THE INTERACTION BETWEEN SPEECH PERCEPTION AND SPEECH PRODUCTION: IMPLICATIONS FOR SPEAKERS WITH DYSARTHRIA

A thesis submitted as fulfilment of the requirements

for the degree of Doctor of Philosophy

at University of Canterbury

by Martina Christina Marion Schaefer

MSLT (distinction)

June 2013

DECLARATION BY AUTHOR

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgement has been made in the text.

LIST OF PRESENTATIONS BY THE AUTHOR RELEVANT TO THE THESIS

1. **Schäfer, M**., McAuliffe, M. (2012). Can the DIVA model of speech perception predict motor speech behaviour in ageing individuals? *NZILBB Ageing Seminar*, Christchurch, New Zealand: 21 June 2012.

2. **Schäfer, M**., McAuliffe, M., Liss, J., Katseff, S.E., O'Beirne, G., & Cai, S. (2012). Responses to manipulations in auditory feedback: The effect of aging. *Motor Speech Conference*, Santa Rosa, California: 1-4 March 2012.

3. **Schäfer, M**., & McAuliffe, M. (2011). How can theoretical models of speech perception inform our clinical practice? *NZSTA Professional Development Seminar*, Dunedin, New Zealand: 30-31 March 2011.

4. **Schäfer, M**., McAuliffe, M., Liss, J., O'Beirne, G., & Cai, S. (2011). Responses of older individuals to manipulations in auditory feedback: Preliminary findings. *Eight Asia Pacific Conference on Speech, Language and Hearing*, Christchurch, New Zealand: 11-14 January 2011.

5. **Schäfer, M**., McAuliffe, M., Liss, J. (2009). The interaction between speech perception and speech production: Implications for speakers with dysarthria. *Postgraduate Research Conference, Department of Communication Disorders*, Christchurch, New Zealand: 12 November 2009.

ACKNOWLEDGEMENTS

I would like to thank Associate Professor Megan McAuliffe for her guidance and advice on the overall structure and direction of this work and her support in bringing Dr. Shanqing Cai from the Massachusetts Institute of Technology (MIT) to the University of Canterbury to set up the perturbation experiment. Special thanks to Dr. Shanqing Cai for his coaching and highly appreciated guidance with the implementation of the programme. I would like to further thank and acknowledge Associate Professor Julie Liss for her mentorship—her student-focused approach to coaching, aiming at empowering the student and helping them to excel, coupled with her enthusiasm, encouragement, and genuine care will always remain a source of inspiration for me. Thanks to Dr. Greg O'Beirne for his technical support with the perturbation study. Thanks to my whole supervisory team for their feedback on my written work. An extensive acknowledgement to Dr. Patrick La Shell who worked many hours on the statistical analyses of the studies presented in this dissertation. Further thanks to Professor Jennifer Hay for sharing her linguistic expertise on New Zealand English front-vowels characteristics. Thanks also to the Motor Speech Disorders Laboratory at Arizona State University, particularly Rene Utianski and Steven Sandoval, for their methodological and statistical guidance. In addition, thanks to Professor Tim Anderson for supporting my participant recruitment. And of course, this research would not have been possible without all the participants who have committed their time and energy. I thank Dr. Tze-Peng Wong, Dr. Dona Jayakody, Dr. Stephanie Borrie, and Jayne Moyle for their ongoing support and friendship throughout the highs and lows of my PhD experience and allowing me to support them in theirs. Additional thanks to Jo and his emotional and editorial support. Last but not least, my deepest gratitude goes to my family and friends whose unconditional love and firm belief in me has helped me to pick myself up and move forward even when I did not think I could walk any further.

ABSTRACT

The purpose of the research presented here was to systematically investigate the role of speech perception on speech production in speakers of different ages and those with PD and hypokinetic dysarthria. For this, the experimental designs of auditory perturbation and mimicry were chosen. The initial research phase established that the magnitude of compensation to auditory vowel perturbation was reduced in 54 speakers of New Zealand English (NZE) when compared to previous studies conducted with speakers of American (AE) and Canadian English (CE). A number of factors were studied to determine possible predictors of compensation and distinguish between potential changes associated with ageing. However, no predictors of compensation were found for the overall group. Post-hoc analyses established an increased variability in response patterns in NZE when compared to previous studies of AE and CE. Subsequent follow-up analyses focused on the response-dependent categories of (1) *big compensators,* (2) *compensators*, (3) *big followers,* and (4) *followers.* Linear mixed-effect modelling revealed that in *big compensators,* the magnitude of compensation was greater in speakers who exhibited larger F1 baseline standard deviation and greater F1 vowel distances of HEAD relative to HEED and HAD. F1 baseline standard deviation was found to have a similar predictive value for the group of *compensators*. No predictors of compensation were found for the other two subgroups. Phase two was set up as a continuation of phase one and examined whether a subset of 16 speakers classified as *big compensators* adapted to auditory vowel perturbation. Linear mixed-effect modelling revealed that in the absence of auditory feedback alterations, *big compensators* maintained their revised speech motor commands for a short period of time until a process of de-adaptation was initiated. No predictors of adaptation were found for the group. Due to the unexpected results from the first two research phases indicating a dominant weighting of somatosensory feedback in NZE compared to auditory-perceptual influences, a different experimental paradigm was selected for phase three—mimicry. The purpose of this study was to determine whether eight speakers with PD and dysarthria and eight age-matched healthy controls (HC) are able to effectively integrate speech perception and speech production when attempting to match an acoustic target. Results revealed that all speakers were able to modify their speech production to approximate the model speaker but the acoustic dimensions of their speech did not move significantly closer to the target over the three mimicry attempts.

Although speakers with moderate levels of dysarthria exhibited greater acoustic distances (except for the dimension of pitch variation), neither the perceptual nor the acoustic analyses found significant differences in mimicry behaviour across the two groups. Overall, these findings were considered preliminary evidence that speech perception and speech production can at least to some extent be effectively integrated to induce error-correction mechanisms and subsequent speech motor learning in these speakers with PD and dysarthria.

KEY WORDS

dysarthria, ageing, error-detection, error-correction, sensory-motor integrational processes, speech perception, auditory perturbation, compensation, adaptation, mimicry, DIVA model

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

A Review of the Literature

"The better we understand a problem the better we are able to manage it."

(Duffy, 2005: p. 446)

1.1 Introduction

Speech perception and production are closely interlinked in the learning and maintenance of accurate speech motor control (S. M. Wilson, Saygin, Sereno, & Iacoboni, 2004). In other words, the way you hear yourself affects the way you speak. Similarly, the DIVA model of speech production proposes that a speech sound must first reliably be perceived and identified before accurate productions of that sound can be learnt (Guenther, Hampson, & Johnson, 1998). Accurate speech production can then be assumed to highly rely on (1) the ability to perceive quantitative differences in one's own speech production, and (2) the ability to effectively integrate speech perception and speech production. But what happens to these processes when hearing is compromised due to naturally occurring ageing declines? And what do we know about auditory-perceptual skills in speakers with neurological speech impairments? Currently, dysarthria management focuses predominantly on improving the neuromuscular deficits of the disorder. However, the transfer of immediate treatment effects has been limited so far (Fox, Morrison, Ramig, & Sapir, 2002; Wenke, Theodoros, & Cornwell, 2008). Thus, the link between speech perception and production becomes an important one to consider. That is, if an individual is not able to detect his or her own speech errors, then how are they going to correct for them?

The purpose of this introductory chapter is to: (1) describe dysarthria and highlight current limitations in dysarthria management, with a particular emphasis on the role of speech perception, (2) provide a review of auditory-perceptual skills in speakers with dysarthria and Parkinson's disease, (3) present a theoretical framework to highlight the significance of speech perception for accurate speech production, (4) introduce auditory perturbation as a method to study perceptual-motor learning and present a review the literature, and (5) detail the research aims of the current thesis.

1.2 Dysarthria

Dysarthria has been described as "a group of speech disorders resulting from disturbances in muscular control over the speech mechanism due to damage of the central or peripheral nervous system" (Darley, Aronson, & Brown, 1969: p. 246) and is most commonly caused by stroke, traumatic brain injury (TBI), and degenerative diseases such as Parkinson's disease (PD) (Duffy, 2005). Weakness, slowness, and incoordination of the neuromuscular system result in impairments to the speed, strength, range, timing, and/or accuracy of speech muscle movement (Duffy, 2005). The salient deviant speech characteristics of dysarthrias then impact the more global aspects of speech: respiration, phonation, resonance, articulation, rate, fluency, and prosody (Duffy, 2007). While the specific speech characteristics of dysarthria differ depending on the aetiology and severity of the underlying neurological problem, overall, individuals with dysarthria are generally likely to exhibit reductions in speech intelligibility (Hustad & Weismer, 2007).

1.2.1 Functional limitation

Reduction in speech intelligibility has commonly been described as the functional limitation of dysarthria, with intelligibility defined as "the extent to which an acoustic signal, generated by a speaker, can be correctly recovered by a listener" (Hustad & Weismer, 2007: p. 265) and functional limitation as a restriction in the performance of an action secondary to an underlying impairment as experienced by an individual (Hustad, Beukelman, & Yorkston, 1998; Yorkston, Strand, & Kennedy, 1996). Consequently, reduced intelligibility negatively affects the success, competence, and effectiveness of communication (Bunton & Weismer, 2001). When communicative effectiveness (i.e., the success of interaction and exchanging information from a speaker's point of view) is limited, a societal limitation is further indicated (Bloch & Wilkinson, 2009; Hustad, et al., 1998). Therefore, decrements in intelligibility may negatively affect both vocational and social participation, and thereby, significantly reduce quality of life for individuals with dysarthria; the effects include changes in self-identity, relationships, emotional disruptions, and feelings of stigmatisation (Dickson, Barbour, Brady, Clark, & Paton, 2008; Hustad, 2008; Margaret Walshe, Peach, & Miller, 2009; Yorkston et al., 2007). On this basis, improvement to speech intelligibility is

commonly the fundamental goal of rehabilitation in dysarthria (Ansel & Kent, 1992; Duffy, 2005; Hustad, 1999; Rosenbek & LaPointe, 1985).

1.3 Behavioural management of dysarthria

There is no one approach to treatment for dysarthria (Duffy, 2005; Hustad, 1999). Its management depends on the type and severity of the dysarthria and the needs and goals of the individual client. That said, a variety of potential treatment techniques exist in the field of dysarthria rehabilitation and management. Traditionally, these have been grouped into three primary categories; medical, prosthetic, and behavioural management. Amongst those, behavioural management is the most common form of treatment. Furthermore, research has indicated that individuals with dysarthria view speech production intervention as the best choice of treatment, thereby reflecting their desire to actively be involved in the treatment process (Yorkston, et al., 1996). To date, improved evidence within the clinical setting exists for the use of compensatory speech strategies such as the modification of rate and loudness to enhance speech intelligibility (Yorkston, Hakel, Beukelman, & Fager, 2007; Yorkston, Spencer, & Duffy, 2003). However, understanding of the long-term effects of these compensatory speech techniques on the effectiveness and efficiency of communication is still scarce (Fox, et al., 2002; Ramig et al., 2001; Solomon, McKee, & Garcia-Barry, 2001; Wenke, et al., 2008).

1.4 The link between speech perception and speech production

"Sensorimotor integration is the key to motor control."

(Brooks, 1986: p. 39, cited from Van der Merwe, 1997: p. 3)

In light of the current lack of generalisation and maintenance of immediate treatment outcomes, an essential question must be raised: *Why is it so difficult in the treatment of dysarthria to achieve functional change?* One possible explanation may be the current lack of

focus on the integrative aspects of speech production as part of a speech rehabilitation plan. So far, research and rehabilitation management are mostly concerned with understanding and treating the speech production characteristics of individuals with dysarthria. However, it has been proposed that sensory systems are invaluable for the regulation of skilled movement. For example, recent brain-imaging studies have found that speech perception and speech production are interrelated (e.g., Davis & Johnsrude, 2007; Dhanjal, Handunnetthi, Patel, & Wise, 2008; Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Grabski et al., 2013; Pulvermüller et al., 2006; Watkins & Paus, 2004; Watkins, Strafella, & Paus, 2003; S. M. Wilson, 2009; S. M. Wilson, et al., 2004; Zheng, Munhall, & Johnsrude, 2009). That is, motor regions of the cortex are active during speech perception and are therefore believed to 'mirror' the perception of an articulatory gesture (Watkins & Paus, 2004). Specifically, somatomotor representations are thought to transform acoustically encoded representations into articulatory representations (Davis & Johnsrude, 2007). A recent review of research investigating the phenomena of mirror neurons further suggests that mirror neurons are established through sensorimotor learning and able to adapt to environmental changes (Catmur, 2013). Importantly, it has been suggested that the purpose of this integrational process is to ensure ongoing consistency and accuracy in speech production (Davis & Johnsrude, 2007). Therefore, it can be assumed that the ability to correctly learn and produce a speech sound is closely linked to the ability to accurately perceive that same sound.

As a theoretical foundation to this neurological finding, researchers have called attention to the role of internal models. Specifically, it has been proposed that the internal model may serve as a control centre for skilled speech motor control (Kent, Kent, Weismer, & Duffy, 2000; Perkell et al., 2000), thereby representing the key component of successful integrational processes of speech production. The internal model has been described as "a representation in the central nervous system of the functional properties of lower-level reflex, muscular, biomechanical and acoustic components of the speech production mechanism" (Perkell et al., 1997: p. 231). In regard to speech sound production, the internal model represents learnt mappings of how auditory-perceptual targets correlate with somatosensory information (e.g., vocal tract configuration and muscle length) (Perkell, et al., 2000). In other words, internal models link movement sequences (i.e., tactile-kinaesthetic information required to produce an acoustic target) with expected outcomes (i.e., speech sounds). Thus, the importance of the integration of speech perception and speech production for accurate speech production becomes clear.

Substantial evidence in support of the theoretical construct of an internal model comes from research in normal hearing and hearing impaired speakers (Cowie & Douglas-Cowie, 1983, 1992; Economou, Tartter, Chute, & Hellman, 1992; Lane, Catania, & Stevens, 1961; Lane & Webster, 1991; Lane et al., 1997; Leder & Spitzer, 1993; Leder et al., 1987; Matthies, Svirsky, Perkell, & Lane, 1996; Oller & Eilers, 1988; Perkell, et al., 2000; Perkell et al., 2007; Perkell, Lane, Svirsky, & Webster, 1992; Perkell, et al., 1997; C. R. Smith, 1975; Svirsky, Lane, Perkell, & Wozniak, 1992; Svirsky & Tobey, 1991; Waldstein, 1990). Findings of these studies suggest that the acquisition, maintenance, and regulation of accurate and consistent speech production are highly reliant on integrational processes. In particular, these studies highlighted that the ability to self-regulate one's own speech production is greatly dependent on the availability of auditory and tactile-kinaesthetic sensory information. For example, when the auditory feedback status is changed via on-off experiments in cochlear implant users, parameters of fundamental frequency, formant frequencies, rate, and loudness among others improve in the presence of auditory feedback, and in contrast, deteriorate in the absence of auditory feedback (Economou, et al., 1992; Perkell, et al., 1992; Svirsky, et al., 1992; Svirsky & Tobey, 1991). The significance of the successful integration of auditory feedback to update internal models for consistent and accurate speech motor control is therefore highlighted.

Taken together, neuro-imaging, the internal model, and hearing research strongly suggests that speech perception and speech production are not separate, but instead, closely interlinked in processes involved in the learning and maintenance of accurate speech motor control. Consequently, it can be assumed that the following two skills are essential for the successful self-regulation of speech motor control: (1) the ability to perceive quantitative differences in one's own speech production (i.e., self-monitoring) and (2) the ability to effectively integrate speech perception and speech production.

1.4.1 Implications for speakers with dysarthria

"Lesions of sensory portions of the sensorimotor system can result in abnormal motor behaviour."

(Duffy, 2005: p. 36)

The potential significance of the integration of sensory and motor systems for appropriate and successful motor speech behaviour in speakers with dysarthria has been acknowledged. For example, Kent (2000) proposed an essential role of sensory function for the regulation of skilled speech production movements and emphasised the potential role of internal models for accurate speech production in individuals with PD. Moreover, Fox et al. (2002) proposed that the reduced ability to perceive effort in individuals with PD may negatively affect self-monitoring and self-regulatory mechanisms to adjust vocal loudness. Importantly, Maas et al. (2008) emphasised that the amount of information that is available and interpretable to the speaker with dysarthria may be crucial for the amount of learning that this person can undergo – specifically referring to the degree to which the individual may benefit from behavioural intervention. Following this line of thought, it could be assumed that if treatment aims at modifying clinical features without recognising possible impairments in the integrational processes of speech production, treatment outcomes may be limited[1]. In light of this interpretation, the key question to inform current theory and management of dysarthria may be: *Do we have a complete understanding of how the integrational processes of speech production are affected in individuals with dysarthria?*

[1] The focus of this dissertation is primarily on motor and sensory subsystems of speech production. However, the existence of cognitive deficits in ageing individuals and those with dysarthria, particularly with basal ganglia and/or cerebellar damage, is fully acknowledged (Duffy, 2005).

1.5 Perceptual abilities of individuals with dysarthria

"The extent of learning depends on the amount of information available and interpretable to the learner."

(Maas, et al., 2008: p. 12-13)

Despite the growing interest in the integrational processes of speech production and particularly the role of speech perception for accurate speech production, only two studies have observed auditory-perceptual abilities in speakers with dysarthria. That is, Walshe et al. (2008) investigated the ability of individuals with different aetiologies and types of dysarthria to perceive intelligibility differences across two more generic tasks—targeting inter-subject and intra-subject perceptual skills. In the inter-subject experiment, the ability to rate intelligibility levels in 10 independent speakers with dysarthria and one healthy control was tested. In contrast, the subsequent intra-subject experiment examined the ability of individuals with dysarthria to compare their own levels of intelligibility to other speakers with dysarthria. In total, 20 individuals with dysarthria and 30 healthy listeners (10 speech-language therapists (SLTs) and 20 naïve listeners)) completed the experiment. Individuals were asked to rate speech intelligibility (i.e., defined as "how easy or difficult it is to understand speech") using a direct magnitude estimation (DME) paradigm. The modulus sentence was taken from a speaker who achieved a score of 50% intelligible at sentence-level on the Assessment of Intelligibility of Dysarthric Speech (AssIDS). The modulus sentence was inserted after each speech sample to allow for direct comparison. In experiment two, speakers with dysarthria listened to the standard modulus again and rated their own intelligibility online relative to the modulus. Listeners rated the speech intelligibility of the 20 individuals with dysarthria relative to the 10 independent speakers with dysarthria, using the same modulus sentence as in the first experiment.

Results of the first component of the study revealed no significant differences in intelligibility ratings of independent speakers with dysarthria, indicating comparable inter-subject perceptual skills across the two groups. Similarly, results from the second experiment showed that individuals with dysarthria do not significantly differ in rating their own speech

intelligibility when compared to equivalent ratings of the other two listener groups, indicating normal intra-subject perceptual skills. It is important to acknowledge, however, that this study examined the ability of individuals with dysarthria to rate quantitative differences of speech production in other speakers as well as the ability to compare speech intelligibility of others to themselves. Unfortunately, the important question, if speakers with dysarthria can discriminate between quantitative differences in their own speech production, which relates to the ability to detect errors and update internal models to adjust articulatory placements, was not addressed.

Secondly, and perhaps more revealing, Beijer, Rietveld and Van Stiphout (2011) investigated auditory speech discrimination skills in 14 Dutch-speaking individuals who acquired dysarthria subsequent to stroke, PD, or encephalitis. Subtests included the discrimination of intensity, overall pitch, speech rate, intonation, and segmental elements of articulation. An auditory discrimination score was allocated on the basis of a forced choice 'same-different' paradigm. Importantly, hearing loss was included as a variable for correlational analysis with percentage scores, number of hits, and reaction times. Results revealed significantly lower performance for all subtests with the exception of intensity for individuals with dysarthria when compared to healthy controls. Moreover, individuals with dysarthria showed significantly fewer hits for the subtests speech rate, intonation, and segmental elements of articulation. Reaction times were significantly longer in the group of individuals with dysarthria for all subtests. When hearing loss was introduced as a covariate, however, the results were less supportive of an auditory discrimination deficit, and instead, more supportive of a hearing loss effect on auditory discrimination skills. Overall, hearing loss was significantly greater in individuals with neurological impairments compared to healthy controls. The authors therefore argued that neurological impairments may be a risk factor for hearing loss, and consequently, defects in the auditory processing and discrimination of segmental elements of articulation, intensity, overall pitch, speech rate, and intonation. While hearing loss was the main covariate to establish possible relationships with performance measures, the authors emphasised that cognitive deficits likely contributed to reduced auditory discrimination skills.

In addition to these dysarthria-specific studies, further information regarding auditory-perceptual abilities can be drawn from research of individuals with PD, even though the presence or absence of dysarthria was not noted in these studies (Adams et al., 2006; Artieda,

Pastor, Lacruz, & Obeso, 1992; Dagenais, Southwood, & Mallonee, 1999; Ho, Bradshaw, & Iansek, 2000; Ho, Bradshaw, Iansek, & Alfredson, 1999; Pell & Leonard, 2003; Troche, Troche, Berkowitz, Grossman, & Reilly, 2012; Vitale et al., 2012). For example, Artieda, Pastor, Lacruz and Obeso (1992) examined the ability of individuals with PD to recognise paired *non-speech* sensory stimuli (i.e., auditory, tactile, visual). For this purpose, temporal discrimination thresholds for the recognition of these three sensory modalities were examined in 44 individuals with PD and 20 controls matched for age and gender. Temporal discrimination thresholds were defined as the minimum time interval required between two consecutive stimuli to be perceived as two separate as opposed to one stimulus. Results revealed significant differences across the two groups for all three modalities. That is, individuals with PD demonstrated limited *non-speech* temporal discrimination skills of tactile, auditory, and visual stimuli. In addition, the correlation between these findings and PD was strengthened when the severity of the disorder increased. Similarly, Troche and colleagues (2012) investigated pure-tone discrimination skills in 12 individuals with PD and 15 age-matched controls. For this, the dimensions of frequency, amplitude and duration were examined. Results revealed that discrimination skills of individuals with PD were significantly reduced on the dimensions of amplitude and duration. In addition, a recent investigation of hearing impairment using pure-tone audiometry in 118 individuals with PD and age-matched controls found a significant hearing loss in these individuals with PD beyond that of normal ageing (Vitale, et al., 2012). Within the group of individuals with PD, hearing loss was most pronounced in older individuals and males. Importantly, all individuals with PD were reportedly unaware of their auditory deficit, highlighting why auditory deficits so far may not have received more attention in dysarthria rehabilitation.

For the following *speech-related* studies, it is perhaps noteworthy that while not specific to dysarthria, the areas of interest regarding the perceptual abilities in individuals with PD relate to common treatment targets in dysarthria rehabilitation: prosody, rate, and loudness. Pell and Leonard (2003) investigated the processing of speech-related emotional tone (i.e., the discrimination of emotional tone, identification of emotional tone, and rating of the presence of emotional features) in 21 individuals with mild PD and 21 age-matched controls. The authors reported significant deficits in decoding emotional prosody. That is, individuals with PD exhibited reduced sensitivity to differences in emotional prosody (according to a same/different paradigm). Furthermore, the ability to identify emotional prosody (e.g., happiness, sadness, anger) was significantly inferior when compared to the control group.

Since individuals with mild PD were chosen to rule-out cognitive and intelligence effects (and no correlations were found between the results and neuropsychological tests), the authors attributed the findings to reduced basal ganglia function due to its role in emotional prosody recognition.

Dagenais, Southwood and Mallonee (1999) experimentally examined the ability of 10 individuals with PD to self-monitor and adjust their speech production (i.e., speech rate, articulation rate, fluency, and intelligibility) to external changes, using delayed auditory feedback (DAF). Results revealed that individuals with PD, age-matched controls, and young adults all showed decreases across speech measures during DAF. However, while young healthy adults exhibited the highest scores across the four outcome measures, individuals with PD showed the highest declines in performance. Importantly, no group differences were observed under normal conditions. The authors interpreted these findings as a specific difficulty in speakers with PD to shift attentional focus between speech perception and speech production tasks, and further, to integrate sensory and motor information. While no specific cognitive assessment was administered to validate this hypothesis, the authors suggested that limited resources in shifting attentional focus may have caused both temporal processing and speech accuracy declines.

Perhaps the most conclusive data regarding auditory self-perception deficits in individuals with PD derive from studies of the perception of loudness. Ho et al. (1999) analysed vocal responses to implicit and explicit cues as measured by the Lombard effect – a largely involuntary phenomenon – in 12 individuals with PD and hypophonia. While conversing, individuals with PD and controls were concurrently exposed to two different stimuli via headphones; (1) their own speech increasing in 5 decibel (dB) steps, and (2) pink noise increasing in 5 dB steps. According to the Lombard effect, the researchers expected to see an increase in loudness when speaking in noise and a decrease in loudness when hearing one's own voice amplified. Findings of the study revealed that the group with PD did not naturally increase or decrease their speech volume across three of the four experimental conditions; however, as expected, the controls did. The authors interpreted this as a resistance to the influence of implicit cues. Specifically, a perceptual mismatch between the internal representation of volume and the level of loudness produced in individuals with PD was proposed.

Similarly, Adams et al., (2006) studied speech intensity responses to conversation in multi-talker noise (50, 55, 60, 65, 70 dB) in 10 individuals with PD and 10 age-matched controls. In contrast to Ho et al. (1999), however, results revealed that while individuals with PD spoke with significantly lower speech intensity in all noise conditions compared to the normal controls, they significantly increased their speech intensity when speaking in background multi-talker noise. Methodological differences may serve as an explanation for these contradictory findings. Firstly, Ho and her colleagues used headphones while Adams and his colleagues used a loudspeaker to expose individuals to the background noise. Consequently, individuals in the latter study were able to listen to both their own voice and the background noise whereas individuals in the former study had reduced capacity to self-monitor their loudness level. Secondly, Ho and her colleagues increased pink noise in 5 dB increments during the reading task, starting from the individual's hearing threshold. However, only two levels (i.e., 10 and 25 dB increase from the individual's hearing threshold) were tested during conversational speech. In contrast, Adams and his colleagues used fixed noise intensities. Additionally, multi-talker noise may have allowed speakers with PD to use F1 as a cueing influence when comparing and adjusting levels of loudness. It is possible, therefore, that the different findings across the two studies are due to these methodological differences.

Based on the findings of the first study, Ho et al. (2000) then specifically investigated the perception of volume in individuals with PD. This study comprised two experiments: (1) examination of immediate perception skills and (2) investigation of playback or external perception skills. For experiment one, 15 individuals with PD were required to indicate their perceived loudness level immediately subsequent to the production tasks of reading and conversing at different loudness levels. This was achieved by presenting individuals with their own speech production immediately after the production task via headphones and asking them to adjust the volume level of a volume control knob to indicate how loud they had just read or conversed. In experiment two, the same 15 individuals were presented with their own speech production stimuli from experiment 1 and asked to adjust the volume level to the loudness level they perceived themselves at while listening to the recordings. Individuals were naïve to the loudness adjustment procedure to ensure that results would not be falsified by facilitation of introspective attention. Results of the two experiments indicated that individuals with PD: (1) consistently overestimated the loudness of their speech production in both conditions (immediate and playback) to a significantly greater degree than healthy

controls, (2) significantly and consistently perceived themselves to be louder than healthy controls, and most importantly, (3) the difference between the actual level of loudness and the perceived level of loudness was significantly greater in individuals with PD for (i) both the immediate and playback condition in the reading task, and (ii) the immediate condition in the conversational task.

While these findings provide clear evidence for a sensory deficit in individuals with PD and hypophonia, Ho and her colleagues raised an essential question regarding the interpretation of these data: Is the mismatch between speech perception and speech production an effect of auditory or a tactile-kinaesthetic deficit? In answering this question, an important variable must be taken into consideration: the definition of loudness. That is, if the assumption of Ladefoged and McKinney (1963) – by which the loudness of a sound is judged by the effort one would exert to produce that sound – is accurate, it could be proposed that the mismatch in perception and production is due to a tactile-kinaesthetic rather than an auditory deficit. Following this line of thought, Ho and her colleagues suggested that the mismatch between perception and production in individuals with PD may be due to an exaggerated self-perception of effort as opposed to a reduced auditory self-perception of loudness.

This assumption is strongly supported by studies of the efficacy of the Lee Silverman Voice Treatment® (LSVT) for the treatment of speech disorders associated with PD (Fox, et al., 2002; Fox & Ramig, 1997; Ramig, et al., 2001). In particular, a mismatch between perceived vocal effort and actual vocal output in individuals with PD has been clinically observed during administrations of the LSVT® (Fox, et al., 2002). Interestingly, numerous research studies in the field of dysarthria have cited reduced oral tactile or kinaesthetic sensation as an explanation for aberrations in speech production (LSVT RESEARCH GROUP; Cheng, Goozee, & Murdoch, 2005; Goozee, Murdoch, & Theodoros, 1999; Kuruvilla, Murdoch, & Goozee, 2008). Furthermore, Schneider, Diamond and Markham (1986) suggested that orofacial kinetic sensitivity is reduced in individuals with PD. Yet, research examining the intra-oral sensation in speakers with dysarthria is scarce.

Coming back to the question about the origin of the mismatch between speech perception and speech production, it is important to note that data on auditory and tactile-kinaesthetic sensory processing deficits are largely based on the same body of research, differing only in the interpretation of the results (i.e., reduced auditory self-perception vs. exaggerated self-

perception of effort). Whether a client presents with reduced auditory or exaggerated tactile-kinaesthetic self-perception, however, has significant implications for the rehabilitation management of individuals with dysarthria. Thus, the importance of examining and differentiating between possible auditory and tactile-kinaesthetic sensory deficits in speakers with dysarthria is highlighted.

1.5.1 Summary

Most research in the field of dysarthria has focused on understanding the speech production characteristics of dysarthria. Consequently, a good source of information exists regarding the deviant speech characteristics of dysarthria and speech production techniques to help improve speech intelligibility. In contrast, it is far from clear how auditory-perceptual abilities of individuals with dysarthria are affected, if at all. Similarly, it is currently unknown if, and how much, speakers with dysarthria are aware of their breakdowns in speech intelligibility related parameters such as prosody, rate, and loudness. Preliminary findings lend support to the notion of an auditory-perceptual deficit in individuals with dysarthria and those with PD. And despite equivocal findings, overall, more data on auditory perception lends support to the notion of a mismatch between speech perception and speech production, at least in individuals with PD. Specific deficits observed include (1) high frequency hearing loss, (2) difficulty with auditory and temporal discrimination, (3) difficulty decoding, differentiating, and identifying emotional prosody, (4) difficulty in shifting attentional focus between self-monitoring and speech production and integrating the two to ensure accurate speech production, (5) resistance to implicit cueing, and (6) a mismatch between the internal representation of loudness and the actual loudness level produced.

Taken together, these data provide some indication for a possible deficit in the sensorimotor integrational processes of speech production in individuals with neurological speech impairments. However, a more comprehensive understanding of the auditory-perceptual abilities of individuals with dysarthria and their relationship to speech output measures is required before such assumptions can be confirmed.

1.6 Ageing effects on integrational processes of speech production

One particular area of research that has received little attention in the field of neurological speech and language impairments so far – but may well contribute to our understanding of potential difficulties in these populations – is that of ageing effects on the integrational processes of speech production. A vast body of research suggests that the integrational processes of speech production may be compromised by age (e.g., Crow & Ship, 1996; Goozee, Stephenson, Murdoch, Darnell, & Lapointe, 2005; Sommers, 2008). That is, ageing is typically associated with physiological, anatomical, and neurological changes in the human body. Loss of motor units (e.g., Brown, 1972), reduced muscle strength and increased muscle fatigability (e.g., Campbell, McComas, & Petito, 1973; Lexell & Taylor, 1991; Wohlert & Smith, 1998), declines and inconsistency in firing rate (e.g., Luschei, Ramig, Baker, & Smith, 1999), increased latency of reflexes (e.g., A. Smith, Weber, Newton, & Denny, 1991), and dopaminergic deficiencies in the basal ganglia (Morgan & Finch, 1988) have been considered as explanations for age-related declines in sensorimotor functions (e.g., reduced velocity and greater variability of movement). Dopaminergic deficiencies in the basal ganglia, in particular, are hypothesised to restrict the modulation of auditory feedback for adjustments in vocalisation, movement selection, and motor learning (Duffy, 2005). While little is known about how any of these age-related changes influence the articulators (Goozee, et al., 2005), there is some indication that tongue mobility and functioning may be reduced in the ageing population (e.g., Ballard, Robin, Woodworth, & Zimba, 2001; Goozee, et al., 2005; Price & Darvell, 1981). Moreover, some declines in tactile and kinaesthetic sensory systems have been reported (e.g., Breustedt, 1983; Ostericher & Hawk, 1982; Sonies, 1991; Wohlert, 1996). However, it seems that speech production is largely preserved from ageing effects, with significant declines only occurring in very old individuals (e.g., Crow & Ship, 1996; Liss, Weismer, & Rosenbek, 1990).

Brain-imaging research and post-mortem studies provide further evidence for a possible effect of ageing on auditory feedback mechanisms. Brain shrinkage and loss of brain structures within prefrontal, parietal, and temporal brain regions are common phenomena in the ageing population (Head, Raz, Gunning-Dixon, Williamson, & Acker, 2002; Pichora-Fuller & Singh, 2006; Raz, 2000). This decline may be detrimental to accurate speech production since self-monitoring and error-detection mechanisms as part of the sensory-motor integrational network are located in Broca's area in the frontal motor cortex, the

posterior parietal lobe, and the superior posterior temporal lobe (Hickok, Okada, & Serences, 2009; Pulvermüller, et al., 2006; Watkins & Paus, 2004; S. M. Wilson & Iacoboni, 2006; S. M. Wilson, et al., 2004; Zheng, et al., 2009).

In regard to declines in speech perception, it is well established that with age, hearing sensitivity deteriorates due to sensorineural hearing loss (SNHL) (e.g., Kiessling et al., 2003) as does auditory processing (e.g., Gordon-Salant, 2008, November; Kiessling, et al., 2003). In addition, recent research has demonstrated that elderly adults exhibit difficulty perceiving frequency, intensity, and temporal differences in *non-speech* stimuli (Fitzgibbons & Gordon-Salant, 1998, 2004; Gordon-Salant, 2008, November; Kiessling, et al., 2003; Sommers, 2008). Ross, Tremblay and Picton (2007) found that older adults exhibited difficulty detecting acoustic changes in sinusoidal amplitude modulated tones. Importantly, *speech-related* processing of information is also slowed (Federmeier, Van Petten, Schwartz, & Kutas, 2003; Gordon-Salant, Yeni-Komshian, Fitzgibbons, & Barrett, 2006; Pichora-Fuller, 2003a, 2003b; Tun, Wingfield, Stine, & Mecsas, 1992).

The combination of delayed general processing and age-related auditory and tactile-kinaesthetic sensory declines may very well negatively affect feedback mechanisms to update internal models, thereby restricting internal cueing strategies to adjust articulatory parameter settings. Hence, it is possible that sensorimotor adaptation is computationally more demanding for older adults compared to younger adults (Bock & Schneider, 2002). Unfortunately, however, research addressing these issues is lacking (Liss, et al., 1990).

1.6.1 Conclusion

Results from individuals with dysarthria and those with PD as outlined in the previous sections make it reasonable to assume that perceptual and internal cueing deficits may negatively affect the ability to self-regulate one's speech output. Evidence from ageing individuals further supports this notion. So far, this theory has only been informed by the interpretation of clinical and research findings. However, a theoretical speech production model incorporating both speech perception and speech production for older individuals and those with dysarthria is currently lacking.

One theoretical framework that highlights the importance of speech perception for accurate speech production and may be of crucial significance to the study of dysarthria is the DIVA model. Although a number of speech production theories have been proposed over the last decades, to date, the DIVA model is arguably the most thoroughly defined and tested neural network model of speech production. DIVA has evolved over nearly two decades now and includes motor, sensory, and cognitive subsystems. Perhaps most significantly, the DIVA model is unique in that it is able to explain various aspects of speech production and that it remains adaptable and changeable to best capture the complexity of speech production mechanisms in healthy individuals and those with impaired speech motor control. It is because of these properties that the DIVA model was chosen as a theoretical foundation to further investigate the link between speech perception and speech production in individuals with dysarthria. However, no computational modelling was undertaken for the current study.

1.7 The DIVA model – A theoretical model of speech production

The DIVA model was firstly invented with the aim to provide a theoretical framework that would be able to explain speech sound acquisition in infants (Guenther, 1995b). More specifically, the aim was to build a computational model that would shed light onto the neural processes involved in speech perception, speech production and the interactions between the two (Guenther, et al., 1998). The key philosophy of the DIVA model is that the learning of a speech sound is based on an auditory-perceptual teaching signal. This philosophy derives from the assumption that the timeframes during which an infant learns how to perceive and how to produce a speech sound, while overlapping, also differ considerably. That is, a target sound must first reliably be perceived and identified before learning of consistent and accurate production can follow (Guenther, 1995b). The DIVA model derives its name from this interaction, which is **D**irections (in auditory planning space) **I**nto **V**elocities of **A**rticulators (Guenther, et al., 1998).

The DIVA model accounts for a number speech acoustic, kinematic, EMG, and fMRI data. Specifically, the DIVA model is currently able to explain (a) speech sound acquisition (b) speech production accuracy in the face of developmental changes in the size and shape of the speech articulators or external alterations, (c) articulatory variability, (d) motor equivalence, (e) coarticulation, (f) economy of effort, (g) rate effects, (h) interactions between

cortical, subcortical, and cerebellar regions within the human brain for accurate integrational processes of speech production, (i) the role of auditory feedback in speech production for healthy individuals as well as individuals with hearing loss and cochlear implant users, (j) underlying causes to AOS and stuttering, and (k) responses to perturbation (Brunner et al., 2011; Civier & Guenther, 2005, June; Civier, Tasko, & Guenther, 2010; Golfinopoulos et al., 2011; Golfinopoulos, Tourville, & Guenther, 2010; Guenther, 1995b; Guenther, Ghosh, & Tourville, 2006; Guenther, et al., 1998; Max, Guenther, Gracco, Ghosh, & Wallace, 2004; Perkell, 2012; Terband, Maassen, Guenther, & Brumberg, 2009; Tourville, Reilly, & Guenther, 2008).

In essence, the DIVA model is based on three main components: (1) an auditory reference frame that guides speech sound acquisition and maintenance, (2) a feedback control subsystem, and (3) a feedforward control subsystem. Feedforward and feedback control subsystems are not independently activated during speech production. Instead, the generation of speech motor commands requires integrational networking between the two subsystems to ensure error-detection, and therefore, accurate speech production. In an attempt to elucidate the conceptual relevance of the DIVA model for speech perception and speech production, and further, for a theoretical investigation of the current lack of generalisation and maintenance in the treatment of dysarthria, speech sound acquisition in infants will be explained on the basis of the theoretical framework embedded in the DIVA model. For a more technical description of the DIVA model, please refer to Appendix A.

Speech sound acquisition in infants is driven by auditory feedback. During the first babbling attempts, auditory-perceptual (and perhaps to some degree also visual) feedback allows the infant to establish relationships between tactile and proprioceptive information and their auditory consequences. For example, when the lips are closed and the vocal folds are vibrating, a /m/ sound is produced. This is a phase of discovery of the articulators. Therefore, the phonemes produced in this babbling phase are universal and not yet representative of an infant's native language inventory. Once an infant is able to reliably perceive and identify an auditory-perceptual target in a speaker (e.g., 'Mom just made a /m/-sound'), this auditory-perceptual target becomes a teaching signal. That is, the infant now compares between the speech production of his/her environment and his/her own speech production, thereby trying to establish a link between the acoustic target and the tactile and proprioceptive information required to produce that sound.

This process also includes establishing speech motor commands that correspond with individual articulators and related muscle groups to produce the auditory target. Many "trials and errors" follow during which the infant learns to use auditory feedback to detect articulatory attempts that do not match the target sound and to make appropriate corrections to better fit the auditory target. Moreover, the infant learns that phonemes are planned in auditory-perceptual space, and therefore, a number of different constriction configurations produce the same auditory-perceptual target sound. That is, as long as a speech sound is realised within the auditory-perceptual target region, the model allows for flexibility in how that target is reached. Consequently, the relationship between the auditory target and its articulatory realisation is one-to-many, which is one of the key differences to construction-based theories of speech production. According to DIVA, therefore, phonemes are produced in multi-dimensional auditory and somatosensory space.

Once the link between an auditory-perceptual target and the tactile and proprioceptive information required to produce a sound is reliably established, this information is stored in an internal representation (i.e., the internal model or forward model). This signifies an essential milestone in speech sound acquisition as from this point onward, the infant can use that information for feedforward commands, and therefore, is no longer reliant on auditory feedback to drive accurate speech production. Hence, the role of feedforward commands is to ensure reliable and consistent speech sound production and to allow for faster processing as required in normal speech. However, auditory and somatosensory feedback control subsystems are available to the speech system at all times and are particularly activated when auditory or somatosensory feedback (i.e., actual sensory input) differs from the intended target region (i.e., expected input). When a mismatch between the expected auditory target and the actual auditory feedback is detected, auditory-sensory feedback is used to guide the corrective articulatory response required to ensure that intelligibility is maintained. Importantly, that information is then stored in the internal model to ensure accurate productions on future attempts.

Examples of when integrative processes of speech perception and speech production are activated to ensure error-detection and error-correction include developmental changes in the articulators[2] and damage to the articulators. That is, developmental changes in the size and

[2] Changes in the morphology of the vocal tract take place until the age of 20, with some continuous changes in specific structures and cavities (Munhall, MacDonald, Byrne, & Johnsrude, 2009).

shape of the articulators cause ongoing changes in the mappings between auditory-perceptual targets and their related somatosensory targets. Thus, close networking between feedback and feedforward control subsystems is required to account for frequent error signals and make the necessary corrections to ensure speech intelligibility. Perhaps more relevant to the immediate effects of dysarthria, feedforward commands are also updated secondary to internal changes of the vocal tract configurations due to damage or changes to the articulators. That is, damage to the articulators also changes the mappings between auditory-perceptual targets and their related somatosensory targets, for example, in the case of a partial tongue resection. The reason why speech intelligibility can be maintained in these two examples refers back to the idea that sounds are planned in auditory-perceptual space, and thus, the relationship between an auditory target and its somatosensory realisation is one-to-many.

In conclusion, the DIVA model highlights the significance of the following abilities for the acquisition and maintenance of accurate speech production: (1) error-detection, (2) error-correction based on the integration of sensory information and subsequent articulatory adjustments, and (3) storage of updated auditory and somatosensory information in the internal model to ensure accurate speech production on future attempts as required for motor learning. Thus, the concept of the DIVA model emphasises that perceptual learning and speech motor learning are not separate, but instead, process together. That said, the DIVA model is based on a healthy and fully functioning speech system. To date, it is not clear how the DIVA model applies to a system that has sensorimotor problems as indicated by factors such as ageing or dysarthria. However, due to the distinct properties outlined above, the theoretical framework of the DIVA model provides a unique foundation for testing specific hypotheses regarding the complexity of speech motor control. That is, a systematic experimentation programme can be developed based on the model's components and their hypothesised functional connectivity. One particular experimental design that is compatible with DIVA-based hypotheses, and therefore, lends itself to the study of sensory and motor systems and their connectivity is that of a perturbation paradigm.

1.8 A methodological proposal of how to test auditory-perceptual abilities in individuals with dysarthria

The perturbation paradigm has been used in a variety of sensory contexts. For example, researchers have investigated sensorimotor learning based on visual (e.g., Field, Shipley, & Cunningham, 1999; Ingram et al., 2000; Welch, 1986), kinaesthetic (e.g., Abbs & Gracco, 1984; Baum & McFarland, 1997; Golfinopoulos, et al., 2011; Jones & Munhall, 2003; Shaiman, 1989; Tremblay, Shiller, & Ostry, 2003; Vennila & Aruin, 2011), and auditory (e.g., Houde & Jordan, 1998; Houde & Jordan, 2002; Keough & Jones, 2009; Shiller, Sato, Gracco, & Baum, 2009; Villacorta, Perkell, & Guenther, 2007) feedback mechanisms. For the purpose of comparing responses to perturbation with DIVA-based hypotheses, however, only somatosensory and auditory perturbations are of interest and will therefore be described in more detail as follows.

1.8.1 Somatosensory perturbation

The motivation of somatosensory perturbation has been two-folded, investigating the importance of both, auditory and somatosensory feedback for accurate speech production. Results of somatosensory perturbation studies that applied an artificial palate (e.g., Baum & McFarland, 1997, 2000; McAuliffe, Robb, & Murdoch, 2007) or a prosthesis (e.g., Jones & Munhall, 2003) revealed that in the absence of somatosensory feedback, accurate speech production is guided by auditory feedback. That is, speakers altered their speech production pattern to ensure speech intelligibility even when somatosensory information was not available. Moreover, these findings have been replicated for bite block perturbation in individuals with cochlear implants, providing further support of the significance of auditory feedback for accurate speech production (Lane et al., 2005).

For example, Brunner and Hoole (2012) investigated articulatory variability in the production of the sibilants /s/ and /ʃ/ under somatosensory perturbation using two types of artificial palates. That is, a palate prosthesis that contained an alveolar ridge, thereby providing a landmark for speech production, and a central palate prosthesis that filled the complete palate, thereby prohibiting the use of landmarks for speech production. Individuals were asked to wear the artificial palate for two weeks and adapt to it as best as possible. It was hypothesised that the provision of a landmark in alveolar palate participants would result

in less articulatory variability compared to the central palate participants who had no somatosensory feedback to guide speech production. However, after the two-week adaptation period, speech production measures indicated that despite different amounts of somatosensory information available to the speakers, articulatory variability was in fact comparable across these two conditions. Thus, these results were taken as evidence that accurate speech production is primarily based on auditory feedback mechanisms and less so on somatosensory feedback mechanisms.

In contrast, evidence also exists in support of a significant role of somatosensory feedback in speech production. For example, mechanically applied perturbations to the jaw revealed compensatory responses even when the perturbation did not affect the acoustic properties of the speech motor command (e.g., Nasir & Ostry, 2006; Tremblay, et al., 2003). Furthermore, Nasir and Ostry (2008) examined responses to jaw perturbations in cochlear implant users when auditory feedback was turned off. Results showed that even in the absence of auditory feedback, cochlear implant users were able to compensate for the jaw perturbation. Moreover, response patterns of cochlear implant users were comparable to those of healthy controls. These findings were taken as proof that compensatory responses are not dependent on the availability of auditory feedback. Hence, these studies suggest that somatosensory targets are also important in accurate speech production.

Perhaps the most revealing results regarding the interaction between auditory and somatosensory feedback systems in maintaining speech intelligibility stem from an investigation of the effect of palatal perturbation in the face of masked auditory and/or somatosensory feedback. That is, Honda, Fujino and Kaburagi (2002) and Honda and Murano (2003) tested responses to palatal perturbation (via an inflatable balloon) under four succinct feedback conditions: (1) normal, (2) auditory feedback blocked with masking noise, (3) tactile feedback blocked with locally applied anaesthesia to the tongue tip, and (4) joint masking of tactile and auditory feedback. Results revealed that compensatory responses were best in the normal condition and during auditory masking, diminished under tactile masking and worst when both feedback modalities were blocked. Consequently, it can firstly be assumed that speech sounds have both auditory and somatosensory targets, and secondly, that auditory and somatosensory feedback are both involved in the maintenance of accurate speech production. A similar finding comes from auditory perturbation research with locally applied anaesthesia. That is, Larson, Altman, Liu and Hain (2008) found that when the vocal

folds were deprived of somatosensory feedback due to administration of local anaesthetic, compensatory responses to auditory perturbation were significantly greater compared to when somatosensory feedback was normal. Importantly, however, while these findings highlight that both auditory and somatosensory feedback systems guide corrective responses to auditory perturbation, the interaction and precedence of the two feedback systems in maintaining accurate speech production in the face of externally applied perturbation is less clear.

In regard to older adults, only one study investigated responses to somatosensory perturbation so far. That is, de Miranda Marzulollo and her colleagues (2010) investigated the effects of lip displacements in 10 healthy young men and compared the results to 10 older men with a mean age of 60 years (SD = 9). Results revealed that compensatory responses were greater and compensatory strategies different for young adults when compared to older adults. Furthermore, the displacements resulting from lip perturbation were larger in older adults compared to young adults. The authors hypothesised that this may be due to weaker perioral reflex responses in older adults and altered anticipatory strategies in overcoming the perturbation due to ageing. No studies have been conducted with individuals with neurological speech impairments so far.

The outline of studies investigating somatosensory perturbation as given above shows that the interpretation of results can vary significantly, depending on parameters such as the experimental design and the selection of dependent variables. In addition, it is likely that responses are not comparable to naturally occurring corrective speech mechanisms due to the rather invasive experimental set-up. However, when choosing a study design to test the perception-production link, it is important to ensure that the experimental test procedure does not interfere with naturally occurring processes involved in maintaining speech production accuracy. An auditory perturbation paradigm therefore appears to be the more preferred and effective choice of method to test the perception-production link; in particular, because it is a non-invasive procedure and allows for direct observation of how auditory feedback affects speech motor control and motor learning (Purcell & Munhall, 2006a; Shiller, et al., 2009; Sivasankar, Bauer, Babu, & Larson, 2005).

1.8.2 Auditory perturbation

Although a relatively young field of research, the auditory perturbation paradigm has been applied to three acoustic speech phenomena so far: (1) fundamental frequency (F0) (e.g., Behroozmand, Korzyukov, & Larson, 2011, 2012; Behroozmand, Korzyukov, Sattler, & Larson, 2012; Burnett, Freedland, Larson, & Hain, 1998; Chen, Liu, Xu, & Larson, 2007; Elman, 1981; Hawco & Jones, 2010; Jones & Munhall, 2000; Keough & Jones, 2009; Korzyukov, Karvelis, Behroozmand, & Larson, 2012; H. Liu, Meshman, Behroozmand, & Larson, 2011; H. Liu, Xu, & Larson, 2009; Parkinson et al., 2012; Scheerer, Behich, Liu, & Jones, 2013; Scheerer & Jones, 2012; Sivasankar, et al., 2005), (2) loudness (e.g., Bauer, Mittal, Larson, & Hain, 2006; Heinks-Maldonado & Houde, 2005), and (3) first and second formant frequencies (F1/F2) (e.g., Katseff & Houde, 2008; Katseff, Houde, & Johnson, 2010; Katseff, Houde, & Johnson, 2012; E. N. MacDonald, Goldberg, & Munhall, 2010; E. N. MacDonald, Purcell, & Munhall, 2011; Munhall, et al., 2009; Pile, Dajani, Purcell, & Munhall, 2007, February; Purcell & Munhall, 2006a, 2006b, 2008; Rochet-Capellan & Ostry, 2011a; Shiller, et al., 2009; Shum, Shiller, Baum, & Gracco, 2011; Tourville, et al., 2008; Villacorta, et al., 2007).

In essence, auditory perturbation studies aim at changing a speaker`s auditory feedback in a different direction from what is expected, based on the intended speech signal. This is usually done by implementing four distinct phases; (1) start phase – where auditory feedback is amplified but otherwise unaltered, (2) ramp phase – where auditory perturbation is introduced in gradual increments, (3) stay phase – where auditory perturbation remains at its maximum, and (4) end phase – where auditory feedback is again amplified but otherwise unaltered. As a consequence of auditory perturbation, speakers commonly shift their production in the direction opposite to the perturbation. For example, when the first formant frequency (F1) of the word “head” is raised to perceptually approximate the word “had”, a speaker typically lowers his or her F1. In American and Canadian English, this lowering of F1 perceptually approaches the word “hid”. This response pattern is referred to as “compensation” and offers the unique opportunity to test the ability of an individual to detect and correct for speech errors in real-time.

When individuals are exposed to auditory perturbation for a prolonged time, compensatory responses persist even when the perturbation is removed or when unperturbed feedback is blocked by masking noise. This is referred to as “adaptation” or “after-effect”. Persistent

changes are then interpreted as motor-perceptual learning. That is, adjustments to the link between a target sound and the proprioceptive information required to produce that particular sound have been made. Thus, the internal model of that particular sound has been updated and the feedforward command for that sound modified. The newly established link between auditory perception and motor output as a result of continuous perturbation is also referred to as remapping or recalibration. Accordingly, the remapping to original articulatory settings that occurs with prolonged availability of normal auditory feedback is called 'de-adaptation'.

Research examining compensatory responses to formant frequency perturbation dates back to the studies of Houde and Jordan (1998; 2002), who investigated the effect of concomitant F0, F1, and F2 perturbation on speech motor control in whispered speech. Their main finding of compensation has since been replicated for conversational speech; particularly, for American English (AE) (Houde & Jordan, 1998; Houde & Jordan, 2002; Katseff & Houde, 2008; Katseff, et al., 2010; Katseff, et al., 2012; Tourville, et al., 2008; Villacorta, et al., 2007), Canadian English (CE) (E. N. MacDonald, et al., 2010; E. N. MacDonald, et al., 2011; Munhall, et al., 2009; Pile, et al., 2007, February; Purcell & Munhall, 2006a, 2006b, 2008; Rochet-Capellan & Ostry, 2011b; Shiller, et al., 2009; Shum, et al., 2011), Mandarin (i.e., Akagi, Dang, Lu, & Uchiyamada, 2006; Cai, Boucek, Ghosh, Guenther, & Perkell, 2008; Cai, Ghosh, Guenther, & Perkell, 2010; P. Liu, Chen, Jones, Huang, & Liu, 2011), Japanese (Akagi, et al., 2006; Mitsuya, MacDonald, Purcell, & Munhall, 2011), French (Mitsuya, Samson, Menard, & Munhall, 2013), and CE speaking children (E. N. MacDonald, Johnson, Forsythe, Plante, & Munhall, 2012; Shiller, Gracco, & Rvachew, 2010).

While a lot less information is available regarding adaptation, evidence to support adaptation has been replicated for American English (AE) (Houde & Jordan, 1998; Houde & Jordan, 2002; Katseff & Houde, 2008; Katseff, et al., 2010; Katseff, et al., 2012; Tourville, et al., 2008; Villacorta, et al., 2007), Canadian English (CE) (E. N. MacDonald, et al., 2010; E. N. MacDonald, et al., 2011; Munhall, et al., 2009; Pile, et al., 2007, February; Purcell & Munhall, 2006a, 2006b, 2008; Rochet-Capellan & Ostry, 2011b; Shiller, et al., 2009; Shum, et al., 2011), Mandarin (i.e., Akagi, et al., 2006; Cai, et al., 2008; Cai, et al., 2010), Japanese (Akagi, et al., 2006; Mitsuya, et al., 2011), and CE speaking children (E. N. MacDonald, et al., 2012; Shiller, et al., 2010).

1.8.3 Building the bridge: How DIVA informs findings of auditory perturbation

The mechanisms of "compensation" and "adaptation" to auditory perturbation can also be described and explained within the theoretical framework of the DIVA model (Guenther, 2006; Guenther, et al., 2006). Firstly, "compensation" is mainly driven by the *Auditory Feedback Control Subsystem*. That is, auditory perturbation manipulates the expected auditory target. Consequently, what is being heard through the headphones differs from what the individual speaker expects to hear. When an individual detects the mismatch between the intended auditory-perceptual target and the actual auditory feedback of that sound production, *Auditory Error Maps* are activated. This activation results in the initiation of corrective speech motor commands to match the intended auditory target and is referred to as the compensatory response in perturbation studies.

Secondly, "adaptation" relies on both the *Feedback* and the *Feedforward Control Subsystem*. That is, the information embedded in the articulatory adjustments to correct for the auditory feedback error is incorporated in the following speech motor command. As a result, feedforward commands are continuously updated as corrective speech motor commands are initiated by the activation of *Auditory Error Maps*. This is also known as remapping or recalibration of the internal model or feedforward model. Thus, when auditory perturbation is removed or blocked by noise, the persistence of altered speech motor commands is referred to as "adaptation".

"De-adaptation" is driven by both the *Auditory Feedback Control Subsystem* and the *Feedforward Control Subsystem* and involves very similar processes as those described above. That is, the removal of the auditory perturbation again results in a mismatch between the expected auditory feedback and the actual auditory input. Hence, *Auditory Error Maps* activate corrective speech motor commands to match the *Speech Sound Map.* Continuous updating of feedforward commands establishes new mappings between reference frames. The timeframe between "adaptation" and complete remapping of the internal model, resulting in consistent speech motor commands that are comparable to baseline productions, is referred to as "de-adaptation". For a more detailed description of the components of the DIVA model, please refer to Appendix A.

1.8.4 Factors influencing compensation and adaptation

A number of studies have investigated factors influencing compensation in healthy young AE and CE speakers, using formant frequency perturbation. However, much less is known about factors influencing adaptation to auditory perturbation. Importantly, a number of limitations have to be taken into account before any conclusions can be drawn. Firstly, research studies so far have only looked at relatively small sample sizes (~ 20 participants), and for the most part, healthy young speakers. Moreover, different research designs have been employed to expand current knowledge, which makes it difficult to compare results. For example, since the early studies of Houde and Jordan, researchers have employed different designs of perturbation (i.e., F1 perturbation only versus concomitant F2 perturbation), different procedures (i.e., epochs versus phases), different normalization procedures, different perturbation measures (i.e., different quantities of Hz, perts, mels, semitones), different amounts of masking noise added to the acoustic signal, and different amounts of trials overall. With this in mind, a summary of common factors and findings regarding compensation and adaptation in AE and CE speakers follows.

1.8.4.1 Predictors of compensation to auditory vowel perturbation

1. Perturbation of auditory feedback typically results in partial compensation.

Partial compensation is a phenomenon that is shared among studies investigating formant frequency perturbation (e.g., Houde & Jordan, 2002; Katseff & Houde, 2008; Katseff, et al., 2012; E. N. MacDonald, et al., 2010; E. N. MacDonald, et al., 2011; Munhall, et al., 2009; Pile, et al., 2007, February; Purcell & Munhall, 2006a, 2006b, 2008; Shiller, et al., 2009; Tourville, et al., 2008; Villacorta, et al., 2007), with a recent meta-analysis of 116 female CE speakers reporting mean compensation values of 53 Hz (E. N. MacDonald, et al., 2011), which is equivalent to 26.5% of the perturbation magnitude.

2. The degree of compensation is related to the magnitude of the perturbation.

Katseff and colleagues (2012) specifically investigated the effect of perturbation magnitude on compensatory response. For this purpose, speakers were exposed to an upward shift in F1 of 250 Hz, which was reached in five distinct perturbation steps. Results showed that compensation was complete at 50 Hz, nearly complete at 100 Hz, and showed a trend

towards complete compensation up until 150 Hz. However, compensation ceased to increase for perturbations greater than 150 Hz.

Similarly, MacDonald et al. (2010) examined the effect of three distinct perturbation amplitudes on compensation. Step one increased F1 by 50 Hz and decreased F2 by 75 Hz. Step two increased F1 by 100 Hz and decreased F2 by 125 Hz. Finally, step three increased F1 by 200 Hz and decreased F2 by 250 Hz. At each step, perturbation was held constant for 50 trials. In a second experiment, perturbation of F1 was ramped upward in 4 Hz increments over a number of 110 trials until a maximum perturbation of 360 Hz was reached. Simultaneously to the F1 shift, F2 was continuously ramped downward in 5 Hz increments to meet the maximum perturbation of 450 Hz. Results were similar to Katseff and colleagues, showing a linear response pattern up to a shift in F1 of 200 Hz/F2 250 Hz. However, when perturbation further increased, the compensatory response pattern became non-linear. Differences in magnitude where the linearity of compensatory response patterns ceases across the two studies (i.e., 150 Hz vs. 200 Hz) may stem from the specific type of formant frequency shift employed (i.e., F1 upward only vs. F1 upward and F2 downward).

3. Compensation is influenced by the somatosensory value of the perturbed vowel.

To investigate more specifically into the interaction between auditory and somatosensory feedback systems during auditory vowel perturbation, Purcell and Munhall (2008) examined the effect of available somatosensory information in different vowels (i.e., the robustness of F1 as a cueing strategy) on compensatory responses. Results revealed that compensation was significantly larger for /ɛ/ compared to /u/ and /i/. These findings suggest that compensation may vary across vowels depending on the contribution of auditory and somatosensory information to a particular vowel. For example, a stronger somatosensory contribution can be expected for the production of /i/ due to the approximation of tongue and palate. Consequently, the authors assumed that the weight of proprioceptive feedback (i.e., the robustness of F1 as a cueing strategy) affects the amount of compensation in different vowel contexts. Likewise, research investigating the effects of magnitude on compensation suggests that the linearity of compensation ceases for perturbations larger than 150 Hz/ 200 Hz because of the information available from somatosensory feedback. In other words, compensation ceases to increase when the productions fall outside the acceptable target region of the perturbed sound (Katseff, et al., 2012; E. N. MacDonald, et al., 2010).

4. Compensation is not affected by the precision of speech motor control.

Perkell and his colleagues (2004a; 2008, December; 2004b) found that individuals with good auditory-perceptual discrimination skills produced speech sounds with greater articulatory contrast and smaller target regions compared to those with poor discrimination skills. Based on these findings, it may be proposed that individuals with better discrimination skills, and therefore smaller articulatory target regions, are better compensators because they are more likely to detect the mismatch between the expected and the actual feedback. On the other hand, individuals with poorer discrimination skills and a naturally increased range of acceptable articulatory realisations would be expected to tolerate auditory perturbations more, and hence, compensate less. MacDonald and his colleagues (2011) tested this hypothesis in their meta-analysis of 116 CE speaking females. For this, the standard deviations of F1 and F2 were computed from the last 15 words of the start phase and correlated with the speaker's mean compensation in the stay phase. No significant correlation in support of the articulatory variance hypothesis was found. Likewise, Mitsuya and colleagues (2011) did not find standard deviation to be significantly correlated with compensatory behaviour.

5. Compensation is not affected by the location of vowel category boundary.

MacDonald and colleagues (2011) explored whether compensation is influenced by the spacing between adjacent vowels. For this purpose, a correlational analysis was undertaken between the speaker's compensation in one formant and that speaker's formant difference between the target vowel /ɛ/ and its neighbouring vowels /æ/ (i.e., direction of perturbation) and /ɪ/ (i.e., direction of expected response). Analyses were run separately for F1 and F2. None of the correlations were significant. Thus, the authors concluded that compensation is not affected by individual differences in vowel category boundary.

6. Compensation is influenced by perceptual sensitivity to F1 differences.

Villacorta and colleagues (2007) specifically investigated this phenomenon by comparing the individual's magnitude of compensatory response to F1 perturbation (as measured in an Adaptive Response Index) with their accuracy to distinguish between differences in F1 frequencies (as indicated by their just noticeable difference). Results of the correlational analysis revealed that the magnitude of the compensatory response was greater in those individuals who were better able to detect differences in F1.

7. The degree of compensation decreases with perceptual adaptation.

Shiller and his colleagues (2009) explored whether speakers adapt to auditory perturbation not only motorically, but also perceptually. For this purpose, the study examined (1) if prolonged auditory perturbation changes the perceptual representation of the target sound /s/, and if so, (2) how changes in the perceptual representation of the target sound influences the degree of compensation observed in an individual.

The study was conducted in three distinct experiments; (1) altered feedback design, (2) same experimental design but with unaltered feedback, and (3) listening experiment of perturbed speech stimuli. Perceptual adaptation was measured through a phoneme identification test, which was administered before and after the auditory perturbation. The difference in the two vowel boundary functions was then compared to the production data during auditory perturbation to see whether a shift in perception had affected the amount of compensation. Results revealed that the auditory feedback group indeed shifted their boundary in the direction of the perturbation, thereby increasing the category range of the sound /s/. This shift was not observed for the unaltered feedback group.

Most importantly, this shift in the identification function for the altered feedback group counteracted the effect of auditory perturbation because the mismatch between the altered target sound and the perceptual representation of that sound was reduced. In an attempt to ensure that the change in identification function was not due to auditory exposure of perturbed stimuli alone, Shiller and his colleagues compared the data of the altered feedback group with that of the passive listening group. Significant differences between groups were in support of a distinct effect of the speech production component implemented in the altered feedback group.

8. Variability in response patterns *within* and *across* speakers is common.

Intra-subject variability has been reported across auditory vowel perturbation studies. More specifically, the meta-analysis of 116 female CE speakers (E. N. MacDonald, et al., 2011) revealed a large amount of variance from token-to-token as indicated by a mean standard deviation of 44 Hz (compared to the mean compensation of 53 Hz). Moreover, while a general trend in the direction and magnitude of the response pattern to auditory vowel perturbation can be seen in the majority of speakers across studies, inter-subject variability regarding these trends is often reported (Houde & Jordan, 2002; Katseff & Houde, 2008;

Munhall, et al., 2009; Purcell & Munhall, 2006a, 2006b; Tourville, et al., 2008; Villacorta, et al., 2007). Although less common, three phenomena besides partial compensation can be distinguished; (1) near complete compensation (Katseff & Houde, 2008; Katseff, et al., 2012; E. N. MacDonald, et al., 2010; Purcell & Munhall, 2006b; Zraick, Gregg, & Whitehouse, 2006), (2) no compensation (Houde & Jordan, 1998; Houde & Jordan, 2002; E. N. MacDonald, et al., 2011; Purcell & Munhall, 2006a, 2006b), and (3) following response (i.e., some individuals change their response pattern in the direction of the perturbation (Katseff & Houde, 2008; Munhall, et al., 2009; Villacorta, et al., 2007).

9. Responses to auditory perturbation seem to be multi-dimensional.

Some studies reported perturbation response patterns on both the F1 and F2 plane (Houde & Jordan, 1998; Houde & Jordan, 2002; Katseff, et al., 2010; Katseff, et al., 2012; E. N. MacDonald, et al., 2010; E. N. MacDonald, et al., 2011; Munhall, et al., 2009; Pile, et al., 2007, February; Purcell & Munhall, 2006b; Villacorta, et al., 2007). However, there is no consistency in the direction of movement of F2 and MacDonald and colleague's meta-analysis (2011) found that only 16% of the variance in compensation in one formant could be explained by compensation in the other. Their study revealed, however, a mean compensation in F2 of 58 Hz (note that F2 was shifted downward by 250 Hz), with a standard deviation of 69 Hz. Some studies also reported changes in F0 as a result of F1 perturbation (Katseff, et al., 2012; Purcell & Munhall, 2006b; Villacorta, et al., 2007) but no systematic investigations have been undertaken yet to shed light onto this phenomenon.

10. Responses to auditory perturbation are free of cognitive strategies.

A recent study by Munhall and colleagues (2009) experimentally tested the effect of awareness on the compensatory response. However, no significant effect of instruction group on compensatory response was found. That is, compensatory response patterns were independent from whether (1) speakers were uninformed about the purpose of the study, (2) informed but asked to ignore any discrepancies in auditory feedback, or (3) explicitly asked NOT to compensate for the auditory mismatch induced by the perturbation paradigm. Thus, it can be assumed that the compensatory response does not seem to be influenced by cognitive intervention, and consequently, can be categorised as an automatic process that predominantly relies on auditory feedback.

11. Compensation is not dependent on how the perturbation is introduced.

A more implicit way to test the effect of awareness on the compensatory response is to compare the effect of introducing perturbation in small increments (which is less noticeable) with the effect of introducing perturbation in a step-wise fashion (which is more noticeable). This study design was employed by MacDonald's et al. (2010) and revealed no effect for rate at which perturbation was introduced on compensation.

12. The amount of compensation is likely not dependent on the shift direction.

It appears to be inconclusive whether the direction of perturbation affects the magnitude of compensatory response. Some studies found larger (Houde & Jordan, 2002; Villacorta, et al., 2007) or more consistent (Katseff & Houde, 2008) compensatory responses when auditory feedback was shifted upward compared to when it was shifted downward while others did not find an effect of shift direction on compensation (E. N. MacDonald, et al., 2011; Mitsuya, et al., 2011; Purcell & Munhall, 2006a; Tourville, et al., 2008). That said, later studies investigating compensatory responses to vowel perturbation have usually preferred to either employ a F1 upshift condition only (e.g., Katseff & Houde, 2008; Katseff, et al., 2010; Katseff, et al., 2012; Shum, et al., 2011) or to simultaneously shift F1 up and F2 down to more closely approximate the neighbouring vowel (e.g., E. N. MacDonald, et al., 2010; E. N. MacDonald, et al., 2012; Munhall, et al., 2009; Pile, et al., 2007, February).

1.8.4.2 Predictors of adaptation to auditory vowel perturbation

1. Adaptation occurs regardless of availability of auditory feedback.

Compensatory response patterns to auditory perturbation equally persist for a short period of time when the perturbation is removed (Shiller, et al., 2009) or unaltered feedback is blocked by masking noise (Houde & Jordan, 2002; Purcell & Munhall, 2006a; Rochet-Capellan & Ostry, 2011b; Villacorta, et al., 2007). That said, Villacorta and colleagues (2007) specifically compared adaptation responses in normal auditory feedback conditions with blocked auditory feedback conditions. Results revealed that adaptation was observed in both conditions but to a lesser degree when auditory feedback was masked with noise compared to the amplified but otherwise unaltered feedback condition.

2. Adaptation is not affected by the number of trials in the Stay phase.

Purcell and Munhall (2006a) explored whether the amount of trials in the stay phase (i.e., full perturbation) would affect de-adaptation during the end phase. For this purpose, 41 speakers were allocated to three groups. The first group was exposed to 20 *baseline tokens*, 50 *ramp tokens*, and 115 *end phase tokens*. The second group was exposed to 20 *baseline tokens*, 50 *ramp tokens*, 15 *hold tokens*, and 115 *end phase tokens*. The third group was exposed to 20 *baseline tokens*, 50 *ramp tokens*, 45 *hold tokens*, and 115 *end phase tokens*. For all groups, auditory perturbation during the ramp phase was increased in 4 Hz increments to reach 200 Hz over the course of 50 trials. Background speech-shaped noise of approximately 50 Hz was added to all tokens. Results showed that all groups had significantly different mean F1 values in the end phase compared to the mean F1 of the baseline tokens. Most importantly, adaptation and de-adaptation was not affected by the amount of hold trials. In fact, an extended ramp phase was enough to induce sensorimotor learning.

3. Adaptation is dependent on vowel and consonant context.

Findings regarding adaptation effects to the same vowel in different consonant contexts or to other vowels – also known as generalisation – have been equivocal, with some evidence to support (Houde & Jordan, 2002; Villacorta, et al., 2007), but more to refute a hypothesis of generalisation (Pile, et al., 2007, February; Rochet-Capellan & Ostry, 2011a; Villacorta, et al., 2007).

4. Exposure to normal auditory feedback results in de-adaptation.

Speakers usually show significant differences in F1 values in the first few trials of the end phase when compared to the start phase. However, this effect ceases over time, with F1 values either returning to baseline or in some cases remaining different but no longer significantly so (Houde & Jordan, 2002; Pile, et al., 2007, February; Purcell & Munhall, 2006a; Shiller, et al., 2009; Villacorta, et al., 2007).

5. De-adaptation requires the production of the target sound.

Munhall's research laboratory specifically examined whether the adaptation effect was time and stimulus dependent (Pile, et al., 2007, February). For this purpose, 59 female speakers were allocated to four different post-perturbation conditions. That is, all speakers were exposed to the same *baseline phase* (45 tokens, all amplified but otherwise unaltered),

eliciting the words ‘head’, ‘hayed’, and ‘hid’. Similarly, all speakers underwent the same *formant shift phase* (40 tokens of ‘head’ presented in full auditory perturbation where F1 was shifted upward and F2 downward to approximate the individual’s production of /æ/). However, for the following *generalisation phase* (40 tokens, all amplified but otherwise unaltered), speakers were allocated to four different conditions. Group one produced the word ‘head’ as in the previous phase, and consequently, was referred to as the trained condition. Group two produced the word ‘hayed’ and was referred to as the generalisation /e/ condition. Group three produced the word ‘hid’ and was referred to as the generalisation /ɪ/ condition. Lastly, group four was exposed to a period of silence. In the final *end phase* (40 tokens, all amplified but otherwise unaltered), all speakers produced the word ‘head’ again as in the formant shift phase.

Comparisons were drawn between the first 15 words of the trained condition and the first 15 words of the end phase of speakers undergoing the other three conditions. No significant difference was found, indicating that adaptation of the vowel /ɛ/ as in ‘head’ was observed even after a time delay due to the production of different vowels during the generalisation phase. In other words, the revised speech motor pattern secondary to auditory perturbation is not affected by productions of other sounds in normal feedback conditions. More importantly, these results suggest that unlearning of the newly established speech motor command during full perturbation requires the production of the target sound under unaltered feedback conditions.

6. Variability in response patterns within and across speakers is common.

While intra-subject and inter-subject variability has been noted, specific information describing adaptation behaviour is currently lacking (Houde & Jordan, 2002; Purcell & Munhall, 2006a; Shiller, et al., 2009; Villacorta, et al., 2007).

1.7.5 Auditory perturbation in ageing populations and communication disorders

To date, only two auditory perturbation studies have looked at the effect of ageing on compensatory response (H. Liu, Russo, & Larson, 2010; P. Liu, et al., 2011). Similarly, only two auditory perturbation studies investigated the effect of neurological impairment on compensation (Kiran & Larson, 2001; H. Liu, Wang, Metman, & Larson, 2012). Although all

three of these studies investigated fundamental frequency as opposed to formant frequency perturbation, results suggest that older adults compensate more compared to younger adults. Furthermore, individuals with PD showed greater magnitudes in compensation (H. Liu, et al., 2012) and more variability and atypical response behaviour compared to age-matched controls (Kiran & Larson, 2001; H. Liu, et al., 2012). In addition, there is growing evidence in both fundamental frequency and formant frequency perturbation research that specific subject groups show different response patterns to auditory perturbation (e.g., singers vs. non-singers, adults vs. children, younger adults vs. older adults, individuals who stutter vs. individuals who do not stutter), indicating that internal cueing strategies, sensory feedback mechanisms, and self-regulatory processes of speech production vary across subject groups (Bauer, Hubbard Seery, LaBonte, & Ruhnke, 2008; Cai et al., 2012; Keough & Jones, 2009; Kiran & Larson, 2001; H. Liu, et al., 2010; Shiller, et al., 2010).

Considering these differences, the absence of perturbation research in communication disorders seems surprising; particularly since clear relationships between self-monitoring, error-detection, error-correction, speech motor learning, and therefore, generalisation of treatment effects can be drawn. Specifically, the following questions can be investigated: (1) Do age-related deficits in auditory perception influence the compensation and adaptation to auditory perturbation? (2) Can speakers with dysarthria detect and correct for alterations in the speech signal as shown in compensatory responses to auditory perturbation? (2) What is the role of auditory feedback in the process of self-monitoring? (3) Can speakers with dysarthria learn new perceptual-motor representations of sounds as shown in the adaptation to auditory perturbation? A number of potential implications and benefits of perturbation research in ageing individuals and those with dysarthria apply.

1.8.6 Limitations of auditory perturbation in ageing populations and communication disorders

While auditory perturbation is a useful tool to investigate error-detection and error-correction mechanisms of speech production, the digital processing and algorithm responsible for the alteration of formant frequencies to modify auditory feedback is highly dependent on the quality of the speech signal. Breathy, harsh, or rough voices and those with poor formant frequency trajectories therefore are prone to erroneous modifications in auditory feedback

(Kiran & Larson, 2001), which can complicate the study of speakers with dysarthria (Berry, North, & Meyers, 2012, March).

1.8.7 Mimicry: An alternative way to study error-detection and error-correction mechanisms of speech production

Imitation is an essential part of speech acquisition and second language learning. As outlined in the DIVA model (see Section 1.7), infants learn to identify a teaching signal (i.e., a sound), and then to compare their own productions with that of the environment. Through repeated imitation of the same sound (or phrase), the mapping of an acoustic target with the proprioceptive information required to consistently produce that sound (or phrase) is established. Recent research investigating mirror neurons provides further support of the notion that sensorimotor experiences create and refine mirror neurons (Catmur, 2013). Specifically, observation of an action in another person activates the same motor programme in the observer, and this motor programme is formed into a mirror neuron when contingent or predictive relationships between the observed and performed action is warranted.

Consequently, it can be assumed that (1) the accurate use of auditory feedback for error-detection and error-correction mechanisms is a vital requisite for the correct imitation and learning of speech production, and (2), similar speech production processes are induced through auditory perturbation and imitation. One of the differences in the two procedures, however, is the focus of attention. Responses to auditory perturbation are free of cognitive processes. In general, imitation is also highly unconscious (M. Wilson & Knoblich, 2005). However, when prompted and instructed to match one's own speech output to that of an acoustic target, imitation may draw, at least in part, on cognitive resources to better match the intended speech target. With that in mind, mimicry tasks seem to lend themselves to the study of error-detection and error-correction mechanisms of speech production in those individuals where the auditory perturbation paradigm is not effective (e.g., due to poor voice quality).

1.9 Summary and future directions

Theory and research have supported a link between speech perception and speech production (e.g., Davis & Johnsrude, 2007; Dhanjal, et al., 2008; Duffy, 2005; Hickok, et al.,

2009; Kent, 2000; Maas, et al., 2008; Perkell, et al., 1997; Pisoni, 1993). This link has also been highlighted in recent brain-imaging studies (e.g., Davis & Johnsrude, 2007; Fadiga, et al., 2002; Pulvermüller, et al., 2006; Watkins & Paus, 2004; Watkins, et al., 2003; S. M. Wilson, 2009; S. M. Wilson, et al., 2004). Motor neurons activated during speech perception are believed to mirror the perception of an articulatory gesture, also referred to as "mirror neurons" (Watkins & Paus, 2004). It is speculated that the transformation of acoustically encoded representations into articulatory representations ensures consistency and accuracy in speech production (Davis & Johnsrude, 2007).

One theoretical framework that particularly highlights the role of auditory and somatosensory feedback for accurate speech production is the **D**irections (in auditory planning space) **I**nto **V**elocities of **A**rticulators model (Golfinopoulos, et al., 2010; Guenther, 1995a; Guenther, et al., 2006; Guenther, et al., 1998; Guenther & Perkell, 2004; Perkell, 2012; Perkell, et al., 2000). Error-detection and error-correction mechanisms are key components of the model and activated whenever the actual sensory feedback differs from the intended acoustic target (Guenther, 1995b, 2003; Guenther & Perkell, 2004; Perkell, et al., 2000). DIVA-driven hypotheses regarding the role of auditory feedback for accurate speech production have successfully been validated using auditory perturbation experimental designs (Brunner, et al., 2011; Villacorta, et al., 2007). Specifically, two major components of speech motor control and motor learning can be studied with auditory perturbation and compared with the parameters and predictions of activation patterns within the DIVA model: compensation and adaptation.

To date, computer simulations in DIVA are based on a healthy and fully functional neural network. Likewise, perturbation studies have mostly focused on young healthy speakers. In other words, research so far has paid little attention to the adaptive flexibility of the sensorimotor systems of older adults and individuals with dysarthria – the ability of an individual to change his or her speech production or update his or her perceptual representations in response to auditory perturbation or imitation. However, it is well known that ageing negatively affects neural structures and their processing. Moreover, current research findings are indicative of auditory-sensory declines in both older adults and individuals with dysarthria. While preliminary results of fundamental frequency perturbation studies in elderly populations and those with PD do not suggest negative effects on error-detection and error-correction mechanisms, it still remains to be systematically investigated

whether the abilities to (1) adjust articulatory settings in response to error-detection mechanisms and (2) update internal models necessary for the learning and maintenance of accurate speech production are limited in these populations.

1.10 Aims of the present thesis

The overall aim of this dissertation was to explore factors influencing feedback mechanisms in healthy speakers of different ages and those with dysarthria and differentiate between the effects of ageing and neurological impairment on speech motor control. It was anticipated that through the theoretical application of the DIVA model to systematic behavioural experimentations underlying auditory perturbation (or, where not possible, through the application of a mimicry task), significant advances in our understanding of auditory-perceptual abilities and sensorimotor control in ageing individuals and those with dysarthria would be made. We further proposed that the information obtained from these investigations would be invaluable for informing the rehabilitation management of individuals with dysarthria. Specifically, it was anticipated that an improved understanding of whether, and how, auditory-perceptual feedback mechanisms are affected in individuals with neurological speech impairments would: (1) shed light onto the interaction between integrational speech processes and treatment strategies, (2) improve our understanding of the relationship between treatment effects and their generalisation, and (3) provide directions for future research in the treatment of dysarthria focusing on ways to better maintain treatment gains.

1.10.1 Phase One: Compensation in healthy speakers

Phase one (Chapter 2) aimed to establish a fundamental understanding of error-detection and error-correction mechanisms to auditory vowel perturbation (i.e., compensation) in healthy speakers of various ages from young to old adults. The aim was to create an inventory of factors influencing compensation in healthy speakers as a baseline against which performances of speakers with dysarthria would then be compared. All age groups were suitable for the perturbation paradigm.

1.10.2 Phase Two: Adaptation in healthy speakers

Phase two (Chapter 3) was a continuation of phase one and examined factors influencing the ability to update feedforward commands (i.e., adaptation) as a consequence of auditory vowel perturbation in healthy speakers of various ages from young to old adults. The aim was to create an inventory of factors influencing adaptation in healthy speakers to allow for comparisons in performance with speakers with dysarthria.

1.10.3 Phase Three: Ability of individuals with dysarthria to mimic suprasegmental features of speech production

Phase three (Chapter 4) required a change in protocol. Thus, a mimicry task was used to examine whether speakers with dysarthria would accurately detect and correct suprasegmental speech errors in their own speech when compared to a model stimulus.

CHAPTER TWO

Predictors of Compensation to Auditory Vowel Perturbation

2.1 Abstract

Compensation to auditory feedback manipulation is a common phenomenon in healthy young speakers of American and Canadian English (e.g., Houde & Jordan, 2002; Katseff, et al., 2012; E. N. MacDonald, et al., 2012; Munhall, et al., 2009; Purcell & Munhall, 2006b; Villacorta, et al., 2007). While past research has shed some light on possible predictors of compensation, little is known regarding relative weighting amongst predictors or how error-correction mechanisms may change with a speaker's age or dialect of the English language. The purpose of the present study was therefore to determine whether, and to what extent, auditory-perceptual and somatosensory measures of speech production predict magnitudes of compensation to auditory vowel perturbation in healthy speakers of New Zealand English (NZE) and distinguish between possible changes associated with ageing. In line with previous research, it was hypothesized that compensation would not be negatively affected by age; however, that auditory-perceptual measures would hold predictive value for magnitudes of compensation. Due to increased somatosensory feedback in NZE front vowels /ɛ/ and /i/, it was further hypothesised that magnitudes of compensation may be reduced compared to previous reports of compensation in American and Canadian English. Fifty-four speakers aged 20-78 years completed the following primary tasks: (1) an auditory vowel perturbation experiment to examine compensatory responses to altered auditory feedback, (2) a single vowel identification task completed before and after the perturbation component to explore perceptual adaptation, and (3) a test of auditory acuity of vowel formant differences based on an adaptive staircase protocol. Examination of the data revealed that NZE speakers exhibited reduced magnitudes of compensation and more variable response patterns relative to previously published data for American and Canadian English speakers (e.g., E. N. MacDonald, et al., 2011; Villacorta, et al., 2007). Linear mixed-effect modelling revealed that there were no significant predictors of compensatory response in the group of 54 speakers. Therefore, follow-up analysis was conducted with speakers allocated to one of four subgroups based on their magnitude and direction of response to perturbation; namely, (1) *big compensators*, (2) *compensators*, (3) *big followers*, and (4) *followers*. Linear mixed-effect modelling indicated that for *big compensators*, F1 baseline standard deviation and F1 vowel distance of HEAD relative to HEED and HAD had a significant effect on the magnitude of compensation. For the subgroup of *compensators*, however, only F1 standard deviation was significant. No predictors could be established for the two other subgroups. Implications of

these results are discussed with a particular focus on characteristics of NZE and the influence of somatosensory feedback.

2.2 Introduction

The ability to make use of auditory and somatosensory feedback for the successful processing of error-detection and error-correction mechanisms is a vital skill for the acquisition and maintenance of fluent speech (e.g., Golfinopoulos, et al., 2011; Golfinopoulos, et al., 2010; Perkell, 2012). Auditory-perceptual feedback enables an infant to establish a relationship between an acoustic target and the tactile/proprioceptive information required to produce that sound (Kent, 2000; Pratt & Tye-Murray, 1997). Furthermore, auditory feedback drives the learning of how to detect and correct articulatory attempts that do not match the target sound (Guenther, 1995b; Guenther & Perkell, 2004).

Importantly, auditory and somatosensory error-detection and error-correction mechanisms learnt during speech acquisition remain available throughout the lifespan. These are particularly activated whenever the *actual* sensory feedback differs from the *intended* acoustic target (Guenther, 1995b, 2003; Guenther & Perkell, 2004; Perkell, et al., 2000); for example, in the light of internal changes (e.g., developmental changes in the size and shape of the articulators, anaesthetics, or damage to the articulators) or external changes (e.g., eating and talking, braces, or speaking in noise). In this instance, these internal and external changes create a mismatch between the expected and the actual auditory feedback of the acoustic target sound. In healthy adult speakers, this interaction has been highlighted in somatosensory perturbation studies that have artificially induced a somatosensory mismatch – for example through the insertion of an artificial palate (Baum & McFarland, 1997, 2000; McAuliffe, et al., 2007), prosthesis (Jones & Munhall, 2003), or bite block (Lindblom, 1990). These studies have shown that when somatosensory feedback is altered or made unavailable through a perturbation device, auditory error-detection and error-correction mechanisms guide articulatory adjustments to ensure accurate speech production.

An alternative way to effectively examine the link between speech perception and speech production and the mechanisms by which auditory feedback guides error-detection and error-correction is through auditory perturbation (Purcell & Munhall, 2006a; Shiller, Sato, Gracco, & Baum, 2009; Sivasankar, Bauer, Babu, & Larson, 2005). Auditory perturbation introduces

an unexpected change to the intended speech signal, as heard by the speaker, in real-time. As a consequence, speakers *compensate* by shifting their production in the direction opposite to the perturbation. For example, when the first formant frequency (F1) of the word “head” is raised to perceptually approximate the word “had”, speakers typically respond by lowering their F1. The ability to detect and correct for formant frequency perturbations in real-time was first observed by Houde and Jordan (1998; 2002) and has since been replicated for American English (AE) (Houde & Jordan, 1998; Houde & Jordan, 2002; Katseff & Houde, 2008; Katseff, et al., 2010; Katseff, et al., 2012; Villacorta, et al., 2007), Canadian English (CE) (E. N. MacDonald, et al., 2010; E. N. MacDonald, et al., 2011; Munhall, et al., 2009; Pile, et al., 2007, February; Purcell & Munhall, 2006a, 2006b, 2008; Rochet-Capellan & Ostry, 2011b; Shiller, et al., 2009; Shum, et al., 2011), French (Mitsuya, et al., 2013) Mandarin (i.e., Akagi, et al., 2006; Cai, et al., 2008; Cai, et al., 2010), Japanese (Akagi, et al., 2006; Mitsuya, et al., 2011), and CE speaking children (E. N. MacDonald, et al., 2012; Shiller, et al., 2010). These studies provide substantial evidence for the role of auditory feedback for accurate speech production. Moreover, the accumulation and analysis of this information has established a number of factors that appear to influence a speaker’s degree of compensation. Table 2.1 contains a summary listing of the common factors and findings regarding compensation in AE and CE speakers (summarised from Houde & Jordan, 2002; Katseff & Houde, 2008; Katseff, et al., 2012; E. N. MacDonald, et al., 2010; E. N. MacDonald, et al., 2011; Munhall, et al., 2009; Pile, et al., 2007, February; Purcell & Munhall, 2006a, 2006b, 2008; Shiller, et al., 2009; Tourville, et al., 2008; Villacorta, et al., 2007). For a more detailed description, please refer to Chapter 1, Section 1.8.4.1.

Table 2.1: *Common findings regarding compensation in AE and CE*

1. Factors influencing compensation	
(a) Magnitude of perturbation	The degree of compensation is related to the magnitude of the perturbation, with linear responses to perturbations up to 150-200 Hz.
(b) Somatosensory value of the vowel	Compensation is influenced by the availability and salience of proprioceptive feedback, such that different vowels elicit different levels of compensation.
(c) Perceptual sensitivity to F1 differences	Compensation is greater in those who are better able to discriminate between F1 differences.
(d) Perceptual adaptation reduces compensation	If perception shifts towards the direction of perturbation, compensation is reduced.
2. Factors that do not influence compensation	
(a) Precision of motor control	as measured in baseline standard deviation.
(b) Location of vowel category boundary	as measured in F1 and F2 distances between adjacent vowels.
(c) Cognitive Strategies	Compensation is impervious to cognitive strategies or active attempts not to compensate.
(d) Direction of perturbation shift	Upward vs. downward
(e) How perturbation is introduced	Gradual vs. immediate
3. Commonly reported compensation phenomena	
(a) Partial compensation	Perturbation of auditory feedback typically results in partial compensation, roughly 25% of the perturbation magnitude.
(b) Inter- and intra-speaker variability is common	With some compensating, some following, and others not responding at all.
(c) Compensation is multi-dimensional	but no consistency in F0 or F2 responses have been established.

From Table 2.1 it is clear that, at least for speakers of North American dialects of English, magnitude of perturbation, somatosensory value of the vowel, perceptual sensitivity to F1 differences, and perceptual adaptation influence the compensatory process. However, most prior investigations employed small numbers of young, healthy speakers (N<20). Therefore, it is not obvious whether the findings of these studies are easily transferable to larger groups of speakers. Similarly, it is currently unknown whether these findings are transferable to other dialects of English – for example, where more somatosensory feedback is available – or speakers across the lifespan. Theoretically, it could be anticipated that ageing effects on hearing, central auditory processing (e.g., Federmeier, et al., 2003; Gordon-Salant, 2008, November; Kiessling, et al., 2003; Pichora-Fuller, 2003b), and working memory capacity (Gazzaley, Sheridan, Cooney, & D'Esposito, 2007; Humes & Floyd, 2005; Kramer, Bherer, Colcombe, Dong, & Greenough, 2004) may well negatively influence the ability of an individual to change his or her speech production in response to auditory perturbation. Indeed it might be predicted that older adults compensate less compared to young adults. However, preliminary findings of pitch perturbation studies in older adults do not suggest negative effects on processes of error-detection and error-correction (H. Liu, et al., 2010; P. Liu, et al., 2011). That said, these results still need to be systematically investigated and validated for formant frequency perturbation.

As a result, this study aimed to investigate predictors of compensation to auditory feedback in a larger pool of speakers from different age groups. Importantly, the unique vowel characteristics in New Zealand English, in which front vowels are elevated relative to AE and CE (MacLagan & Hay, 2007), particularly in young female speakers, required an additional set of questions. These were linked to the potential influences of somatosensory feedback on compensation. Consequently, the three main questions of the present study were: (1) Do speakers of NZE show decreased magnitudes of compensation compared to previous reports of AE and CE, (2) Which factors, or combination of factors, predict compensatory response in speakers of NZE, and (3) Does the weighting of somatosensory information in NZE influence commonly known predictors of compensation? For questions two and three, the four main factors previously investigated in compensation research were targeted: (1) somatosensory feedback (Katseff, et al., 2012; E. N. MacDonald, et al., 2010; E. N. MacDonald, et al., 2011; Mitsuya, et al., 2011; Purcell & Munhall, 2008), (2) vowel boundary (E. N. MacDonald, et al., 2011), (3) vowel discrimination (Villacorta, et al., 2007), and (4) perceptual adaptation (Shiller, et al., 2009).

It was hypothesized that if somatosensory feedback plays a predominant role in compensation, then compensation may be reduced to some extent in NZE compared to previous reports of AE and CE. Specifically, it was hypothesized that compensation may be most reduced in young adults (particularly females), given reports that their /ɛ/-vowel is more elevated than that of older adults and assuming that this higher front vowel provides increased somatosensory feedback, thereby counteracting auditory feedback manipulations. Consistent with previous findings, it was hypothesised that ageing would exhibit no effect on magnitude of compensation. However, it was hypothesised that auditory-perceptual measures would be statistically linked to magnitudes of compensation. Taking into account the possibility of increased somatosensory feedback in NZE, it was hypothesised that vowel boundary category may predict compensation; that is, smaller distances between /ɛ/ and /i/ in NZE were thought to potentially increase the availability of somatosensory feedback, and thus, hypothesised to be associated with reduced magnitudes of compensation. As with the hypothesis regarding compensation overall, it was hypothesised that this effect would be most pronounced in young adults and particularly in young female adults. In regard to the weighting of predictors, it was hypothesised that stronger somatosensory cues as potentially experienced at least in some NZE speakers, would compromise the influence of vowel discrimination and perceptual adaptation previously found to be strongly associated with compensation due to the conflicting sources of feedback. Again, this effect was hypothesised to be particularly present in young (female) speakers (provided they did indeed exhibit the comparatively more elevated /ɛ/).

2.3 Method

All procedures were reviewed and approved by the University of Canterbury Human Ethics Committee and the New Zealand Human Disability Ethics Committee. Written consent was obtained from all participants, each of whom received a $10 voucher as compensation for participation.

2.3.1 Participants

Sixty healthy NZE speakers aged between 20 and 78 years were recruited for the study. This included 12 speakers aged 20 to 29 years, ten aged 31 to 37 years, eight aged 40 to 49 years, nine aged 52 to 58 years, 12 aged 60 to 67 years, and nine aged 70 to 78 years. Within each decade, half of the participants were male and the other half female to a total of 30 male and 30 female participants. All participants reported that they were native speakers of NZE and exhibited no history of neurological or cognitive impairment (including dementia and/or any medical condition affecting attention or concentration), or speech and language impairment (including aided hearing and/or prior history of surgery involving lip and/or tongue).

2.3.2 Procedure

The experiment was conducted at a sound-attenuated testing room within the Communication Disorders Research Facility of the University of Canterbury, Christchurch, New Zealand. Participants were informed that the purpose of the study was to examine how the way speakers hear themselves affects the way they speak. They were further informed that their results would be compared to individuals with neurological speech impairments to advance current clinical practices. That said, participants were blinded to the perturbation aspect of the study but were told that they would hear themselves speaking through earphones. Participants completed five tasks: (1) a hearing screen, (2) a cognitive screen, (3) a vowel identification task, (4) the perturbation experiment, and (5) a hearing acuity task (full details of these tasks are given in the following sections). A full debrief regarding the purpose of the study was undertaken upon completion of the experimental tasks. Although participants were offered to complete testing in one or two sessions, depending on their level

of fatigue and ability to concentrate, all chose to complete testing in one single session, which took approximately two hours.

2.3.2.1 Pre-Perturbation Procedure

Prior to commencement of the perturbation experiment, all speakers underwent a standard pure-tone auditory hearing screening. Participants were screened binaurally for hearing loss at 500 Hz, 1 kHz, 2 kHz, and 4 kHz. In accordance with previous perturbation research in older adults, the screening threshold was set at 25dB HL (Kiran & Larson, 2001; H. Liu, et al., 2010; P. Liu, et al., 2011). It is well known that hearing deteriorates as a function of age (e.g., Cruickshanks et al., 1998; Gates, Cooper, Kannel, & Miller, 1990; Kiessling, et al., 2003); therefore, exploring perceptual abilities across a wide range of age groups naturally imposes limitations regarding the compliance of hearing thresholds. In the present study, all participants passed the screening at 25dB for 500 Hz and 1 kHz, with the exception of two participants (i.e., 35(L)/25(R) dB and 35dB (both ears) for 500 Hz and 35(L)/30(R) dB and 50(L)/25(R) dB for 1 kHz, respectively). High frequency loss was evident for some participants at 2k Hz and 4k Hz. Given that frequencies most relevant to the production and perception tasks employed in the current study design were within the range of 300 to 700 Hz and that participants spoke at approximately 75 dB, with their auditory feedback being amplified by 15 dB, the presence of age-related high frequency hearing loss was not a significant concern for data analysis. Thus, none of the participants were excluded from the study based on their hearing thresholds.

The pure-tone auditory hearing screening was followed by collection of connected speech samples (i.e., the grandfather passage and a map task) and a cognitive screening, containing of a letter-number sequencing task, a digital span task (both subparts of the Wechsler Memory Scale, 1997) and the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005). The order of tasks within these two domains (i.e., cognition and connected speech) was randomised. Finally, speakers were asked to read 10 repetitions of a total of 11 /hVd/ words, all of which were presented in randomized order. The first 5 repetitions were read under normal feedback conditions (i.e., without earphones) to familiarise speakers with the task. The latter 5 repetitions were read under amplified but otherwise unaltered feedback

conditions (i.e. with insert earphones) (see more details regarding the audio recording procedure below).

2.3.2.2 Audio Recording

For all speech recordings, speakers were fitted with a self-mounted headset with an adjustable microphone (Audio-Technica **AT803**) attached. The microphone was positioned at a fixed distance of approximately 10 cm from the speaker's mouth. The Grandfather Passage and Map Task were recorded and digitized into .wav files at a sampling rate of 44.1 kHz (24-bit resolution), using a TASCAM recorder (HD-P2). All remaining speech utterances were digitized and recorded at a sampling rate of 48 kHz (24-bit resolution) and a frame size of 64, and subsequently, down sampled to 12 kHz and a frame size of 16, using the M-Audio Delta-44 sound card (M-Audio, Irwindale, CA, USA).

Down sampling was undertaken to reduce the computational load for audio signal processing, and thus, meet the requirements for the successful manipulation and playback of auditory feedback. The recording and digitization of these remaining tokens was controlled from a PC, running the Audapter software (Cai, et al., 2012; Cai, et al., 2010; Cai, Ghosh, Guenther, & Perkell, 2011). Audapter is a custom real-time digital speech signal processing software written in Visual C++ (Microsoft, Redmond, WA, USA) and called from MATLAB (MathWorks, Natick, MA, USA). For tokens that were presented with auditory feedback, Audapter used two channels to record the different signals relevant to this study. The first channel recorded the speaker's actual spoken utterance, and the second channel recorded the signal that was generated by Audapter and fed back to the speaker through the earphones. Depending on the phase of the experiment, the Audapter generated signal was either (1) identical to that of the first channel recording, or (2) an F1-altered version of the speech sound according to the experimental protocol (see below). Microphone calibration for a 1kHz tone at 80 db SPL was also conducted via the Audapter software and MATLAB.

Insert earphones (ER-3A, Etymotic Research Ltd., Elk Grove Village, IL, USA) were used for the provision of auditory feedback during the second half of the /hVd/ recordings (see Section 2.3.2.1) and for the perturbation experiment (see Section 2.3.2.3). Due to the presence of noise above 5 kHz, the speech signal was low-pass filtered at 6.3 kHz using a graphic equalizer (GQX-3101 Mono Graphic Equalizer, Ashly Audio Inc., Webster, NY,

USA). With the graphic equalizer in place, the output levels of the insert earphones were measured using a Verifit (Audioscan, Dorchester, ON, Canada). The gain of the preamplifier (**Behringer Xenyx 802**) and graphic equalizer were then adjusted to ensure that the acoustic signal in both ears was 15 dB higher than the speech output signal at the microphone level. Both insertion earphones and the 15 dB difference in auditory feedback served to mask the air- and bone-conducted feedback. To ensure stability and consistency in the speech signal within and across speakers, speakers were encouraged to keep their speech intensity within a range of 76±8 dB SPL. This was achieved by providing speakers with visual feedback. After each production, one of the following visual cues was displayed on a monitor: (1) not loud enough, (2) target loudness achieved, and (3) too loud. Similarly, an attempt was made to control for the vowel duration in the CVC stimulus words, using the same visual cues and feedback options. Target vowel duration was set at 250±50 ms. Visual feedback regarding target loudness and duration was given for the entire duration of the experiment.

2.3.2.3 Vowel Perturbation: A Real-time Formant Shift Procedure

Similar to the audio recording, the formant shift procedure was also controlled from a PC, running the Audapter software (Cai, et al., 2012; Cai, et al., 2010; Cai, et al., 2011). During the production of the target words HEAD and FED, vowel productions were detected in near-real-time by employing a short-time root-mean-square (RMS) threshold. The first formant frequency (F1) of the vowel /ɛ/ was estimated, using a linear prediction coding (LPC) analysis. In addition to the LPC analysis mentioned above, low-pass cepstral liftering and dynamic-programming for formant tracking (Xia & Espy-Wilson, 2000) were employed to improve the quality of formant estimation, especially for high-pitched speakers. Subsequent to this F1 estimation and tracking, the identified F1 tracks were then smoothed online with a 10.67-ms window. This smoothing used a weighting of the samples with the instantaneous RMS amplitude of the signal, which effectively emphasized the closed phase of the glottal cycles and reduced the impact of the sub-glottal resonances on the formant estimates. A formant shifting algorithm generated an upward shift of F1 for each word production in the Ramp, Stay, and Stay2 phases of the experiment. This was done on a moment-to-moment basis, generating F1 shifts specific to each individual's F1 production. Thus, while the degree of perturbation was pre-determined (see Section 2.3.2.6), the amount of F1 perturbation dynamically changed from trial to trial. This particular perturbation method allowed to

preserve the glottal cycles of the original vowel, and therefore, sounded more natural and more like the individual's productions than a traditionally synthesized vowel token would have.

2.3.2.4 Vowel Identification Task

The Vowel Identification Task served to determine the vowel boundary between the vowels /ɛ/ and /æ/ for each speaker. Arithmetic mean F1 and F2 values of the individual's unaltered HEAD and HAD productions (extracted from Pre Phase and Start Phase) served to employ a nine-step continuum of synthesised vowel formants, using the KLATT synthesizer. To ensure accurate vowel synthesis, all formant productions were visually examined prior to the offline computation of the nine-step vowel continuum and trials with erroneous formant tracking discarded from the vowel synthesis code. Participants were then asked to complete a forced-choice paradigm, identifying each stimulus as either "Ed" or "Add". Each step of the continuum comprised of 10 repetitions, which were presented to the participants in randomized order, making a total of 90 tokens. Participants underwent two sets of Vowel Identification Tasks (i.e., Pre-Test and Post-Test), which served to examine whether the location of the vowel boundary changed as a consequence of extended exposure of altered auditory feedback. The order of tokens in the Pre-Test and Post-Test was randomized separately.

2.3.2.5 Hearing Acuity Test

The Hearing Acuity Test was carried out in order to determine the just noticeable difference (JND) in F1 for each participant. The task employed a set of synthesised vowel formants, which were based on the individual's own formant frequencies. The information required for speech synthesis was derived from the production data of the Start, Ramp, and Stay Phase. That is, following the production tasks, the arithmetic mean F1 and F2 values of the word productions of HEAD and FED in the above-mentioned three phrases were calculated. The F1 and F2 values of the base token was determined as those of the speaker's production in the three phases closest to this mean on the F1-F2 plane. This was done offline, using the identical perturbation algorithm employed for the production task. The administration of six identical adaptive, one-up, two-down staircase protocols then aimed to generate a mean estimate of the JND in F1. For this, participants were presented with three

consecutive synthesized vowels. The first one always represented the base token. Of the two following tokens, one was a repetition of the base token (i.e., base token 2) and one had greater F1 values compared to the base token (i.e., test token). The order of the base token 2 and the test token was randomized across trials. A forced-choice paradigm was put in place to identify either the second or third token to be different. After a selection was made by operating the computer mouse, immediate feedback was given regarding the correct answer to this particular sequence of sounds. This was done to help ensure that participants stayed focused throughout the six staircase protocols, which took approximately 15 minutes to complete. In line with the perturbation employed in the production task, participants were first exposed to a 50% difference in F1. Any incorrect response resulted in increased distances and two consecutive correct responses in decreased distances between the base token and the test token. For a more detailed description of the adaptive one-up, two-down staircase protocol, please refer to Levitt (Levitt, 1971).

2.3.2.6 Experimental Protocol

The main experimental protocol was divided into eleven phases (see Fig. 2.1 for schematic) and took approximately 90 minutes to complete. All tasks were displayed on the same computer monitor and controlled by a custom-written MATLAB script. The Practice, Pre, Start, Ramp, Stay, Stay2 and End Phases represented production tasks and comprised of a total of 205 distinct consonant-vowel-consonant (CVC) stimulus words (i.e., heed, hid, head, fed, had, heard, hud, who'd. hood, hod, hoard, hard). The randomization of these stimulus words was carried out in a block-by-block manner within each phase of the experiment. In each trial, the speaker had approximately 2.5 seconds to read aloud each word displayed on the computer screen. In contrast to the production tasks, the Vowel Identification Task (Pre-Test and Post-Test) and Hearing Acuity Task were perceptual in nature and required the use of a computer mouse. Some older adults were not familiar with using a mouse and felt uncomfortable to do so. In this case, speakers were asked to point with their fingers on their preferred choice displayed on the computer monitor. The researcher then executed the task with the mouse accordingly.

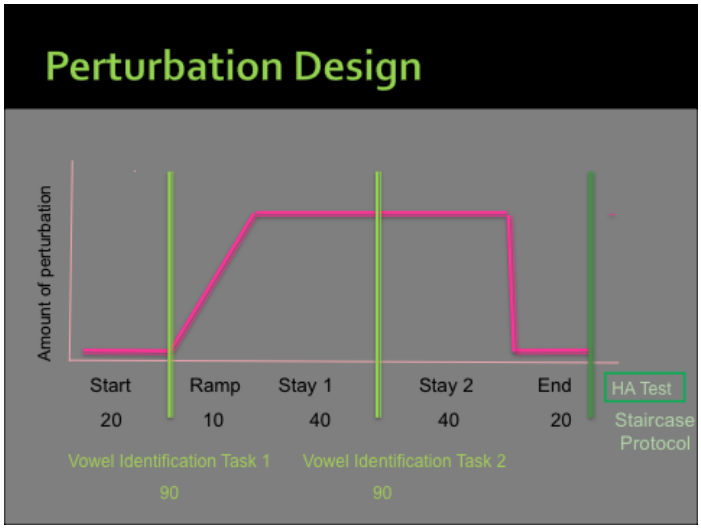

Figure 2.1: Auditory perturbation design employed in the current study. Numbers indicate the amount of trials in the specified phase.

Upon completion of the experimental procedures, speakers completed a short questionnaire, which aimed at investigating whether they were aware of any changes in auditory feedback. A full description of each phase is provided below.

(1) Familiarisation Phase

Speakers were asked to produce 5 repetitions of eleven /hVd/ words. This served to familiarize speakers with the experimental set-up.

(2) Practice-1 Phase

Speakers were trained to produce utterances at a loudness of 76±8 dB SPL.

(3) Practice-2 Phase

While maintaining the loudness target mentioned above, speakers were trained to produce utterances at a duration of 250±50 ms.

(4) Pre Phase

Speakers were asked to produce 5 repetitions of eleven /hVd/ words. This served to familiarize speakers with the experimental procedure and collect data regarding vowel space and vowel separation for post analysis.

(5) Start Phase

The Start Phase served to collect F1 baseline samples of the training words HEAD and FED. Twenty utterances were collected under amplified but otherwise unaltered auditory feedback conditions (i.e., 10 tokens of HEAD and 10 tokens of FED). These F1 baseline measures for the vowel /ɛ/ then served to compute speaker-specific perturbation fields for phases (6) and (7).

(6) Vowel Identification Task Pre-Test

Phase (6) represented the first of two Vowel Identification Tasks. Participants were presented with synthetic speech stimuli through insertion earphones and asked to identify each stimulus as either "Ed" or "Add", using a computer mouse. Stimuli were given on a nine-step vowel continuum with each step comprising of 10 representations, making a total of 90 tokens to identify.

(7) Ramp Phase

In the Ramp Phase, auditory feedback was gradually shifted upward from zero perturbation to maximum perturbation. This was done over the period of 10 utterances (i.e., 5 repetitions of HEAD and 5 repetitions of FED).

(8) Stay Phase

In the Stay Phase, auditory perturbation was held at a maximum of 50% of each individual's F1 production. Forty productions of HEAD and FED were collected (i.e., 20 repetitions of HEAD and 20 repetitions of FED) and used for the analysis of whether or not speakers compensated for F1 perturbation.

(9) Vowel Identification Task Post-Test

Following the compensation experiment, participants underwent a second series of the Vowel Identification Task.

(10) Stay2 Phase

The Stay2 Phase was a replication of the Stay Phase (8) but the stimuli words were presented in a different randomized order. The replication served to ensure that the speaker had a chance to recover from any potential decay of the compensation in the Stay Phase due

to the ensuing Vowel Identification Task Post-Test. This was important for the analysis of adaptation (see below).

(11) End Phase

In the End Phase, the perturbation to auditory feedback was removed. This phase was similar to the baseline phase (but with a different randomized order of production tokens) and served the analysis of whether or not speakers maintained any vowel production changes in the absence of auditory perturbation, indicating remapping of the auditory-motor representation of the target sound (i.e., adaptation). The results of this phase will be presented in the Chapter 3.

(12) Hearing Acuity Test

The experiment finished with a Hearing Acuity Test, whereby participants were presented with three consecutive sounds and asked to identify which one of the last two tokens sounded different compared to the first one. A forced-choice paradigm was offered on the monitor and participants were given a mouse to click on the appropriate answer. This task served to determine the smallest F1 difference that participants were able to detect.

2.3.3 Data Analysis

2.3.3.1 Spectral Analysis of Tokens

For the spectral analysis of each recorded token, the produced F1 and F2 tracks were first smoothed by 41.3-ms Hamming windows. Then, an automatic algorithm extracted the first two formant frequencies based on the arithmetic mean F1 inside the interval in which the RMS intensity exceeded 75 percent of the word's maximum RMS. Hence, F1 and F2 analyses for all utterances used in the present study were based on 75 percent of the word's peak RMS. With this automatic algorithm in place, each individual word production was manually screened for production errors and/or formant tracking errors. While doing so, the researcher was blinded towards the phase of the experiment. When a tracking error was detected, the parameters of the formant tracker were manually adjusted in an attempt to correct for the error. Trials where this was not possible were discarded from subsequent analyses, as were those with production errors.

In addition, trials with F1 or F2 values outside 3 standard deviations from mean F1 or F2 values were removed from the analysis to reduce the impact of outliers on the overall results. For this, F1 and F2 means and standard deviations were calculated separately for each phase (i.e., Pre, Start, Ramp, Stay, Stay 2, End). Thus, the exclusion criterion of single trials was phase-dependent. Participants for which more than 25% of tokens had to be discarded in either one of the Start, Stay or End Phase were removed from the analysis. Two participants did not meet the less than 25% of tokens discarded per phase requirement (see Section 2.3.3.1 for more details), and thus, were removed from the analysis. Both were female aged 58 and 78 years with predominantly formant tracking errors. Four additional participants were removed from the analysis due to substantial variability in the Start Phase (both females, aged 34 and 78 years) or response patterns that were atypical from the rest of the participants (both females, aged 47 and 49 years). For the remaining 54 participants, a total of 121 out of 3510 FED tokens (i.e., 3.5%) , 141 out of 3510 HEAD tokens (i.e., 4%) were excluded from subsequent computations and analyses of dependent and independent variables.

Finally, a number of vowel measures were computed as independent variables. These were based on productions of HAD, HEED, and HEAD from the Pre Phase. Although it is known that some vowel realizations are closer together or further apart in word lists compared to conversational speech, the focus for this study was on single word productions elicited from the main experiment to ensure for most exact comparisons within and between speakers.

2.3.3.2 Computation of Independent Vowel Space Variables

To explore the effect of vowel distance on compensation and adaptation, the distance of /ɛ/ in relationship to the two neighbouring vowels /æ/ and /i/ for F1 were computed for each individual. These computations were derived from F1 medians (to allow for a more stable measure due to possibly unequal number of words) of HEAD, HAD, and HEED tokens from the Pre Phase. The distance value was derived by dividing the F1 distance between HEED and HEAD by the overall F1 range of HEED, HEAD, and HAD. This calculation was preferred over Pythagorean Euclidean Distance measurements to account for positive and negative distances between HEAD and HEED. That is, HEAD can be either below or above HEED in NZE vowel space. Furthermore, one-dimensional distances ensure equal weighting of F1 and F2, and hence, are a more accurate measure of the influence of vowel distance on

compensation and adaptation. Similarly to the perturbation data, each /hVd/ token of the Pre Phase was first manually screened for production errors and/or formant tracking errors. The same procedures for production and formant tracking errors applied as described above. Based on these exclusion criteria, a total of 13 out of 270 HAD tokens (i.e., 4.8%), 16 out of 270 HEAD (i.e., 5.9%) and 9 out of 270 HEED tokens (i.e., 3.3%) were excluded from the analysis. To further examine whether precision of speech motor control affects compensation and adaptation, the mean F1 standard deviation of the vowels of interest to the perturbation design (i.e., HEAD, HEED, HAD) were computed.

2.3.3.3 Perceptual Analysis

Perceptual data obtained during the experiment included the Vowel Identification Tasks (conducted pre Ramp Phase and post Stay Phase) and the Hearing Acuity Test. Analysis of both parameters is described below.

2.3.3.3.1 Vowel Identification Function

To identify whether perceptual learning had occurred secondary to extended auditory vowel perturbation, a vowel identification function was computed for each of the two Vowel Identification Tasks. First, the percentage of /æ/-responses (in a set of 10 tokens) for each of the nine stimuli of the /ɛ/ - /æ/ vowel continuum was calculated. Then, a four-parameter logistic (i.e., sigmoid) function was fitted to the resulting nine data points which ranged from 0 (i.e., 100% /ɛ/-responses) to 1 (i.e., 100% /æ/-responses). The stimulus step at which participants identified half of the tokens as /ɛ/ and half of the token as /æ/ was integrated into the sigmoid function and represented the vowel boundary. Finally, a difference score was computed by subtracting the Pre-Test estimate of the sigmoid vowel boundary function from the Post-Test estimate for each participant. Values greater than zero represented a vowel boundary shift in the direction of the perturbation (i.e., participants identified more tokens as /æ/). Conversely, negative values represented a shift in the opposite direction of the perturbation (i.e., participants identified more tokens as /ɛ/).

2.3.3.3.2 Determining Auditory Acuity

To determine the smallest perceivable F1 vowel difference for each individual, a mean JND score was computed based on the six endpoints of the staircase protocols. Note that the staircases started at a 50% difference in F1 frequency (=1.0) and then generally increased in difficulty (<1). Therefore, smaller numbers represent better hearing acuity.

2.3.3.4 Cognitive Analysis

The cognitive procedures undertaken for this study were scored according to the test manuals (Nasreddine, et al., 2005; Wechsler, 1997). All participants passed the MoCA with a score equal to or higher than 26. A combined value of the Forward Digit Span, Backward Digit Span, and Letter-Number-Sequencing Test of the Wechsler Memory Scale (Wechsler, 1997) was computed as an independent variable for later analyses.

2.3.3.5 Data Processing

Prior to statistical analysis, two separate data normalisation procedures were completed. To ensure that the comparison of F1 and F2 responses to auditory feedback alterations was independent of age and gender differences in NZE front vowel characteristics, F1 and F2 were normalised using (a) the division method (Villacorta, et al., 2007), and (b) the subtraction method (e.g., E. N. MacDonald, et al., 2011). Note that separate normalisations were carried out for HEAD and FED to account for any differences in formant frequency by word. Correlational analysis between the two normalisation procedures for the dependent variable F1 revealed a high correlation coefficient (r=0.99). Consequently, all subsequent analyses were carried out with the subtraction method to allow for interpretations in Hz units. For this, the mean F1 from the last 15 tokens of the Start Phase was subtracted from each of the last 30 trials of the Stay Phase, with each word token (i.e., HEAD and FED) being normalised to its specific mean F1 of the Start Phase (i.e., either HEAD or FED).

2.3.3.6 Statistical Procedure

Predictors of response behaviour to auditory vowel perturbation were analysed with linear mixed-effect models using the lmer function from the lme4 library (Bates, Maechler, & Bolker, 2011) in the R programming language. Linear mixed-effect modelling was primarily chosen to account for between-subject variability in front vowel productions typically observed in NZE (MacLagan & Hay, 2007). A series of linear mixed-effect models were fit to the data. Analysis began with a full model including the fixed effects of hearing acuity (JND), vowel shift (Vowel Identification Pre and Post-Test difference score), F1 distance of HEAD in relationship to HEED and HAD, mean F1 standard deviation of front vowels, age, gender, and working memory. In addition, the random effect of subject was included. Note that item did not need to be included as a random effect as analysis was based on HEAD tokens only (see Section 2.4.1). Models were fit using the residual maximum likelihood criterion. The best-fitting model was established through systematic mixed forward and backward step-wise model comparisons using likelihood ratio tests. The random effect structure of subject was evaluated with random slopes for each fixed within-subject effect using the maximal random effect structure justified by the data. However, the more complicated random effects were not found to qualitatively change the final model. Consequently, they were dropped from the reported models below.

2.4 Results

2.4.1 Preliminary Analyses

Visual inspection of the data appeared to indicate differences in response magnitude by word production (i.e., HEAD versus FED)—a finding unexpected based on prior research (e.g., Houde & Jordan, 2002; Shiller, et al., 2009; Villacorta, et al., 2007). To ensure the robustness of any subsequent analyses, this visual trend was subject to analysis. In the Start Phase, F1 productions of FED were on average 7.7 Hz higher than HEAD, a difference that was statistically significant [$t(1,53)=3.7$, $p<.001$]. Due to reports of dialectal differences in NZE across age and gender (MacLagan & Hay, 2007), it was therefore investigated whether these differences were gender specific. A linear mixed-effect model fit to the data revealed significant main effects of both gender [$\beta=-49.2$, SE=11.2, $t=-4.4$, $p<.0001$]—revealing lower F1 in males relative to females—and word [$\beta=-7.7$, SE=2.1, $t=-3.7$, $p<.001$], indicating

significantly lower F1s in HEAD relative to FED. However, the gender-by-word interaction was non-significant. Hence, it can be assumed that males and females treat HEAD and FED similarly.

Given the differences between HEAD and FED in the Start Phase, further preliminary analysis was conducted within the Stay Phase to ensure accurate later analyses of compensatory responses. The normalised results for HEAD and FED in the Stay Phase were compared using a paired *t*-test and results revealed that speakers' normalized F1 of FED was significantly higher in the Stay Phase compared to that of HEAD [$t(1,53)=-2.04$, $p=.047$], with on average 3.8 Hz more compensation in FED compared to HEAD. As a result, separate t-tests were performed to examine compensation in HEAD and FED. Results revealed significant compensation for HEAD [$t(1,53)=-2.4$, $p=.02$] and FED [$t(1,53)=-3.4$, $p=.001$], with a mean compensation of 6.6 Hz (SD=13 Hz) for HEAD (i.e., approximately 4.8% compensation relative to the amount of perturbation) and 10.4 Hz (SD=13.3) for FED (i.e., approximately 7.4% compensation relative to the amount of perturbation). Furthermore, response patterns for HEAD and FED were frequently inconsistent across phases and individuals, with some individuals compensating in one word and following in the other (see Figure 2.2 for an example). Given the differences in magnitude and pattern of compensatory responses to the two speech tokens, the primary predictive analysis of compensation was conducted based on trials of HEAD only. This approach is consistent with prior research studies (Katseff, et al., 2012; E. N. MacDonald, et al., 2010; E. N. MacDonald, et al., 2011; Munhall, et al., 2009; Purcell & Munhall, 2006a, 2006b; Shum, et al., 2011).

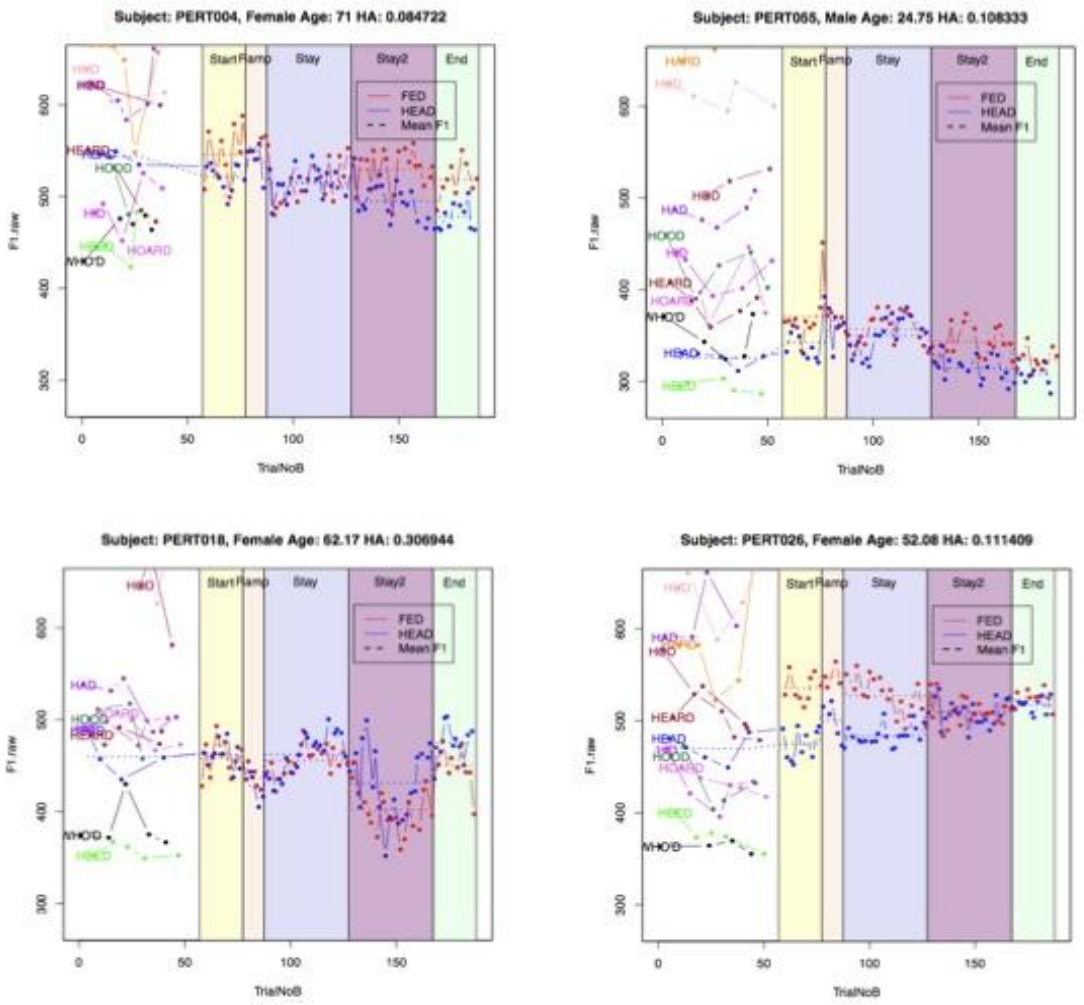

Figure 2.2: Examples of inter- and intra-phase variability in HEAD and FED.

2.4.1.1 Awareness of Compensation

Analysis of post-test questionnaire data revealed that six participants (10%) noticed their voice sounded artificial, with two specifically noting the upward shift. Another five participants (8%) reported that "something was different" but could not identify the source or reason for the difference. Data from those participants who exhibited some degree of awareness of the perturbation was visually examined and no systematic differences were observed. Of these participants, eight speakers shifted their F1 in the direction opposite to the

perturbation (i.e., compensating) and three in the direction of the perturbation (i.e., following).

2.4.2 Subsequent Analysis

2.4.2.1 Compensation to Auditory Vowel Perturbation in NZE

On average, speakers shifted their F1 by 8.5 Hz (SD=14.9 Hz) in the direction opposite to the perturbation (i.e., approximately 6.1% compensation relative to the amount of perturbation). One-sample *t*-tests revealed that this shift was significant for the group [$t(1,53) = -3.1$, $p = .003$), indicating that as a group, speakers compensated for F1 auditory feedback perturbations. A closer look at response type (i.e., whether a speaker exhibited a compensatory or following response) revealed that 34 out of the 54 speakers (i.e., 63%) showed the expected compensatory response (i.e., showed a negative mean normalised F1 value in the Stay Phase); however, 20 speakers (i.e., 37%) exhibited a following response (i.e., shifted their F1 in the same direction as the stimulus). Hence, the current study exhibited a larger number of speakers who evidenced following responses relative to the approximately 11.1% (Villacorta, et al., 2007) and 22.4% of speakers (E. N. MacDonald, et al., 2011) in prior research of AE and CE speakers. Note that the latter study was based on 116 speakers and followers comprised of individuals who either followed in F1 or F2 and those who followed in both F1 and F2.

2.4.2.2 Predictors of Compensation to Auditory Vowel Perturbation

The linear mixed-effect model fit to the entire group of speakers was unable to establish significant predictors of response pattern to auditory vowel perturbation. Hence, it appeared that none of the fixed effects entered into the statistical model were predictive of speakers compensatory response. However, a distinction by gender approached significance [$\beta = -9.9$, SE=5.2, $t = -1.9$, $p = .057$], indicating that males (N=30) exhibited a greater trend towards shifting their F1 in the opposite direction to the perturbation compared to females. That is, males on average shifted their F1 downward by 10.9 Hz (SD=25 Hz), which equates to approximately 9.4% compensation relative to the amount of perturbation. In contrast, females

shifted their F1 downward by -.9 Hz (SD=21 Hz), which equates to approximately 2% compensation relative to the amount of perturbation.

2.4.2.3 Predictors of Compensation: A Subgroup Analysis

Given the lack of significant findings reported in Section 2.4.2.2, further analysis of the data was conducted within subgroups of response types (i.e., whether a speaker exhibited a compensatory or following response). Four response categories were developed; (1) *compensators*: those who exhibited a negative normalised mean F1 value in the Stay Phase (N=34); (2) *big compensators:* those who showed a normalised mean F1 in the Stay Phase greater than -1 standard deviation from the mean F1 of the Start Phase (N=20); (3) *followers:* those who demonstrated a positive normalised mean F1 value in the Stay Phase (N=20); and (4) *big followers:* those with a normalised mean F1 in the Stay Phase > 1 standard deviation from the mean F1 of the Start Phase (N=4). Follow-up analyses were conducted within each of the four groups. Variables within each group were examined for normal distributions and a 3 standard deviation (from the mean of that variable) cut-off was put in place for variables that possessed extreme values to ensure that these outliers were eliminated from data analysis.

2.4.2.3.1 Predictors of Compensation in the Group of "Compensators"

For the group of *compensators* (N=34), a 3 standard deviation cut-off was carried out for the two independent variables of mean F1 standard deviation of front vowels and hearing acuity. As a result, the final analysis was reduced to 29 speakers. For these 29 speakers, the average downward shift in F1 was 18.9 Hz (SD=12.5 Hz), which is equal to approximately 13.9% compensation relative to the amount of perturbation. A linear mixed-effect model was fit to the *compensators* data only, using similar fixed and random effects structures to those described in Section 2.3.3.6. Analysis revealed a significant effect of F1 standard deviation (as measured in the overall mean of the front vowels HEED, HEAD and HAD) on compensation [β =-1.2, SE=.48, t=-2.5, p=.01], indicating that speakers with more variable response patterns in the Pre Phase exhibited significantly increased magnitudes of compensation to perturbation. The remaining factors tested were non-significant.

2.4.2.3.2 Predictors of Compensation in the Group of "Big Compensators"

Similar to the group of *compensators*, a 3 standard deviation cut-off was carried out for the two independent variables of mean F1 standard deviation of front vowels and hearing acuity to ensure that outliers did not affect the overall results. As a result, the initial group of *big compensators* (N=20) was reduced to 17 speakers. The mean downward shift of this reduced group of 17 speakers was 25.3 Hz (SD=18.8 Hz), which equates to approximately 18.5% compensation relative to the amount of perturbation. As with the group of *compensators*, a linear mixed-effect model was fit to the *big compensator* data only. Analysis revealed two significant effects on compensation; namely (1) the mean F1 standard deviation of the front vowels HEED, HEAD and HAD [β =-2.32, SE=.35, t=-6.7, p<.0001], and (2) the F1 distance of HEAD relative to HEED and HAD [β =-40.83, SE=7.7, t=-5.3, p<.0001], indicating that speakers with more variable response patterns and more centralised F1 HEAD productions in the Pre Phase exhibited significantly larger magnitudes of compensation to auditory perturbation. Moreover, speakers with more centralised F1 productions of HEAD in relationship to HEED and HAD compensated more compared to speakers whose F1 HEAD productions were closer to HEED.

2.4.2.3.3 Predictors of Compensation in the Group of "Followers"

For the group of *followers* (N=20), a 3 standard deviation cut-off was carried out for the independent variable of hearing acuity. Thus, the group was reduced to 17 speakers. The mean upward shift of this reduced group was 14.3 Hz (SD=14.4 Hz), which equates to approximately 10.5% following relative to the amount of perturbation. Importantly, no significant predictors for the group of *followers* were found using linear mixed-effect modelling.

2.4.2.3.4 Predictors of Compensation in the Group of "Big Followers"

For the group of *big followers* (N=6), the mean F1 upward shift was 25.7 Hz (SD=15.2 Hz), which equates to approximately 19.5% following relative to the amount of perturbation. Linear mixed-effect modelling could not be carried out due to the limited number of

speakers. Visual inspection of the data did not reveal any obvious effects influencing following for the group.

2.4.3 Vowel Space Characteristics of NZE

Interestingly, speakers in the current study did not evidence the expected age and gender differences in their production of NZE front vowels (i.e., based on prior research by MacLagan and Hay (2007)—which may have played some role in the lack of significant findings. As a result, further analysis of the vowel space of speakers within the current study was conducted to shed light on the specific front vowel characteristics of this cohort of speakers. Overall, the current data set suggests a trend towards centralisation of HEAD with increased age. There was a significant correlation between the two variables (r=.46, p<.001), in the expected direction (i.e., older adults exhibiting larger distances between /ɛ/ and /i/ compared to younger adults). However, from Figure 2.3 it appeared that this trend was predominantly driven by female speakers. That is, females' front vowel distances changed with age whereas males' front vowel distances did not show an age trend. This is in contrast to previous reports from MacLagan and Hay (2007), which indicated an age-related trend of front vowel elevation for both males and females.

Moreover, the current study suggests that male speakers overall showed more centralised HEAD vowels compared to female speakers (see Figure 2.4). That said, a few young female speakers behaved similar to young male speakers (with larger distances between HEAD and HEED) and vice versa. Similarly, some young speakers showed vowel space characteristics of older speakers (with larger distances between HEAD and HEED) and vice versa. Figures 2.5 and 2.6 show some vowel space differences and similarities in male and female speakers of different ages, emphasising the difficulty in establishing gender and age dependent predictions.

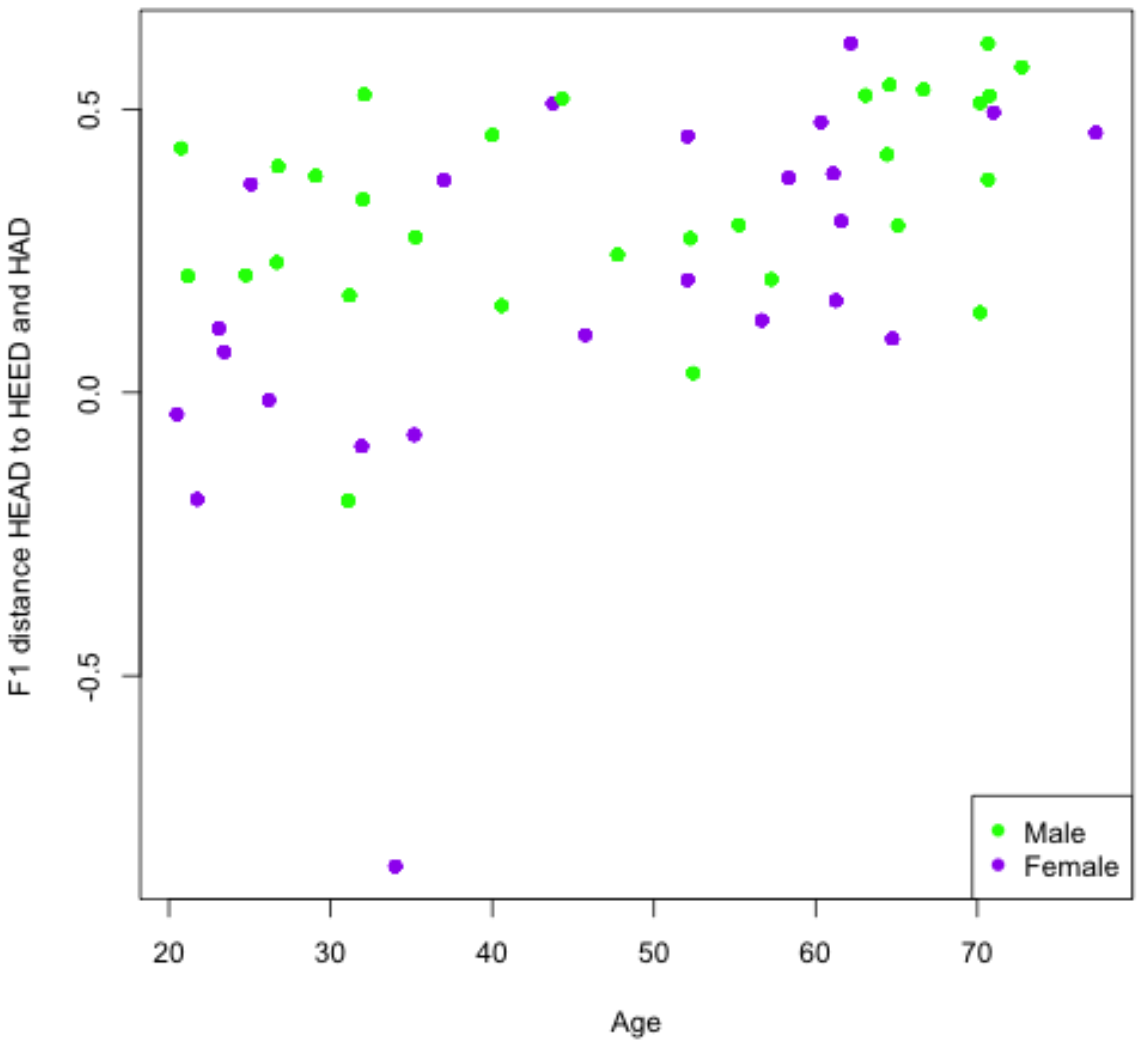

Figure 2.3: Changes in F1 HEAD distance relative to HEED and HAD with age and gender. The value of .5 indicated a centralised HEAD, whereas smaller numbers indicated HEAD moving closer to HEED and negative values indicated that HEAD was located above HEED.

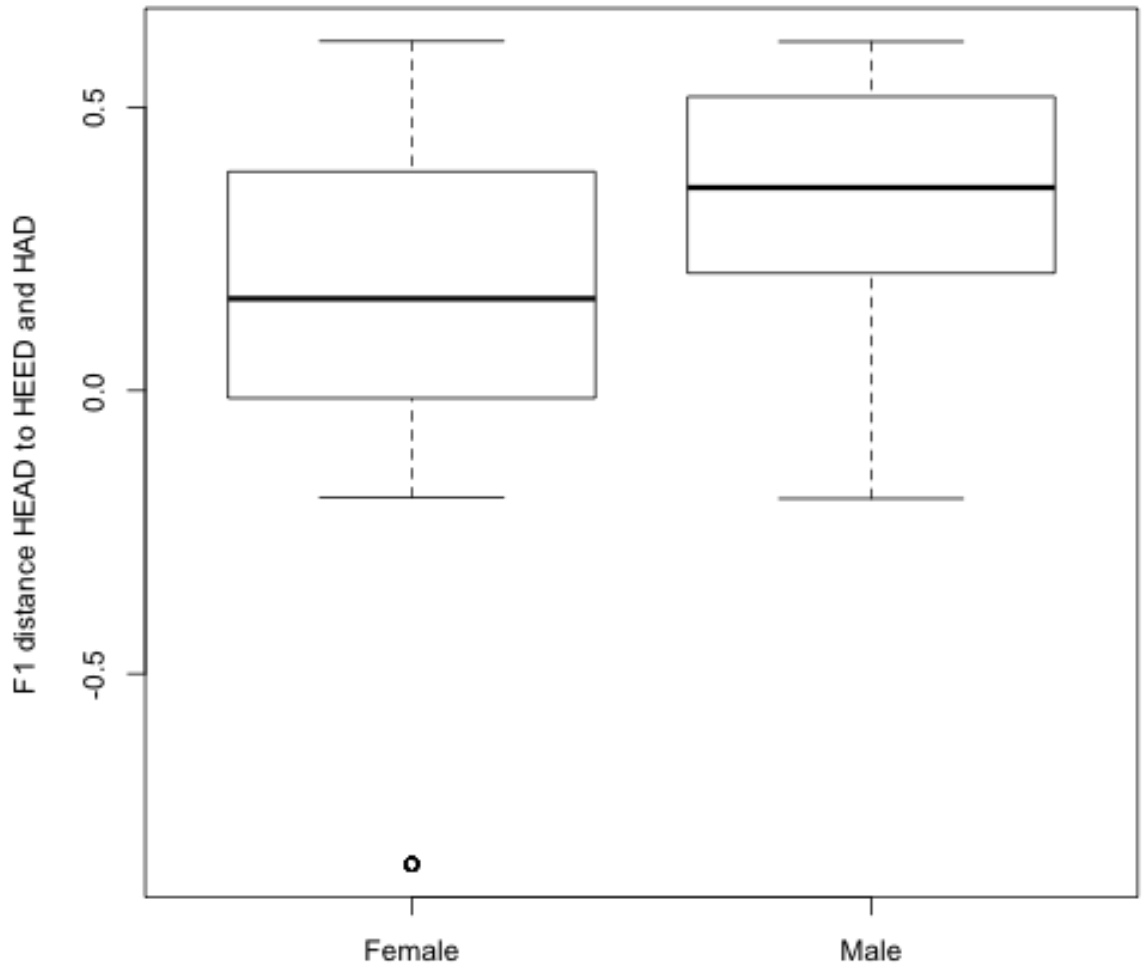

Figure 2.4: Boxplot showing the effect of gender on F1 HEAD distance relative to HEED and HAD.

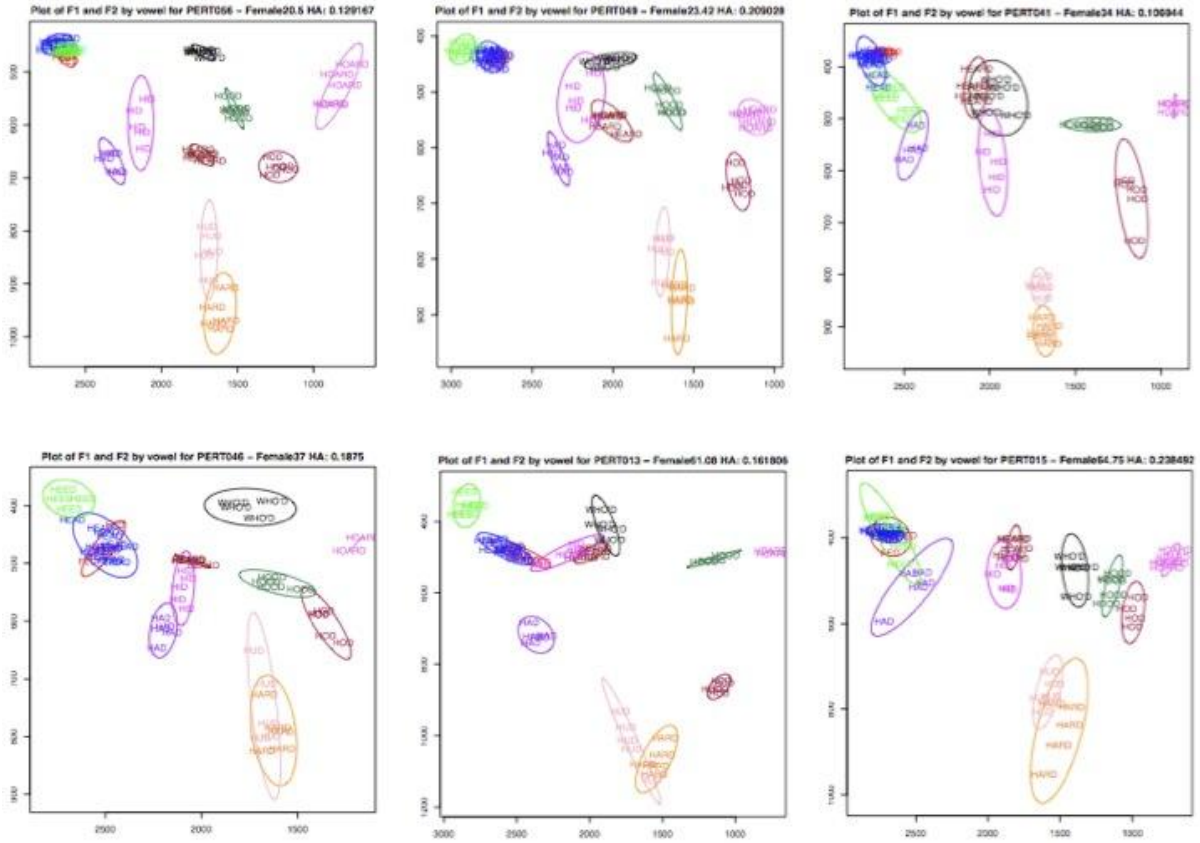

Figure 2.5: Vowel space characteristics of female speakers. The number after the gender indicates the age and "HA" represents the JND score of the hearing acuity test for that individual speaker. Points of interest are the blue (HEAD), light green (HEED) and purple (HAD) front vowel characteristics, which illustrate the considerable amount of variability across speakers of different ages.

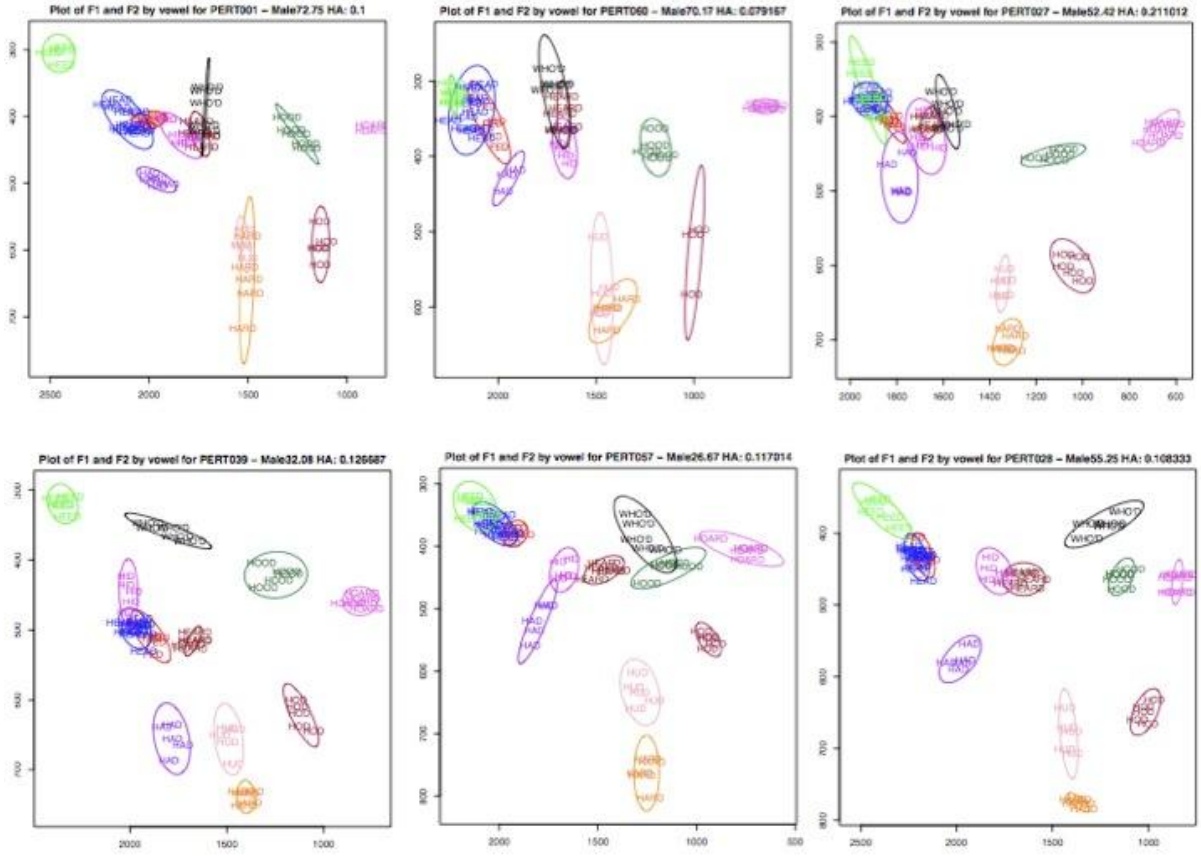

Figure 2.6: Vowel space characteristics of male speakers. The number after the gender indicates the age and "HA" represents the JND score of the hearing acuity test for that individual speaker. Points of interest are the blue (HEAD), light green (HEED) and purple (HAD) front vowel characteristics, which illustrate inconsistencies across speakers of different ages; although these are less pronounced compared to female speakers.

2.5 Discussion

The current study aimed to determine (1) magnitudes of compensation in NZE compared to previous reports of AE and CE, (2) predictors of compensation in NZE, and (3) the effect of potentially increased somatosensory feedback in the NZE front vowel /ɛ/ on commonly known factors to (not) influence compensation in AE and CE. It was hypothesised that if somatosensory information was an important factor influencing the magnitude of compensation, then speakers of NZE may compensate less compared to previous reports of compensation in AE and CE. This was found to be true; that is, while speakers overall

compensated to auditory vowel perturbation, the amount of compensation—approximately 6.1%—was less than the approximately 25% previously reported for AE and CE (E. N. MacDonald, et al., 2011; Villacorta, et al., 2007). Further, it was hypothesised that young female speakers would potentially compensate less due to closer approximations of HEAD and HEED vowels. Gender approached significance for the overall group (p=.057), with males compensating more than females. However, there was no significant effect of age, contrary to the hypothesis. As per previous research, it was hypothesised that ageing would not negatively affect compensation, which held true for the current data set. In contrast, the results did not support the study's hypothesis that auditory-perceptual measures would predict magnitudes of compensation. Due to NZE front vowel characteristics, it was hypothesised that the F1 HEAD distance relative to HEED and HAD would predict the magnitude of compensation, with young females exhibiting the smallest magnitudes of compensation due to the assumed closer approximation of HEAD relative to HEED. Again, this hypothesis was not supported—no predictors of response pattern to auditory vowel perturbation could be established for the overall group. However, a follow-up subgroups analysis (i.e., *compensators, big compensators, followers, big followers)* revealed that within the *big compensator* group, speakers with greater distances of F1 HEAD relative to HEED and HAD compensated significantly more compared to those with closer approximations of HEAD to HEED. A gender-by-age interaction was not found to be a predictor as were none of the auditory-perceptual measures. Interestingly, within both groups of compensators (i.e., *compensators* and *big compensators*), a higher baseline F1 standard deviation was associated with increased magnitudes of response. No predictors could be established for *followers* or *big followers*. Thus, the study hypothesis regarding the effect of somatosensory feedback on compensation was partially supported by our current subgroup findings; as was the hypothesis that in speakers with dominant somatosensory weights (particularly young females), sensory auditory predictors would be less pronounced. In the following paragraphs, each of the main findings will be discussed in turn.

The current results are in support of previous research investigating the influence of proprioceptive feedback on compensation. Purcell & Munhall (2008) found that the magnitude of compensation varied across vowels depending on the somatosensory information available. Vowels with more somatosensory information resulted in reduced compensation. Similarly, the reduced magnitude of compensation in speakers of NZE may be explained by an increased availability of somatosensory feedback in the front vowel /ɛ/

counteracting induced auditory feedback errors. That said, a number of additional factors influencing compensation in NZE should be considered. First of all, it is possible that due to the increased amount of acceptable realisations of HEAD and HEED secondary to generational shifts in front vowel realisations, speakers were perhaps less prompted to correct for auditory feedback manipulations in a consistent manner. This may also explain why young females did not exhibit the smallest magnitudes of compensation for the group as somatosensory feedback may not have been the only factor influencing these results. That said, the variability in front vowel productions (in particular, the distance and location of HEAD relative to HEED) across speakers may also have contributed to the absence of an age-by-gender effect for young females in any of our hypotheses. In addition, speakers may not have been able to detect the direction of the perturbation (perhaps also due to the variability of front vowel realisations in NZE), and thus, responded inappropriately by either following the shift direction or showing little or no response to auditory feedback alterations. Third, the influence regarding F1 HEAD distances relative to HEED and HAD could also be interpreted in a perceptual manner. That is, smaller distances between HEAD and HEED indicated greater distances between HEAD and HAD. From a perceptual point of view, it could be argued that speakers compensated less because auditory perturbation did not reach the vowel boundary of HAD, and thus, did not activate error-detection mechanisms to correct the feedback error. However, given the magnitude of perturbation employed in the current study and the somatosensory weight of HEAD and HEED approximations, this explanation appears less likely.

The fact that no predictors were found for the overall group of speakers is not surprising given the large variability in response magnitude and direction. Subgroup linear mixed-effect analyses of *compensators* and *big compensators* (who showed 13.9% and 18.5% compensation, respectively), however, afforded additional support for somatosensory influences on compensation. Specifically, compensation was reduced in speakers with greater somatosensory weights as defined by their F1 HEAD distance relative to HEED and HAD in the subgroup of *big compensators*. While this contrasts MacDonald and colleagues' (2011) finding that vowel category boundary did not significantly influence compensation, the pool of speakers in that study with their specific vowel space characteristics of CE did not allow for that differentiation. Thus, it can be assumed that in English dialects with relatively equal distances of HEAD to adjacent vowels relevant to the perturbation, vowel boundary does not affect compensation. However, in dialects with closer distances between adjacent vowels as

in NZE, associated somatosensory weights result in significant effects of vowel boundary on compensation.

Perhaps more interesting are the results regarding the effect of standard deviation on compensation. Based on research of Perkell and colleagues (Perkell, Guenther, et al., 2004; Perkell, et al., 2008, December; Perkell, Matthies, et al., 2004), MacDonald et al. (2011) proposed that speakers with small articulatory target regions – as a consequence of good auditory discrimination skills – should compensate more to auditory perturbation compared to speakers with larger articulatory target regions. However, their meta-analysis did not find evidence to support this hypothesis. The present results suggest differently, and more importantly, they suggest an effect in the opposite direction to the proposed theory. This has once more significant implications for auditory and somatosensory weights on compensation. In particular, this would suggest that speakers with larger standard deviations were less likely to detect somatosensory errors. An alternative explanation for the influence of increased baseline standard deviation on compensation is given by a fundamental frequency (F0) study investigating the effect of vocal accuracy on compensation. Scheerer and Jones (2012) suggest that individuals with poor motor control (as indicated by increased standard deviations) relied more on auditory feedback mechanisms, and thus, showed increased compensation compared to speakers with good motor control. Given the availability of somatosensory information in NZE, however, the explanatory value of this interpretation appears to be limited for the present findings.

Past research has acknowledged that different speakers may place different weights on somatosensory and auditory feedback mechanisms (Houde & Jordan, 2002; E. N. MacDonald, et al., 2010; Purcell & Munhall, 2006a, 2008). Thus, speakers may respond differently to induced auditory errors depending on which sensory error-detection mechanism is strongest within a particular speaker. In AE and CE, for example, the absence of a saturation effect in the production of /ɛ/ (Purcell & Munhall, 2008; Villacorta, et al., 2007) may explain why auditory-perceptual measures were found to be predictors of compensation. In contrast, raised /ɛ/-production in NZE benefits from a saturation effect for most speakers, which may be responsible for the dominant weighting of somatosensory predictors on compensation. Taken together, it can be assumed that some speakers have stronger auditory and other more somatosensory weights on error-detection mechanisms, which may influence predictors of compensation and their weighting amongst each other. Based on the current

knowledge, it is likely that these weightings relate to the availability of proprioceptive information both across and within languages.

Most importantly, the current results provide evidence that error-detection and error-correction mechanisms as measured by first formant frequency perturbations are not affected by age – at least not in regard to response magnitude and direction. This finding is in line with fundamental frequency perturbation research in the elderly population (Kiran & Larson, 2001; H. Liu, et al., 2010; P. Liu, et al., 2011). Thus, the present results may be taken as preliminary evidence that auditory vowel perturbation research may yield important insights in speakers with dysarthria and PD as direct comparisons to healthy individuals can be drawn. However, given the vowel space characteristics of NZE, a replication of the present study for AE and/or CE where somatosensory weights do not dominate possible effects of ageing (i.e., declines in auditory, somatosensory, cognitive systems) on predictors of compensation would be beneficial to validate and expand on the current findings.

It is important to consider some limitations of the present results. First of all, while initially equal numbers of male and female speakers were collected, subgroup analyses were based on a relatively small amount of (particularly female) speakers. Thus, any assumptions made from this data set have to be taken with caution until these preliminary results have been replicated in a larger cohort of speakers. Moreover, existing evidence suggests that the magnitude of perturbation affects compensation (Katseff, et al., 2012; E. N. MacDonald, et al., 2010). The present study employed a perturbation magnitude of 50%, which for most speakers, would have resulted in a shift greater than 200 Hz. This magnitude was chosen to ensure that even older speakers with possible hearing declines would be able to detect the induced auditory feedback error. However, given the relatively close approximation of HEAD and HEED in NZE, it is possible that smaller magnitudes of perturbation would result in different overall results due to the reduced weight of somatosensory error-detection mechanisms. Additionally, the current perturbation study was designed to shift F1 frequencies on a trial-to-trial basis. The overall shift was therefore dynamic rather than constant in nature. From a motor learning point of view, this could explain the increased response variability within speakers and phases as no consistent cause-effect relationships could be established. Lastly, the unusual variability observed in the current study made it necessary to reduce the analysis to a limited amount of trials. While previous studies have

used a similar amount of trials overall, an effect of the reduced number of trials on the overall results in the present study cannot be completely eliminated.

With these limitations in mind, the results of the present study have some important implications for theory and research. First of all, this study lends further support for the necessity of the DIVA model to incorporate and computationally test differing weights in somato-auditory error maps to better represent differences in sensory error-detection and error-correction mechanism. Another important finding of the present study is the inconsistency regarding vowel space characteristics across age and gender, indicating continuous changes in the front vowel characteristics of NZE. It is perhaps not surprising that perceptual and articulatory representations are adaptable as shown in EXEMPLAR theory (Goldstone, 1998) and well-reported research analysing changes in the Queen's speech in response to generational changes in the British dialect (Harrington, 2006). However, this means that gender and age do not serve as accurate predictors of vowel space characteristics in NZE. Consequently, future studies may aim to group speakers according to vowel space characteristics to examine possible effects on compensation. In regard to NZE speakers with dysarthria, implications are two-fold. On the one hand, the present findings suggest that there may be difficulties in studying auditory-perceptual influences on compensation in this population due to the unexpected increased elevation of the /ɛ/-vowel, and hence, somatosensory information available in at least some older adults. On the other hand, when carefully controlled for vowel space characteristics, perturbation research in NZE may shed light onto the differentiation between auditory and tactile-kinaesthetic declines as explanations for reduced maintenance and generalisation effects, particularly in individuals with PD. Specifically, the availability and use of auditory and somatosensory information for error-detection and error-correction can be investigated in this population. However, given the within and across speaker variability in response patterns, more research has to be conducted first to allow for more consistent results against which data from individuals with dysarthria can then be compared. That said, it would be warranted to conduct pitch or loudness perturbation studies with speakers with dysarthria and PD and compare these results to the present findings as well as with previous studies in AE and CE to further shed light onto the potential benefits of perturbation research in this population.

In addition, a number of improvements in the experimental design based on the present findings may be considered. First, male speakers generally seemed to show greater and more

consistent responses to auditory perturbation. A reduced cohort of male speakers may therefore be preferred to achieve more consistent results overall. Alternatively, a larger study focusing on young males and females may shed some light into the influence of HEAD to HEED approximations on compensation and predictors of compensation. Second, the administration of a discrimination test between HEED and HEAD may yield useful information to further explain why some individuals compensated more than others. MacDonald and colleagues (2010) proposed that increased discrimination skills between HEAD and HAD may result in larger magnitudes of compensation because auditory errors would be detected earlier. In contrast, it is possible that improved discrimination skills between HEAD and HEED (given the somatosensory feedback available) may result in decreased compensation. Yet another scenario may be that for some speakers, HEAD and HEED are perceptually (and kinaesthetically) the same, which may result in increased compensation due to reduced vowel boundary violations. Third, a tonal discrimination task aiming at distinguishing between upward and downward shifts may help to further distinguish possible reasons for the increased variability in response type (i.e., followers and little or no response) in NZE. Additionally, an exclusion criterion based on musicality may further help to reduce variability. Finally, comparing experimental designs with different quantities of perturbation and trials using only one test word (i.e., HEAD instead of HEAD and FED) may shed further light into the validity of the present results.

In summary, speakers of NZE compensate to auditory vowel perturbation although an increased variability in response magnitude and direction is noted. The current study provides strong evidence for somatosensory weights on error-detection and error-correction mechanisms, which seem to dominate auditory-perceptual influences on compensation. That said, these vowel space specific weights do not appear to be reliably predictable by age and gender. As no clear relationships between auditory speech perception and speech production can be drawn based on these findings, alternative methods should be explored to study auditory error-detection and error-correction mechanisms in speakers with neurological speech impairments—particularly for NZE speakers. Overall, the present results were consistent with the following existing characteristics and predictors of compensation to auditory vowel perturbation: partial compensation, salience of proprioceptive feedback for compensation, variability in response pattern within and across speakers, multi-dimensional responses to auditory perturbation, and independence of compensation from cognitive strategies.

CHAPTER THREE

Predictors of Adaptation to Auditory Vowel Perturbation

3.1 Abstract

Chapter 2 reported that speakers of NZE compensated less to auditory perturbation compared to previous reports of AE and CE speakers. No clear predictors of compensation could be established for the group as a whole. However, follow-up subgroup analysis revealed that for those labelled *big compensators*, a higher variability of F1 and more centralised F1 HEAD productions in the Pre Phase were predictive of larger compensatory responses. Earlier studies of AE and CE speakers have shown that when a compensatory response is elicited following auditory perturbation, the compensatory vowel articulation persists for a short period of time after auditory perturbation is removed (e.g., Houde & Jordan, 2002; E. N. MacDonald, et al., 2011; Munhall, et al., 2009; Villacorta, et al., 2007). It is thought that remapping of the forward command as a result of motor-perceptual learning is responsible for this phenomenon referred to as adaptation. While potentially an important aspect in speech motor learning, little research has investigated adaptation and no predictors of adaptation have been established to date. Likewise, it is unknown whether adaptation is negatively affected by age. The purpose of the current study was therefore to determine whether speakers of NZE who compensated to auditory vowel perturbation also show this pattern of adaptation, and if so, whether factors previously investigated in compensation research have predictive value for adaptation, including possible changes associated with ageing. Sixteen speakers aged 21 to 77 years from the subgroup labelled *big compensators* were included within this analysis. A continuation of the auditory perturbation experiment employed in Chapter 2, adaptation was examined by comparing mean F1 values of the Start Phase with mean F1 values of the End Phase of the experiment. Specifically, the End Phase was split into an Adaptation Period and a De-adaptation Period. The Adaptation Period determined adaptation and was used for linear mixed-effect modelling to establish possible predictors of adaptation. The De-adaptation Period determined if normal auditory feedback had re-set the forward model towards its original setting from the Start Phase. For this, the De-adaptation Period was compared against both the Start Phase of the experiment and the Adaptation Period preceding the De-adaptation Period. In line with previous research, it was hypothesised that *big compensators* would adapt for auditory vowel perturbation. Further, it was hypothesised that normal auditory feedback would induce the process of de-adaptation in the De-adaptation Period of the experiment. In the absence of previously established predictors of adaptation and based on previous reports that the magnitude of compensation is positively correlated with adaptation, it was further hypothesised that the predictors of

magnitude of compensation established in Chapter 2 would transfer to adaptation but none of the other fixed effects would turn out to be significant. Linear mixed-effect modelling revealed that speakers who compensated also adapted and de-adapted to auditory vowel perturbation. However, no speaker-specific predictors of adaptation were established. Adaptation and de-adaptation are discussed in light of the DIVA model of speech production and perceptual-motor learning. Possible reasons for the absence of predictors of adaptation and directions for future research are discussed.

3.2 Introduction

Speech perception and speech production are closely linked—and this relationship is changeable with altered sensory experiences (Catmur, 2013; Watkins & Paus, 2004). In other words, the mapping between an acoustic target and the tactile/proprioceptive information required to produce that sound may adapt following either internal or external sensory changes (Catmur, 2013; Guenther, et al., 1998; Perkell, 2012; Perkell, et al., 2000). Developmental changes in the size and shape of the articulators, for example, require the continuous remapping of internal speech sound maps to ensure speech intelligibility. Another example highlighting the salience of remapping internal speech sound maps for successful speech motor learning is that of cochlear implant users switching between on and off settings. That is, through the use of auditory feedback in the on-condition, adjustments to segmental and suprasegmental features of speech production are made, thereby approximating the speech of normal hearing speakers (Economou, et al., 1992; Perkell, et al., 1992; Svirsky, et al., 1992; Svirsky & Tobey, 1991).

One way of testing this phenomenon of remapping internal speech sound maps in healthy and normal hearing individuals is through an auditory perturbation paradigm as described in Chapter 2 of this thesis. That is, Chapter 2 examined the ability to *compensate* for auditory perturbation and focused on the analysis of speech production in the Stay Phase where maximum feedback alterations were employed. This prolonged exposure to auditory feedback modifications simultaneously offers the basis of remapping—an altered sensory experience that over time influences the relationship between speech perception and speech production. The remapping of speech sounds, termed as *adaptation*, is witnessed in the

persistence of the adjusted speech sound map over a period of time immediately after auditory perturbation is removed and feedback returns to normal.

One theoretical framework that explains the phenomena of compensation and adaptation to altered auditory feedback is that of the DIVA model of speech production (e.g., Guenther, 2006; Guenther, et al., 1998; Perkell, 2012; Villacorta, et al., 2007). When the intended feedback differs from the actual feedback, articulatory adjustments are made to better fit the intended target sound (i.e., compensation). If this mismatch between speech perception and speech production persists, the corrective speech motor commands are stored in the internal model (or feedforward command) of that sound, thereby altering the speech map for following articulation attempts (i.e., adaptation). In other words, consistent sensory experiences alter the mapping between the acoustic target and its articulatory realisation, and thus, build the foundation for speech motor learning. When the source of the mismatch (i.e., auditory perturbation) is removed, recovered auditory feedback induces a process of de-adaptation by which the speech sound map is re-tuned to its original mapping.

Adaptation to auditory formant frequency perturbation has so far been reported for AE (Houde & Jordan, 1998; Houde & Jordan, 2002; Katseff & Houde, 2008; Katseff, et al., 2010; Katseff, et al., 2012; Villacorta, et al., 2007), CE (E. N. MacDonald, et al., 2010; E. N. MacDonald, et al., 2011; Munhall, et al., 2009; Pile, et al., 2007, February; Purcell & Munhall, 2006a, 2006b, 2008; Rochet-Capellan & Ostry, 2011b; Shiller, et al., 2009; Shum, et al., 2011), Mandarin (i.e., Akagi, et al., 2006; Cai, et al., 2008; Cai, et al., 2010), Japanese (Akagi, et al., 2006; Mitsuya, et al., 2011), and CE speaking children (E. N. MacDonald, et al., 2012; Shiller, et al., 2010). While adaptation and speech motor learning have been less investigated compared to the phenomenon of compensation, a few commonalities across adaptation research in AE and CE speakers can be established and are listed in Table 3.1 (summarised from Houde & Jordan, 2002; Mitsuya, et al., 2011; Pile, et al., 2007, February; Purcell & Munhall, 2006a; Rochet-Capellan & Ostry, 2011a; Shiller, et al., 2009; Villacorta, et al., 2007). For a more detailed description, please refer to Chapter 1, Section 1.8.4.2.

Table 3.1: *Common findings regarding adaptation in AE and CE*

1. Factors influencing adaptation	
(a) Adaptation is dependent on vowel and consonant context	Generalisation to other phoneme environments is unlikely.
(b) Magnitude of compensation	Adaptation is positively correlated with magnitudes of compensation
2. Factors that do not influence adaptation	
(a) Masking of auditory feedback	Adaptation is observed regardless of whether feedback is normal or blocked by noise.
(b) Amount of Stay trials	Adaptation is present with a ramp phase only or as little as 15 trials of fully perturbed stimuli.
3. Commonly reported adaptation phenomena	
(a) De-adaptation	Unperturbed feedback over time results in de-adaptation. This process requires the production of the target sound.
(b) Inter- and intra-speaker variability is common	with different magnitudes and perseverance of adaptation over time.

Chapter 2 of this thesis showed that responses to auditory feedback perturbation for speakers of NZE were not consistent with previous reports from AE and CE speakers. In particular, the percentage of speakers who demonstrated a compensatory response was reduced relative to prior studies, and in general, the magnitude of this response was also. Consequently, speakers were grouped according to their response type and magnitude to further investigate possible predictors within each of these groups. The group of *big compensators*, which showed results closest to previous research in AE and CE, revealed that somatosensory feedback influences compensatory responses in NZE speakers. Specifically, speakers with higher F1 variability in front vowels and more centralised F1 vowel distance of

HEAD relative to HEED and HAD compensated most, presumably due to the reduced likelihood of somatosensory error signals counteracting auditory feedback modifications.

Compensation for auditory perturbation highlights the significance of error-correction for accurate speech production as outlined in the DIVA model of speech production. However, speech motor learning – as witnessed in the ability to remap the internal speech sound map – is dependent on the successful storage of these corrective motor commands into the speech sound map (i.e., feedforward model). Therefore, the study of adaptation, and particularly predictors of adaptation, yields the potential of significantly enhancing current knowledge of factors influencing speech motor learning. Such findings may merit the speech rehabilitation of many speakers including those of the elderly population. As previously mentioned in sections 1.6 and 2.2, ageing naturally is associated with declines in hearing and cognition. Moreover, Bock and Schneider (2002) specifically noted that sensorimotor adaptation may be more demanding for older adults due to these age-related changes. While the compensation study employed in Chapter 2 did not find negative effects of ageing on compensation, this still needs to be confirmed for adaptation.

Given the preliminary findings of compensation in NZE speakers, the focus of the current study was therefore to examine whether (1) speakers of NZE who exhibited similar compensatory patterns to those previously reported for AE and CE speakers would evidence similar adaptation (and de-adaptation) patterns when auditory perturbation is removed, and if so, whether (2) predictors of adaptation could be established for this group, including those associated with ageing. Therefore, only speakers who qualified as *big compensators* in Chapter 2 were considered for the examination of adaptation and possible factors influencing speech motor learning. In line with previous research, it was hypothesised that speakers of NZE who classified as *big compensators* would also adapt to auditory perturbation. It was further hypothesised that after a period of exposure to unperturbed auditory feedback, a process of de-adaptation would be initiated, whereby speakers' F1 productions would be re-tuned, and therefore, shift again towards their Start Phase measures. In the absence of knowledge regarding factors influencing adaptation in AE and CE, hypothesis building regarding possible predictors of adaptation was not straightforward. However, given the findings of previous studies that the magnitude of adaptation is positively correlated with the magnitude of compensation, it was hypothesised that predictors of compensation may transfer to adaptation. Theoretical support for an effect of more centralised F1 HEAD productions

relative to HEED and HAD and higher F1 variability in front vowels on adaptation may derive from the reduced chance of somatosensory error signals counteracting the revision of speech motor commands. As per previous findings regarding compensation in NZE speakers, age was not hypothesised to affect adaptation.

3.3 Method

As stated in Section 2.3.2 (ff.), all speakers completed the auditory perturbation experimental protocol. This was designed to ensure that both compensation and adaptation to auditory feedback could be examined within a single experiment. The current analysis focuses on the later sections of the experiment, which are used to analyse adaptation to auditory perturbation through the comparison of the Start Phase with the End Phase (see Section 3.3.2). A schematic overview of the experimental phases is provided in Figure 2.1.

3.3.1 Participants

Since adaptation is a phenomenon closely linked to compensation, only speakers termed *big compensators* in Section 2.4.2.3 (i.e., speakers who showed a normalised mean F1 in the Stay Phase greater than -1 standard deviation from the mean F1 of the Start Phase) qualified for the analysis of adaptation. Given the variability within subjects and across phases (see Section 2.4.1), an additional analysis was conducted to ensure that compensatory responses were consistent and comparable across Stay and Stay2 phases of the experiment. Moreover, the Stay2 Phase immediately preceded the End Phase, and thus, was directly linked to the analysis of adaptation. Final analysis of adaptation was conducted on 16 speakers (out of the 20 speakers classified as *big compensators* in the Stay Phase). Note that *big compensators* were equally distributed in regard to age, with either two or three speakers in each decade between the 20s and the 70s.

3.3.2 Data Analysis

Data analysis was divided into three subanalyses, all based on the Start and End phases of the perturbation experiment (see Figure 2.1). For the analyses of adaptation and de-

adaptation, the End Phase was divided into two equal halves. The first 10 productions of the End Phase were defined as the *Adaptation* Period and the last 10 productions of that same End Phase as the *De-adaptation* Period of the experiment. Note, however, that similar to the analysis of compensation (see Section 2.4.1), only HEAD tokens were used for the analyses of adaptation and de-adaptation.

To determine whether *big compensators* adapted and de-adapted to auditory vowel perturbation, linear mixed-effect modelling using the lmer function (see Section 2.3.3.6) was conducted on raw F1values in Hz. The fixed factor of phase included single mean F1 values from the Start Phase, Adaptation Period and De-adaptation Period, and subject was included as a random effect. A second series of model fitting was conducted to determine speaker-specific predictors of adaptation. Similar to the analysis of compensation, F1 raw values of the Adaptation Period were normalised prior to statistical analysis. For this, the mean F1 from the last 15 tokens of the Start Phase was subtracted from each of the tokens in the Adaptation Period. Linear effects-modelling, including fixed and random effects, was identical to that described for the analysis of compensation in Section 2.3.3.6.

3.4 Results

3.4.1 Adaptation and De-adaptation

Linear mixed-effect regression revealed that both the Adaptation Period [b=-24.7, SE=3.4, t=-7.2, p<.0001] and De-adaptation Period [β=-16.9, SE=3.4, t=-5, p<.0001] were significantly different from the Start Phase, indicating that speakers adapted their F1 productions as a result of auditory perturbation. Moreover, these results suggest that when compared to the Start Phase, adaptation is measured in both the Adaptation and De-Adaptation Period. To determine whether De-Adaptation occurred over the End Phase, the model was re-levelled to make the Adaptation Period the baseline. Results showed that the Adaptation Period was significantly different from the De-adaptation Period [β=7.8, SE=2.2, t=3.6, p<.001], indicating that F1 productions were higher (i.e., shifted in the direction of the Start Phase) in De-adaptation Period compared to the Adaptation Period. For an overview regarding the results in Hz units, please refer to Figure 3.2.

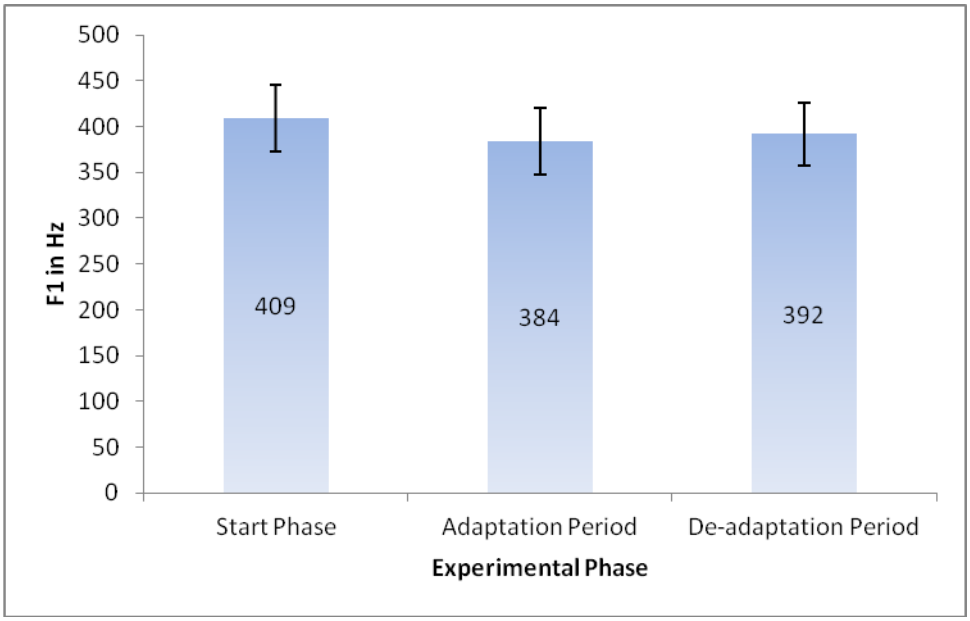

Figure 3.2: Mean F1 values and standard deviations across the Start Phase and the Adaptation and De-adaptation Periods of the End Phase.

3.4.2 Predictors of Adaptation

As outlined in sections 2.3.3.6 and 3.3.2, a series of linear mixed-effect models were fit to the data using the same fixed and random effects as in the analysis of compensation. However, no predictors of adaptation to auditory vowel perturbation were found for the group.

3.5 Discussion

The purpose of the current study was to determine whether speakers of NZE termed *big compensators* as per the preceding study of compensation also adapt to auditory vowel perturbation, and if so, establish possible predictors of adaptation. Moreover, the study aimed to determine whether a process of de-adaptation is initiated in this group following the return of normal feedback. It was hypothesised that both adaptation and de-adaptation would be evidenced in this group of speakers. In line with the study hypothesis, results showed that *big compensators* indeed adapted to auditory vowel perturbation. That is, when auditory

perturbation was removed after prolonged auditory perturbation, F1 values throughout the End Phase remained significantly different from the Start Phase. Significant differences between the Adaptation Period (i.e., first half of the End Phase) and the De-adaptation Period (i.e., second half of the End Phase) provided support of the second study hypothesis, indicating that a process of de-adaptation had been initiated. Lastly, it was hypothesised that predictors of compensation would transfer to the adaptation experiment due to the established correlation between magnitudes of compensation and adaptation in previous research. However, this hypothesis was rejected as no predictors of adaptation to auditory vowel perturbation could be identified for this group of *big compensators*. In line with the study hypothesis, age was not found to influence adaptation. Following each of these results will be discussed in turn.

Overall, the current results regarding adaptation and de-adaptation are comparable to previous research in AE and CE (Houde & Jordan, 2002; Mitsuya, et al., 2011; Pile, et al., 2007, February; Purcell & Munhall, 2006a; Rochet-Capellan & Ostry, 2011a; Shiller, et al., 2009; Shum, et al., 2011; Villacorta, et al., 2007). Bearing in mind the small number of speakers who qualified for the study of adaptation, these results provide further support for the theory underlying the DIVA model of speech production (Guenther, et al., 1998; Perkell, 2012). According to DIVA, speech sounds are planned in auditory-perceptual space based on an auditory teaching signal. The link between the acoustic target sound and proprioceptive information required to produce that sound is stored in feedforward commands, which in turn, are adaptable to sensory changes. In the current study, altered auditory-sensory feedback changed the mapping between the acoustic target sound and the articulatory realisation of that sound, thereby inducing motor-perceptual learning. This mapping was reversed when auditory feedback returned to normal. The current findings therefore provide support for the theory-based notion that motor and perceptual learning are not separate but indeed process together. Computational modelling and testing is required, however, to confirm DIVA-driven hypotheses regarding integrational processes of speech production in older adults.

More interesting perhaps is the rejection of the third study hypothesis that the predictors of compensation for this group (namely more centralised F1 HEAD productions relative to HEED and HAD and higher variability in F1 in the Pre Phase) would carry over to adaptation. Previous research had established a positive link between the magnitude of compensation and adaptation. Hence, it was hypothesised that predictors associated with

higher magnitudes of compensation would likewise predict adaptation. From a theoretical point of view, it was proposed that a reduced likelihood of mismatching error signals in somatosensory feedback would enhance speech motor learning. The current study suggests, however, that these predictors held more value for the direction and magnitude of corrective articulatory movements (i.e., compensation) as opposed to maintaining the revised speech motor pattern. One possible explanation for this finding is that somatosensory feedback was directly related to the toleration of altered speech motor commands, and thus, highlighted in the compensation experiment. Since adaptation and de-adaptation are articulatory processes operating towards the opposite direction of compensation, the influence of somatosensory feedback may therefore be less pronounced. In other words, recovered auditory feedback induces a continuous progression of the revised speech motor plan to move towards the original speech motor plan. Therefore, it may be harder to establish strong relationships between factors influencing adaptation. This would also explain the lack of reported influences on adaptation compared to compensation in past research.

An obvious limitation to the interpretability of the absence of predictors in the current study is the number of speakers available to explore the phenomenon of adaptation. Given the number of independent variables, a larger group of speakers is needed to confirm the current finding and shed light onto the theory that predictors of adaptation are perhaps not as easy to establish compared to predictors of compensation. This also applies for the finding that age was not a factor influencing adaptation. Another limitation is concerned with the interpretability of the current results against previous reports of AE and CE. For these to be validated, it is essential for future studies to directly compare responses of NZE with those of AE and/or CE speakers to allow for distinct evaluations and interpretations of the results. Additionally, it would be worth exploring possible reasons for the increased variability in responses across Stay and Stay2 phases of the experiment. This phenomenon was observed across all four subgroups but specifically in four of the 20 speakers within the group of *big compensators*, which consequently were removed from the analysis. It appears reasonable to assume that the Vowel Identification Task at the end of the Stay Phase may have affected responses to auditory perturbation in the Stay2 Phase. Shiller and colleagues (2010; 2009), who employed a similar perturbation procedure, did not report any negative effects of the vowel identification task on compensation, adaptation or response pattern across Stay and Stay2 phases. Due to the NZE front vowel shift, however, it could be hypothesised that NZE speakers are more vulnerable to vowel identification tasks in that exposure of a vowel

continuum may for some speakers, result in perceptual confusion or even changes to the acoustic target sound. Consequently, auditory perturbation following vowel identification tasks may yield less consistent and predictive responses. Future research, therefore, should focus on conducting experiments with and without vowel identification tasks and comparing variability in response pattern across the two conditions. That said, the increased variability in response patterns in the compensation study would suggest a more in-depth focus on factors stabilising the perturbation design for compensation in NZE speakers first before further investigations into adaptation follow.

In summary, the present study of NZE speakers revealed that compensatory response pattern secondary to auditory vowel perturbation persisted even when the source of auditory feedback manipulation was removed. However, no factors influencing adaptation were found. Given the small fraction of speakers who qualified for the adaptation study, however, conclusions and interpretations remain tentative for NZE until more research has been conducted to support the present findings. Moreover, factors employed to investigate speech motor learning were not exhaustive and more suitable discriminators may still be established, particularly in the context of ageing. Taken together, results of the first two phases of this study were less clear than expected, which may be due to a number of methodological and dialectal reasons. Continuous research of factors influencing speech motor learning is required to inform and enhance speech rehabilitation in the elderly and those with neurological impairments.

CHAPTER FOUR

Error-correction Mechanisms in Speakers with Dysarthria:

A Mimicry Study

4.1 Abstract

An increasing body of research suggests a decline in auditory-perceptual abilities in speakers with dysarthria (e.g., Beijer, et al., 2011) and Parkinson's disease (PD) (e.g., Artieda, et al., 1992; Troche, et al., 2012; Vitale, et al., 2012). While speakers with dysarthria can modify their speech to approximate external cues (e.g., Wang, Kent, Duffy, & Thomas, 2005; Wertz, Henschel, Auther, Ashford, & Kirshner, 1998), it is currently unknown if, and how, auditory-perceptual declines affect the accuracy of cued speech production. The purpose of the present study was therefore to determine whether speakers with PD and dysarthria are able to effectively integrate speech perception and speech production to match an acoustic target. Eight speakers with PD and hypokinetic dysarthria and eight healthy age-matched controls (HC) completed a mimicry task, which consisted of three consecutive attempts to approximate the speech signal of a model speaker. Following this, speakers perceptually evaluated how well their third mimicry attempt matched the acoustic target, using a linear scale from 'not at all' to 'identical'. Finally, speakers re-produced the target speech pattern with the auditory stimulus removed. Four acoustic distance measures were calculated for each mimicry attempt to indicate how closely the speakers' productions matched the model speaker. It was hypothesised that speakers with dysarthria would consistently exhibit greater acoustic distances between their production and the target, and further, evaluate their performance less accurately when compared to HC. Linear mixed-effect modelling revealed that all speakers modified their speech production to approximate the model speaker but did not move significantly closer to the target over the three mimicry attempts. Perceptual self-ratings were inconsistent with the acoustic evaluations of mimicry 'goodness', indicating a possible mismatch between speech perception and speech production. Speakers successfully re-produced the target speech acoustics in the absence of the stimulus except for the acoustic dimension of duration, which deteriorated in the direction of speakers' baseline productions. No difference in mimicry behaviour both perceptually and acoustically was observed across the two groups—within their speech motor abilities, all speakers showed the same response patterns. The present findings are discussed with a particular focus on evaluating the use and benefit of mimicry tasks for future research in dysarthria management.

4.2 Introduction

Dysarthria is a neurological speech disorder commonly caused by stroke, traumatic brain injury (TBI), and degenerative diseases such as Parkinson's disease (PD) (Duffy, 2005). Neuromuscular deficits resulting in weakness, slowness, and incoordination negatively affect the speed, strength, range, timing, and/or accuracy of speech muscle movement (Duffy, 2005). Thus, the primary focus of research and rehabilitation management has been on the speech production characteristics of individuals with dysarthria. While evidence exists that speech intelligibility can be enhanced through the modification of rate and loudness (Yorkston, Hakel, et al., 2007; Yorkston, et al., 2003), evidence for the maintenance and generalisation of these immediate treatment outcomes is scarce (Fox, et al., 2002; Ramig, et al., 2001; Solomon, et al., 2001; Wenke, et al., 2008).

In light of the growing body of research emphasising the significance of error-detection and error-correction mechanisms for accurate speech production (e.g., Guenther, et al., 1998; Perkell, 2012; Perkell, et al., 1997; Villacorta, et al., 2007), an important question should be asked: Are individuals with dysarthria able to use auditory-perceptual resources to inform and modify their own speech production? This question is particularly important from a motor learning point of view as research suggests that sensory experiences help shape skilled motor movement (Catmur, 2013; Kent, 2000; Maas, et al., 2008). Maas and colleagues (2008) particularly emphasised that the amount of information that is available and interpretable to a speaker may predict the amount of learning and the degree to which an individual may benefit from behavioural intervention. Similarly, the DIVA model of speech production proposes that a sound must first reliably be perceived and identified before accurate and consistent productions of that sound can be learnt (Guenther, 1995b). Consequently, the use of speech motor learning, and thus, the success of speech rehabilitation may be limited if treatment – aimed at modifying clinical features – failed to recognise possible impairments in the integrational processes of speech perception and speech production.

Current research lends support to the notion that auditory-perceptual abilities may be negatively affected in individuals with dysarthria (Beijer, et al., 2011) and those with PD (Abbs & Gracco, 1984; Adams, et al., 2006; Artieda, et al., 1992; Dagenais, et al., 1999; Ho, et al., 2000; Ho, et al., 1999; Pell & Leonard, 2003; Troche, et al., 2012; Vitale, et al., 2012). Moreover, the Lee Silverman Voice Treatment® (LSVT) has acknowledged a mismatch between speech perception and speech production in speakers with PD and dysarthria, and

further, emphasised that reduced self-perception of loudness may negatively affect the intended speech output (e.g., Fox, et al., 2002). Importantly, recent research suggests that PD is also associated with hearing loss and that auditory-sensory impairments require further evaluation in speakers with PD (Vitale, et al., 2012). Similar results have been found in a study investigating speakers with acquired dysarthria secondary to stroke, PD, or encephalitis (Beijer, et al., 2011).

Error-detection and error-correction mechanisms as outlined in the DIVA model have typically been tested and validated through experimental procedures underlying an auditory perturbation paradigm. However, the implementation of an auditory perturbation design in phase one and two studies of the current thesis has resulted in unexpected findings, which may be related to a combination of methodological and dialectal reasons. The DIVA model of speech production proposes that speech sounds are learnt by comparing an external signal to that of one's own production. When mismatches between these two signals are detected, articulatory adjustments are made to match the target sound. Thus, DIVA proposes that mimicry is a natural part of speech motor learning through which error-detection and error-correction mechanisms are learnt and applied. Moreover, mimicry, modelling, and demonstration are commonly applied in the rehabilitation of speech motor disorders (Duffy, 2005; McNeil, 2009). Thus, mimicry represents an alternative approach to auditory perturbation in exploring the link between speech perception and speech production and may further inform and enhance current clinical practices. Research that used mimicry as a cueing mechanism to prompt modifications in rate (D'Innocenzo, Tjaden, & Greenman, 2006; Weismer, Laures, Jeng, Kent, & Kent, 2000), loudness (Adams, et al., 2006), and prosody (Wang, et al., 2005; Wertz, et al., 1998) suggests that individuals with dysarthria can integrate auditory-perceptual information to modify their speech towards that of the target stimulus although less so than age-matched HC. However, these findings are only based on comparisons of selected output measures between speakers with dysarthria and HC, independent and unrelated to the acoustic target. That is, none of these studies has specifically investigated the ability of speakers with dysarthria to approximate the speech acoustics of a model speaker.

The purpose of the current study was to determine the effect of neurological disease on error-detection and error-correction mechanisms. In particular, this investigation focused on speakers with PD as basal ganglia deficiencies have specifically been hypothesised to

negatively affect the use of auditory feedback for skilled motor movement (e.g., Duffy, 2005). The DIVA model of speech production (Guenther, et al., 1998) was applied as a theoretical foundation to systematic behavioural experimentations underlying mimicry, investigating whether speakers with dysarthria are able to (1) perceive quantitative differences in their own speech production when compared to a target stimulus (i.e., error-detection), and as a result, (2) effectively integrate speech perception and speech production as observed in modifications in speech production to better match the target (i.e., error-correction). Specifically, the following questions were investigated: Are speakers with PD and dysarthria and age-matched controls able to (1) modify their speech to approximate the acoustic target of a model speaker, (2) show continuous improvements in approximating the target stimulus, (3) make accurate judgements comparing their own speech to that of the target stimulus, and (4) maintain the mimicry speech pattern in the absence of acoustic cueing. Note that no computational modelling was undertaken—DIVA was solely applied as a theoretical framework to the development and discussion of the presented study.

In regard to the DIVA model, it was hypothesised that all speakers would be able to use the target stimulus as a teaching signal and adjust their speech production to approximate the target, at least to some extent. Furthermore, it was hypothesised that speakers would approximate the acoustic target more closely over the three consecutive trials due to continuous error-corrections and stored information of articulatory adjustments from previous attempts into feedforward models. Likewise, it was hypothesised that even with the stimulus removed, these updated feedforward models would be maintained. The DIVA model does not allow for hypotheses regarding the effect of dysarthria as it is based on young healthy individuals. In line with previous research, however, it was hypothesised that speakers with dysarthria would show greater acoustic distances to the model speaker and less accurate self-ratings compared to age-matched controls.

4.3 Method

All procedures were reviewed and approved by the University of Canterbury Human Ethics Committee and the New Zealand Human Disability Ethics Committee and written consent was obtained from all participants. All speakers received a $10 voucher as compensation for participation.

4.3.1.1 Participants: Talkers

A total of eight speakers with dysarthria and eight healthy age-matched controls completed the study. The eight speakers with PD were aged 65-83 years (mean=72.5, SD=5.54, female=4, male=4, NZE speakers =7, British English =1). Speakers were recruited through the New Zealand Institute of Language, Brain and Behaviour database (N=6) and personal contacts of the researcher (N=2). Table 4.1 contains biographical information for the speakers with PD. Dysarthria diagnosis, severity of the dysarthria, and deviant perceptual speech characteristics were defined through consensus judgement of three speech-language therapists (SLTs) based on audio recordings of a conversational speech sample, "The Rainbow Passage", diadochokinesis, and phoneme prolongations.

A group of eight neurologically healthy speakers (HC) matched for age and gender were studied for comparison, all of whom spoke NZE as their first language. They were recruited from the New Zealand Institute of Language, Brain and Behaviour database (N=5) and friends (N=3). A standard pure-tone auditory hearing screen was used to binaurally screen speakers for hearing loss at 500 Hz, 1 kHz, 2 kHz, and 4 kHz. In line with previous research (e.g., Vitale, et al., 2012), hearing loss was apparent for at least one frequency in all but one speaker with PD. Similarly, healthy controls revealed hearing loss for at least one frequency. Since the current study was conducted in a free field with loudspeakers, and all speakers had the opportunity to adjust the presentation level to a comfortable listening loudness, none of the 16 speakers was excluded based on his or her hearing screen results. In addition to the hearing screen, healthy controls were screened for mild cognitive impairment using the Montreal Cognitive Assessment (MoCA) (Nasreddine, et al., 2005). All healthy speakers passed the MoCA with scores higher than 26. The MoCA was set aside for speakers with PD due to the high incidence of cognitive impairment in this group (Ehrt & Aarsland, 2005).

Table 4.1: *Characteristics of speakers with PD*

Speaker	Age (PD)	Age (HC)	Years post diagnosis	Hoehn and Yahr (1967)	Dysarthria diagnosis	Dysarthria severity	Deviant perceptual speech characteristics
1	70	71	8	II-III	hypokinetic	moderate	hoarseness, breathiness, monoloudness, reduced stress and loudness, low pitch, variable rate, inappropriate silences, phoneme and syllable repetition, vocal tremor, imprecise articulation
2	83	82	1	I	hypokinetic	mild	inappropriate pauses, phoneme and syllable repetition, imprecise consonants, breathiness, voice tremor
3	67	66	6	I	hypokinetic	mild	increased rate, inappropriate pauses, phoneme and syllable repetition, harsh/strained voice, breathiness
4	77	79	10	I-II	hypokinetic	mild	breathiness, low pitch, pitch breaks
5	73	73	- *	IV	hypokinetic	moderate	increased rate with short rushes of speech, mono/reduced loudness, monotone, breathiness, phoneme repetition, imprecise articulation, voice tremor
6	72	72	4	III-IV	hypokinetic	mild	strained voice, monotone, monoloudness, lengthened and shortened syllables, imprecise consonants
7	73	74	10	III	hypokinetic	mild-moderate	monoloudness, reduced stress, breathiness, imprecise consonants, phoneme and syllable repetitions, inappropriate pauses, short rushes of speech, vocal tremor
8	66	65	6	II-III	hypokinetic	mild-moderate	increased rate with short rushes of speech, mono/reduced loudness, breathiness, reduced stress, inappropriate pauses, phoneme and syllable repetitions, imprecise articulation

*not obtained

4.3.1.2 Participants: Listeners

Twenty native NZE listeners aged 21 to 43 years (5 male, 15 female, mean=30, SD=7) with no reported history of neurological, cognitive, or speech and language impairment were recruited for the perceptual evaluation of the mimicry task to complement the acoustic analysis. Listeners were recruited from the New Zealand Institute of Language, Brain and Behaviour database (N=6), speech-language therapy and audiology classes at the University of Canterbury, New Zealand (N=8), and the author's friends (N=5). Prior to undertaking the experiment, all listeners were binaurally tested for hearing loss and screened for mild cognitive impairment (i.e., MoCA). All listeners passed the standard pure-tone auditory hearing screen at 25 dB for 500 Hz and 20 dB for 1 kHz, 2 kHz, and 4 kHz. No mild cognitive impairments were identified for the group, with MoCA scores higher than 26.

4.3.2 Speech Stimuli

One female and one male NZE speaker, aged 59 years and 41 years respectively, with no reported neurological, cognitive, or speech and language impairment, provided the speech stimuli for the mimicry task. Both of these speakers were chosen based on their animated reading styles when reading "The Grandfather Passage" for the phase one and two studies of this body of research. The recordings were completed in a sound-attenuated testing room within the Communication Disorders Research Facility at the University of Canterbury, Christchurch, New Zealand. A self-mounted headset with an adjustable microphone (Audio-Technica AT803) attached and positioned at a fixed distance of approximately 10 cm from the speaker's mouth was used for all audio recordings. All speech samples were recorded with a TASCAM recorder (HD-P2) and then digitized into .wav files at a sampling rate of 44.1 kHz (24-bit resolution).

Research has shown that different verbal instructions elicit changes in selected acoustic measures of a reading task (Lam, Tjaden, & Wilding, 2012). Thus, model speakers were first asked to read "The Grandfather Passage" and "The Rainbow Passage" in a normal conversational rate and loudness, as if speaking to a friend or family member. And then, model speakers were instructed to read "The Rainbow Passage" again but in an animated and engaging way, as if reading it to a child. This was done to obtain maximal varied intonation contours for the mimicry task. The recordings of "The Grandfather Passage" were scaled and

then calibrated to an equated loudness level of 65 dB and presented for volume adjustment purposes at the beginning of the mimicry task (see Section 4.3.3) and the perceptual task (see Section 4.3.4). The purpose of reading "The Rainbow Passage" in a normal conversational rate and loudness was to familiarise the model speakers with the passage, and thus, was not further used for any of the mimicry or perceptual tasks. The animated reading style of "The Rainbow Passage" served as the basis for creating the mimicry stimuli. The digital speech files of these recordings were segmented into 32 single phrases (16 for each speaker), scaled and then calibrated to an equated loudness level of 65 dB using Praat software (Boersma, 2001). Note that no attempts were made to ensure that male and female models were comparable regarding prosodic variation.

4.3.3 Mimicry Task

The mimicry task was conducted at a sound-attenuated testing room within the Communication Disorders Research Facility at the University of Canterbury, Christchurch, New Zealand. Voice recording procedures were identical to those described in Section 4.3.2. Before the actual mimicry task, speakers were asked to read "The Rainbow Passage" in a normal conversational rate and loudness, as if speaking to a friend or family member. Following this, speakers listened to the recording of "The Grandfather Passage" through loudspeakers (Logitech Stereo Speakers x140) and were asked to adjust the volume knob to a comfortable listening level. For each experimental run, male speakers listened to the male model speaker and female speakers listened to the female model speaker. All sound files were played from a laptop (Toshiba, Portege M 600) using the Microsoft Windows Media Player for Windows XP. The 16 phrases of "The Rainbow Passage" used for the main mimicry experiment (see Appendix B) were played in sequential order to allow for contextual and intonational flow. For this, speakers were told that they would hear sentences produced by a model speaker and were asked to repeat them exactly as they heard them. Research has shown that the imitation of pitch was enhanced when speakers were given orthographic information (Mantell & Pfordresher, 2013). Hence, speakers were able to read each of the 16 sentences of "The Rainbow Passage" on a monitor placed in front of them while listening to the relevant speech stimulus. This was also done to reduce any negative effects of cognitive load on mimicry performance. Altogether, speakers were given three attempts to mimic exactly how the model speaker sounded. That is, speakers repeated the sentence immediately

after hearing the stimulus sentence and this procedure was repeated three times. After the third attempt, they were asked to indicate on a scale from ‘not at all’ to ‘identical’, using a computer mouse, how much they thought their last attempt sounded like the target sentence. The display of each of the sentences and the scale for the perceptual judgements of their third attempt to mimic the model speaker was custom-programmed, using ALVIN software (Hillenbrand & Gayvert, 2005). Lastly, speakers were asked to re-produce how the model speaker sounded without listening to the stimulus sentence but still with the written sentence provided. This final repetition was used to analyse how much speakers were able to maintain the mimicry speech pattern in the absence of an audio stimulus. Altogether, the mimicry task therefore involved four productions of each of the 16 sentences (three with and one without an auditory stimulus) and one perceptual judgment of their third performance and took approximately one hour to complete.

4.3.4 Perception Task

From the recordings of the mimicry task, a custom-written and -programmed perception task for listeners was created to complement the acoustic analysis. The intent was to account for potential differences in acoustic and perceptual evaluations of mimicry ‘goodness’. The procedure was programmed, using ALVIN software (Hillenbrand & Gayvert, 2005). For this, three of the 16 phrases produced in the mimicry task were randomly selected. Hence, the perception task consisted of the same three sentences produced by each of the sixteen speakers. Each of these sentences then included a number of productions from each speaker; first, speakers’ productions from the mimicry task (N=4 per speaker per sentence), second, their normal conversational readings of “The Rainbow Passage” (N=1 per speaker per sentence), and finally for means of comparisons, productions of the model speakers (N=1 per model speaker per sentence). All productions were scaled and calibrated to an equated loudness level of 65 dB and then combined audio files of the model speaker followed by the speakers’ productions were created, using Praat software (Boersma, 2001). Twenty-five percent of these 240 audio files (5 productions x 3 sentences x 16 speakers) were randomly duplicated to account for intra-listener reliability testing, generating a total of 300 listening files. From these 300 files, twenty distinct protocols were established, all of which were randomised for speakers and sentences. The order of the five representations of each sentence

was randomised once for each speaker and was then used consistently for all twenty protocols to ensure that within-speaker evaluations were comparable across listeners.

Similar to the mimicry task (see Section 4.3.3), listeners were first presented with "The Grandfather Passage" and asked to adjust the volume knob of the loudspeakers to a comfortable listening level before commencing with the perception task. Then listeners were told that they would hear a number of individuals who were asked to mimic a model speaker and attempt to sound as much as the target sentence as possible. Listeners were instructed to rate on a scale from 'not at all' to 'identical', how well the two sentences matched, using a computer mouse. To introduce the listeners to the scale, they were given one audio example of a close to 'identical' and one of a close to 'not at all' audio file. These were selected by the main investigator of this study based on her subjective perception from listening to all of the paired sentences multiple times. In total, the perceptual task took one hour to complete.

4.3.5 Data Analysis

4.3.5.1 Data Processing

Given the preliminary nature of the current study, the acoustic analysis of mimicry performance was based on the more dominant features of speech, rate and prosody. Four automated spectral values were extracted on a frame-by-frame basis from the two model speakers' and each of the 16 speakers' production files, using Praat software (Boersma, 2001). These included: (1) wave duration in seconds – as an indicator of speaking rate, (2) percent of total frames that were voiced – as an indicator of speech rhythm, (3) mean fundamental frequency (F0) for voiced frames – as an indicator of habitual pitch, and (4) standard deviation of F0 for voiced frames – as an indicator for pitch variation. In order to evaluate speakers' productions relative to the target stimuli, one distance measure per sentence and speaker was computed for each of these four acoustic values. For this, the male speakers' values were subtracted from the male model speaker and female speakers' values were subtracted from the female model speaker. The absolute values of these distance measures were used as the dependent variable in all subsequent statistical analyses. In addition, all distance measures were coded to a time reference. The normal passage reading served as the baseline, and the four mimicry attempts were coded M1 to M4.

A severity measure of the dysarthria was chosen to distinguish between speakers with PD and healthy controls. This was preferred over a simple between-group distinguisher because it was considered a more informative measure of the actual speech impairment, and thus, a more valuable measure in exploring possible effects of speech impairment on mimicry performance. Four levels of dysarthria severity were applied; (1) normal (N=8), (2) mild (N=4), (3) mild-moderate (N=2), and (4) moderate (N=2).

The perceptual self-ratings of how well speakers' third mimicry attempt matched the target sentence were based on a ten-centimetre scale, zero referring to 'not at all' and 1000 referring to 'identical'. For standardisation purposes, all 16 ratings per speaker were z-scored within speakers prior to statistical analysis. Similarly, perceptual ratings of the 20 listeners were z-scored based on the average and standard deviation of the individual listener's ratings. For the statistical analysis, however, an average across all listeners was taken for each production, resulting in a single z-score to match the acoustic analysis of that specific production. The test-retest reliability coefficient for listeners perceptual judgements (based on 25% repeated items) was .77, which represents a good-to-acceptable test-retest reliability score according to Cronbach's alpha internal consistency table (Portney & Watkins, 2009).

4.3.5.2 Statistical Procedure

Speakers' mimicry performance was evaluated on acoustic measures and perceptual ratings. The acoustic analyses were performed on the dimensions of duration, voicing, mean F0, and F0 standard deviation. For each of these dimensions, the differences between the model's utterances and the speaker's utterances were used to calculate distance measures (see Section 4.3.5.1). The perceptual analysis focused on how well perceptual ratings of both speakers and listeners matched these distance scores (for a more detailed description of these perceptual ratings see Section 4.3.5.1). Statistical analysis of the mimicry behaviour of speakers with dysarthria and HC was performed using linear mixed-effect models (Baayen, Davidson, & Bates, 2008). The analyses were performed using the lmer function from the lme4 library (Bates, et al., 2011) in the R programming language. Liner mixed-effect analyses allow for inclusion of multiple random effects—in this case, allowing for the simultaneous analysis of subject, item, and pair (i.e., speaker with PD and age/gender matched HC). Additionally, mixed-effect models can incorporate individual differences in

random effects using random slopes—in the present analysis, a random slope of time per subject was included, thereby allowing each subject to have a unique effect of time on performance. A series of mixed-effect models were fit in each of the analyses and evaluated by model fitness tests using the Bayesian Information Criterion (Schwarz, 1978). Analyses began with the fixed factors of time (i.e., baseline to fourth mimicry) and severity of the dysarthria (i.e., 1 to 4); however, fixed factors of the final models varied across analyses and are reported in the relevant result sections. Model fitness tests favoured a simple random effect structure of subject and item over the inclusion of a random slope of time in all of the reported models. Further random slopes were not found to qualitatively change the final models, and hence, were dropped from the reported models.

4.4 Results

4.4.1 Can speakers with PD and HC modify their speech to approximate the target stimulus speech acoustics?

Linear mixed-effect modelling for three of the four dependent variables included the fixed effects of time and severity of dysarthria, with severity of dysarthria not qualifying as a significant predictor of distance (i.e., mimicry accuracy) in the F0 standard deviation model. The results were consistent across the four resulting models, indicating that speakers with dysarthria and healthy controls both successfully modified their speech in the mimicry task to approximate the acoustic signal of the model speaker. That is, the distance between the speakers' productions and the model was significantly greater in the speakers' baseline productions compared to their mimicry productions for the analysis of duration, percent voiced, and mean F0. The analysis of the standard deviation of F0 revealed that speakers had significantly less variation in their F0 contour in the baseline compared to the mimicry productions, indicating that speakers successfully increased their pitch variation to mimic the model speaker. In addition, three of the models revealed that distances between the speakers' productions and the model speaker increased with severity of dysarthria for the acoustic elements of duration, percent voiced, and mean F0. This suggests that while overall speakers were able to change their speech, speakers with moderate levels of dysarthria were consistently further away from the model speaker for all production trials. Table 4.2 comprises of the linear mixed-effect results for the four models.

Table 4.2: *Linear mixed-effect results for the four models comparing the absolute distance of the four mimicry attempts relative to the model speaker to the absolute distance of the baseline reading relative to the model speaker.*

Fixed effects	Estimate (β)	Standard Error	*t*-value	*p*-value
Duration				
(Intercept)	.7	.11	6.58	<.0001
Mimic 1	-.39	.03	-12.87	<.0001
Mimic 2	-.4	.03	-13.09	<.0001
Mimic 3	-.35	.03	-11.56	<.0001
Mimic 4	-.3	.03	-9.7	<.0001
Dysarthria Severity	.08	.04	2.35	<.02
Percent Voiced				
(Intercept)	.08	.01	6.3	<.0001
Mimic 1	-.02	.01	-5.22	<.0001
Mimic 2	-.02	.01	-5.09	<.0001
Mimic 3	-.02	.01	-4.28	<.0001
Mimic 4	-.02	.01	-3.8	<.001
Dysarthria Severity	.01	.005	2.68	<.01
F0 mean				
(Intercept)	13.34	3.82	3.5	<.001
Mimic 1	-2.94	.89	-3.31	.001
Mimic 2	-1.79	.89	-2.01	.04
Mimic 3	-3.27	.89	-3.67	<.001
Mimic 4	-2.53	.89	-2.84	<.01
Dysarthria Severity	3.46	1.45	2.38	<.02
F0 standard deviation				
(Intercept)	21.97	4.14	5.31	<.0001
Mimic 1	3.45	1.57	2.2	<.03
Mimic 2	5.19	1.57	3.32	<.001
Mimic 3	4.38	1.57	2.8	<.01
Mimic 4	3.06	1.57	1.95	.05
Dysarthria Severity	-	-	-	-

4.4.2 Do speakers with PD and HC show improvement in mimicry performance over time?

The linear mixed-effect models described above were re-levelled to make the first mimicry attempt (M1) the baseline for comparisons with the second and third mimicry attempts. All four models were consistent in their finding that mimicry performance did not significantly change over time. Similarly, when the models were re-levelled again to make the second mimicry attempt (M2) the baseline for comparisons with the third mimicry attempt, results showed no significant difference in mimicry performance between the second and third mimicry attempt. In addition, both of the re-levelled models were consistent regarding the predictive value of severity of dysarthria. That is, modelling of the dependent variables of duration with M1 as the baseline [β=.07, SE=.03, t=2.02, p=.04] and M2 as the baseline [β=.07, SE=.03, t=1.93, p=.05], modelling of the dependent variable of percent voiced with M1 as the baseline [β=.01, SE=.005, t=2.42, p=.02] and M2 as the baseline [β=.01, SE=.005, t=2.44, p<.02], and modelling of the dependent variable of mean F0 with M1 as the baseline [β=3.34, SE=1.48, t=2.25, p<.03] and M2 as the baseline [β=3.22, SE=1.46, t=2.2, p<.03] revealed that as severity of dysarthria increased, so did the distance of the speakers' productions to the model speaker. Severity of dysarthria did not affect F0 standard deviation.

4.4.3 Do speakers with PD and HC accurately judge their success in mimicking the target sentence?

Analysis was based on speakers' perceptual judgements of their third attempt to mimic the model speaker (z-scored) and the distance measures of speakers' third mimicry attempt relative to the model speaker. Linear mixed-effect modelling revealed that perceptual judgments could not be predicted by any of the four acoustic distance measures, indicating that perceptual ratings were inconsistent with the acoustic distance evaluations of how closely the speakers matched the target sentence. Severity of dysarthria was non-significant, indicating that the inconsistency between perceptual and acoustic evaluations of mimicry success was not affected by the presence or degree of dysarthria.

4.4.4 Can speakers with PD and HC maintain the mimicry speech pattern in the absence of the acoustic stimulus?

The linear mixed-effect models described in Section 4.4.1 were re-levelled to make the third mimicry attempts (M3) the baseline for comparisons with the fourth mimicry attempts (M4). Only the model of duration showed a significant difference between M3 and M4 [β=.06, SE=.03, *t*=2.14, *p*=.03], indicating that the distance between the speakers' productions and the model speaker increased in M4 compared to M3. For the other three dependent variables, no significant differences between M3 and M4 were found, indicating that the distance between the speakers' productions and the model speaker did not change when speakers re-produced the mimicry speech pattern without listening to the target stimulus. In addition, comparisons of the slopes of Mimic 3 and Mimic 4 in Table 4.2 revealed the direction in which M4 was moving relative to the baseline. For the dependent variable of duration, the slope of M4 indicated that duration was moving closer to the baseline, and hence, away from the model speaker. Severity of dysarthria was predictive for the models of duration [β=.1, SE=.04, *t*=2.66, *p*<.01], percent voiced [β=.01, SE=.005, *t*=2.74, *p*<.01], and mean F0 [β=3.31, SE=1.43, *t*=2.31, *p*=.02], indicating that as severity of dysarthria increased, so did the distance of speakers' productions to the model speaker. Severity of dysarthria did not affect F0 standard deviation.

4.4.5 Do listeners' judgements of mimicry quality match acoustic distance evaluations?

Linear mixed-effect modelling revealed that listeners' perceptual judgements were predictable by the distance measure of duration [β=-.61, SE=.06, *t*=-9.8, *p*<.0001] and percent voiced [β=-1.43, SE=.066, *t*=-2.18, *p*=.03], indicating that as distances between the acoustic measures of the speakers' productions and the model speaker increased, z-scores decreased. In other words, the poorer the mimicry (defined as acoustic distance from the target), the less identical the judged perception. No predictive value was found for the distance measure of mean F0 and F0 standard deviation. Moreover, listeners perceptual ratings were predictable by the severity of dysarthria for the analyses of duration [β=-.19, SE=.03, *t*=-6.78, *p*=.0001], percent voiced [β=-.24, SE=.04, *t*=-6.68, *p*=<.0001], and mean F0 [β=-.26, SE=.03, *t*=-8.5, *p*<.0001], and F0 standard deviation [β=-.26, SE=.03, *t*=-8.5, *p*<.0001], indicating that as severity increased, z-scores decreased. In other words, listeners

rated speakers with moderate dysarthria as the least perceptually identical to the model speaker.

4.5 Discussion

The aim of the present study was to determine whether speakers with PD and dysarthria and HC were able to effectively integrate speech perception and speech production. Particularly, the current study focused on determining whether speakers with PD and dysarthria and HC are able to (1) modify their speech to approximate the speech acoustics of a target stimulus, (2) show improvements in mimicry performance over time, (3) make accurate judgements of their success in mimicking the target stimulus, and (4) maintain the target speech pattern in absence of the stimulus. It was hypothesised that speakers with PD and dysarthria would consistently show greater difficulty in accomplishing these tasks compared to HC. This hypothesis was partially supported. Results were consistent for speakers with PD and HC in that (a) both successfully modified their speech production to approximate the model speaker on all four acoustic dimensions investigated in the current study, (b) no significant change or improvement was found over the three mimicry attempts, (c) perceptual self-ratings did not match acoustic evaluations of distances between speakers' productions and the model speaker, and (d) speakers' third mimicry attempts were not found to significantly differ from their fourth independent mimicry attempts—except for the acoustic dimension of duration, which deteriorated in the direction of speakers' baseline productions. In addition, the productions of speakers with moderate levels of dysarthria were consistently found to be further away from the model speaker when compared to HC and speakers with PD and mild dysarthria. That was found to be true for all of the acoustic dimensions, except for pitch contour. Similarly, analyses of listeners' perceptual ratings revealed that speakers with moderate levels of dysarthria were rated as significantly less identical to the target stimulus on baseline and mimicry production trials compared to speakers without dysarthria and those with mild dysarthria. Inconsistencies in the perceptual and acoustic evaluations of 'mimicry goodness', however, were equally true for speakers with dysarthria and HC. Each of these findings will be discussed in the following paragraphs.

The current finding regarding the ability of speakers with dysarthria and HC to modify their speech towards the acoustic target of a model speaker is in line with previous research in

dysarthria, using mimicry to elicit modifications in specific aspects of speech production, such as rate (D'Innocenzo, et al., 2006; Weismer, et al., 2000), loudness (Adams, et al., 2006), and prosody (Wang, et al., 2005; Wertz, et al., 1998). Note, however, that the comparability of these studies is limited—on the one hand, due to differing aetiologies and types of the dysarthrias, and on the other hand, due to differing acoustic dimensions and analyses employed. From a theoretical point of view, the significance of this finding is that speakers with PD and dysarthria appear to successfully integrate the auditory-perceptual information of a teaching signal (i.e., the mimicry target) into feedforward speech motor commands. Thus, it can be assumed that auditory-perceptual abilities and sensorimotor control required for error-detection and error-correction mechanisms as outlined in the DIVA model are at least partially intact and functioning in these speakers with PD.

Perhaps surprising is the finding that speakers did not show significant changes in mimicry performance over time, suggesting limitations in the flexibility and accuracy of error-detection and error-correction mechanisms. In accordance with the theoretical framework of the DIVA model, at least three levels of possible breakdown can be anticipated to explain the current results. On a primary level, the ability to identify and distinguish between various segmental and suprasegmental features of the speech production of a model speaker can be considered. The fact that speakers modified rate, rhythm, pitch, and intonation to at least some extent suggests a certain level of awareness regarding what aspects of speech production required change to approximate the model speaker. However, it is possible that speakers still lacked attentional focus, and hence, consistency towards the correct assessment of which finer auditory-perceptual details to attend to in their attempt to mimic the model speaker. Additionally, previous research in speakers with dysarthria (Beijer, et al., 2011), PD (e.g., Artieda, et al., 1992; Troche, et al., 2012), and older adults (e.g., Fitzgibbons & Gordon-Salant, 1998; Fitzgibbons & Gordon-Salant, 2004; Gordon-Salant, 2008, November; Kiessling, et al., 2003; Sommers, 2008) lends support to the notion that speakers capacity to detect and distinguish finer differences in the acoustic signal may have been limited. In line with the theory underlying the DIVA model that speech sounds are planned in auditory-perceptual space, it can therefore be assumed that any auditory-perceptual declines are likely to negatively affect speech motor control. On a secondary level, it is important to take into account the ability to compare one's own speech output to that of the model speaker and accurately evaluate any differences in speech acoustics. The current finding that speakers' perceptual ratings were not in agreement with the acoustic evaluations of how well their

speech matched the model speaker may suggest a breakdown on this level. Additional confirmation of this theory is provided by past research suggesting difficulties in speakers with PD and hypophonia to accurately judge their own speech output (Fox, et al., 2002; Ho, et al., 2000). On a final level, it is possible that speakers were able to detect the change but unable to adjust their articulators with the required level of fine motor skills to meet the acoustic target. From a neuromuscular perspective, this would not be surprising for speakers with PD and dysarthria since deficits in adjusting speech muscle movement to modify rate, rhythm, and elements of prosody are anticipated. While less is known about these processes in older adults, there is some research that suggests at least some declines in motor speech control (e.g., Ballard, et al., 2001; Goozee, et al., 2005; Price & Darvell, 1981; Smiljanic, 2013). Each of these aspects may explain why the current study did not find significant changes in mimicry performance over time. However, it is also possible that the absence of improvement is not necessarily an indicator of a problem. The current findings could also highlight the ability of speakers with PD to make significant changes, and most importantly, maintain these over all of the cued and non-cued iterations. Given the growing evidence of auditory deficits in speakers with PD and those with dysarthria, further investigations are required, however, to clarify if, and how, these deficits affect error-detection and error-correction mechanisms in speakers with dysarthria.

The finding that speakers' perceptual ratings did not match the results of the acoustic analysis requires careful assessment of the methodology employed before further assumptions can be made. It is possible that speakers were uncertain what dimensions to judge their perceptual evaluation by since their attention was not focused towards anything specific. If speakers used different modes of comparisons to the dimensions employed in the acoustic analysis, inconsistencies in 'goodness' evaluations, therefore, would not be surprising. An additional factor to consider is the use and design of the rating scale employed in the current study. Having a continuous scale without marks may have given the perceptual raters less orientation and points of references over consecutive trials. That said, it has been suggested that continuous scales allow for finer discriminations and provide a more uniform distribution of scores over the entire length of the scale (Jilka et al., 2007). Also there appears to be a risk of systematic bias in perceptual raters when using equal-appearing interval scaling (Whitehill, Lee, & Chun, 2002). Further investigations are necessary to shed light onto a possible effect of a lack of attentional focus and rating scale on the current results. With these possible limitations in mind, it is important to note that inconsistencies between

acoustic measures of speech production and self-perception ratings have previously been reported for speakers with PD (Ho, et al., 2000), and thus, warrant further investigations. One way of exploring whether the current results are a reflection of speakers' self-monitoring skills or a limitation of the rating scale would be to compare speaker's evaluations of themselves with those of independent listeners. However, the current study did not permit for such comparisons as the listening task was only based on three of the 16 sentences employed.

A positive finding of the current study is that performance remained constant when the auditory stimulus was removed for all acoustic dimensions except duration. In line with the theory of the DIVA model, this indicates that speech perception and speech production can effectively be integrated to induce speech motor learning, at least for some acoustic measures. However, it is important to note that this assumption is made on a single production trial that occurred immediately after speakers' perceptual rating of their third mimicry attempt. Further research is required to examine retention and generalisation of these newly acquired speech patterns. At this stage, it is not clear why speakers maintained mimicry performance on all of the acoustic dimensions except duration. One possible explanation may be that speaking rate is highly habitual whereas aspects of speech rhythm, habitual pitch and pitch variation may be more consciously controllable. Hence, speaking rate may be more difficult to maintain in the absence of a target stimulus.

Perhaps the most unexpected finding of the current study is that speakers with dysarthria and HC did not differ in their ability to modify, evaluate and maintain their speech production. While acoustic analyses generally found greater distances between speakers with moderate levels of dysarthria and the model speaker compared to HC and speakers with mild dysarthria, a time-by-severity of speech impairment interaction would have been required to qualify actual differences in mimicry behaviour. The same applies to the perceptual analyses, which revealed that speakers with moderate levels of dysarthria were generally less identical to the model speaker but not different in their ability to mimic the model speaker. Preliminary findings of auditory perturbation research in older adults suggest that the ability to modify speech as measured by unconscious corrections to induced speech errors is maintained with age (H. Liu, et al., 2010; P. Liu, et al., 2011). Similarly, compensation and adaptation results of the present thesis are not suggestive of an ageing effect on these mechanisms of speech motor control and subsequent speech motor learning. However, it cannot be entirely ruled out that the present mimicry finding may have been affected by ageing effects. Interestingly,

similar speech characteristics have been reported for individuals with hearing impairments and those with dysarthria—for example monopitch, excessive pitch variation, producing contextually appropriate levels of loudness (Pratt & Tye-Murray, 2009). This may potentially explain why speakers with dysarthria and HC did not differ in their mimicry performance. Further investigations are necessary, however, comparing mimicry performance of older HC to young HC to shed light onto this.

Also, it is perhaps surprising that speakers with moderate levels of dysarthria exhibited greater distances to the model speaker on all acoustic dimensions except for pitch variation. Artieda and colleagues (1992) found that the ability to discriminate temporal non-speech stimuli was significantly worse in speakers with PD compared to age-matched HC and that discrimination skills deteriorated with severity of PD. Furthermore, Beijer and colleagues (2011) found that auditory speech discrimination of speaking rate, overall pitch, intonation, and segmental elements of articulation was significantly affected in speakers who acquired dysarthria subsequent to stroke, PD, or encephalitis. In addition, Vitale and colleagues (2012) found a hearing loss independent and beyond that of ageing in speakers with PD, which was positively correlated with the severity of PD. In regard to the present finding, this may explain why speakers with moderate levels of dysarthria consistently had greater distances to the model speaker on the acoustic dimensions of speaking rate, speech rhythm and habitual pitch. It is unclear, however, why pitch variation was not affected in a similar way. Theoretically, this may indicate that overall, speakers with dysarthria in the current study were more able to discriminate and modify pitch variation compared to the other acoustic dimensions employed. That said, perhaps their attention was most focused on pitch contour while attempting to mimic the model speaker. However, without further investigations into the effect of auditory deficiencies and attentional focus on mimicry performance, no clear conclusions can be drawn.

Several limitations apply to the current study. First of all, small numbers of participating speakers with PD and dysarthria ask for caution when interpreting the results, particularly in regard to differentiating between mimicry performance of speakers with dysarthria and HC. That said, perhaps the acoustic and perceptual measures employed in the current study were not sensitive and/or extensive enough to differentiate between mimicry performance of speakers with dysarthria and HC. Also, using different phrase lengths may have affected the results although this was tended to by using item as a random effect in the linear mixed-effect

models. Moreover, an equal amount of speakers in each of the severity subgroups is generally preferable when evaluating the effect of dysarthria in each of the subanalyses. In addition, it wasn't investigated whether, how long, and what type, of speech intervention speakers with dysarthria had received. However, it is important to acknowledge that the speakers' attentional focus on specific aspects of speech production and speakers' ability to modify their speech accordingly may have been affected by previous experience and exposure to modelling techniques. Importantly, without orthographical listener transcriptions, it is difficult to evaluate if, and how, speakers' modifications in the mimicry task relate to actual changes in their speech intelligibility. Lastly, caution has to be taken in regard to applying the theory of the DIVA model to older speakers and those with dysarthria. The DIVA model is currently based on the research of young healthy speakers and computational modelling on a neurologically fully functioning system. Furthermore, DIVA does currently not distinguish between speech production that is internally generated and speech production that is externally cued (Guenther, et al., 2006). While applying the theoretical framework may inform theory and research of integrational processes of speech production in speakers with dysarthria and HC, in-depth production and neuro-imaging analyses as well as computational modelling of these populations and tasks within the DIVA research laboratory are warranted before these can be confirmed.

A number of future directions can be pursued to further inform the results of the present study. Error-detection and error-correction mechanisms are important for the successful rehabilitation of any type of dysarthria. Thus, future research investigating acquired dysarthrias of different origins is warranted. In addition, the current study only focused on the acoustic dimensions related to rate and prosody. However, dysarthria management typically involves articulation, rate, loudness, and prosody. Clinical practices may therefore benefit from a more detailed acoustic and perceptual analysis. Most importantly perhaps, future research needs to establish the level of functioning on the three main tasks involved in error-detection and error-correction mechanisms, and in addition, investigate if, and how, these can be trained or augmented to effectively reduce acoustic distances to the model speaker. On a primary level, it is vital to explore and distinguish between speech-related discrimination skills and hearing loss and their effect on error-detection. On a secondary level, it needs to be investigated whether speakers with dysarthria can accurately compare and evaluate their own speech production to that of a model speaker in different acoustic dimensions. In this same context, improving self-perception and self-evaluation tools require further exploration—in

particular, for training purposes. That is, the DIVA model suggests that the integration of self-monitoring is an essential component for the accurate and successful remapping of internal models. Finally, it is crucial to continue to focus on how to improve the finer speech motor skills necessary to accurately execute intended speech motor commands. A particular field that lends itself to this study and can easily be combined with mimicry tasks is that of motor learning—a subject that recently has received notable attention in the field of motor speech disorders (Maas, et al., 2008; Rosenbek & Jones, 2009). Along the same line, the effect of mental rehearsal and model speaker on mimicry performance – both of which have been reported to affect performance (Morris, Iansek, Matyas, & Summers, 1996; M. Wilson & Knoblich, 2005) – could be examined. Importantly, future research would benefit from exploring the effectiveness of error-detection and error-correction mechanisms in the context of speech intelligibility as it may significantly contribute to current theory and practice of dysarthria management. This may be particularly valuable for evaluating motor learning principles in the context of mimicry performance and their retention and generalisation to untrained stimuli.

In summary, the current results suggest that individuals with PD and dysarthria and HC are able to (1) use auditory-perceptual resources to inform and modify their own speech production to some extent, and (2) re-produce these acquired speech patterns in the absence of the acoustic stimulus. Thus, it can be assumed that speech perception and speech production can effectively be integrated to induce speech motor learning in these speakers. However, finer adjustments to more closely approximate the acoustic target have not been found. In addition, the present findings suggest a mismatch between speakers' perceptions of mimicry 'goodness' and the acoustic evaluations of distances between speakers and the model. Further research is necessary to determine if, and how, auditory-perceptual and speech motor limitations affect mimicry performance, and consequently, how to improve error-detection and error-correction mechanisms to maximise mimicry performance.

CHAPTER FIVE

Summary, Clinical Implications, Limitations, and Future Directions

5.1 Summary

The present study provides preliminary evidence that error-detection and error-correction may be important mechanisms to consider in the development of theory and management of dysarthria. Chapter 1 highlighted the significance of improving speech intelligibility in speakers with dysarthria and called for a closer examination of whether, and how, auditory-perceptual abilities and their integration to skilled motor movement may be affected in this population. Support of this notion has previously been provided by brain research, proposing that mirror neurons are established and adapted through the integration of sensory and motor learning. Furthermore, theoretical relevance was emphasised by an in-depth account of the DIVA model of speech production and its assumption that speech sounds must first be reliably perceived before accurate learning of their speech production can follow. A review of the literature revealed a number of resources suggesting auditory-perceptual declines in speakers with dysarthria, and particularly, those with PD. On the other hand, relatively few studies have addressed the integration of speech perception and speech production in this population. Anecdotal reports from the LSVT® had suggested a possible mismatch between speech perception and speech production in speakers with PD, but only one independent study provides support for this theory. Chapter 1 therefore called to attention the need for a comprehensive assessment of the integration of error-detection and error-correction mechanisms, and in addition, to determine the value of distinguishing between auditory declines due to ageing and those perhaps inherent to the characteristics of dysarthria. Two methods of investigating error-detection and error-correction mechanisms as outlined in the theoretical framework of the DIVA model were presented: auditory perturbation and mimicry.

Chapter 2 contained the first of two perturbation studies, examining compensation to auditory vowel perturbation in speakers of various ages. In brief summary, fifty-four NZE speakers aged 20-78 years completed the following experimental procedure: (1) an auditory vowel perturbation experiment, (2) a single vowel identification task completed before and after the perturbation component, and (3) a test of auditory acuity of vowel formant differences based on an adaptive staircase protocol. This was the first study of its kind to assess compensation in NZE speakers as well as in various age groups. The examination of an array of factors previously reported in compensation research has enabled a more thorough perspective on mechanisms involved in error-detection and error-correction of perturbed

vowel productions. For this group of NZE speakers, it was found that the magnitude of compensation was generally reduced and variability in response patterns increased when compared to previous studies of AE and CE. No predictors of compensation were found for the overall group. Subsequent follow-up analyses focused on the response-dependent categories of (1) *big compensators*, (2) *compensators*, (3) *big followers*, and (4) *followers*. Results revealed a predictive value of both (a) F1 baseline standard deviation and (b) F1 vowel distance of HEAD relative to HEED and HAD on compensation for *big compensators*. F1 baseline standard deviation was also found to have predictive value for the group of *compensators*. No predictors of compensation were found for the other two subgroups. The findings of phase one of this study provided preliminary support for the notion that error-detection and error-correction mechanisms are not affected by age. Moreover, the study of NZE afforded new conclusions regarding the effect of somatosensory feedback on compensation and its weighting against auditory-perceptual influences on error-detection and error-correction due to somatosensory cues available in mid-to-high front vowels.

A continuation of Chapter 2, Chapter 3 aimed to establish whether the 16 speakers who had been categorised as big compensators in phase one and had consistent responses across Stay and Stay2 phases also showed adaptation to auditory vowel perturbation, and if so, whether predictors of adaptation could be determined. Results revealed that in the absence of auditory perturbation, speech motor commands adapted for a short period of time until a process of de-adaptation was initiated. No predictors of adaptation could be established for the group. Thus, it was speculated that ageing did not affect processes underlying speech motor learning.

Collectively, phases one and two of this research identified that commonly applied auditory perturbation resulted in unexpected findings for speakers of NZE. Therefore, Chapter 4 employed a mimicry task to determine whether speakers with dysarthria and HC are able to effectively integrate speech perception and speech production when attempting to match an acoustic target. In brief summary, eight speakers with PD and hypokinetic dysarthria and eight HC were given three consecutive attempts to approximate the speech signal of a model speaker. After the third attempt, speakers perceptually evaluated their mimicry performance, using a linear scale from ‘not identical’ to ‘identical’. Finally, speakers re-produced the target speech pattern with the auditory stimulus removed. Mimicry performance was then analysed based on how closely the speakers’ productions matched the

model speaker, using the acoustic distance measures of duration, percent voiced, mean F0, and F0 standard deviation. Phase three revealed that all speakers were able to modify their speech production to approximate the model speaker. However, speakers' productions did not move significantly closer to the target over the three mimicry attempts. In addition, phase three found inconsistencies between perceptual self-evaluations and acoustic distance measures of mimicry 'goodness'. The target speech acoustics were successfully re-produced in the absence of the stimulus except for the acoustic dimension of duration, indicating successful perceptual-motor learning. While speakers with moderate levels of dysarthria exhibited greater acoustic distances overall compared to HC and speakers with mild dysarthria (except for the dimension of pitch variation), speakers with dysarthria and HC were not found to differ in their mimicry performance, both perceptually and acoustically. Taken together, these findings were considered preliminary evidence that speech perception and speech production can at least to some extent be effectively integrated to induce error-correction mechanisms and subsequent speech motor learning in these speakers with PD and dysarthria. The third phase of the presented research acknowledged, however, that auditory-perceptual declines cannot be eliminated as a possible explanation for some of the observed limitations in the flexibility and accuracy of error-detection and error-correction mechanisms. Thus, Chapter 4 concluded with a call for additional research to determine if, and how, auditory-perceptual limitations affect mimicry performance, and further, whether mimicry performance can be improved through the systematic training of error-detection and error-correction mechanisms.

Overall, the present thesis aimed to offer a theoretically-based perspective on processes involved in error-detection and error-correction—using the DIVA model of speech production as an interpretive framework. While additional studies are required to address some of the limitations in these studies, this research provides preliminary evidence for the potential use of the DIVA model for speakers with dysarthria. Although in-depth analyses of the DIVA research laboratory are necessary to confirm and further consult and computationally test and validate the present findings, the application of the DIVA model may provide a foundation to enable the integration of error-detection and error-correction processes to the assessment and management of dysarthria.

5.2 Clinical implications

Chapter 1 acknowledged that if treatment focuses on modifying speech production alone, without assessing and managing possible auditory-perceptual deficits, then the maintenance and generalisation of immediate treatment outcomes may be limited. Therefore, research that informs about such possible auditory-perceptual declines holds considerable clinical value. The fact that this study did not find improvements in mimicry performance over three consecutive trials – while not necessarily pointing towards a problem – may suggest perceptual difficulties in either a) identifying and distinguishing between different acoustic targets of the stimulus, or b) comparing one's own speech output to that of a model speaker. The observed mismatch between speakers' self-evaluations and their actual performances on the mimicry task may provide additional support for such a notion. Although more research is required to confirm or refute these results for different types and severities of dysarthria, auditory-perceptual deficits are important to consider in dysarthria management—particularly in light of the growing evidence of a dysarthria-specific hearing loss, independent of age-related hearing loss, as well as difficulties in speech-specific discrimination tasks. This is further highlighted by preliminary evidence of a lack of awareness in speakers with PD towards any auditory deficiencies. Thus, it may be beneficial for current clinical practices in dysarthria rehabilitation to incorporate hearing assessments, speech-specific discrimination tasks, and client's evaluations of their own speech production into the assessment and management plan of speakers with dysarthria. Further research is required, however, to determine the value of auditory-perceptual training modules for the finer modifications of speech production, and consequently, on speech intelligibility before such claims can be validated. Besides these auditory-perceptual implications, the present findings offered additional evidence for the use of mimicry, demonstration, and modelling in the management of the dysarthrias, targeting the modification of specific acoustic characteristics in the speech signal. Moreover, the observed mismatch in speech perception and speech production supports a possible benefit in enhancing the self-evaluation of sensory experiences to guide speech production when using such techniques. This is in line with recent suggestions in the motor skill learning literature (e.g., Rosenbek & Jones, 2009), and thus, may need to be considered more carefully in the speech rehabilitation of speakers with dysarthria.

5.3 Limitations and future directions

5.3.1 Auditory vowel perturbation

The variability and inconsistency in responses to auditory perturbation observed in Chapter 2 and Chapter 3 are indicative of a number of limitations, which may be addressed in subsequent research. For this, a number of variables will need to be assessed in order to establish how best to control for some of the characteristics of NZE and their effect on compensatory behaviour. Some of these may include magnitude of perturbation, constant shifts of feedback, number of trials in each phase, choice of the target word, and inclusion of a vowel identification task. In addition, future research may investigate if, and how, NZE front vowel characteristics affect perceptual elements crucial for successful compensation and adaptation, such as the discrimination of HEAD and HEED as well as the ability to distinguish between a tonal upward shift and downward shift. In general, it may be worth investigating larger subgroups first (e.g., young male healthy speakers) before other age groups and gender are considered. That said, it was acknowledged that generation-dependent acoustic variations in the production of the target sound across speakers of NZE may have made the target sound an ambiguous teaching signal. If true, the perturbation of NZE front vowels may be accompanied with variability and inconsistency regardless of any attempts to control the variables described above. In addition, the current findings suggest that determining predictors of compensation, and particularly, auditory-perceptual factors in the context of ageing, require further investigations in the dialects of AE and CE where somatosensory weights are less dominant. For NZE, on the other hand, a similar study design investigating fundamental frequency or loudness perturbation may shed light onto potential dialectal influences on compensation and adaptation.

5.3.2 Mimicry

Perhaps one of the greatest limitations of the third phase of the research presented here was the limited number of speakers and the lack of a control group of young healthy speakers. Likewise, future research would benefit from employing speakers with different aetiologies, types and severities of dysarthria. Moreover, analyses of mimicry performance in the context of speech intelligibility would have allowed for a more clinically applicable interpretation of the results. In any case, thorough auditory-perceptual assessments are

required to identify and distinguish between possible breakdowns in the processes involved in error-detection and error-correction. In addition, future research may seek to determine the effect of focused attention on mimicry performance to complement auditory-perceptual assessment tools and allow for a more holistic view of the potential breakdowns in mimicry performance. While this study focused on rate and prosody, future research may consider implementing additional acoustic measures such as loudness and segmental features of speech production. Importantly, some interesting extensions to the current body of research may arise from using auditory stimuli that are overtly different from speakers' every-day language use. Methodologies of such undertakings may vary anywhere from atypical prosodic speech samples to foreign language use. A joint-approach, based on the findings of the perturbation and mimicry studies presented here, may yet ask older adults and speakers with dysarthria to mimic /ɛ/-productions of young healthy NZE speakers who present with distinctly elevated front vowels compared to typical older adults. Any of these studies have the potential to broaden current knowledge of the integration of error-detection mechanisms into articulatory error-correction. Lastly, future research would benefit from a comprehensive investigation of how motor learning principles and their modifiable variables may improve mimicry performance and speech intelligibility. Similarly, a well-researched protocol for facilitating auditory-perceptual learning and subsequent gains in mimicry performance may enhance current clinical practices in the management of dysarthria.

5.4 Conclusions

This thesis identified that the study of error-detection and error-correction mechanisms in speakers with dysarthria requires further development and research. The series of experiments presented in this research have provided a theoretical account of error-detection and error-correction mechanisms as outlined in the DIVA model and their relevance for older adults and speakers with dysarthria. As a result, this is the first study of its kind to investigate a combination of factors previously reported in perturbation research and their (combined) influence on compensation and adaption in healthy speakers of different age groups. Moreover, this thesis represents an original attempt to investigate the integrational processes of speech perception and speech production in speakers with dysarthria. While a number of limitations have to be taken into account, this research has offered a number of future directions for the study of error-detection and error-correction mechanisms in speakers with

dysarthria. It has been emphasised that the amount of learning an individual can undergo may be limited if current clinical practices fail to address potential auditory-perceptual deficits in speakers with dysarthria. The current results have highlighted a potential deficit in speakers with PD and dysarthria to discriminate and modify the more refined aspects of speech production. Moreover, a potential mismatch between speech perception and speech production cannot be ruled out. Although current dysarthria assessment and management does not typically consider auditory-perceptual declines and their effect on speech production, the present body of research suggests that this may be a promising area of research to examine and improve error-detection and error-correction mechanisms in this population. It is anticipated that ongoing research in this area will enhance clinical practices in dysarthria management and enable greater rehabilitative benefits for speakers with dysarthria.

APPENDIX A

The DIVA model of speech production

1. DIVA – A theoretical framework of speech production

The following explanation of the latest version of the DIVA model is based on the accumulation of information processed over the last two decades (Golfinopoulos, et al., 2010; Guenther, 1994, 1995a, 1995b, 2003, 2006; Guenther, et al., 2006; Guenther, et al., 1998; Guenther & Perkell, 2004; Perkell, 2012). The model contains of the following components, which will be described in detail below: (1) a Speech Sound Map, (2) an Articulatory Velocity and Position Map, (3) a Feedforward Control Subsystem, and (4) a Feedback Control Subsystem. Together, the Feedforward and Feedback Control Subsystems build the key component in ensuring accurate speech sound production. To establish a more thorough understanding of how the different components of the DIVA model are interrelated, each component will entail its relationship with other parts of the model. An overview of the DIVA model is provided in Figure 1.

The schematic illustration shows the latest version of the DIVA model (Golfinopoulos, et al., 2010). Errors in the figure connecting different components of DIVA resemble synaptic projections and a ‘map’ refers to a set of neurons. The following abbreviations are used to refer to the neural substrates involved in the individual processes: **aSMg**=anterior supramarginal gyrus; **Cau**=caudate; **Pal**=pallidum; **Hg**=Heschl's gyrus; **pIFg**=posterior inferior frontal gyrus; **pSTg**=posterior superior temporal gyrus; **PT**=planum temporale; **Put**=Putamen; **slCB**=superior lateral cerebellum; **smCB**=superior medial cerebellum; **SMA**=supplementary motor area; **Tha**=thalamus; **VA**=ventral anterior nucleus of the cerebellum; **VL**=ventral lateral nucleus of the thalamus; **vMC**=ventral motor cortex; **vPMC**=ventral premotor cortex; **vSC**=ventral somatosensory cortex.

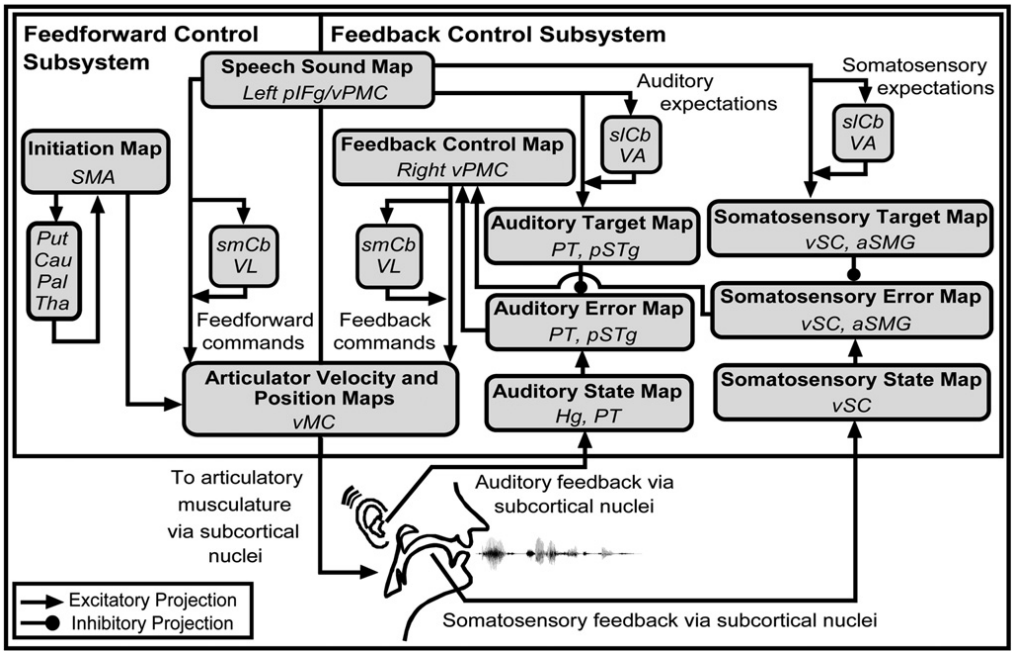

Fig. 1. Schematic of the DIVA neural network model. The neural substrates underlying this integrated control scheme include the premotor and primary motor cortices, somatosensory cortices, auditory cortices, the cerebellum, and the basal ganglia. (Figure and text copied from Golfinopoulos, et al., 2010).

Speech Sound Map.

The *Speech Sound Map* refers to a set of cells, each one representing a different speech sound. In DIVA, a speech sound can be either a single phoneme or a high frequency multi-phonemic syllable or word. These cells typically only become active after the speech recognition system of the model is able to identify a signal as an auditory-perceptual target. Repeated activation of the *Speech Sound Map* upon recognition of a speech sound then tunes the auditory target region of the teaching signal.

Once the auditory target region of a speech sound has been tuned, reliable mappings between reference frames for the accurate production of the target sound can be established. As a result, the *Speech Sound Map* can also be activated to produce the target sound. Thus, it is hypothesised that a *Speech Sound Map* for a particular sound is activated during two processes. First, when an auditory-perceptual target (i.e., speech sound) is recognised as a teaching signal and auditory expectations of that sound are tuned. Second, when the same

auditory-perceptual target is realised; that is, feedforward commands are initiated for the production of that speech sound.

Neuro-physiologically, the *Speech Sound Map* is hypothesised to be located in the left ventral premotor cortex (i.e., the left posterior and inferior portion of Broca's area), also referred to as the frontal operculum. In neuroscience, neurons with the specific quality of being activated both during speech perception and speech production of a particular speech sound have been referred to as 'mirror neurons' (Watkins & Paus, 2004).

Articulatory Velocity and Position Maps.

The *Articulatory Velocity Map* cells are responsible for the execution of muscle contraction, which consequently results in articulatory movement. In the DIVA model, both the *Articulatory Velocity Map* and the *Articulatory Position Map* cells are based on the primary movement degrees of freedom of the vocal tract (Maeda, 1990). These are antagonistic in nature. The *Articulatory Position Map* cells account for jaw height, tongue shape, tongue body position, lip protrusion, larynx height, upper lip height, and lower lip height. Close communication between the *Articulatory Position Map* cells and the *Articulatory Velocity Map* cells ensures that the model knows the current articulatory positions. This information guides the computation of articulatory movement adjustments to match an intended auditory-perceptual target (as represented in the *Speech Sound Map*).

When the model first learns that it can produce sound, *Articulatory Velocity Maps* are randomly activated (i.e., without an auditory-perceptual target to guide the motor command). Directional activation of *Articulatory Velocity Maps* for speech sound production can only be learnt once an auditory target has been identified as the teaching signal. To tune *Articulatory and Position Maps*, the model requires continuous auditory and somatosensory feedback to guide the learning process. These feedback channels guide the learning of consistent mappings between reference frames for accurate speech sound production. Namely, the model has to learn to consistently map (1) the tactile and proprioceptive information, (2) the muscle length, and (3) the articulatory movement required to produce the auditory-perceptual target sound.

Feedforward and Feedback Control Subsystems work very closely together during this time. Consequently, with each consecutive trial, the feedforward command of a given speech sound more accurately represents the auditory (and analogously the somatosensory) target.

Once the speech motor command for an auditory-perceptual target has been fully established (including the somatosensory target region representative of that auditory target), the DIVA model can reliably produce that speech sound.

Neuro-physiologically, *Articulatory Velocity and Position Maps* are hypothesised to be located in the ventral motor cortex (i.e., the caudoventral portion of the precentral gyrus).

Feedforward Control Subsystem.

The Feedforward Control Subsystem works two-directional. Feedforward commands are initiated by the *Speech Sound Map* and project information from the *Speech Sound Map* to the *Articulatory Velocity and Position Maps*. Forward projections ensure that the motor program is ready to be executed. In contrast, backward projections ensure accuracy of all components before releasing the speech motor command. An important component for this process is the *Initiation Map*. Each cell in the *Initiation Map* has a corresponding *Speech Sound Map*. Moreover, *Articulatory Velocity and Position Maps* also correspond with the *Initiation Map*. Therefore, the *Initiation Map* links the information of the *Articulatory Velocity and Position Map* back to the *Speech Sound Map* to ensure congruency in the two information subsystems. Congruency in the variables required to accurately produce a target sound results in the activation of the *Initiation Map,* and consequently, the initiation of the speech motor command.

The essence of the *Feedforward Control Subsystem* is the establishment of a forward model (also referred to as “internal model”; see page 9 for a more thorough explanation of the internal model) that drives the forward command. For this purpose, it corresponds with the *Feedback Control Subsystem* through projections from the *Speech Sound Map* to *Somatosensory* and *Auditory Target Maps* (see Feedback Control Subsystem below). These projections hold the information regarding the expected sensory consequences of the sound to be produced. In DIVA, the relationship between a teaching signal and its sensory realisation is tuned on each production attempt during the babbling cycles of speech sound acquisition. During these learning phases, feedforward commands do not contribute greatly to motor commands. As a consequence, the system heavily relies on auditory feedback until continuous updating of a feedforward command establishes accurate speech sound production.

Once feedforward commands can be generated without errors, the *Feedback Control Subsystem* does no longer serve a primary role during speech production. However, while *Auditory Feedback Control Subsystems* are used less often once a motor command of a speech sound has been established, it continuously remains available to the system. External influences that may influence accurate speech production (e.g., adjustment to background noise or a bad phone line) will therefore automatically activate the *Feedback Control Subsystems*, and generated information will be integrated into the feedforward command. This shows that accurate speech production is ensured by close networking between feedforward and feedback control subsystems.

Neuro-physiologically, the learning and maintenance of feedforward commands is hypothesised to be supported by projections from the premotor cortex (the left frontal operculum) to the primary motor cortex. These projections are further supplemented by cerebellar projections. The timing of feedforward commands is hypothesised to be supported by bilateral projections from the cerebellum (i.e., cells in the anterior paravermal cerebellar cortex). Furthermore, the tuning of feedforward commands is hypothesised to be supported by the superior paravermal regions of the cerebellum. *Initiation Maps* are hypothesised to lie bilaterally within the supplementary motor areas. Reciprocal connections with the basal ganglia (including bilateral caudate, putamen, pallidum, and thalamus) are believed to influence the timing of their activation, and thus, the release of speech motor commands. Projections from the *Speech Sound Map* to the *Somatosensory and Auditory Target Maps* are hypothesised to lie bilaterally in the superior lateral cerebellar cortex.

Feedback Control Subsystem.

There are two *Feedback Control Subsystems*; namely, an *Auditory Feedback* and a *Somatosensory Feedback Control Subsystem*. The cortical mechanisms of both feedback subsystems are similar to each other, but separate in their sensory modality. Auditory feedback is based on the first three formant frequencies, whereas somatosensory feedback corresponds to tactile and proprioceptive feedback from the articulators. Next, the two feedback mechanism will be explained in more detail.

Auditory Feedback Control Subsystem.

As highlighted earlier, the *Auditory Feedback Control Subsystem* plays an important role in the tuning of *Articulatory Velocity and Position Maps* and the *Feedforward Control*

Subsystem (i.e., feedforward commands). The way the *Auditory Feedback Control Subsystem* works is that an auditory target sound (i.e., the expected auditory feedback) is compared with the auditory feedback of the produced target sound. If an error between these two signals is detected, articulatory changes are being made to better match the auditory-perceptual target. These corrections are then used to update the current speech motor program of the feedforward command.

The expected sensory feedback is represented in the *Auditory Target Map* and the feedforward projections received from the *Speech Sound Map*. It is important to understand that in DIVA, an auditory target does not correspond to a single point but rather a target region. This implies that a number of articulatory variations can produce the same auditory target sound. In other words, the relationship between an auditory target and its articulatory realisation is one-to-many.

The *Auditory State Map* on the other hand, corresponds to the actual sensory feedback that is generated during the production of a speech sound. In regard to the error signal, which is represented in the *Auditory Error Map*, the *Auditory Target Map* sends inhibitory signals whereas the *Auditory State Map* sends excitatory signals. When the realised auditory feedback matches the expected auditory feedback, inhibitory and excitatory signals cancel each other out and the *Auditory Error Map* remains deactivated. If, however, there is a mismatch between the realised auditory feedback and the expected feedback, the excitatory signal will be greater than the inhibitory signal, therefore resulting in an activation of the *Auditory Error Map*. In an attempt to correct for the detected error, the *Auditory Error Map* sends a signal to the articulatory velocity cells within the *Feedback Control Map*, which computes the corrective error command.

Neuro-physiologically, *Auditory State Maps* are hypothesised to be located along the supratemporal plane and *Auditory Target Maps* in the bilateral planum temporale and superior temporal gyrus. In addition, *Auditory Target Maps* are fed by a trans-cerebellar pathway (i.e., cells in both hemispheres of the superior lateral cerebellar cortex). *Auditory Error Maps* lie in the same region as *Auditory Target Maps* but are topographically superimposed.

Somatosensory Feedback Control Subsystem.

Somatosensory feedback includes information regarding the pressure receptors on the surfaces of the speech articulators (i.e., tactile input) and the degree of constriction of the articulators at different points along the vocal tract as indicated by muscle length adjustments (i.e., combined tactile and proprioceptive input). During the babbling cycles, the DIVA model learns to differentiate between important and unimportant tactile and proprioceptive cues for the production of a given sound. This process then results in the establishment of a somatosensory target region of a given sound, which embodies the expected tactile and proprioceptive sensations associated with the target sound. Thus, somatosensory feedback – analogous to auditory feedback – provides important information regarding the activation of appropriate articulator movements. The *Somatosensory Feedback Control Subsystem* follows the same rules and interactions as the *Auditory Feedback Control Subsystem*. In essence, *Somatosensory Error Maps* help detect somatosensory errors, which are then transformed into corrective motor commands, and consequently, serve to update feedforward commands.

Neuro-physiologically, *Somatosensory State Maps* are hypothesised to lie along the inferior postcentral gyrus and *Somatosensory Target Maps* in the ventral somatosensory cortex (bilaterally) and the anterior supramarginal gyrus. Similarly to the *Auditory Target Map*, the *Somatosensory Target Map* also receives information regarding the expected somatosensory target region via a trans-cerebellar pathway (i.e., cells in both hemispheres of the superior lateral cerebellar cortex).

2. Application of DIVA to Communication Disorders

Because the DIVA model has been described and tested so thoroughly, it allows the study of some communication disorders associated with lesions to particular neural regions or systems. That is, the DIVA model provides a theoretical framework to investigate the behavioural consequences often observed in communication disorders as a result of neural lesions (Guenther, 2006; Jacks, 2008). For example, the programming difficulty associated with Apraxia of Speech (AOS) can be related to an impaired *Feedforward Control Subsystem*.

Jack (2008) tested this hypothesis in five individuals with AOS plus aphasia (aged between 49 and 75 years) and five healthy matched controls. Participants underwent two experiments; one requiring speech production in an unconstraint condition and one requiring speech production in a bite block condition. Since *Feedback Control Subsystems* are hypothesised to be sufficiently accessible in individuals with AOS, it was suggested that the bite block condition would show relatively normal performance in comparison with healthy controls. On the other hand, during normal speech it was hypothesised that speech would significantly deteriorate in the AOS group due to the reliance on the *Feedforward Control Subsystem*. Results were in support of this hypothesis. Furthermore, Jacks suggested that speech characteristics in AOS could be explained by an overreliance on *Feedback Control Subsystems*, specifically the increased processing time required for the use of feedback systems when compared to feedforward systems.

Terband et al. (2009) computationally simulated this hypothesis for children with AOS using DIVA. Results revealed that not all speech characteristics of AOS could be simulated by an increased dependency on *Feedback Control Subsystems*. However, coarticulation, speech sound distortion, articulatory groping, and inconsistency of articulatory behaviour increased when the model had to rely more on the *Feedback Control Subsystem* during speech production.

Similarly, Max et al. (2004) hypothesised that stuttering results from either an unstable or insufficiently activated internal model or from feedback-biased motor control. Computer simulations of this hypothesis found that stuttering characteristics and repair attempts for these errors can both be simulated in the DIVA model by inducing a higher dependency on the *Auditory Feedback Control Subsystem* (Civier, Bullock, Max, & Guenther, 2009, August; Civier & Guenther, 2005, June; Civier, et al., 2010). These results were taken as in support of the hypothesis that stuttering may result from an overreliance on auditory feedback.

3. Limitations

While the DIVA model has been continuously improved over the last two decades, currently, the model does not account for the following aspects of speech perception and speech production. First, the model in its current form does not imply if, and how, auditory and somatosensory reference frames may be interrelated (e.g., in differing weights

represented in somato-auditory cells and somato-auditory error maps). Second, auditory representations in the model are non-adaptive, and thus, do not allow for perceptual adaptation. Third, the model currently doesn't distinguish between speech production that is internally generated and speech production that is externally cued (e.g., by an imitation task). Forth, the DIVA model does not fully account for the neural bases of speech production. That is, fMRI studies suggest the contribution of inferior cerebellum, anterior insula, and anterior cingulate gyrus for sensory error-detection during speech production. However, these brain structures are currently not included in the DIVA model. Fifth, the DIVA model does not consider phonology or linguistic meaning in its computation of speech production and speech perception. Last, higher-level sequence planning (e.g., the selection, initiation, and sequencing of speech movements) and its associated brain areas are not integrated into the DIVA model in its present form (Golfinopoulos, et al., 2010; Guenther, 1995b, 2006; Guenther, et al., 2006).

4. Future Directions

A separate but integrated model has recently been established, called **G**radient **O**rder DIVA (Bohland, Bullock, & Guenther, 2010). GODIVA was inspired by the aim to incorporate phonological and linguistic meaning into the DIVA model. Similarly to the DIVA model, GODIVA is associated with specific neuro-anatomical substrates that account for phonological and phonetic encoding and the ongoing changes within the speech plan as it moves through the different stages of speech production.

Currently, DIVA and GODIVA are separate models. However, the DIVA and GODIVA research laboratory aims to integrate these two models into one to give a more complete account of speech planning and production. In integrating DIVA and GODIVA, it is also anticipated that a more thorough account for the underlying pathology in AOS and stuttering can be established (Bohland, et al., 2010). Moreover, future work aims to incorporate higher-level linguistic processes into GODIVA, and subsequently, fuse all three versions into one unified model of speech production. In summary, while the DIVA model clearly has limitations, ongoing improvements and the integration of GODIVA aim to more closely resemble natural speech planning and speech production processes, including neurolinguistic parameters.

APPENDIX B

Mimicry Stimuli

The rainbow passage

1. When the sunlight strikes raindrops in the air,
2. they act as a prism and form a rainbow.
3. The rainbow is a division of white light into many beautiful colours.
4. These take the shape of a long round arch, with its path high above,
5. and its two ends apparently beyond the horizon.
6. There is, according to legend, a boiling pot of gold at one end.
7. People look, but no one ever finds it.
8. When a man looks for something beyond his reach,
9. his friends say he is looking for the pot of gold at the end of the rainbow.
10. Throughout the centuries people have explained the rainbow in various ways.
11. Some have accepted it as a miracle without physical explanation.
12. To the Hebrews it was a token that there would be no more universal floods.
13. The Greeks used to imagine that it was a sign from the gods
14. to foretell war or heavy rain.
15. The Norsemen considered the rainbow as a bridge
16. over which the gods passed from earth to their home in the sky.

Fairbanks, G. (1960). Voice and articulation drillbook (2nd ed.). New York: Haper & Row.

REFERENCES

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures - orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology, 51*, 705-723.

Adams, S., Moon, B. H., Dykstra, A., Abrams, K., Jenkins, M., & Jog, M. (2006). Effects of multitalker noise on conversational speech intensity in Parkinson's disease. *Journal of Medical Speech-Language Pathology, 14*(4), 221-228.

Akagi, M., Dang, J., Lu, X., & Uchiyamada, T. (2006). Investigation of interaction between speech perception and production using auditory feedback. *Journal of the Acoustical Society of America (Proceedings of the Fourth Joint Meeting of ASA and ASJ), 120*(5, Part II), 3253.

Ansel, B. M., & Kent, R. D. (1992). Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *Journal of Speech and Hearing Research, 35*(2), 296-308.

Artieda, J., Pastor, M. A., Lacruz, F., & Obeso, J. A. (1992). Temporal discrimination is abnormal in Parkinson's disease. *Brain, 115*, 199-210.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Ballard, K., Robin, D., Woodworth, G., & Zimba, L. (2001). Age-related changes in motor control during articulator visuomotor tracking. *Journal of Speech, Language, and Hearing Research, 44*(4), 763-777.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42. Retrieved from http://CRAN.R-project.org/package=lme4

Bauer, J. J., Hubbard Seery, C., LaBonte, R., & Ruhnke, L. (2008). Voice F0 responses elicited by perturbations in pitch of auditory feedback in persons who stutter and controls. *Proceedings of Meetings on Acoustics: The 153rd Meeting of the Acoustical Society of America, Vol. 1* (p. 060004). doi: http://dx.doi.org/10.1121/1.2959144

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

Baum, S. R., & McFarland, D. H. (1997). The development of speech adaptation to an artificial palate. *Journal of the Acoustical Society of America, 102*(4), 2353-2359.

Baum, S. R., & McFarland, D. H. (2000). Individual differences in speech adaptation to an artificial palate. *Journal of the Acoustical Society of America, 107*(6), 3572-3575.

Behroozmand, R., Korzyukov, O., & Larson, C. R. (2011). Effects of voice harmonic complexity on ERP responses to pitch-shifted auditory feedback. *Clinical Neurophysiology, 122*(12), 2408-2417.

Behroozmand, R., Korzyukov, O., & Larson, C. R. (2012). ERP correlates of pitch error detection in complex tone and voice auditory feedback with missing fundamental. *Brain Research, 1448*, 89-100.

Behroozmand, R., Korzyukov, O., Sattler, L., & Larson, C. R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control. *Journal of the Acoustical Society of America, 132*(4), 2468-2477.

Beijer, L. J., Rietveld, A. C. M., & Van Stiphout, A. (2011). Auditory discrimination as a condition for E-learning based speech therapy: A proposal for an auditory discrimination test (ADT) for adult dysarthric speakers. *Journal of Communication Disorders, 44*(6), 701-718.

Berry, J., North, C., & Meyers, B. (2012, March). *Sensorimotor adaptation in severe dysarthria*. Paper presented at the Conference on Motor Speech: Motor Speech Disorders and Speech Motor Control, Santa Rosa, CA.

Bloch, S., & Wilkinson, R. (2009). Acquired dysarthria in conversation: Identifying sources of understandability problems. *International Journal of Language & Communication Disorders, 44*(5), 769-783.

Bock, O., & Schneider, S. (2002). Sensorimotor adaptation in young and elderly humans. *Neuroscience and Biobehavioral Reviews, 26*(7), 761-767.

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

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience, 22*(7), 1504-1529.

Breustedt, A. (1983). Age-induced changes in the oral-mucosa and their therapeutic consequences. *International Dental Journal, 33*(3), 272-280.

Brooks, V. B. (1986). *The neural basis of motor control*. New York: Oxford University Press.

Brown, W. F. (1972). A method for estimating the number of motor units and the changes in motor unit count with aging. *Journal of Neurology, Neurosurgery, and Psychiatry, 35*, 845-852.

Brunner, J., Ghosh, S., Hoole, P., Matthies, M., Tiede, M., & Perkell, J. (2011). The influence of auditory acuity on acoustic variability and the use of motor equivalence during adaptation to a perturbation. *Journal of Speech, Language, and Hearing Research, 54*(3), 727-739.

Brunner, J., & Hoole, P. (2012). Motor equivalent strategies in the production of German /esh/ under perturbation. *Language and Speech, 55*(4), 457-476.

Bunton, K., & Weismer, G. (2001). The relationship between perception and acoustics for a high-low vowel contrast produced by speakers with dysarthria. *Journal of Speech, Language, and Hearing Research, 44*(6), 1215-1228.

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

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory-motor transformation. *PLoS ONE, 7*(7), e41830. doi: 10.1371/journal.pone.0041830

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2008). A system for online dynamic perturbation of formant frequencies and results from perturbation of the Mandarin triphthong /iau/. In R. Sock, S. Fuchs & Y. Laprie (Eds.), *Eighth International Seminar on Speech Production* (pp. 65-68). Strasbourg, France: INRIA. Available at http://issp2008.loria.fr/proceedings.html.

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2010). Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization. *Journal of the Acoustical Society of America, 128*(4), 2033-2048.

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *Journal of Neuroscience, 31*(45), 16483-16490.

Campbell, M. J., McComas, A. J., & Petito, F. (1973). Physiological changes in aging muscles. *Journal of Neurology, Neurosurgery, and Psychiatry, 36*(2), 174-182.

Catmur, C. (2013). Sensorimotor learning and the ontogeny of the mirror neuron system. *Neuroscience Letters, 540*, 21-27.

Chen, S. F., Liu, L., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *Journal of the Acoustical Society of America, 121*(2), 1157-1163.

Cheng, H. Y., Goozee, J. V., & Murdoch, B. E. (2005). Analysis of articulatory dynamics in dysarthria following brain injury in childhood using electromagnetic articulography and electropalatography. *Journal of Medical Speech-Language Pathology, 13*(1), 15-35.

Civier, O., Bullock, D., Max, L., & Guenther, F. H. (2009, August). *Simulating neural impairments to syllable-level command generation in stuttering*. Paper presented at the 6th World Congress on Fluency Disorders, Rio de Janeiro, Brazil.

Civier, O., & Guenther, F. H. (2005, June). *Simulations of feedback and feedforward control in stuttering*. Paper presented at the Proceedings of the 7th Oxford Dysfluency Conference, St. Catherine's College, Oxford University, England. Available at http://www.bu.edu/speechlab/files/2011/10/Oren-paper.pdf

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

Cowie, R., & Douglas-Cowie, E. (1983). Speech production in profound postlingual deafness. In M. Lutman & M. Haggard (Eds.), *Hearing science and hearing disorders* (pp. 183-230). London: Academic Press.

Cowie, R., & Douglas-Cowie, E. (1992). *Postlingually acquired deafness: Speech deterioration and the wider consequences*. New York: Mouton de Gruyter.

Crow, H. C., & Ship, J. A. (1996). Tongue strength and endurance in different aged individuals. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 51*(5), M247-M250.

Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E., Klein, R., Mares-Perlman, J. A., & Nondahl, D. M. (1998). Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin: The epidemiology of hearing loss study. *American Journal of Epidemiology, 148*(9), 879-886.

D'Innocenzo, J., Tjaden, K., & Greenman, G. (2006). Intelligibility in dysarthria: Effects of listener familiarity and speaking condition. *Clinical Linguistics & Phonetics, 20*(9), 659-675.

Dagenais, P. A., Southwood, M. H., & Mallonee, K. O. (1999). Assessing processing skills in speakers with Parkinson's disease using delayed auditory feedback. *Journal of Medical Speech-Language Pathology, 7*(4), 297-313.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12*(2), 246-269.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, 229*(1-2), 132-147.

de Miranda Marzullo, A. C., Neto, O. P., Ballard, K. J., Robin, D. A., Chaitow, L., & Christou, E. A. (2010). Neural control of the lips differs for young and older adults following a perturbation. *Experimental Brain Research, 206*(3), 319-327.

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.

Dickson, S., Barbour, R. S., Brady, M., Clark, A. M., & Paton, G. (2008). Patients' experiences of disruptions associated with post-stroke dysarthria. *International Journal of Language & Communication Disorders, 43*(2), 135-153.

Duffy, J. R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and management* (2nd ed.). St Louis, Missouri: Elsevier Mosby.

Duffy, J. R. (2007). Motor speech disorders: History, current practice, future trends and goals. In G. Weismer (Ed.), *Motor speech disorders: Essays for Ray Kent* (pp. 7-56). San Diego: Plural Publishing.

Economou, A., Tartter, V. C., Chute, P. M., & Hellman, S. A. (1992). Speech changes following reimplantation from a single-channel to a multichannel cochlear implant. *Journal of the Acoustical Society of America, 92*(3), 1310-1323.

Ehrt, U., & Aarsland, D. (2005). Psychiatric aspects of Parkinson's disease. *Current Opinion in Psychiatry, 18*(3), 335-341.

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

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience, 15*, 399-402.

Federmeier, K. D., Van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, words, sentences: Age-related changes across levels of language processing. *Psychology and Aging, 18*(4), 858-872.

Field, D., Shipley, T. F., & Cunningham, D. W. (1999). Generalization gradients in prism adaptation to dynamic stimuli. *Perception and Psychophysics, 61*(1), 161-176.

Fitzgibbons, P. J., & Gordon-Salant, S. (1998). Auditory temporal order perception in younger and older adults. *Journal of Speech, Language, and Hearing Research, 41*(5), 1052-1060.

Fitzgibbons, P. J., & Gordon-Salant, S. (2004). Age effects on discrimination of timing in auditory sequences. *Journal of the Acoustical Society of America, 116*(2), 1126-1134.

Fox, C. M., Morrison, C. E., Ramig, L. O., & Sapir, S. (2002). Current perspectives on the Lee Silverman Voice Treatment (LSVT) for individuals with idiopathic Parkinson disease. *American Journal of Speech-Language Pathology, 11*(2), 111-123.

Fox, C. M., & Ramig, L. O. (1997). Vocal sound pressure level and self-perception of speech and voice in men and women with idiopathic Parkinson disease. *American Journal of Speech-Language Pathology, 6*, 85-94.

Gates, G. A., Cooper, J. C., Kannel, W. B., & Miller, N. J. (1990). Hearing in the elderly: The Framingham cohort, 1983-1985: Part I. Basic audiometric test results. *Ear and Hearing, 11*(4), 247-256.

Gazzaley, A., Sheridan, M. A., Cooney, J. W., & D'Esposito, M. (2007). Age-related deficits in component processes of working memory. *Neuropsychology, 21*(5), 532-539.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612.

Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *Neuroimage, 55*(3), 1324-1338.

Golfinopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage, 52*(3), 862-874.

Goozee, J. V., Murdoch, B. E., & Theodoros, D. G. (1999). Electropalatographic assessment of articulatory timing characteristics in dysarthria following traumatic brain injury. *Journal of Medical Speech-Language Pathology, 7*(3), 209-222.

Goozee, J. V., Stephenson, D. K., Murdoch, B. E., Darnell, R. E., & Lapointe, L. L. (2005). Lingual kinematic strategies used to increase speech rate: Comparison between younger and older adults. *Clinical Linguistics & Phonetics, 19*(4), 319-334.

Gordon-Salant, S. (2008, November). *Auditory temporal processing limitations in older adult listeners: Implications for everyday speech perception tasks*. Paper presented at the Annual Conference of the American Speech-Language and Hearing Association, Chicago, IL.

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Barrett, J. (2006). Age-related differences in identification and discrimination of temporal cues in speech segments. *Journal of the Acoustical Society of America, 119*(4), 2455-2466.

Grabski, K., Schwartz, J.-L., Lamalle, L., Vilain, C., Vallée, N., Baciu, M., . . . Sato, M. (2013). Shared and distinct neural correlates of vowel perception and production. *Journal of Neurolinguistics, 26*(3), 384-408.

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. (1995a). A modeling framework for speech motor development and kinematic articulator control. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 2, pp. 92-99). Stockholm, Sweden: Stockholm University. Available at http://www.cns.bu.edu/~speech/publications.php.

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

Guenther, F. H. (2003). Neural control of speech movements. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 209-240). Berline: Mouton de Gruyter.

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

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

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

Guenther, F. H., & Perkell, J. S. (2004). A neural model of speech production and its application to studies of the role of auditory feedback in speech. In B. Maassen, R. Kent, H. Peters, P. Van Lieshout & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 29-49). Oxford: Oxford University Press.

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcast. *Journal of Phonetics, 34*(4), 439-457.

Hawco, C. S., & Jones, J. A. (2010). Multiple instances of vocal sensorimotor adaptation to frequency-altered feedback within a single experimental session. *Journal of the Acoustical Society of America, 127*(1), EL13-EL18.

Head, D., Raz, N., Gunning-Dixon, F., Williamson, A., & Acker, J. D. (2002). Age-related differences in the course of cognitive skill acquisition: The role of regional cortical shrinkage and cognitive resources. *Psychology and Aging, 17*(1), 72-84.

Heinks-Maldonado, T. H., & Houde, J. F. (2005). Compensatory responses to brief perturbations of speech amplitude. *Acoustics Research Letters Online, 6*(3), 131-137. doi: http://dx.doi.org/10.1121/1.1931747

Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology, 101*(5), 2725-2732.

Hillenbrand, J. M., & Gayvert, R. T. (2005). Open source software for experiment design and control. *Journal of Speech, Language, and Hearing Research, 48*(1), 45-60.

Ho, A. K., Bradshaw, J. L., & Iansek, R. (2000). Volume perception in Parkinsonian speech. *Movement Disorders, 15*(6), 1125-1131.

Ho, A. K., Bradshaw, J. L., Iansek, R., & Alfredson, R. (1999). Speech volume regulation in Parkinson's disease: Effects of implicit cues and explicit instructions. *Neuropsychologia, 37*(13), 1453-1460.

Hoehn, M. M., & Yahr, M. D. (1967). Parkinsonism: Onset, progression, and mortality. *Neurology, 17*(5), 427-442.

Honda, M., Fujino, A., & Kaburagi, T. (2002). Compensatory responses of articulators to unexpected perturbation of the palate shape. *Journal of Phonetics, 30*, 281-302.

Honda, M., & Murano, E. Z. (2003). Effects of tactile & auditory feedback on compensatory articulatory response to an unexpected palatal perturbation. In S. Palethorpe & M. Tabain (Eds.), *Proceedings of the 6th International Seminar on Speech Production* (pp. 97-100). Sydney, Australia: Macquarie University.

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

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research, 45*(2), 295-310.

Humes, L. E., & Floyd, S. S. (2005). Measures of working memory, sequence learning, and speech recognition in the elderly. *Journal of Speech, Language, and Hearing Research, 48*(1), 224-235.

Hustad, K. C. (1999). Optimizing communicative effectiveness: Bringing it together. In K. M. Yorkston, D. Beukelman, E. Strand & K. Bell (Eds.), *Management of motor speech disorders in children and adults* (2nd ed., pp. 483-539). Austin: Pro-Ed.

Hustad, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language, and Hearing Research, 51*(3), 562-573.

Hustad, K. C., Beukelman, D. R., & Yorkston, K. M. (1998). Functional outcome assessment in dysarthria. *Seminars in Speech and Language, 19*(3), 291-302.

Hustad, K. C., & Weismer, G. (2007). Interventions to improve intelligibility and communicative success for speakers with dysarthria. In G. Weismer (Ed.), *Motor speech disorders: Essays for Ray Kent* (pp. 261-303). San Diego: Plural Publishing.

Ingram, H. A., van Donkelaar, P., Cole, J., Vercher, J. L., Gauthier, G. M., & Miall, R. C. (2000). The role of proprioception and attention in a visuomotor adaptation task. *Experimental Brain Research, 132*(1), 114-126.

Jacks, A. (2008). Bite block vowel production in apraxia of speech. *Journal of Speech, Language, and Hearing Research, 51*(4), 898-913.

Jilka, M., Anufryk, V., Baumotte, H., Lewandowski, N., Rota, G., & Reiterer, S. (2007). Assessing indiviual talent in second language production and perception. In A. S.

Rauber, M. A. Watkins & B. O. Baptista (Eds.), *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech* (pp. 243-258). Florianopolis, Brazil: Federal University of Santa Catarina. Available at http://www.nupffale.ufsc.br/newsounds/proceedings.htm.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of *F*0 production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America, 108*(3), 1247-1251.

Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *Journal of the Acoustical Society of America, 113*(1), 532-543.

Katseff, S., & Houde, J. F. (2008). Partial compensation in speech adaptation. *2008 Annual Report of the UC Berkeley Phonology Lab*, from Univesity of California Berkeley, Department of Linguistics website: http://linguistics.berkeley.edu/~shira/katseff_qp1.pdf

Katseff, S., Houde, J. F., & Johnson, K. (2010). Auditory feedback shifts in one formant cause multi-formant responses. *Journal of the Acoustical Society of America, 127*(3), 1955-1955.

Katseff, S., Houde, J. F., & Johnson, K. (2012). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech, 55*(2), 295-308.

Kent, R. D. (2000). Research on speech motor control and its disorders: A review and prospective. *Journal of Communication Disorders, 33*(5), 391-428.

Kent, R. D., Kent, J. F., Weismer, G., & Duffy, J. R. (2000). What dysarthrias can tell us about the neural control of speech. *Journal of Phonetics, 28*(3), 273-302.

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

Kiessling, J., Pichora-Fuller, M. K., Gatehouse, S., Stephens, D., Arlinger, S., Chisolm, T., . . . von Wedel, H. (2003). Candidature for and delivery of audiological services: Special needs of older people. *International Journal of Audiology, 42*, S92-S101.

Kiran, S., & Larson, C. R. (2001). Effect of duration of pitch-shifted feedback on vocal responses in patients with Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 44*(5), 975-987.

Korzyukov, O., Karvelis, L., Behroozmand, R., & Larson, C. R. (2012). ERP correlates of auditory processing during automatic correction of unexpected perturbations in voice auditory feedback. *International Journal of Psychophysiology, 83*, 71-78.

Kramer, A. F., Bherer, L., Colcombe, S. J., Dong, W., & Greenough, W. T. (2004). Environmental influences on cognitive and brain plasticity during aging. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 59*(9), 940-957.

Kuruvilla, M. S., Murdoch, B. E., & Goozee, J. V. (2008). Electropalatographic (EPG) assessment of tongue-to-palate contacts in dysarthric speakers following TBI. *Clinical Linguistics & Phonetics, 22*(9), 703-725.

Ladefoged, P., & McKinney, N. P. (1963). Loudness, sound pressure, and subglottal pressure in speech. *Journal of the Acoustical Society of America, 35*(4), 454-460.

Lam, J., Tjaden, K., & Wilding, G. (2012). Acoustics of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research, 55*(6), 1807-1821.

Lane, H., Catania, A. C., & Stevens, S. S. (1961). Voice level: Autophonic scale, perceived loudness, and effects of sidetone. *Journal of the Acoustical Society of America, 33*(2), 160-167.

Lane, H., Denny, M., Guenther, F. H., Matthies, M. L., Menard, L., Perkell, J. S., . . . Zandipour, M. (2005). Effects of bite blocks and hearing status on vowel production. *Journal of the Acoustical Society of America, 118*(3), 1636-1646.

Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *Journal of the Acoustical Society of America, 89*(2), 859-866.

Lane, H., Wozniak, J., Matthies, M., Svirsky, M., Perkell, J., Oconnell, M., & Manzella, J. (1997). Changes in sound pressure and fundamental frequency contours following changes in hearing status. *Journal of the Acoustical Society of America, 101*(4), 2244-2252.

Larson, C. R., Altman, K. W., Liu, H., & Hain, T. C. Y. J. (2008). Interactions between auditory and somatosensory feedback for voice F0 control. *Experimental Brain Research, 187*(4), 613-621.

Leder, S. B., & Spitzer, J. B. (1993). Speaking fundamental frequency, intensity, and rate of adventitiously profoundly hearing-impaired adult women. *Journal of the Acoustical Society of America, 93*(4), 2146-2151.

Leder, S. B., Spitzer, J. B., Kirchner, J. C., Flevaris-Phillips, C., Milner, P., & Richardson, F. (1987). Speaking rate of adventitiously deaf male cochlear implant candidates. *Journal of the Acoustical Society of America, 82*(3), 843-846.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal the Acoustical Society of America, 49*(2), 467-477.

Lexell, J., & Taylor, C. C. (1991). Variability in muscle fibre areas studied in whole human quadriceps muscle: Effects of increasing age. *Journal of Anatomy, 174*, 239-249.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (Vol. 55, pp. 403-439). Dordrecht: Kluwer Academic Publ.

Liss, J. M., Weismer, G., & Rosenbek, J. C. (1990). Selected acoustic characteristics of speech production in very old males. *Journals of Gerontology, 45*(2), P35-P45.

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.

Liu, H., Russo, N. M., & Larson, C. R. (2010). Age-related differences in vocal responses to pitch feedback perturbations: A preliminary study. *Journal of the Acoustical Society of America, 127*(2), 1042-1046.

Liu, H., Wang, E. Q., Metman, L. V., & Larson, C. R. (2012). Vocal responses to perturbations in voice auditory feedback in individuals with Parkinson's disease. *PLoS ONE, 7*(3), e33629. doi: 10.1371/journal.pone.0033629

Liu, H., Xu, Y., & Larson, C. R. (2009). Attenuation of vocal responses to pitch perturbations during Mandarin speech. *Journal of the Acoustical Society of America, 125*(4), 2299-2306.

Liu, P., Chen, Z., Jones, J. A., Huang, D., & Liu, H. (2011). Auditory feedback control of vocal pitch during sustained vocalisation: A cross-sectional study of adult aging. *PLoS ONE, 6*(7), e22791. doi: 10.1371/journal.pone.0022791

Luschei, E. S., Ramig, L. O., Baker, K. L., & Smith, M. E. (1999). Discharge characteristics of laryngeal single motor units during phonation in young and older adults and in persons with Parkinson disease. *Journal of Neurophysiology, 81*, 2131-2139.

Maas, E., Robin, D. A., Hula, S. N. A., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17*(3), 277-298.

MacDonald, E. N., Goldberg, R., & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *Journal of the Acoustical Society of America, 127*(2), 1059-1068.

MacDonald, E. N., Johnson, Elizabeth K., Forsythe, J., Plante, P., & Munhall, Kevin G. (2012). Children's development of self-regulation in speech production. *Current Biology, 22*(2), 113-117.

MacDonald, E. N., Purcell, D. W., & Munhall, K. G. (2011). Probing the independence of formant control using altered auditory feedback. *Journal of the Acoustical Society of America, 129*(2), 955-965.

MacLagan, M., & Hay, J. (2007). Getting fed up with our feet: Contrast maintenance and the New Zealand English "short" front vowel shift. *Language Variation and Change, 19*, 1-25.

Maeda, S. (1990). Compensatory articulation durch speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling* (pp. 131-149). Boston: Kluwer Academic Publishers.

Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition, 127*(2), 177-202.

Matthies, M. L., Svirsky, M., Perkell, J., & Lane, H. (1996). Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant. *Journal of Speech and Hearing Research, 39*(5), 936-946.

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. *Contemprorary Issues in Communication Science and Disorders, 21*, 105-122.

McAuliffe, M. J., Robb, M. P., & Murdoch, B. E. (2007). Acoustic and perceptual analysis of speech adaptation to an artificial palate. *Clinical Linguistics & Phonetics, 21*(11-12), 885-894.

McNeil, M. R. (Ed.). (2009). *Clinical management of sensorimotor speech disorders*. New York: Thieme.

Mitsuya, T., MacDonald, E. N., Purcell, D. W., & Munhall, Kevin G. (2011). A cross-language study of compensation in response to real-time formant perturbation. *Journal of the Acoustical Society of America, 130*(5), 2978-2986.

Mitsuya, T., Samson, F., Menard, L., & Munhall, K. G. (2013). Language dependent vowel representation in speech production. *Journal of the Acoustical Society of America, 133*(5), 2993-3003.

Morgan, D. G., & Finch, C. E. (1988). Dopaminergic changes in the basal ganglia: A generalized phenomenon of aging in mammals. *Annals of the New York Academy of Sciences, 515*, 145-160.

Morris, M. E., Iansek, R., Matyas, T. A., & Summers, J. J. (1996). Stride length regulation in Parkinson's disease normalization strategies and underlying mechanisms. *Brain, 119*, 551-568.

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. *Journal of the Acoustical Society of America, 125*(1), 384-390.

Nasir, S. M., & Ostry, D. J. (2006). Somatosensory precision in speech production. *Current Biology, 16*(19), 1918-1923.

Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience, 11*, 1217-1222.

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatric Society, 53*, 695–699.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development, 59*(2), 441-449.

Ostericher, H. J., & Hawk, A. M. (1982). Patterns of performance for two groups of normal adults on a test of oral form discrimination. *Journal of Communication Disorders, 15*, 329-335.

Parkinson, A. L., Flagmeier, S., G., Manes, J., L., Larson, C. R., Rogers, B., & Robin, D., A. (2012). Understanding the neural mechanisms involved in sensory control of voice production. *Neuroimage, 61*, 314-322.

Pell, M. D., & Leonard, C. L. (2003). Processing emotional tone from speech in Parkinson's disease: A role for the basal ganglia. *Cognitive, Affective, & Behavioral Neuroscience, 3*(4), 275-288.

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

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M., Perrier, P., Vick, J., . . . Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics, 28*, 233-272.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America, 116*(4), 2338-2344.

Perkell, J. S., Lane, H., Denny, M., Matthies, M. L., Tiede, M., Zandipour, M., . . . Burton, E. (2007). Time course of speech changes in response to unanticipated short-term changes in hearing state. *Journal of the Acoustical Society of America, 121*(4), 2296-2311.

Perkell, J. S., Lane, H., Gosh, S. S., Matthies, M. L., Tiede, M., & Guenther, F. H. (2008, December). *Mechanisms of vowel production: Auditory goals and speaker acuity*. Paper presented at the 8th International Seminar on Speech Production, Strasbourg, France.

Perkell, J. S., Lane, H., Svirsky, M., & Webster, J. (1992). Speech of cochlear implant patients: A longitudinal-study of vowel production. *Journal of the Acoustical Society of America, 91*(5), 2961-2978.

Perkell, J. S., 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*(2-3), 227-250.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., . . . Guenther, F. H. (2004). The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language, and Hearing Research, 47*, 1259-1269.

Pichora-Fuller, M. K. (2003a). Cognitive aging and auditory information processing. *International Journal of Audiology, 42*, S26-S32.

Pichora-Fuller, M. K. (2003b). Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension. *International Journal of Audiology, 42*, S59-S67.

Pichora-Fuller, M. K., & Singh, G. (2006). Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation. *Trends in Amplification, 10*(1), 29-59.

Pile, E. J. S., Dajani, H. R., Purcell, D. W., & Munhall, K. G. (2007, February). *Talking under conditions of altered auditory feedback: Does adaptation of one vowel generalize to other vowels?* Paper presented at the The 16th International Conference of Phonetics Sciences, Saarbrücken, Germany.

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication, 13*, 109-125.

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice* (3rd ed.). Upper Saddle River, New Jersey: Pearson Education.

Pratt, S. R., & Tye-Murray, N. (1997). Speech impairment secondary to hearing loss. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (pp. 345-387). New York: Thieme Medical Publishers.

Pratt, S. R., & Tye-Murray, N. (2009). Speech impairment secondary to hearing loss. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (2nd ed., pp. 204-234). New York: Thieme.

Price, P. A. S., & Darvell, B. W. (1981). Force and mobility in the aging human tongue. *Medical Journal of Australia, 1*(2), 75-78.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America, 103*(20), 7865-7870. doi: 10.1073/pnas.0509989103

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

Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America, 119*(4), 2288-2297.

Purcell, D. W., & Munhall, K. G. (2008). Weighting of auditory feedback across the English vowel space. In R. Sock, S. Fuchs & Y. Laprie (Eds.), *Eighth International Seminar on Speech Production* (pp. 389-392). Strasbourg, France: INRIA. Available at http://issp2008.loria.fr/proceedings.html.

Ramig, L. O., Sapir, S., Countryman, S., Pawlas, A. A., O'Brien, C., Hoehn, M., & Thompson, L. L. (2001). Intensive voice treatment (LSVT (R)) for patients with Parkinson's disease: A 2-year follow up. *Journal of Neurology, Neurosurgery, and Psychiatry, 71*(4), 493-498.

Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (Vol. 2, pp. 1-90). Mahwah, NJ: Lawrence Erlbaum.

Rochet-Capellan, A., & Ostry, D. J. (2011a). Simultaneous acquisition of multiple auditory-motor transformations in speech. *The Journal of Neuroscience, 31*(7), 2657-2662.

Rochet-Capellan, A., & Ostry, D. J. (2011b). Simultaneous Acquisition of Multiple Auditory–Motor Transformations in Speech. *The Journal of Neuroscience, 31*(7), 2657-2662. doi: 10.1523/jneurosci.6020-10.2011

Rosenbek, J. C., & Jones, H. N. (2009). Principles of treatment for sensorimotor speech disorders. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (2nd ed., pp. 269-286). New York: Thieme.

Rosenbek, J. C., & LaPointe, L. L. (1985). The dysarthrias: Description, diagnosis, and treatment. In D. F. Johns (Ed.), *Clinical management of neurogenic communication disorders* (2nd ed., pp. 97-152). Boston: Little, Brown, & Company.

Ross, B., Tremblay, K., & Picton, T. (2007). Physiological detection of interaural phase changes. *Journal of the Acoustical Society of America, 121*(2), 1017-1027.

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

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

Schneider, J. S., Diamond, S. G., & Markham, C. H. (1986). Deficits in orofacial sensorimotor function in Parkinson's disease. *Annals of Neurology, 19*, 275-282.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461-464.

Shaiman, S. (1989). Kinematic and electromyographic responses to perturbation of the jaw. *Journal of the Acoustical Society of America, 86*, 78-88.

Shiller, D. M., Gracco, V. L., & Rvachew, S. (2010). Auditory-motor learning during speech production in 9-11-year-old children. *PLoS One, 5*(9), e12975. doi: 10.1371/journal.pone.0012975

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *Journal of the Acoustical Society of America, 125*(2), 1103-1113.

Shum, M., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2011). Sensorimotor integration for speech motor learning involves the inferior parietal cortex. *European Journal of Neuroscience, 34*(11), 1817-1822.

Sivasankar, M., Bauer, J. J., Babu, T., & Larson, C. R. (2005). Voice responses to changes in pitch of voice or tone auditory feedback. *Journal of the Acoustical Society of America, 117*(2), 850-857.

Smiljanic, R. (2013). Can older adults enhance the intelligibility of their speech? *Journal of the Acoustical Society of America, 133*(2), EL129-EL135.

Smith, A., Weber, C. M., Newton, J., & Denny, M. (1991). Developmental and age-related changes in reflexes of the human jaw-closing system. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 81*(2), 118-128.

Smith, C. R. (1975). Residual hearing and speech production in deaf children. *Journal of Speech and Hearing Research, 18*(4), 795-811.

Solomon, N. P., McKee, A. S., & Garcia-Barry, S. (2001). Intensive voice treatment and respiration treatment for hypokinetic-spastic dysarthria after traumatic brain injury. *American Journal of Speech-Language Pathology, 10*(1), 51-64.

Sommers, M. S. (2008). Age-related changes in spoken word recognition. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (2nd ed., pp. 469-493). Malden, MA: Blackwell Publishing.

Sonies, B. C. (1991). The aging oropharyngeal system. In D. N. Ripich (Ed.), *Handbook of geriatric communication disorders* (pp. 187-203). Austin, TX: Pro-Ed.

Svirsky, M. A., Lane, H., Perkell, J. S., & Wozniak, J. (1992). Effects of short-term auditory deprivation on speech production in adult cochlear implant users. *Journal of the Acoustical Society of America, 92*(3), 1284-1300.

Svirsky, M. A., & Tobey, E. A. (1991). Effect of different types of auditory stimulation on vowel formant frequencies in multichannel cochlear implant users. *Journal of the Acoustical Society of America, 89*(6), 2895-2904.

Terband, H., Maassen, B., Guenther, F. H., & Brumberg, J. (2009). Computational neural modeling of speech motor control in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 52*, 1595-1609.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage, 29*, 1429-1443.

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature, 423*(6942), 866-869.

Troche, J., Troche, M. S., Berkowitz, R., Grossman, M., & Reilly, J. (2012). Tone discrimination as a window into acoustic perceptual deficits in Parkinson's disease. *American Journal of Speech-Language Pathology, 21*, 258-263.

Tun, P. A., Wingfield, A., Stine, E. A. L., & Mecsas, C. (1992). Rapid speech processing and divided attention: Processing rate versus processing resources as an explanation of age effects. *Psychology and Aging, 7*(4), 546-550.

Vennila, K., & Aruin, A. S. (2011). Postural control in response to a perturbation: Role of vision and additional support. *Experimental Brain Research, 212*, 385-397.

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

Vitale, C., Marcelli, V., Allocca, R., Santangelo, G., Riccardi, P., Erro, R., . . . Barone, P. (2012). Hearing impairment in Parkinson's disease: Expanding the nonmotor phenotype. *Movement Disorders, 27*(12), 1530-1535.

Waldstein, R. S. (1990). Effects of postlingual deafness on speech production: Implications for the role of auditory feedback. *Journal of the Acoustical Society of America, 88*(5), 2099-2114.

Walshe, M., Miller, N., Leahy, M., & Murray, A. (2008). Intelligibility of dysarthric speech: Perceptions of speakers and listeners. *International Journal of Language & Communication Disorders, 43*(6), 633-648.

Walshe, M., Peach, R. K., & Miller, N. (2009). Dysarthria impact profile: Development of a scale to measure psychosocial effects. *International Journal of Language & Communication Disorders, 44*(5), 693-715.

Wang, Y. T., Kent, R. D., Duffy, J. R., & Thomas, J. E. (2005). Dysarthria associated with traumatic brain injury: Speaking rate and emphatic stress. *Journal of Communication Disorders, 38*(3), 231-260.

Watkins, K. E., & Paus, T. (2004). Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience, 16*(6), 978-987.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41*(8), 989-994.

Wechsler, D. (1997). *Wechsler Memory Scale (WMS-III)* (3rd ed.). San Antonio: The Psychological corporation.

Weismer, G., Laures, J. S., Jeng, J. Y., Kent, R. D., & Kent, J. F. (2000). Effect of speaking rate manipulations on acoustic and perceptual aspects of the dysarthria in amyotrophic lateral sclerosis. *Folia Phoniatrica et Logopaedica, 52*(5), 201-219.

Welch, R. B. (1986). Adaptation of space perception. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, pp. 1-45). New York: John Wiley and Sons.

Wenke, R. J., Theodoros, D., & Cornwell, P. (2008). The short- and long-term effectiveness of the LSVT (R) for dysarthria following TBI and stroke. *Brain Injury, 22*(4), 339-352.

Wertz, R. T., Henschel, C. R., Auther, L. L., Ashford, J. R., & Kirshner, H. S. (1998). Affective prosodic disturbance subsequent to right hemisphere stroke: A clinical application. *Journal of Neurolinguistics, 11*(1-2), 89-102.

Whitehill, T. L., Lee, A. S. Y., & Chun, J. C. (2002). Direct magnitude estimation and interval scaling of hypernasality. *Journal of Speech, Language, and Hearing Research, 45*, 80-88.

Wilson, M., & Knoblich, G. (2005). The case of motor involvement in perceiving conspecifics. *Psychological Bulletin, 131*(3), 460-473.

Wilson, S. M. (2009). Speech perception when the motor system is compromised. *Trends in Cognitive Sciences, 13*(8), 329-330.

Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Neuroimage, 33*(1), 316-325.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7*(7), 701-702.

Wohlert, A. B. (1996). Tactile perception of spatial stimuli on the lip surface by young and older adults. *Journal of Speech and Hearing Research, 39*(6), 1191-1198.

Wohlert, A. B., & Smith, A. (1998). Spatiotemporal stability of lip movements in older adult speakers. *Journal of Speech, Language, and Hearing Research, 41*, 41-50.

Xia, K., & Espy-Wilson, C. (2000). A new strategy of formant tracking based on dynamic programming. In ICSLP-2000 (Ed.), *Proceedings of the Sixth International Conference on Spoken Language Processing* (Vol. 3, pp. 55-58). Beijing, China. Available at http://www.isca-speech.org/archive/icslp_2000/i00_3055.html.

Yorkston, K. M., Baylor, C. R., Klasner, E. R., Deitz, J., Dudgeon, B. J., Eadie, T., . . . Arntmann, D. (2007). Satisfaction with communicative participation as defined by adults with multiple sclerosis: A qualitative study. *Journal of Communication Disorders, 40*(6), 433-451.

Yorkston, K. M., Hakel, M., Beukelman, D. R., & Fager, S. (2007). Evidence for effectiveness of treatment of loudness, rate, or prosody in dysarthria: A systematic review. *Journal of Medical Speech-Language Pathology, 15*(2), xi-xxxvi.

Yorkston, K. M., Spencer, K. A., & Duffy, J. R. (2003). Behavioral management of respiratory/phonatory dysfunction from dysarthria: A systematic review of the evidence. *Journal of Medical Speech-Language Pathology, 11*(2), xiii-xxxviii.

Yorkston, K. M., Strand, E. A., & Kennedy, M. R. T. (1996). Comprehensibility of dysarthric speech: Implications for assessment and treatment planning. *American Journal of Speech-Language Pathology, 5*(1), 55-66.

Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. (2009). 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.

Zraick, R. I., Gregg, B. A., & Whitehouse, E. L. (2006). Speech and voice characteristics of geriatric speakers: A review of the literature and a call for research and training. *Journal of Medical Speech-Language Pathology, 14*(3), 133-142.