to 110 nV) than those in the current study. Additionally, in contrast to the current study where detection rates were low (approximately 7 to 43%), Krishnan and colleagues (2002) had a 100% detection rate. The differences observed between the two studies may be a result of different stimuli. Krishnan and colleagues (2002) used less complex synthetic speech stimuli to elicit the FFR whereas in the current study more complex natural vowels were used to elicit the FFR. A discrete Fourier transform can be employed when using synthetic speech as opposed to natural speech, which is less noisy than the Fourier analyzer used in the current study. Further, in the study by Krishnan and colleagues (2002), stimuli were presented 2000 times over 2 hours compared to 500 times and approximately 1 hour in the present study. This difference in recording time could influence the detection rate because longer recording times increase the SNR.

## 4.5 Compensation versus EFR and FFR

It was hypothesized that the differences observed in compensation might be due to differences in the auditory information available to the cortex about the F1 of the vowel / $\epsilon$ / in “head”. The EFR and FFR are evoked potentials with sources in the upper brainstem and the neurons responsible for them can carry this type of complex auditory information to the cortex. The EFR has been found to follow the envelope of natural English vowels (Aiken & Picton, 2006). As mentioned above, compensation to F1 perturbations was highly variable between participants.

In order to relate the compensation results from the F1 shift (/ $\epsilon$ / to / $\text{æ}$ /) to the EFRs elicited by the vowels / $\epsilon$ / and / $\text{æ}$ /, specific metrics of encoding were determined. The first was the absolute amplitude of the EFR to the vowel / $\epsilon$ / in “head”, which was intended to be an overall measure of the robustness of vowel encoding. Laroche et al. (2013) and Choi et al. (2013) have suggested that the EFR elicited by broadband vowels is dominated by the EFR response to harmonics near F1. The second was the change in EFR amplitude (i.e. no use of phase) from “head” to “had”, and finally the vector change from “head” to “had” (which includes both phase and amplitude). The motivation to use these change metrics was the hypothesis that changes in the EFR might reflect

information about changes in speech that are available to the vocal control and therefore be related to the consequent changes in production (a reduction of F1) observed during the F1 formant manipulation paradigm. Studies such as Krishnan (1999) show that the brainstem FFR robustly represents F1 and F2 of two-tone vowel approximations. The same comparisons used for the EFR were completed for the FFR. Again, the intention was to determine if changes in the FFR from the harmonic closest to the F1 of the vowel might reflect information available to the speech controller and thus the observed changes in production.

Contrary to our hypotheses, no significant relationships were observed between the metrics of encoding and compensation magnitude. Both the speech compensation and the change in amplitude measures had adequate ranges, however there was no obvious relationship between the variables. Although studies demonstrate that English vowels are well represented at the level of the brainstem (Aiken & Picton, 2006; Aiken & Picton, 2008; Krishnan, 2002; Krishnan, 1999), this does not appear to influence how we change production to remedy errors in auditory feedback. The process of speech perception, recognition and production is not linear and therefore is likely to be influenced by a number of complex mechanisms.

Studies show that compensation to feedback perturbations is automatic or unconscious (Munhall, MacDonald, Byrne, & Johnsrude, 2009; Elman, 1981; Keough, Hawco, & Jones, 2013), therefore beginning our investigation at the subcortical level seemed appropriate. Brainstem potentials are driven by the auditory signals they receive from the cochleae (vocal feedback in the present case) and reflect information that travels to the cortex. A recent study by Song, Skoe, Banai, & Kraus (2012) found that after training on speech-in-noise perceptual tasks, subjects significantly improved their speech-in-noise perceptual ability and subcortical processing was enhanced for pitch-related cues. Higher amplitude response peaks to the  $f_0$  were interpreted as identifying enhanced encoding in the transition period of the stimulus, and enhanced phase locking to the periodicity of the vowel in noise. Changes to brainstem AEPs appear to represent how effectively auditory information is being encoded during perceptual tasks. Similarly, Russo, Nicol, Zecker, Hayes, & Kraus (2005) trained young participants with known language-based learning

problems on auditory perceptual software. The software included training in phonological awareness, auditory processing and language processing skills. Participant's FFRs were analyzed. Results showed improved encoding of the stimulus [da] in noise compared to those who were not trained. Again, changes in perceptual ability appear to coincide with more robust encoding at the level of the brainstem. This result has been demonstrated in a number of other studies (Cunningham, Nicol, Zecker, Bradlow, & Kraus, 2001; King, Warrier, Hayes, & Kraus, 2002; Hayes, Warrier, Nicol, Zecker, & Kraus, 2003). Finally, in a study by Krizman, Skoe, & Kraus (2012), young Spanish-English bilinguals showed enhanced encoding of the  $f_0$  of the speech syllable [da] compared to their monolingual peers. These changes in encoding were linked to improvements in attention, a behaviour associated with the prefrontal cortex (Miller & Cohen, 2001). Although these examples do not directly address the questions we sought to answer in the current study, they demonstrate that the quality of auditory brainstem potentials in normal hearing individuals, which represent activity in brainstem neurons, reflects enhanced perceptual ability. Characteristics of auditory encoding such as amplitude appear to reflect the information available to the auditory cortex. One might reasonably predict that changes in brainstem encoding would influence accuracy of speech production, however this does not appear to be the limiting factor.

Another possibility is that brainstem encoding of the voice is a wholly sufficient input to the cortex for normal hearing individuals. The appropriate speech production responses to an apparent acoustic error may begin entirely above the brainstem. Gockel, Carlyon, Mehta, & Plack (2012) recorded FFRs to complex tones with altered spectral profiles (e.g. pitch-shifted harmonics, missing harmonics) presented to either one or both ears. Results indicated that the FFRs maintained monaural temporal information but demonstrated no additional processing beyond what is present in the peripheral auditory system. This supports the idea that processing may take place above the level of the brainstem.

Guenther's (2006) DIVA model of speech production fits nicely with the altered auditory feedback results of the current study. He postulates that speech production starts with the activation of a *speech sounds map* cell, which is located in the cortex in Broca's area, or



the frontal operculum. The cells in this area are active when an individual is both producing and perceiving a sound, like during vowel production. The talker uses the feed-forward mechanism to produce sounds and eventually the feedback mechanism to update and incorporate information from the feed-forward model. As the system uses this information, it develops both an auditory target region and a somatosensory target region. These regions consist of the expected auditory and tactile/proprioceptive sensations, respectively and could act as references when remedying speech errors. Areas in the cortex making up the *speech sound map* have potential for future investigation because of the role they play in producing accurate acoustic output. Further, they are integrated in both the feedback control subsystem with projection to both the somatosensory and auditory target regions and the feed-forward system with projections to both the cerebellum and the motor cortex. All of these areas provide viable avenues for future research.

A number of studies have investigated what cortical regions are active during tasks in which auditory feedback is altered or is incongruent with expected feedback. In a verbal self-monitoring task, participants read aloud while their auditory feedback was experimentally modified (Fu, 2005). Once in the scanner, participants made a button press to identify the source of the auditory feedback as either their own voice or another voice. Subjects made more misattributions when the feedback was an altered version of their own voice, and this condition displayed greater superior temporal activation relative to hearing their own voice undistorted or another person's voice (Fu, 2005). In a similar study, individuals produced the vowel /a/ for 5 s while the feedback was frequency shifted up, down, or held constant (Toyomura et al., 2007). In the pitch-shift conditions, participants were instructed to alter their production to keep the pitch constant. Activation was seen in the supramarginal gyrus, prefrontal areas, anterior insula, the superior temporal area and the intraparietal sulcus in the right hemisphere (Toyomura et al., 2007). In a further study, the superior temporal sulcus and superior temporal gyrus bilaterally were found to be most active when participants heard unpredicted auditory feedback while talking (Zheng, Munhall, & Johnsrude, 2010). These results are consistent with the DIVA model of speech production that postulates both somatosensory and auditory

information are required to produce accurate speech.

Another technique employed to investigate altered auditory feedback and how the cortex uses it to remedy speech errors, is measurement of cortical event related potentials (ERP). The N1 P2 is a cortical AEP that reflects changes in the acoustic environment. In a pitch-shifted auditory feedback study, participants'  $f_0$  was shifted either up or down at three different magnitudes. Results showed that greater compensation to pitch-shifts was related to larger amplitude N1 P2 responses (Liu, Meshman, Behroozmand, & Larson, 2011). In a similar study, ERPs were recorded to upward pitch-shifted auditory feedback of five different magnitudes at voice onset during production and during passive listening. Results indicate that the N1 component is maximally suppressed during active speech production with unaltered auditory feedback and becomes greater in amplitude with increasing shift magnitude (Behroozmand & Larson, 2011). The suppression of the N1 response happens because the motor system suppresses the auditory feedback response of active vocalization that is anticipated by the motor system. As the feedback becomes more and more incongruent, the auditory feedback signal becomes more important. These studies provide evidence for the importance of auditory processing that takes place above the brainstem in alterations to auditory feedback.

While providing many avenues for future research, these studies outline the highly complex nature of the human auditory system. The auditory brainstem plays a role in encoding vowel sounds; this finding is demonstrated by our results and supported by past literature. However, the quality of the human EFR and FFR does not appear to influence the degree of compensation observed during a real-time formant manipulation paradigm (hypotheses 1 to 3). In normal hearing individuals, the information provided to cortex by the brainstem may be sufficient and therefore not a factor in the amount of compensation.

## 4.6 Tuning and Compensation

Stimulus frequency otoacoustic emission latency reflects BM travelling wave and filtering delays and can provide an estimate of cochlear tuning: longer latency corresponds to more narrow auditory filters (Shera & Guinan, 2003). Contrary to our 4<sup>th</sup>

hypothesis, a narrower auditory filter bandwidth, measured by SFOAEs, did not correspond with greater compensation magnitude. The aim with the SFOAE group delay measurement was to determine if the sharpness of cochlear tuning was related to the encoding of the spectral changes occurring in the auditory feedback (increasing F1) and affecting the resulting production changes (decreased F1). Human cochlear tuning has been found to be significantly sharper than previously thought (Shera & Guinan, 2003), and therefore was considered to have the potential to encode small changes in the spectrum of auditory feedback. However, the variability observed in the F1 compensation results cannot be explained by the variability in cochlear filter bandwidth. Spectral changes in F1 reach the cochlea and are clearly encoded as shown by the perceptual and AEP data. However, the control of speech production must involve other structures in the auditory and vocal controller pathways (i.e. primary auditory cortex, auditory association areas), which contribute to the variability seen in compensation.

One observation of the SFOAE tuning results that may contribute to the lack of a relationship with speech compensation was that the estimated filter bandwidths had low variability. This finding suggests that all participants had typical filter bandwidths and therefore similar cochlear frequency selectivity. Although the likelihood of a speech compensation mechanism being controlled at the level of the cochlea was not high, it was an important avenue to explore.

Behavioural measures of cochlear tuning were recorded using the SWPTC program to compliment physiological measures. Like the SFOAE measure, behavioural measures of cochlear tuning did not predict the variability observed in the speech compensation data. Unexpectedly and contrary to research conducted by Sek and colleagues (2005), there was no relationship between physiological measures of cochlear tuning and behavioural measures of cochlear tuning. The differences observed could be explained by the extent to which the auditory system is employed in each task. In the behavioural task, much more of the auditory system is being recruited to complete the task, whereas in the physiological measure, results are from the cochlea and OHCs. Further independent validation of this new psychophysical method is required before any firm conclusions can be drawn from the results.

## 4.7 Perception & Compensation

A robust relationship was found between vowel goodness ratings (perception) and compensation magnitude per ramp step (production). This result was anticipated as it is in line with previous data from our laboratory (Nguyen, 2012). This finding is also related to research that finds the perceptual organization of vowel space, vowel categories, and vowel goodness all influence formant control (Mitsuya, 2011). It also fits with the idea of auditory target regions outlined in Guenther's DIVA model of speech production (2006). Guenther uses the term target regions to accommodate the variability seen in speech production, which was evident in the F1 productions observed in the formant-shift paradigm results. The results from the current study provide evidence for the existence of auditory perceptual targets and provide a direct link between these targets and how they are related to production. On average, the farther away the vowel feedback was from the exemplar, the greater the compensatory response to remedy that error. No other studies to our knowledge demonstrate such a relationship.

## 4.8 Closing remarks, limitations, and future work

This study aimed to determine if brainstem and peripheral acoustic encoding play a role in the control of formant production during real-time auditory feedback manipulation. On average, participants opposed the F1 manipulation. Further, participants own vowel utterances (/ɛ/ and /æ/) and those from a standard talker produced detectable EFRs and FFRs. Contrary to the hypotheses, no significant linear relationships were observed between results from the formant-shifting paradigm and the AEPs. Cochlear tuning measures produced similar non-significant results. Despite the non-significance of the results, determining that the control of speech production responses to an acoustic error may not originate in the brainstem is a valuable contribution to the field. Further, the discovery of polarity response asymmetry has important methodological implications moving forward with the speech elicited EFR.

This study is not without limitations. One limitation to our design that may have influenced the electrophysiology results was the relatively short duration of each vowel production. As mentioned above, increased stimulus duration is directly related to higher SNR values and therefore higher detection rates. The challenge with using participants' own vowel productions was the variability of vowel length between participants. Although this may have influenced our response detection rate, it provided an ecologically valid method to investigate how participants encode their own voice and the voice of a standard talker. One way to improve SNR while still using participants' own vowel productions would be to have participants increase the duration of their vowel productions. Although this may not provide an entirely accurate representation of natural production, it may increase SNR values and response detection. Another way to increase SNR value would be to record more sweeps during the EFR and FFR measurements.

Another limitation was the results of the SWPTC program as a behavioural measure of cochlear tuning. Findings were not consistent with physiological measures of tuning obtained from the SFOAEs. Further validation of this program alongside physiological measures is recommended before concrete conclusions can be drawn about its effectiveness as a behavioural measure of cochlear tuning.

Future studies should focus on the role of cortical AEPs in encoding and remedying speech errors. The current study examined only brainstem AEPs. It would be interesting to investigate the cortical N1 P2 response, a cortical AEP that occurs in response to a change in the auditory environment, and compensation. Martin & Boothroyd (2000) investigated the acoustic change complex (ACC) to vowels during a change in F2 and discovered the behavioural threshold for detecting change in F2 was similar to the threshold for detecting the ACC. It would be interesting to do a similar comparison with the N1 P1 response and F1.

It would also be interesting to further investigate the role of the cortex in the perception and production of speech errors and how they are remedied using functional imaging (Christoffels, Formisano, & Schiller, 2007). Despite the methodological challenges with

real-time speech production and functional imaging, it could provide another avenue for comparing behaviour and brain activation.

In conclusion, subcortical processing of speech sounds does not appear to control speech production changes to remedy perceived errors in auditory feedback. Future research examining the role of cortical AEPs is required. A better understanding of this mechanism may have clinical benefits for the fields of speech pathology and hearing rehabilitation.

## References

- Aiken, S. J., & Picton, T. W. (2006). Envelope following responses to natural vowels. *Audiology and Neurotology, 11*(4), 213–232. doi:10.1159/000092589
- Aiken, S. J., & Picton, T. W. (2008a). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research, 245*, 35–47. Elsevier B.V. doi:10.1016/j.heares.2008.08.004
- Aiken, S. J., & Picton, T. W. (2008b). Human cortical responses to the speech envelope. *Ear and Hearing, 29*(2), 139–157.
- Batra, R., Kuwada, S., & Maher, V. L. (1986). The frequency-following response to continuous tones in humans. *Hearing Research, 21*(2), 167–177.
- Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience, 12*(1), 54. doi:10.1186/1471-2202-12-54.
- Bentsen, T., Harte, J. M., & Dau, T. (2011). Human cochlear tuning estimates from stimulus-frequency otoacoustic emissions. *The Journal of the Acoustical Society of America, 129*(6), 3797-3807. doi:10.1121/1.3575596
- Berger, J. R., & Blum, A. S. (2006). Brainstem Auditory Evoked Potentials. In A. S. Blum & S. B. Rutkove (Eds.), *The Clinical Neurophysiology Primer*. Totowa: Humana Press Inc.
- Boberg, C. (2000). Geolinguistic diffusion and the US-Canada border. *Language, Variation and Change, 12*, 1-24.
- Burkard, R. (1994). Gerbil brain-stem auditory-evoked responses to maximum length sequences. *The Journal of the Acoustical Society of America, 95*(4), 2126–2135.
- Burkard, R.F., Don, M., & Eggermont, J. J. (2007). Auditory Evoked Potentials: basic principles and clinical application. Philadelphia: Lippincott Williams & Wilkins.
- 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.

- 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*(3), 721–736.
- Choi, J.M., Purcell, D.W., Coyne, J., A. & Aiken, S.J. (2013). Envelope Following Responses Elicited By English Sentences. *Ear & Hearing*, in press
- 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. doi:10.1002/hbm.20315
- Clarke, S., Elms, F., & Youssef, A. (1995). The third dialect of English: Some Canadian evidence. *Language, Variation and Change, 7*, 209–228.
- Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clinical neurophysiology, 112*(5), 758–767.
- Dajani, H. R., Wong, W., & Kunov, H. (2005). Fine structure spectrography and its application in speech. *The Journal of the Acoustical Society of America, 117*(6), 3902. doi:10.1121/1.1896365
- Davis, R. L., & Britt, R. H. (1984). Analysis of the frequency following response in the cat. *Hearing Research, 15*(1), 29–37.
- Dhanjal, N. S., Handunnetthi, L., Patel, M. C., & Wise, R. J. S. (2008). Perceptual systems controlling speech production. *Journal of Neuroscience, 28*(40), 9969–9975. doi:10.1523/JNEUROSCI.2607-08.2008
- Dhar, S., Abel, R., Hornickel, J., Nicol, T., Skoe, E., Zhao, W., & Kraus, N. (2009). Exploring the relationship between physiological measures of cochlear and brainstem function. *Clinical Neurophysiology, 120*(5), 959–966. doi:10.1016/j.clinph.2009.02.172
- Elman, J. L. (1981). Effects of Frequency-shifted Feedback on the Pitch of Vocal Productions. *Journal of Acoustical Society of America, 70*(1), 45–50.
- Flanagan, J. R., & Wing, A. M. (1993). Modulation of grip force with load force during point-to-point arm movements. *Experimental Brain Research, 95*(1), 131–143.



- Fu, C. H. Y. (2005). An fMRI Study of verbal self-monitoring: neural correlates of auditory verbal feedback. *Cerebral Cortex*, *16*(7), 969–977.  
doi:10.1093/cercor/bhj039
- Garber, S. R., Siegel, G. M., & Pick, H. L. (1980). The effects of feedback filtering on speaker intelligibility. *Journal of Communication Disorders*, *13*(4), 289–294.
- Gardi, J., Merzenich, M., & McKean, C. (1979). Origins of the scalp recorded frequency-following response in the cat. *Audiology*, *18*(5), 358–381.
- Gerken, G. M., Moushegian, G., Stillman, R. D., & Rupert, A. L. (1975). Human frequency-following responses to monaural and binaural stimuli. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 379–386.
- Glaser, E. M., Suter, C. M., Dasheiff, R., & Goldberg, A. (1976). The human frequency-following response: its behavior during continuous tone and tone burst stimulation. *Electroencephalography and Clinical Neurophysiology*, *40*, 25–32.
- Gockel, H. E., Carlyon, R. P., Mehta, A., & Plack, C. J. (2011). The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. *Journal of the Association for Research in Otolaryngology*, *12*(6), 767–782. doi:10.1007/s10162-011-0284-1
- Greenberg, S., Marsh, J. T., Brown, W. S., & Smith, J. C. (1987). Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hearing Research*, *25*(2-3), 91–114.
- 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. (2006). Cortical interactions underlying the production of speech sounds. *Journal of communication disorders*, *39*(5), 350–365.  
doi:10.1016/j.jcomdis.2006.06.013
- 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., Ghosh, S., & Tourville, J. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*, 280–301.

- Hagiwara, R. E. (2006). Vowel production in Winnipeg. *Canadian Journal of Linguistics*, 51, 127–141.
- Hall, J. W. (1979). Auditory brainstem frequency following responses to waveform envelope periodicity. *Science*, 205(4412), 1297–1299.
- Hayes, E. A., Warrier, C. M., Nicol, T. G., Zecker, S. G., & Kraus, N. (2003). Neural plasticity following auditory training in children with learning problems. *Clinical Neurophysiology*, 114(4), 673–684. doi:10.1016/S1388-2457(02)00414-5
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216.
- Iverson, P., & Kuhl, P. (1996). Influences of phonetic identification and category goodness on American listeners' perception of r and l. *The Journal of the Acoustical Society of America*, 99(2), 1130–1140.
- Jackson, C. P. T., & Miall, R. C. (2007). Contralateral manual compensation for velocity-dependent force perturbations. *Experimental Brain Research*, 184(2), 261–267. doi:10.1007/s00221-007-1179-6
- John, S, & Picton, T. W. (200). MASTER: A window program for recording multiple auditory steady-state responses. *Computer Methods and Programming in Biomedicine*, 61, 125-150.
- Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, 108(3 Pt 1), 1246–1251.
- Jones, J., & Kevin, M. (2002). The role of auditory feedback during phonation: studies of Mandarin tone production. *Journal of Phonetics*, 30(3), 303–320. doi:10.1006/jpho.2001.0160
- Kalluri, R., & Shera, C., S. (2007). Near equivalence of human click-evoked and stimulus-frequency otoacoustic emissions. *Journal of the Acoustical Society of America*, 121(4), 2097-2110.
- Kandel, E., Schwartz, J., & Jessel, T. (2000). Principles of Neural Science, 5<sup>th</sup> Ed. Siegelbaum, S., & Hudspeth, A. J. (Eds). McGraw Hill, New York.
- Kemp, D. T. (1978). Stimulated acoustic emissions from within the human auditory system. *The Journal of the Acoustical Society of America*, 64(5), 1386–1391.

- Kemp, D. T. (1979). Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea. *Archives of Oto-rhino-laryngology*, 224(1-2), 37–45.
- Kemp, D. T. (2007). Otoacoustic emissions: Clinical applications 3<sup>rd</sup> Ed. Robinette, M., & Glatke, T. (Eds). Thieme, New York.
- 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), 1–1. doi:10.1186/1471-2202-14-25
- King, C., Warrier, C. M., Hayes, E., & Kraus, N. (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neuroscience Letters*, 319(2), 111–115.
- Kraus, N. (1999). Speech sound perception, neurophysiology, and plasticity. *International journal of pediatric otorhinolaryngology*, 47(2), 123–129.
- Krishnan, A. (1999). Human frequency-following Responses to two-tone approximations of steady-state vowels. *Audiology & Neurotology*, 4, 95–103.
- Krishnan, A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hearing Research*, 166(1-2), 192–201.
- Krishnan, A., Xu, Yisheng, Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: representation of pitch contours in Chinese tones. *Hearing Research*, 189(1-2), 1–12. doi:10.1016/S0378-5955(03)00402-7
- Krizman, J., Skoe, E., & Kraus, N. (2012). Sex differences in auditory subcortical function. *Clinical Neurophysiology*, 123(3), 590–597. International Federation of Clinical Neurophysiology. doi:10.1016/j.clinph.2011.07.037
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50, 93–107.
- Kuwada, S., Batra, R., & Maher, V. L. (1986). Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones. *Hearing Research*, 21(2), 179–192.
- Laroche, M., Dajani, H. R., Prévost, F., & Marcoux, A. M. (2013). Brainstem auditory responses to resolved and unresolved harmonics of a synthetic vowel in quiet and noise. *Ear and Hearing*, 34(1), 63–74. doi:10.1097/AUD.0b013e31826119a1

- Larson, C. R., Burnett, T. A., Kiran, S., & Hain, T. C. (2000). Effects of pitch-shift velocity on voice F0 responses. *The Journal of the Acoustical Society of America*, *107*(1), 559–564.
- Lee, B. (1950). Effects of delayed speech feedback. *Journal of the Acoustical Society of America*, *22*(6), 824–826.
- Levi, E. C., Folsom, R. C., & Dobie, R. A. (1993). Amplitude-modulation following response (AMFR): effects of modulation rate, carrier frequency, age, and state. *Hearing Research*, *68*(1), 42–52.
- Levi, E. C., Folsom, R. C., & Dobie, R. A. (1995). Coherence analysis of envelope-following responses (EFRs) and frequency-following responses (FFRs) in infants and adults. *Hearing Research*, *89*(1-2), 21–27.
- Lindblom, B. E., & Sundberg, J. E. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, *50*(4), 1166–1179.
- 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.  
doi:10.1016/j.clinph.2010.08.010
- MacDonald, E. N., Goldberg, R., & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, *127*(2), 1059. doi:10.1121/1.3278606
- MacDonald, E. N., Purcell, D. W., & Munhall, K. G. (2011). Probing the independence of formant control using altered auditory feedback. *The Journal of the Acoustical Society of America*, *129*(2), 955. doi:10.1121/1.3531932
- Malicka, A. N., Munro, K. J., & Baker, R. J. (2009). Fast method for psychophysical tuning curve measurement in school-age children. *International Journal of Audiology*, *48*(8), 546–553. doi:10.1080/14992020902845899
- Marsh, J. T., Brown, W. S., & Smith, J. C. (1974). Differential brainstem pathways for the conduction of auditory frequency-following responses  
*Electroencephalography and Clinical Neurophysiology*, *36*(4), 415–424.
- Martin, B. A., & Boothroyd, A. (2000). Cortical, auditory, evoked potentials in response

- to changes of spectrum and amplitude. *Journal of the Acoustical Society of America*, 107(4), 2155–2161.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.  
doi:10.1146/annurev.neuro.24.1.167
- Mitsuya, T., MacDonald, E. N., Purcell, D. W., & Munhall, K. G. (2011). A cross-language study of compensation in response to real-time formant perturbation. *The Journal of the Acoustical Society of America*, 130(5), 2978.  
doi:10.1121/1.3643826
- Moller, A. R. (2006) Hearing: anatomy, physiology and disorders of the auditory system, 2<sup>nd</sup> ed. Academic Press, UK: London.
- Moore, B. C. (1978). Psychophysical tuning curves measured in simultaneous and forward masking. *The Journal of the Acoustical Society of America*, 63(2), 524–532.
- Moushegian, G., Rupert, A. L., & Stillman, R. D. (1973). Laboratory note. Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalography and Clinical Neurophysiology*, 35(6), 665–667.
- Munhall, K. G., MacDonald, E. N., Byrne, S. K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America*, 125(1), 384. doi:10.1121/1.3035829
- Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., & Guiod, P. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication*, 22, 227–250.
- Peterson, G., & Barney, H. (1952). Control Methods Used in a Study of the Vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184.
- Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials. II. Effects of attention. *Electroencephalography and Clinical Neurophysiology*, 36(2), 191–199.
- Picton, T. W., Hillyard, S. A., Krausz, H. I., & Galambos, R. (1974). Human Auditory Evoked Potentials I: Evaluation of components. *Electroencephalography and*

- Clinical Neurophysiology*, 36, 179–190.
- 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. doi: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. doi:10.1121/1.2173514
- Purcell, D. W., John, S. M., Schneider, B. A., & Picton, T. W. (2004). Human temporal auditory acuity as assessed by envelope following responses. *The Journal of the Acoustical Society of America*, 116(6), 3581. doi:10.1121/1.1798354
- Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research*, 156(1), 95–103. doi:10.1016/j.bbr.2004.05.012
- Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology*, 115(9), 2021–2030. doi:10.1016/j.clinph.2004.04.003
- Sachs, M. B., & Young, E. D. (1979). Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. *The Journal of the Acoustical Society of America*, 66(2), 470–479.
- Sachs, M. B., Voigt, H. F., & Young, E. D. (1983). Auditory nerve representation of vowels in background noise. *Journal of Neurophysiology*, 50(1), 27–45.
- Schairer, K. S., Ellison, J. C., Fitzpatrick, D., & Keefe, D. H. (2006). Use of stimulus-frequency otoacoustic emission latency and level to investigate cochlear mechanics in human ears. *The Journal of the Acoustical Society of America*, 120(2), 901. doi:10.1121/1.2214147
- Seikle, A., King, D., & Drumriht, D. (2010). *Anatomy and physiology for speech, language and hearing*, 3<sup>rd</sup> Ed. Delmar, Clifton Park, NY.
- Şek, A., Alcántara, J., Moore, B. C. J., Kluk, K., & Wicher, A. (2005). Development of a fast method for determining psychophysical tuning curves. *International Journal of Audiology*, 44(7), 408–420. doi:10.1080/14992020500060800
- Shera, C. A. (2003). Mammalian spontaneous otoacoustic emissions are amplitude-

- stabilized cochlear standing waves. *The Journal of the Acoustical Society of America*, 114(1), 244. doi:10.1121/1.1575750
- Shera, C. A., & Guinan, J. J. (2003). Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning. *The Journal of the Acoustical Society of America*, 113(5), 2762. doi:10.1121/1.1557211
- Siegel, G. M., & Pick, H. L. (1974). Auditory feedback in the regulation of voice. *The Journal of the Acoustical Society of America*, 56(5), 1618–1624.
- Small, A. M. (1959). Pure-Tone Masking. *The Journal of the Acoustical Society of America*, 11, 1619–1625.
- Smith, J. C., Marsh, J. T., & Brown, W. S. (1975). Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalography and Clinical Neurophysiology*, 39(5), 465–472.
- Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2012). Training to Improve Hearing Speech in Noise: Biological Mechanisms. *Cerebral Cortex*, 22(5), 1180–1190. doi:10.1093/cercor/bhr196
- Souter, M. (1995). Stimulus frequency otoacoustic emissions from guinea pig and human subjects. *Hearing Research*, 90(1-2), 1–11.
- Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, 84(3), 917–928.
- Tobey, E. A., & Murchison, C. (1989). Vowel formant frequencies produced with and without auditory feedback. *The Journal of the Acoustical Society of America*, 86, S97–S98.
- Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., & Kuriki, S. (2007). Neural correlates of auditory feedback control in human. *Neuroscience*, 146(2), 499–503. doi:10.1016/j.neuroscience.2007.02.023
- Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, 423(6942), 866–869. doi:10.1038/nature01710
- 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.


doi:10.1121/1.2773966

- Waldstein, R. S. (1991). A response to Sapir and Canter [S. Sair and G. J. Canter, J. Acoust. Soc. Am. 90, 1672 (1991)]. *The Journal of the Acoustical Society of America*, 90, 1673.
- Ward, W. D. (1955). Tonal monaural diplacusis. *The Journal of the Acoustical Society of America*, 27, 365–372.
- Wegel, R. L. (1931). A study of Tinnitus, *Archives of Otolaryngology*, 14(2), 158-165.  
doi:10.1001/archotol.1931.00630020182004
- Worden, F. G., & Marsh, J. T. (1968). Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalography and Clinical Neurophysiology*, 25(1), 42–52.
- Xu, Y., Larson, C. R., Bauer, J. J., & Hain, T. C. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *The Journal of the Acoustical Society of America*, 116(2), 1168.  
doi:10.1121/1.1763952
- Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, 60, 213–232.
- Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, 22(8), 1770–1781.  
doi:10.1162/jocn.2009.21324



# Appendices

## Appendix A: Ethics approval notice



**Western  
Research**

Research Ethics

Use of Human Participants - Ethics Approval Notice

**Principal Investigator:** Dr. David Purcell  
**File Number:** 102066  
**Review Level:** Delegated  
**Approved Local Adult Participants:** 50  
**Approved Local Minor Participants:** 0  
**Protocol Title:** The Influence of Speech Encoding on the use of Auditory Feedback during Speech Production - 18673E  
**Department & Institution:** Health Sciences/Communication Sciences & Disorders, Western University  
**Sponsor:** Ontario Early Researcher Award

**Ethics Approval Date:** August 14, 2012 **Expiry Date:** August 31, 2013  
**Documents Reviewed & Approved & Documents Received for Information:**

Document Name	Comments	Version Date
Revised Western University Protocol	There has been a revision in the testing that will be done to provide a different stimulus.	

---

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the University of Western Ontario Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Signature \_\_\_\_\_

**Ethics Office to Contact for Further Information**

Janice Sutherland 	Grace Kelly 	Shantel Walcott 
-----------------------	-----------------	---------------------

*This is an official document. Please retain the original in your files.*

Western University, Support Services Bldg., Rm. 5150  
 1393 Western Rd., London, ON, N6G 1G9 t: www.uwo.ca/research/ethics

## Appendix B: Forms and questionnaires for participant



November 29, 2011

### The Influence of Speech Encoding on the use of Auditory Feedback during Speech Production

**David Purcell, Ph.D., Assistant Professor**  
**National Centre for Audiology**  
**School of Communication Sciences and Disorders**  
**University of Western Ontario**

#### LETTER OF INFORMATION

##### Study Background

You are being invited to participate in a study, which investigates how speech production is affected by differences in neural encoding when hearing our own voice, also known as auditory feedback. All measurements will take place in the Electrophysiology Laboratory of the National Centre for Audiology in Elborn College at the University of Western Ontario.

Hearing your own voice is crucial for learning and maintaining accurate speech production. Deafness early in life hampers the development of normal speech in children, and in adults the onset of hearing impairment can cause deterioration in many aspects of speech. This study will attempt to determine how the auditory system encodes auditory feedback and how this shapes articulation patterns for vowels where English is the first language.

##### Questionnaire and Hearing Assessment

This study will include a total of 50 individuals. If you agree to participate in the study, you will take part in a brief questionnaire and a brief assessment of your hearing. This will be followed by the main experiment, which will be conducted over two different testing sessions. In the first part of the experiment, you will be prompted to speak into a microphone and listen to speech through a set of headphones.

The hearing assessment will include three assessments called otoscopy, tympanometry, and pure-tone audiometry which together take about 20 minutes to complete. Otoscopy is a brief visual examination of your ears with an instrument called an otoscope. Tympanometry involves placing earphones in your ears while sounds are played and the pressure in your ear canal is varied. For audiometry, you will hear tones one at a time

Initials \_\_\_\_\_

Faculty of Health Sciences • The University of Western Ontario  
 National Centre for Audiology

Page 1 of 4

November 29, 2011



through headphones, and you will signal when you detect each tone. The tones will progressively become quieter until you are no longer able to hear them. This procedure is repeated for several different pitches and for each ear.

#### **Speech Production and Auditory Feedback**

Following the hearing assessment, you will be seated in front of a video monitor that displays words. You will be asked to say the words on the screen. Your voice will be recorded with a microphone, and played to you through headphones. The microphone and headphones are connected to a computer that can analyze and change the speech sounds. This part of the experiment normally takes less than 45 minutes to complete. You may also be asked to listen to speech sounds and to indicate what you think of those sounds with a button press. This part of the experiment also normally takes less than 30 minutes to complete.

#### **Electroencephalogram (EEG) and Otoacoustic Emissions (OAE)**

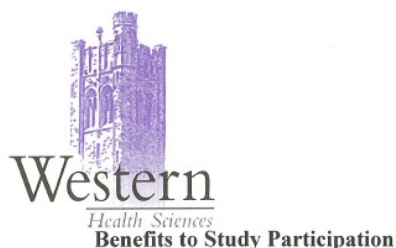
In a second experimental testing session you will take part in a measurement of your inner ears response to sound. This requires the placement of earphone inserts in the ear canal. Additionally, an electrical measurement of your brain's response to sound will be taken. This requires the placement of three electrodes on the surface of your skin. The sites for these will be cleaned using an alcohol pad and a gentle scrub pad to improve electrical contact. One electrode will be placed on your collarbone and the other two will be placed on your head. A conductive gel and light adhesive will hold them in place. After the experiment, the electrodes will be gently removed and the gel cleaned away with a damp cloth. Sometimes people may temporarily experience redness where the surface electrodes were placed due to the skin cleaning procedure. Usually this redness fades within an hour or two, but sometimes may be present up to a day or two later. If you are concerned after this time, you should seek medical advice from your family doctor or a walk-in clinic.

During the measurement, you will lie comfortably in a reclined easy chair and are encouraged to sleep. Sounds will be presented at a comfortable loudness and measurement time will be approximately 50 minutes.

#### **Risks**

These methods are widely used in laboratories studying hearing. There are no known risks associated with this technology.

November 29, 2011



Participation in the study is voluntary. You may refuse to participate, or withdraw from the study at any time. The procedures to be used in this study are designed for research purposes and are not intended to provide you with any direct benefit. It may contribute to our understanding of the role of hearing on speech production and the use of auditory feedback, which is of benefit to society in the long term. There may be the possibility that the brief hearing assessment could identify a previously unknown hearing impairment. If this were to occur, we will encourage you to seek professional assessment from your family practitioner or audiologist. We may also provide information about obtaining an assessment at the UWO audiology clinic in Elborn College.

All information obtained in this study will be held in strict confidence and participant anonymity will be maintained. Your name will not appear in any publications or presentations of the findings of this study. Your personal and background information will be kept separately from all data. If you would like to receive copies of these publications, please contact Dr. Purcell at the telephone number below.

On the Consent Form that follows, you will be given the opportunity to indicate that you would be interested in receiving invitations to participate in future research studies conducted by Dr. Purcell. This opportunity is unrelated to the present research study.

If you have any questions or would like additional information about this study, please contact Dr. David Purcell, National Centre for Audiology, School of Communication Sciences and Disorders, University of Western Ontario, London, Ontario, N6G 1H1 (telephone: [REDACTED])

If you have questions regarding the conduct of this study or your rights as a research participant, you may contact the Office of Research Ethics at [REDACTED] or via electronic mail at [REDACTED]

### **Compensation**

Participants in this study are reimbursed for the time committed to the study and the inconveniences associated with participation in the study at the rate of \$5/half-hour or part-thereof.

### **Signing of Consent Form**

---

Initials

Faculty of Health Sciences • The University of Western Ontario  
National Centre for Audiology

Page 3 of 4

November 29, 2011



If you agree to participate in this study, please sign the consent form. You do not waive any legal rights by signing the consent form. You will be given a copy of this Letter of Information for your records.

Sincerely, 

\_\_\_\_\_  
Initials

Faculty of Health Sciences • The University of Western Ontario  
National Centre for Audiology

Page 4 of 4



**The Influence of Speech Encoding in the use of Auditory Feedback**

**David Purcell, Ph.D., Assistant Professor  
National Centre for Audiology  
School of Communication Sciences and Disorders  
University of Western Ontario**

**CONSENT FORM**

I have read the Letter of Information, have had the nature of the study explained to me, and I agree to participate. All the questions have been answered to my satisfaction.

Research Participant (please print): \_\_\_\_\_

Signature: \_\_\_\_\_ Date: \_\_\_\_\_

Signature of Person Responsible for Obtaining Signed Consent

Signature: \_\_\_\_\_ Date: \_\_\_\_\_

\_\_\_\_/\_\_\_\_/\_\_\_\_

Date:

### Background Information

Participant ID: \_\_\_\_\_

Birth year (mm/yyyy): \_\_\_\_/\_\_\_\_ Age: \_\_\_\_\_ Sex: Male/Female

Handedness: Right/Left

Vision status: Glasses/Contacts/None

Any known problems with:

i) Hearing:

\_\_\_\_\_

ii) Speech and Language:

\_\_\_\_\_

iii) Vision:

\_\_\_\_\_

iv) Other:

\_\_\_\_\_

What is your country of birth?

\_\_\_\_\_

List the languages you know in order

a) in which you learned  
them: \_\_\_\_\_

b) from the one who know best to the one you know least:

\_\_\_\_\_

Father's 1<sup>st</sup> Language: \_\_\_\_\_

Father's 2<sup>nd</sup> Languages: \_\_\_\_\_

Mother's 1<sup>st</sup> Language: \_\_\_\_\_

Mother's 2<sup>nd</sup> Language: \_\_\_\_\_

**Music History**

Participant ID: \_\_\_\_\_ Date: \_\_\_\_\_

Have you had vocal (singing) lessons:                    YES                    NO

If yes, what type of training did you receive:

\_\_\_\_\_

If yes, what was your highest completion of grade/level/number of years:

\_\_\_\_\_

Have you had instrument lessons:                    YES                    NO

If yes, what was your  
instrument(s): \_\_\_\_\_If yes, what was your highest completion of grade/level/number of  
years: \_\_\_\_\_



## Curriculum Vitae

**Name:** Laura Beamish

**Post-secondary Education and Degrees:** Queen's University  
Kingston, Ontario, Canada  
2007-2011 B.Sc.H

The University of Western Ontario  
London, Ontario, Canada  
2011-2013 M.Sc.

**Related Work Experience** Teaching Assistant  
The University of Western Ontario  
2011- 2012