Predicting Treatment Outcomes in Dysarthria through Speech Feature Analysis

---

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

by Annalise Rebecca Fletcher

Department of Communication Disorders

University of Canterbury

2016

DECLRATION BY AUTHOR

I declare that this thesis is my own original work. It does not contain material written or published by a third party, nor material which has been submitted for any other degree or diploma at a university or other institution of higher learning.

SUBMITTED WORKS BY AUTHOR INCORPERATED INTO THE THESIS

The following referred journal submissions originated from the work presented in this thesis.

Fletcher, A. R., McAuliffe, M. J., Lansford, K. L., & Liss, J. M. (2015). The relationship between speech segment duration and vowel centralization in a group of older speakers. *The Journal of the Acoustical Society of America, 138*(4), 2132-2139.

This article is incorporated into chapter two

Fletcher, A. R., McAuliffe, M. J., Lansford, K., & Liss, J. M. (in press). Assessing vowel centralization in dysarthria: A comparison of methods. *Journal of Speech, Language and Hearing Research.*

This article is incorporated into chapter three

Fletcher, A. R., McAuliffe, M. J., Lansford, K., D. G. Sinex., & Liss, J. M. (under review). Predicting Intelligibility Gains in Individuals with Dysarthria from Baseline Speech Features. *Journal of Speech, Language and Hearing Research.*

This article is incorporated into chapter four

RELEVANT PUBLICATIONS NOT INCLUDED IN THE THESIS

McAuliffe, M. J., **Fletcher, A. R.**, Kerr, S. E., Anderson, T., & O'Beirne, G. (in press). Effect of dysarthria type, speaking condition and listener age on speech intelligibility. *American Journal of Speech-Language Pathology*.

LIST OF PRESENTATIONS BY THE AUTHOR RELEVANT TO THE THESIS

Aspects of this thesis were presented at the following conferences:

Fletcher, A., McAuliffe, M., Lansford, K., (2014). Changes in vowel articulation across older speakers: The interaction of speech rate and precision. *American Speech-Language-Hearing Association Convention,* Orlando, United States: November, 2014.

Fletcher, A., McAuliffe, M., Lansford, K., Liss, J., (2014). Distinguishing dysarthric speech: Vowel acoustics and measurement. *Australasian International Speech Science and Technology Conference*, Christchurch, New Zealand: December, 2014.

Fletcher, A., McAuliffe, M., Lansford, K., Liss, J., (2015). Assessing vowel centralization in dysarthria: A comparison of methods, *170th Meeting of the Acoustical Society of America,* Jacksonville, United States: November, 2015.

***Fletcher, A.,** McAuliffe, M., Lansford, K., Sinex, D., Liss, J., (2015). Predictors of intelligibility improvement in dysarthria: A treatment simulation study. *American Speech-Language-Hearing Association Convention,* Denver, United States: November, 2015

Fletcher, A., McAuliffe, M., Lansford, K., Sinex, D., & Liss, J. (2016). Comparing treatment strategies: Rate control and increased loudness. *Conference on Motor Speech,* Newport Beach, United States: March, 2016.

*Awarded an ASHA Convention Student Research Travel Award for top student submission in topic area

CONTENTS

DECLRATION BY AUTHOR ii
SUBMITTED WORKS BY AUTHOR INCORPERATED INTO THE THESIS iii
RELEVANT PUBLICATIONS NOT INCLUDED IN THE THESIS iv
LIST OF PRESENTATIONS BY THE AUTHOR RELEVANT TO THE THESIS v
ACKNOWLEDGMENTS 1
ABSTRACT 4
LIST OF TABLES 7
LIST OF FIGURES 8
Chapter One 9
1.1 INTRODUCTION 10
1.2 DYSARTHRIA 11
1.2.1 Prevalence 12
1.3 TREATMENT OF DYSARTHRIA 13
1.3.1 Cued Speech Studies 15
1.3.2 Variability in Intelligibility Gains 15
1.3.3 Selection of Participants for Treatment Studies 17
1.4 NEW DIRECTIONS IN DYSARTHRIA ASSESSMENT 19
1.4.1 Measurement of Speech Features in Dysarthria 20
1.5 SUMMARY AND AIMS OF THE PRESENT THESIS 22
Chapter Two 25
2.1 PREFACE 26
2.2 INTRODUCTION TO VOWEL ARTICULATION 26
2.3 VOWEL CHANGES IN OLDER SPEAKERS 27
2.4 EFFECT OF SPEAKING RATE ON VOWEL CENTRALIZATION 28
2.5 SUMMARY AND AIMS OF THE CURRENT STUDY 30
2.6 METHODS 30
2.6.1 Participants 30
2.6.2 Speech Stimuli 31
2.6.3 Extraction of Acoustic Data 31
2.6.3.1 Segmentation of the data set 31
2.6.3.2 Measurement of mean vowel duration 32
2.6.3.3 Measurements of formant frequencies 32
2.6.3.4 Vowel space area 33

2.6.3.5 Reliability of acoustic measures 34
2.7 RESULTS 34
2.7.1 Comparison of VSA from Articulatory and Midpoint Values 34
2.7.2 Relationship between Average Vowel Duration and VSA 35
2.7.2.1 Predictors of midpoint VSA 36
2.7.2.2 Predictors of articulatory point VSA 36
2.7.3 Vowel Duration as a Function of Age 36
2.7.4 Relationship between Formant Values and Age 37
2.8 DISCUSSION 38
2.8.1 Measurement of Vowel Articulation in NZE 38
2.8.2 Effect of Average Vowel Duration on Vowel Space Area 40
2.8.3 Changes in Vowel Articulation with Age 41
2.8.3.1 Changes in vowel duration with age 42
2.8.3.2 Changes in vowel formants with age 43
2.9 CONCLUSIONS 43
2.9.1 Summary 45
Chapter Three 46
3.1 PREFACE 47
3.2 INTRODUCTION 48
3.2.1 Time-point of Formant Extraction 49
3.2.2 Calculation of Vowel Centralization 49
3.2.3 Perceptual Measurement of Speech Disorder 51
3.3 METHOD 52
3.3.1 Speakers 52
3.3.2 Speech Stimuli 54
3.3.3 Extraction of Acoustic Data 55
3.3.3.1 Segmentation of the data set 55
3.3.3.2 Extraction of formant values 55
3.3.3.3 Midpoint formant values 56
3.3.3.4 Articulatory point formant values 56
3.3.4 Description of the Acoustic Metrics 57
3.3.4.1 Vowel space area 57
3.3.4.2 Formant centralization ratio 58
3.3.5 Reliability of Acoustic Measures 58
3.3.6 Perceptual Task 59

3.3.6.1 Listeners 59
3.3.6.2 Listening stimuli 59
3.3.6.3 Procedure 59
3.3.6.4 Reliability of the perceptual task 60
3.4 RESULTS 61
3.4.1 Differences in Measurements 61
3.4.1.1 Method of formant extraction 61
3.4.1.2 Unit of measurement and vowel centralization metric 61
3.4.1.3 Perceptual correlates of dysarthric speech 62
3.4.2 Measurement Differences between Speakers with and without Dysarthria 63
3.4.3 Relationship between Acoustic and Perceptual Measures 65
3.5 DISCUSSION 69
3.5.1 Method of Formant Extraction 69
3.5.2 Unit of Measurement and Vowel Centralization Metric 70
3.5.3 Perceptual Correlates 71
3.5.4 Limitations 72
3.5.4.1 Use of peripheral vowel space to index articulatory disorder 72
3.5.4.2 Generalization to other dialects 73
3.5.4.3 Effects of speaker sex 74
3.5.5 Summary 75
Chapter Four 77
4.1 PREFACE 78
4.2 INTRODUCTION 78
4.2.1 Issues with Categorizing Speakers who have Dysarthria 79
4.2.2 The Relationship between Speech Features and Treatment Outcomes 80
4.2.3 Summary 82
4.3 METHOD 83
4.3.1 Speakers 83
4.3.2 Speech Stimuli 84
4.3.3 Procedures 84
4.3.4 Step One: Determining Intelligibility Gains 84
4.3.4.1 Reliability of measurements of intelligibility gain 86
4.3.5 Step Two: Analyses of Baseline Speech Features 87
4.3.5.1 Perception task: Speech severity. 87
4.3.5.2 Acoustic analysis 87

4.4 RESULTS....89
4.4.1 Effect of Loud and Slow Cued Speech on Intelligibility....89
4.4.2 Acoustic Measures of Dysarthric Speech....90
4.4.3 Predicting Intelligibility Improvement....91
4.4.3.1 Model one: Intelligibility gains in response to cues to speak slower....92
4.4.3.2 Model two: Intelligibility gains in response to cues to speak louder....93
4.4.4 Choosing between Treatment Cues....93
4.4.4.1 Model three: Speakers' most successful strategy....94
4.5 Discussion....94
4.5.1 Intelligibility Gains Following Cues to Speak Louder and Slow Rate of Speech 95
4.5.2 Predictors of Intelligibility Gain in the Slow Condition....96
4.5.3 Predictors of Intelligibility Gain in the Loud Condition....97
4.5.4 Comparing Treatment Cues in the Same Speakers....98
4.5.5 Summary....98
Chapter 5....101
5.1 Preface....102
5.2 Introduction....102
5.2.1 Automated Measurements of the Dysarthric Speech Signal....103
5.2.2 Summary and Aims of the Current Study....105
5.3 Method....105
5.3.1 Speakers and Speech Stimuli....105
5.3.2 Perceptual Experiment to Determine Intelligibility Gain....106
5.3.3 Baseline Speech Analysis....106
5.3.3.1 Mel-frequency cepstral coefficients....106
5.3.3.2 Envelope modulation....107
5.3.3.3 Long-term average spectra....107
5.3.4 Statistical Analyses....107
5.4 Results....108
5.4.1 Intelligibility Gain in Response to Cueing....108
5.4.2 Stepwise Regression and Cross-Validation on Untrained Speakers....109
5.4.3 Final Models of Intelligibility Gain....111
5.4.3.1 Predicting intelligibility gains in the loud condition....111
5.4.3.2 Predicting gains in the slow condition....111
5.5 Discussion....111

5.5.1 Cross Validation of Models 112
5.5.2 Final Model Selected using Automated Baseline Speech Analyses 113
Chapter 6 115
6.1 Summary 116
6.2 Limitations 119
6.2.1 Sample Size 119
6.2.2 Therapist Cueing Strategies 120
6.2.3 Measurements of Intelligibility Gain 121
6.3 Future Directions 122
6.4 Conclusions 123
REFERENCES 124
Appendix A: The Grandfather Passage 137
Appendix B: Listener Rating Instructions 138

ACKNOWLEDGMENTS

When I started my PhD candidacy, the idea of producing a thesis was daunting. Although I was passionate about my field and eager to continuing learning, the work required felt enormous, uncertain and unstructured. When I look back, it is sometimes difficult to recognise how far I have travelled. I have learnt so much in my PhD studies. This learning comes not just from the development of new technical skills and knowledge, but from the confidence, curiosity, and determination I've gained in the process. I have enjoyed my PhD journey so very much, and I owe thanks to many people for guiding me along the way.

Firstly, I am forever grateful to my team of supervisors for their advice, support and friendship. My primary advisor, Megan McAuliffe, has taught me so much about being an effective researcher. But perhaps more fundamentally, she gave me the encouragement that I needed to begin this PhD. For all my achievements, I credit Megan's counselling. Megan helped me maintain enthusiasm and good-humour throughout my time as a PhD candidate—while teaching me to approach problems in a structured way, seek help when needed, and persevere. She has been unceasingly optimistic, approachable and open to my ideas. I owe her my biggest thanks.

I am also very fortunate to have the support of two associate advisors: Kaitlin Lansford and Julie Liss. Kaity and Julie made time to meet and advise me throughout this PhD—despite being external to both my University and country. I am particularly appreciative of their support in my endeavour to apply for a Fulbright scholarship and study in the United States (US). Under ordinary circumstances, moving to another country while completing a PhD dissertation would be stressful. However, thanks to the friendship and practical support I received from them, the transition was remarkably smooth. I am privileged to have experienced their kindness and generosity. They gave me exposure to new ideas, helped me gain insight into different university systems, and—above all—made me feel at home in Tallahassee and Phoenix, respectively. My PhD experience has been so much richer thanks to their support.

There are several other individuals who assisted me in developing new skills during my PhD. I owe special thanks to Patrick LaShell for helping me to conceptualise how to measure and model 'treatment outcomes'. Pat was a fantastic teacher, and he inspired me to persevere in

developing my knowledge of statistics. I remember him every time I try to code something new in R! I am also very grateful to Donal Sinex. Don assisted me in planning and programming acoustic measurements of voice quality. He was always willing to discuss and trial different methods with me and has been a pleasure to work with. Finally, I am thankful to have worked with Alan Wisler. Alan introduced me to new automated acoustic measures, and spent considerable time helping me figure out how to apply them to my work. Alan has been a wonderful source of new ideas and I am excited to continue our collaboration.

There are so many other individuals who have brightened my time as a PhD student that I am afraid I cannot list them all individually. I am grateful for the support of the Department of Communication Disorders, and my many friends within it. I am also grateful to the inclusive and welcoming group of staff and students within the New Zealand Institute of Language Brain and Behaviour (NZILBB). I have benefited so much from your seminars, stats groups, and social events. The NZILBB houses a group of researchers who I especially admire for their openness to new ideas, and willingness to share knowledge and data. I have many memorable experiences of my studies because of you all! I am also appreciative of the wider group of students and faculty at Florida and Arizona State Universities, who included me in their labs, seminars and social events. I am fortunate to have met so many talented people around the world!

I am also fortunate to have had financial support throughout my PhD. My research was primarily funded by a Sir Don Beaven Doctoral Scholarship. I am honoured to hold an award that bears the name of such an influential medical researcher. I am also thankful for the Fulbright General Graduate Award which funded my studies in the US. The Fulbright organization supported my application for a US visa and provided orientations and seminars to aid me in adjusting to US culture. Above all, being a ‘Fulbrighter’ gave me access to a wonderful network of friendly scholars who I am proud to be associated with.

Being based in New Zealand during the majority of my PhD made presenting at international conferences a particular challenge. For this reason, I am thankful for all the travel awards I received over the last three years. Specifically, I appreciate the support of the Academy of Neurologic Communication Disorders and Sciences Conference Fellowship Award, ASHA Student Travel Award, Claude McCarthy Fellowship and New Zealand Federation of Graduate Women Travel Grant.

To conclude, I must acknowledge my loving family and longstanding, supportive friends in Christchurch. I would not be who I am without your influence.

ABSTRACT

Increasing loudness and reducing speech rate are common behavioural speech modifications used in the treatment of dysarthria. Both strategies provide a promising basis for intervention, but studies have reported considerable inter-participant variability in the intelligibility benefit gained. While neurologic aetiology and the Mayo classification system are generally used to group participants with dysarthria in treatment studies, there is little evidence that this approach provides meaningful insight into which individuals benefit from particular speech modification strategies. This thesis posited that differences in baseline speech symptoms between speakers could underlie significant disparities in treatment outcomes. Hence the overall aim, addressed in the final phase of this thesis, was to identify whether measurements of individuals' baseline speech could be used to predict their intelligibility gains in response to cues to speak louder and reduce speech rate. To begin, the first two phases of this thesis addressed methodological issues in the assessment of speech features. The purpose of these investigations was to refine the methods of acoustic and perceptual analysis employed in the final phase, and test their application on New Zealand speakers with and without dysarthria.

Phase one of this thesis focused on vowel dispersion and speech duration in healthy, older speakers of New Zealand English (NZE), as it was unknown how NZE's unique dialect might affect commonly used metrics of vowel articulation. A group of 149 NZE speakers aged 65 to 90 years read a standard passage. Two formant measurements, from selected [iː], [ɐː], and [oː] vowels, were used to calculate two measurements of Vowel Space Area (VSA) for each speaker. Average vowel duration was calculated from segments across the passage. Results demonstrated that measures of VSA, adapted for NZE, produced a similar range of values to those reported in previous studies of speakers from the United States. In addition, a statistically significant relationship existed between speakers' average vowel durations and VSA measurements indicating that, on average, speakers with slower speech rates produced more acoustically distinct speech segments. As expected, increases in average vowel duration were found with advancing age. However, speakers' formant values remained unchanged.

The second phase explored the ability of different acoustic and perceptual measures to index speakers' baseline (i.e., habitual) dysarthria severity. Sixty-one speakers (17 healthy individuals and 44 speakers with dysarthria) read a standard passage. To obtain acoustic data, different formant extraction points and frequency measures were trialled. VSA and an adapted Formant Centralization Ratio (FCR) were calculated using first and second formants of the speakers' [ɐː], [iː] and [oː] vowels. Twenty-eight listeners completed separate ratings

of intelligibility and speech precision. Perceptually, listener ratings of speech precision provided the best index of acoustic change. Acoustically, the combined use of an articulatory-based formant extraction point, Bark frequency units, and the FCR was most effective in explaining perceptual ratings. Based on this investigation, perceptual ratings of speech precision and acoustic measurements of FCR (derived from a flexible extraction point and calculated in Bark) were selected for use in phase three, to model speakers' responses to treatment cues.

Phase three addressed the central aim of the thesis by exploring whether targeted acoustic and perceptual features of participants' habitual speech could predict their degree of intelligibility improvement in response to cues to speak louder and reduce speech rate. Fifty speakers of NZE participated (aged between 43 and 89 years), 43 of whom were diagnosed with dysarthria. The remaining speakers acted as healthy controls. All participants read the Grandfather Passage in habitual, loud and slow speaking modes. The study was conducted in two parts. Firstly, a perceptual experiment was completed, where 18 listeners rated the intelligibility of speakers' habitual, loud and slow speech recordings. Secondly, an acoustic analysis was completed to measure a range of baseline speech features. This speech analysis employed measurements from the phase two investigation, in conjunction with acoustic measures of articulatory rate, vowel harmonics, cepstral peak prominences, and variability in speakers' vowel durations, pitch and speech intensity. Statistical analyses revealed that intelligibility gains in the loud condition were best predicted by speakers' baseline articulatory rate and listener ratings of baseline speech precision. Intelligibility gains in the slow condition were best modelled by ratings of speech precision and variations in speakers' vowel durations. Overall, these models were able to account for approximately 1/3 of the variance in speakers' intelligibility gains. These findings were promising, but the time required to analyse the acoustic data limited the clinical applicability of the models.

A follow up study investigated whether time-efficient, automated measurements of the baseline speech signal could similarly account for differences in speakers' responses to treatment cues. The performance of features derived from Mel-frequency cepstral coefficients, long term average spectra and envelope modulation spectra was compared against the targeted measurements extracted in the previous study. Cross-validation techniques were used to determine how well models of intelligibility gain could perform on speakers they had not been trained on. When the optimal number of speech features were included in a forward regression model, 17% of the variance in speakers' responses to cues to speak slower could be accounted for by the targeted speech features. The automated

measurements accounted for around 10%. In contrast, in the loud condition, both feature sets exhibited a stronger performance. The automated features were able to account for up to 25% of the variance in speakers' intelligibility gains, while the targeted measures accounted for 19%. Thus this final investigation offered evidence that automated feature sets, which are time efficient and require no subjective judgments of researchers, could be used diagnostically to guide treatment decisions.

Overall, this thesis demonstrated that certain features of speakers' baseline speech could account for significant variation in their intelligibility gains. The ability to classify speakers likely to achieve positive treatment outcomes based on their presenting speech features has the potential to facilitate clinical decision making within an evidence-based framework and, ultimately, promote stronger group treatment outcomes. Future studies that utilize larger participant groups and a wider range of treatment strategies are needed to develop more personalized and targeted approaches to speech therapy for people with dysarthria.

LIST OF TABLES

Table 3.1 Demographic Information for Speakers with Dysarthria 53
Table 3.2 Combinations of Acoustic Vowel Metrics.....58
1. Table 3.3 Measurement Differences Between Males and Females 62
Table 3.4 Differences in Perceptual Ratings and Vowel Dispersion Metrics in Participants with and without Dysarthria.....64
Table 3.5 Relationships Between Acoustic Vowel Metrics and Perceptual Measures 66
Table 4.1 Average Values of Acoustic Measures Across Speakers with Dysarthria and Healthy Controls 91
Table 4.2 Across Speaker Correlations between Targeted Acoustic Measurements.....92

LIST OF FIGURES

Figure 2.1. A comparison of midpoint and articulatory point VSA values for male and female speakers. 35
Figure 2.2. A comparison of midpoint and articulatory point VSA values for male and female speakers. Shaded area indicates 95% confidence interval of regression estimate. 35
Figure 2.3. Average vowel duration as a function of age. 37
Figure 3.1. Example of the two extraction points within a speaker's [oː] vowel. 57
Figure 3.2. Relationship between listeners' ratings of intelligibility and speech precision. 63
Figure 3.3. A comparison of the relationships between acoustic vowel metrics and perceptual measures, plotted by sex. 68
Figure 4.1. Visual analogue scale presented to listeners in the experiment to determine intelligibility gains. 86
Figure 4.2. The relationship between speakers' intelligibility gain indices in the loud and slow cueing conditions. 90
Figure 5.1. Relationship between test speakers' baseline speech features and their intelligibility gain in the loud condition as given by forward stepwise regression models 110
Figure 5.2. Relationship between test speakers' baseline speech features and their intelligibility gain in the slow condition as given by forward stepwise regression models 110

Chapter One

Literature Review

1.1 INTRODUCTION

Spoken language is the primary means through which humans communicate thought. But for individuals with neurological impairment or disease, the ability to produce intelligible and natural-sounding speech can be significantly affected. Reduced intelligibility alters how feelings or ideas are expressed and perceived (Dickson, Barbour, Brady, Clark, & Paton, 2008). This diminished ability to communicate effectively can impact a person's willingness to participate in conversations, leading to changes in self-perception and feelings of social isolation (Miller, Noble, Jones, & Burn, 2006). Consequently, the onset of a speech disorder following neurological injury or disease often triggers significant negative changes in a person's quality of life (Dickson et al., 2008; Miller et al., 2006).

Dysarthria is an umbrella term for neurological impairment of motor speech control. There are a range of causes of dysarthria, including degenerative disorders (e.g. Parkinson's disease, motor neuron disease), stroke and traumatic brain injury (Yorkston, 1996). No one person with dysarthria "sounds" identical to another. Substantial variability lies in the presentation of speech symptoms, with the "sound" of a person's dysarthria dependent on both injury-based factors (e.g., the wide variety of possible sites of lesion, severity of injury) and individual differences in the indexical features of their speech (Duffy, 2013). Thus, even when two people's dysarthria results from a similar neurogenic origin, their speech features may vary considerably across speech rate, lexical stress, articulation of vowels and consonants, and voicing characteristics (Y. Kim, Kent, & Weismer, 2011).

Judgements of dysarthria type – through analysis of neurological presentation and key perceptual speech features – provide the basis for selection of rehabilitation strategies in speech therapy (Simmons & Mayo, 1997). However, even within groups of a uniform dysarthria subtype, quality treatment studies often report insignificant treatment effects, with significant intelligibility gains for some participants and little measurable change for others (e.g. Lowit, Dobinson, Timmins, Howell, & Kröger, 2010; Mahler & Ramig, 2012).

Two common strategies employed in the clinical remediation of speech deficits in dysarthria are increased loudness and reduced speech rate, and there has been significant attention in the research literature to the effects of these speech modifications on intelligibility (e.g. Hammen, Yorkston, & Minifie, 1994; Neel, 2009; Pilon, McIntosh, & Thaut, 1998; Tjaden & Wilding, 2004; Turner, Tjaden, & Weismer, 1995; Van Nuffelen, De Bodt, Vanderwegen, Van de Heyning, & Wuyts, 2010; Van Nuffelen, De Bodt, Wuyts, & Van de Heyning, 2009; Yorkston, Hammen, Beukelman, & Traynor, 1990). While both

techniques provide a promising basis for intervention, there are notable issues with the evidence behind these strategies. The first issue is that studies of these techniques have often been restricted to, and dominated by, specific dysarthria aetiologies. This limits the clinician's ability to make inferences about appropriate treatment strategies when their clients do not fit these moulds. The second issue is that, even amongst groups with the same dysarthria aetiologies, studies have reported considerable inter-participant variability in clinical outcomes, with some speakers demonstrating improved speech production with cueing, while others exhibit no significant intelligibility gain (e.g. Neel, 2009; Pilon et al., 1998; Turner et al., 1995). Similarly, in cases when more than one cueing strategy is trialled, it remains unclear why certain participants achieve greater intelligibility gains in one condition over another (e.g. McAuliffe, Fletcher, Kerr, Anderson, & O'Beirne, in press; Tjaden & Wilding, 2004).

To combat these issues, we need to develop better profiles of the speakers who benefit from different treatment techniques. The current thesis approaches this task by examining the acoustic and perceptual features of participants' baseline speech. These data are used to model the variation observed in speakers' intelligibility gains following two behavioural speech modification strategies. The resulting models provide a theoretically-motivated basis for the selection of participants for future studies. Furthermore, these data represent a first step in the development of more individualised frameworks for selecting treatment programs.

The current chapter serves as an introduction to issues that exist in the assessment and classification of speakers with dysarthria. The purpose of this chapter is to: 1) provide an introduction to dysarthria and the behavioural speech strategies used in its remediation 2) discuss issues faced in the diagnosis and classification of dysarthria, and 3) introduce methods for examining speech features in dysarthria.

1.2 DYSARTHRIA

Dysarthria is a speech disorder arising from impairments to the central or peripheral nervous system that reduces the control or execution of motor speech movements. Dysarthria can affect the activation and coordination of muscles necessary for respiratory, phonatory, resonatory, articulatory or prosodic aspects of speech production (Duffy, 2013). The result is speech that, to the layperson, is often described as "slurred," "rough," "mumbled" or "slow" (Mackenzie, 2011; Miller et al., 2006). Reduced intelligibility is a hallmark feature of dysarthria. Deficits range in severity from mild—with increased attention required by the

listener to understand speech—through to profound disorder and unintelligible speech. Intelligibility impairment is considered one of the most clinically important aspects of dysarthria, as it directly impacts functional communicative performance (Yorkston & Beukelman, 1978). Intelligibility impairment can restrict people's participation in everyday activities and have significant deleterious effects on their quality of life. However, even when speech impairment is the primary cause of a person's disability, the extent to which it affects participation can be highly variable. This variation stems from differences in peoples' personalities, coping mechanisms, and the demands of their social or vocational environment (Dykstra, Hakel, & Adams, 2007).

Finally, while dysarthria is generally considered a movement disorder, it rarely occurs in isolation. Associated neurological aetiologies frequently cause co-occurring impairment to language and cognitive skills. As a result, people with dysarthria often also present with difficulties in managing their attention, memory and mood, and sometimes lack full self-awareness of their disorder (Duffy, 2013; Fox, Morrison, Ramig, & Sapir, 2002). These complexities provide additional challenges in managing the speech disorder.

1.2.1 Prevalence

The overall incidence of dysarthria is difficult to quantify, but it is thought to be one of the most prevalent acquired communication disorders (Duffy, 2013). Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), stroke, multiple sclerosis, traumatic brain injury and cerebral palsy are all cited as clinical aetiologies in which dysarthria occurs as a frequent and prominent symptom (Yorkston, 1996). Part of the difficulty in quantifying dysarthria's incidence is that, while many people present with stable, persisting speech disorders, some naturally recover to varying degrees (e.g. in the initial months following stroke or traumatic brain injury), while others experience progressively degenerative symptoms (as seen in PD and ALS) (Duffy, 2013).

Nevertheless, dysarthria is commonly associated with ageing—and the ageing population means that the incidence of dysarthria is increasing. Within New Zealand, individuals aged over 65 years account for approximately 12.3% of the national population and form one of the fastest growing population sectors (Statistics New Zealand, 2007). The most common dysarthria aetiologies are stroke — which affects approximately 5-7% of people aged 65 and older — and PD — which is estimated to affect 1% of the population

over 60 (de Lau & Breteler, 2006; Feigin, Lawes, Bennett, & Anderson, 2003). Hence, an ageing population contributes to the increasing prevalence of dysarthria.

1.3 TREATMENT OF DYSARTHRIA

There are a wide range of treatment options available for speakers with dysarthria. Some impairments associated with disorder can be treated medically. For example, in certain cases, pharmacological management or surgical procedures can alleviate aspects of the underlying disease processes that contribute to the speech impairment. In other cases, surgical implants offer a more direct way of compensating for weak or paralysed muscles (e.g. a thyroplasty or pharyngeal flap). Prosthetic devices (e.g. palatal lift prosthesis), or assistive communication devices (e.g. voice amplifiers) can also provide some compensation in cases where particular muscles are weak or paralysed (Duffy, 2013). However, while these options can significantly help speakers with dysarthria, their application remains limited. Pharmacological and surgical interventions are usually unable to cure or completely halt the progression of dysarthria. Implants and assistive devices tend to only address impairments within certain speech subsystems, and only aid specific types of muscle impairment. For these reasons, there are many speakers with dysarthria for which none of the aforementioned treatment options are appropriate (Spencer & Beukelman, 2001).

In contrast, behavioural intervention can be utilized by speakers with a range of dysarthrias—and is the primary focus of speech therapy. Speech therapy, in its broadest sense, aims to improve the quality of life of speakers with dysarthria by enhancing their ability to communicate in everyday situations. Various approaches and strategies are used to try and achieve these improvements. For example, there is evidence that certain behavioural alterations, like changes to posture and breath control (Pennington, Smallman, & Farrier, 2006) or practicing specific articulatory targets (Marchant, McAuliffe, & Huckabee, 2008; Robertson, 2001) may aid speakers in producing more intelligible speech. At present, these exercises are regularly incorporated into ‘traditional dysarthria therapy’ programmes (e.g. Palmer, Enderby, & Hawley, 2007; Wenke, Theodoros, & Cornwell, 2011). Additionally, in recent years, there has been increased consideration of how the listener and communicative environment might contribute to a person’s disability (Howe, 2008). As a result, strategies for the communication partner, to help reduce communicative breakdowns (e.g. Yorkston, Strand, & Kennedy, 1996), are also being utilized in speech therapy.

Unfortunately, evidence of improved outcomes following dysarthria therapy is still lacking. Recent Cochrane reviews have questioned the literature underpinning behavioural dysarthria intervention, finding no high quality, large-scale studies to support or refute the efficacy of treatment (Herd et al., 2012; Sellars, Hughes, & Langhorne, 2005). One issue in evaluating the efficacy of speech therapy is that most of the techniques reviewed have only been trialled on small numbers of individuals—and have varied considerably in their implementation from one study to another. This makes it difficult to evaluate their effectiveness for other speakers with dysarthria.

Fortunately, there is growing evidence behind some behavioural approaches (Fox et al., 2006; Yorkston, Hakel, Beukelman, & Fager, 2007). As a first step towards assessing treatment efficacy, this thesis focuses on behavioural strategies that have been trialled on larger treatment study groups and have a rapidly growing evidence base. Two primary examples are programs aimed at reducing speech rate and increasing loudness (e.g. Cannito et al., 2012; Lowit et al., 2010; Mahler & Ramig, 2012). These speech modifications have been shown to result not only in changes to rate and loudness parameters, but also in more diffuse global acoustic changes to articulation, prosody and voice characteristics (Baumgartner, Sapir, & Ramig, 2001; Tjaden & Wilding, 2004). For this reason, loud and slow cued speech are the foundation of many well-established treatment programs (i.e. the Lee Silverman Voice Treatment program (LSVT)) and strategies (e.g. pacing boards, delayed auditory feedback).

In addition, there are practical reasons why these programs may be particularly successful in eliciting positive treatment changes. As mentioned earlier, cognitive impairment is a common co-occurring issue for people with dysarthria. Concepts of ‘loudness’ and ‘rate’ are concrete and simple to understand. This makes louder and slower speech easier to elicit than complex changes of intonation, or changes to more abstract speech qualities (such as cues to speak ‘clearer’). To best facilitate generalisation of new motor patterns, procedures in dysarthria therapy need be simple, with multiple repetitions of target speech behaviour within each treatment session (Fox et al., 2002; Maas et al., 2008). Loudness and rate can be easily monitored during therapy, allowing clinicians to provide accurate and frequent feedback vital for establishing new motor patterns (Maas et al., 2008). Furthermore, cues to speak louder and slower can easily be modelled by a clinician—making the desired behaviour more salient (Fox et al., 2002).

In summary, techniques aimed at increasing loudness and reducing speech rate have been promoted for speakers with a range of dysarthria aetiologies (Fox & Boliek, 2012; Sapir

et al., 2003; Van Nuffelen et al., 2010), which makes them an important tool in the management of dysarthria. However, the success of these techniques is not entirely clear. Not all speakers with dysarthria gain intelligibility benefit from these treatment programs (Cannito et al., 2012). Partly as a result of this variation in individual response, treatment studies often fail to demonstrate intelligibility improvements across speaker groups (e.g. Lowit et al., 2010; Mahler & Ramig, 2012). Being able to identify the types of speakers who are likely to make intelligibility gains may allow us to better target our treatment strategies, promoting stronger outcomes in future studies.

1.3.1 Cued Speech Studies

Cued speech studies provide valuable insight into how individuals may respond to a larger program of treatment. Thus far, cueing strategies applied directly to dysarthric speech have shown promising improvements in blinded listeners' understanding of dysarthric phrases (Hammen et al., 1994; Neel, 2009; Patel, 2002; Patel & Campellone, 2009; Pilon et al., 1998; Tjaden & Wilding, 2004; Turner et al., 1995; Van Nuffelen et al., 2010; Van Nuffelen et al., 2009; Yorkston et al., 1990). Such research provides a foundation for building treatment studies to support the efficacy of dysarthria intervention.

However, while cues to both reduce rate and speak louder have shown promising effects on intelligibility in speakers with dysarthria, it is not yet clear in what instances clinicians should favour a particular strategy. Current research has shown mixed treatment outcomes for speakers with dysarthria, with some individuals demonstrating improved intelligibility with slow speech, some improving with loud speech, and others demonstrating no significant treatment effects (Hammen et al., 1994; Neel, 2009; Pilon et al., 1998; Tjaden & Wilding, 2004; Turner et al., 1995; Van Nuffelen et al., 2010; Van Nuffelen et al., 2009; Yorkston et al., 1990). This issue is considered in more detail in the following section.

1.3.2 Variability in Intelligibility Gains

Increasing loudness and slowing rate of speech are well established as behavioural speech remediation techniques in dysarthria. However, significant variation exists in the effect of these cues on speech intelligibility across speakers (Hammen et al., 1994; Neel, 2009; Pilon et al., 1998; Tjaden & Wilding, 2004; Turner et al., 1995; Van Nuffelen et al., 2010; Van Nuffelen et al., 2009; Yorkston et al., 1990). For example, studies examining the effects of loud speech on intelligibility have mainly focused on hypokinetic dysarthria, which is closely

associated with Parkinson's disease. Neel (2009) examined the effect of loud cued speech on intelligibility in five people with PD and hypokinetic dysarthria. Four of these speakers exhibited significantly improved intelligibility when cued to speak louder. Tjaden and Wilding (2004) also found statistically significant intelligibility gains in a group of 12 speakers with PD. However, for their 15 participants with multiple sclerosis, the cue to speak loudly did not significantly improve speech intelligibility (Tjaden & Wilding, 2004).

Although these results seem promising for those with PD, even within this group, it seems treatment decisions may not be straightforward. While Tjaden and Wilding (2004) put forth results suggesting loud speech exacts greater intelligibility than slow cued speech, McAuliffe, Kerr, Gibson, Anderson, and LaShell (2014) present opposing findings, suggesting it is not always the most effective strategy for speakers with Parkinson's disease. Indeed, this study of speakers with Parkinson's disease found reduced speech rate resulted in a significantly higher proportion of correct listener responses, as compared with increased vocal loudness. In summary, while results tend to demonstrate positive group treatment effects for speakers with PD, this is not true of all individuals. Beyond PD, there remains very little evidence to support or refute the efficacy of increasing loudness as a treatment strategy.

The second cueing strategy, reduced speech rate, can be enacted using a variety of elicitation techniques. Across techniques, studies have demonstrated significant variation in intelligibility outcomes. For example, Turner et al. (1995) examined the impact of reduced speech rate in nine speakers with Amyotrophic Lateral Sclerosis (ALS) and dysarthria, finding improved intelligibility in only four participants. In contrast, across paced reading conditions, Yorkston et al. (1990) found consistent intelligibility improvements amongst participants with hypokinetic (n = 4) and ataxic (n = 4) dysarthria. Consistent improvement was also found by Hammen et al. (1994) in five speakers with hypokinetic dysarthria. However, although Hammen et al. (1994) replicated one of the pacing techniques used by Yorkston et al. (1990), they noted large differences in level of improvement observed between the two studies. When speakers with hypokinetic dysarthria were prompted to reduce their speech to 60% of their habitual rate, the average increase in intelligibility was 26% in the Yorkston et al. (1990) study, but only 9% in Hammen et al. (1994). Other studies have also demonstrated conflicting findings for speakers with hypokinetic dysarthria. For example, McAuliffe et al. (2014) found statistically significant change in their participants with PD, while Tjaden and Wilding (2004) did not find statistically significant outcomes for speakers with PD or multiple sclerosis. Hence, it appears that little consistent picture has

emerged regarding the effects of reduced speech rate on speech outcomes in different groups with dysarthria.

To summarize, although there is promising evidence that speech treatment targeting rate and loudness may be effective in improving the speech of individuals with dysarthria, the lack of consistent results for specific techniques indicates that there is no simple solution that can be used for all speakers.

1.3.3 Selection of Participants for Treatment Studies

The methodological approaches employed in treatment studies may be one key reason behind inconsistent outcomes reviewed in the previous section. The literature reviewed so far has, most often, selected speakers for treatment on the basis of their Mayo System subtype (Darley, Aronson, & Brown, 1969a, 1969b) or neurogenic aetiology, with participants with matching aetiologies or subtypes grouped together in studies (for example, placing all participants with PD and hypokinetic dysarthria within a single treatment group). As homogeneity of speech characteristics is generally assumed within each dysarthria classification, detailed examination of the perceptual or acoustic features of participants' baseline speech patterns is not usually included.

The Mayo classification system (Darley et al., 1969a, 1969b), was developed over 40 years ago and has, since that time, been the basis of the only widely accepted dysarthria classification framework (Duffy, 2013). It originated from a study of 212 people, each with clearly defined neurologic impairment to their lower motor neurons, upper motor neurons, cerebellum or extrapyramidal systems. From these groups, dimensions of pitch, loudness, vocal quality, prosody and articulation were perceptually rated and five distinct clusters of symptoms derived (Darley et al., 1969a, 1969b). These clusters of symptoms describe the dysarthria subtypes of the Mayo System: flaccid, spastic, ataxic, hypokinetic, hyperkinetic, and those considered 'mixed' combinations. These six subtypes of dysarthria form the basis for many research and clinical decisions in our field.

However, recent studies have highlighted some limitations in this method of classification (Y. Kim et al., 2011; Weismer & Kim, 2010). A defining aspect of the Mayo approach is the assumption of relative homogeneity within, as opposed to between, groups.[1]

[1] It should be noted, however, that many of the perceptual features characteristic of dysarthria are said to be common across multiple Mayo system subtypes (e.g. imprecise consonants, slow rate of speech).

Unique clusters of perceptual characteristics are said to co-occur within speakers of each subtype, making each group perceptually distinct to a trained listener (Duffy, 2013). However, this homogeneity within groups does not translate to treatment outcomes. Van Nuffelen et al. (2010), for example, examined rate control across six dysarthria subtypes (including unspecified mixed dysarthrias), noting clinically significant intelligibility improvements in approximately 50% of speakers. There was no subtype in which all speakers improved. Furthermore, there was no indication that one subtype was more likely to be successful than any other. Additional studies have shown that when loud and slow treatment strategies are directly compared, not all participants within a Mayo System grouping demonstrate greatest improvement using the same strategy (McAuliffe et al., in press; Tjaden & Wilding, 2004).

A further concern is the poor reliability evidenced between the clinical diagnosis of dysarthria and perceptual classification of professionals blinded to its aetiology (Fonville et al., 2008; Van der Graaff et al., 2009). Such findings challenge the core notion that each subtype consists of distinct and recognisable speech symptoms. Liss et al. (2009) and Liss, LeGendre, and Lotto (2010) also demonstrate that, for such a well-accepted framework, the Mayo System has been only minimally validated with large-scale studies of acoustic speech characteristics.

Similar issues to those identified with the Mayo approach arise with classification by neurogenic aetiology. This classification is based on the theory that similar brain lesions will have predictable patterns of underlying muscle disorder within the speech subsystems (e.g. impaired strength, coordination or spasticity) (Duffy, 2013). Classification by aetiology was the foundation for the development of the Mayo system, and while grouping speakers by aetiology may provide additional categories of dysarthria, there is no evidence to suggest this technique it is any more valid. It has long been presumed that certain patterns of muscle disorder will contribute to similar speech symptoms, thus making it easier to generalise speech treatments to a group with "the same" type of neurological impairment. However, the underlying rationale of inferring speech symptoms based on differences in the strength or steadiness of muscles in non-speech tasks remains unfounded (Weismer, 2006). Classification by neurogenic aetiology has been critiqued for focusing researchers' attention on the isolated assessment and training of muscles within the impaired speech subsystems – training which has been found to lack any carryover to speech (Weismer, 2006).

In summary, group classification through the Mayo System or dysarthria aetiology permeates the dysarthria literature (Weismer & Kim, 2010) and underpins almost all current

research into dysarthria treatment (e.g. Cannito et al., 2012; Lowit et al., 2010; Mahler & Ramig, 2012; Wenke et al., 2011). However, there is little evidence that grouping participants by the Mayo System or their dysarthria aetiology provides any meaningful insight into whether individuals benefit from speech modification strategies, or why certain participants achieve best results in one treatment condition over another. Baseline variability in speech symptoms could underlie significant disparities in outcome measures within a treatment group, and it is posited that the common finding of statistically insignificant results in trials of dysarthria treatment may be related to this.

1.4 NEW DIRECTIONS IN DYSARTHRIA ASSESSMENT

The lack of clear acoustic evidence behind the Mayo System has begun to prompt examination of whether differences in the acoustic features of dysarthric speech are better accounted for by other classification methods. For example, Y. Kim et al. (2011) examined eight acoustic metrics suggested to differentiate speakers with dysarthria from healthy controls. Classifications by speech severity, neurological aetiology and Mayo System dysarthria subtype were evaluated by comparing each grouping system against these metrics. Across acoustic parameters, dysarthria subtypes were considerably less desirable in producing homogeneous groups than both speaker severity and aetiology. Y. Kim et al. (2011) therefore suggest that grouping people by dysarthria subtype must rely on a combination of a small number of perceptual speech characteristics. These differing combinations of perceptual characteristics do not appear to significantly alter the acoustic signal. Thus, it is questionable whether they provide important information about what makes speech sound disordered.

As discussed in the previous section, aetiology and Mayo-based classifications do not provide a consistent one-to-one mapping with the success of specific rehabilitation strategies. Based on the results of Y. Kim et al. (2011), it seems that, in order to determine the appropriateness of a particular treatment technique, we need to utilize more information about the unique features of an individual's speech. Of course, it should be acknowledged that there are also many other factors—beyond the underlying dysarthria—that could contribute to differences in the way speakers' respond to treatment. Indeed, it is likely that many of the factors influencing speakers' treatment outcomes are altogether independent of the dysarthric speech signal. However, the degree to which cognitive abilities, motivation and fatigue affect how speakers with dysarthria respond to clinician prompts remains unclear

(Fox et al., 2002). This is true even of speakers with the same Mayo system subtype. For example, Ramig, Countryman, Thompson, and Horii (1995) examined 45 speakers with PD and reported that they were unable to find any significant correlations between the sound pressure level changes made following an LSVT program and participants' age, stage of disease, speech severity rating, depression rating, or cognitive functioning (as determined through a battery of neuropsychological tests). For this reason, it is important that we first determine to what degree speakers' baseline dysarthria is impacting their treatment success—before attempting to define the effects of any additional variables. To achieve this, further studies of the speech signal in people with dysarthria are needed.

1.4.1 Measurement of Speech Features in Dysarthria

Acoustic analysis offers an objective tool for describing speech differences amongst people with dysarthria. As previously discussed, impairment to different groups of muscles can differentially affect a speaker's ability to form an adequate airstream, produce clear voicing, control nasal emission, and alter the shape of their vocal tract. However, the effects that these impairments have on the speech signal can be difficult to quantify. While listeners may be able to detect the presence of numerous distortions to the speech signal, questions have been raised regarding their ability to isolate and independently rate impairments to different speech subsystems. For example, Sheard, Adams, and Davis (1991) investigated speech pathologists' ratings of ataxic dysarthria along a seven point scale. Five perceptual dimensions were rated. Three were related to speech articulation, one to prosody (excess and equal stress), and one to voice quality (harsh voice). The study examined the level of agreement across each dimension (with agreement defined as ratings within 1 scale point of each other). Although overall agreement ranged from between 66 to 76 percent within each feature, ratings of the five dimensions were closely linked. For example, imprecise consonant ratings had a .79 correlation with ratings of voice quality and a .88 correlation with ratings of speech prosody. Given these correlations, the authors suggested that it was unlikely that listeners' ratings of each feature were completely independent. This idea was further supported by statements from the speech pathologists, who noted that the presence of abnormal prosody and nasal emission confused their judgments of articulatory precision and accuracy.

Sheard et al. (1991) concluded that there were two potential issues with high correlations between ratings of different features. The first is that they confound any attempt

to ascertain the ‘true’ reliability with which listeners’ can judge a specific speech feature. That is to say, if a speaker is difficult to understand and listeners rate all speech features as severely impaired, the agreement between listeners will always be high, regardless of the particular feature examined. The second issue relates to the relevance of the rating. If we cannot tell to what extent listener ratings of one feature are biased by the perception of others, it is difficult to interpret what any one rating might mean.

Acoustic analyses offer a clear advantage. It allows us to examine features in an unbiased manner, knowing that—while these features may co-occur within a speaker—the measurement of one characteristic is not directly affected by the presence of others. Acoustic analyses can also provide more meaningful units of comparison between speakers. For example, a difference in speech rate from four to five syllables per second can be easily interpreted, modelled and replicated across studies.

Despite these advantages, the extent to which speech disorder—and the things that make one dysarthria sound ‘different’ to another—can be acoustically indexed is limited. There is no single acoustic measure that can detect dysarthria as effectively as the human ear (an issue further explored in chapter three) (Lansford & Liss, 2014b; Liss et al., 2010; Sapir, Ramig, Spielman, & Fox, 2010). Furthermore, some physiological processes, such as increased nasal emission and vocal fold spasticity, remain difficult to isolate via any single acoustic measurement (Kent, Weismer, Kent, Vorperian, & Duffy, 1999; Maryn, Corthals, Van Cauwenberge, Roy, & De Bodt, 2010). For this reason, this thesis considers both the overall perceptual severity of a disorder—from the perspective of the listener—as well as a range of targeted acoustic metrics.

The metrics that will be investigated have shown sensitivity to speech changes across dysarthrias. They include measurements of vowel articulation (Lansford & Liss, 2014b; Sapir et al., 2010), voice quality (Awan, Roy, Zhang, & Cohen, 2015), speech rate and rhythm (Liss et al., 2009; Niimi & Nishio, 2001), as well as variations in pitch and intensity (Bunton, Kent, Kent, & Rosenbek, 2000; Rosen, Kent, Delaney, & Duffy, 2006). As these measurements have all shown an ability to differentiate healthy and dysarthric speech, they present an excellent starting point for understanding variations in different aspects of speech motor control.

However, there are further considerations that must be addressed when applying these measurements to new datasets of dysarthric speech. People have unique acoustic speech signals that can be affected by many factors unconnected to the presence or severity of dysarthria. Within the second and third chapter of this thesis, several factors that may

systematically influence our acoustic measurements are considered: including sex, age and speech dialect. Speech dialect is of particular interest, given that—outside of the dataset discussed in this thesis—there have been no extensive investigations of dysarthric speech acoustics within a New Zealand (NZ) English context. Specifically, dialect presents an issue when measuring any aspect of vowel production, as it is well known that acoustic vowel properties vary widely across different English speaking countries (Maclagan, 2009).

Acoustic measurements of vowel dispersion are commonly used to provide information about speakers' articulatory precision and intelligibility (Lansford & Liss, 2014a). However, the vowel production of healthy NZ speakers needs to be considered before adapting these vowel dispersion measures to NZ speakers with dysarthria. For example, it has been frequently reported that the vowel found in the word "goose", commonly transcribed as the /u/ phoneme, is produced in a much more anterior position by NZ speakers than by speakers of other English dialects (see Easton & Bauer, 2000 for examples). In the dysarthria literature this phoneme is commonly used to represent a 'back' vowel in measurements of vowel dispersion. Given its production by NZ speakers, it seems logical to substitute the phoneme for another more representative vowel sound. However, measurements of vowel dispersion using different phonemes are not commonly examined in the dysarthria literature.[2] Thus, it is not known how the substitution of different vowel sounds in our calculations might affect the magnitude of these measurements and the variation across them. For this reason, the adaptation of acoustic vowel measurements—and their application to the NZ English dialect—is examined in chapters two and three.

1.5 SUMMARY AND AIMS OF THE PRESENT THESIS

In summary, this literature review has highlighted the variable patterns of speech intelligibility gains observed in studies of dysarthria treatment. In examining possible sources of this variation, this chapter reviewed issues in classification of dysarthria—which may result in participants who exhibit significant differences in the baseline features of their dysarthria being grouped together for treatment studies. It was posited that speaker-specific variations in baseline speech characteristics could have a marked influence on the success or

[2] Although it is acknowledged that measures of 'lax vowel space' (using an entirely different set of phonemes) have been reported in several studies of dysarthric speech (e.g. H. Kim et al., 2011; Tjaden, Rivera, Wilding, & Turner, 2005).

otherwise of treatment techniques. Thus, an alternative approach, based on characteristics of an individual's baseline speech pattern, may be able to provide more direct insight into the variation observed in participants' treatment outcomes. The last section of this chapter discussed how to quantify these differences between participants, as a first step towards understanding why some speakers make greater treatment gains than others.

To summarize the purpose of this thesis, two problems with current investigations of dysarthria treatment strategies are presented. The first is that these studies are typically limited to participants with a specific aetiology and dysarthria subtype. This means that, while the literature is dominated by examples of treatment strategies for some groups (e.g. speakers with PD and hypokinetic dysarthria), there are little data available to infer whether these strategies are appropriate for speakers who do not fit these moulds. The second problem (as addressed in section 1.3.1) is that, even amongst participants with the same aetiology and dysarthria subtype, there has been considerable variability in treatment effects observed across studies. To identify the types of participants who will achieve success with certain behavioural strategies—in addition to the types of participants who will not—we need to know more than simply their Mayo System subtype. To determine whether differences in a speakers' underlying dysarthria can affect their treatment outcomes, we need a deeper understanding of these participants' baseline speech features. Ideally, these features should be measurable across participants with dysarthria, so that we can infer information about appropriate treatment strategies for any speaker—regardless of whether their dysarthria aetiology and subtype have been studied before.

Broadly speaking, this thesis had two main aims. The first was to compare how cues to speak louder and reduce speech rate change speakers' intelligibility. The second was to determine whether measurements of speakers' baseline speech features were able to account for the variation observed in their intelligibility gains. These aims are addressed directly in chapter four. However, before exploring these questions, it was important for us to develop methods of speech feature extraction that could be applied to ageing New Zealand speakers with dysarthria.

To achieve this, a considerable amount of normative data was necessary to understand the effect of dialect on measurements of vowel articulation and speech duration. This issue is discussed within chapter two. Additionally, it was important to explore the variability in correlations between acoustic and perceptual measures of dysarthria. Chapter three compares acoustic and perceptual methods used to index articulatory impairment and dysarthria severity. Baseline speech severity has been reported to be an important factor in explaining

acoustic variability between speakers (Y. Kim et al., 2011). It has also been speculated that differences in baseline severity may explain some of the inter-speaker variations in intelligibility gains that occur following dysarthria treatment (Hammen et al., 1994; Pilon et al., 1998). For this reason, special attention is given to the methods used in this thesis to measure this feature (see chapter three for further details).

Chapter Two

Acoustic Features in Older New Zealand Speakers

Chapter two is an adaptation of the article titled "The relationship between speech segment duration and vowel centralization in a group of older speakers", which was recently published in the Journal of the Acoustical Society of America. In some sections the text has been modified and additional information has been provided to ensure consistency and relevance to the current chapter and thesis.

2.1 PREFACE

This thesis focuses on speakers of New Zealand English (NZE) with dysarthria. In the dysarthria literature, the vast majority of acoustic data have involved speakers from the United States (US). However, from analysis of healthy talkers, we know that speech rate and vowel formant frequencies vary considerably between English dialects—with significantly different values found for NZE speakers compared with published data from the US (Maclagan, 2009; Robb, Maclagan, & Chen, 2004; Sapir et al., 2010; Turner et al., 1995). Indeed, vowel articulation is particularly sensitive to systematic changes in speaking style. For example, speakers will alter their acoustic production of vowels when trying to convey different emotions (C. M. Lee et al., 2004), or when told to speak louder, slower or clearer (Tjaden, Lam, & Wilding, 2013). Thus far, studies which have explored the spectral properties of NZE vowels have typically focused on historical trends (Langstrof, 2006; Watson, Maclagan, & Harrington, 2000)—or relatively young groups of speakers (Watson, Harrington, & Evans, 1998). As discussed in the first chapter, the most common dysarthria aetiologies are more prevalent in speakers over 60. Hence, one of the issues in measuring the speech of New Zealanders with dysarthria is that we have little sense of what naturally occurring variations we should expect in these older speakers.

In addition to examining formant frequencies, data on NZ speakers' vowel durations are important for several reasons. Firstly, amongst healthy speakers, measures of temporal speech rhythm and rate are linked to vowel duration in direct and predictable ways (Kessinger & Blumstein, 1998; Nokes & Hay, 2012). Hence, if there are differences in the vowel durations of older NZ speakers, we will also know what types of variation to expect in other measurements of temporal prosody. Secondly, it has been speculated that there may be a relationship between speakers' spectral vowel production and their natural speech rate—although evidence of this association has not been substantiated (Robb et al., 2004; Tsao, Weismer, & Iqbal, 2006). For this reason, the current chapter also explores the association between measures of vowel duration and spectral vowel dispersion amongst healthy speakers. Broadly speaking, the aim of the chapter was to better understand the interplay between age, vowel duration and spectral vowel dispersion amongst older speakers of NZE.

2.2 INTRODUCTION TO VOWEL ARTICULATION

When hearing and classifying vowels, listeners attend primarily to movements of the first two formant frequencies (F1 and F2) (Hillenbrand, Getty, Clark, & Wheeler, 1995). It is theorized that each vowel has a distinguishing target position in the F1/F2 space that represents its steady state realization: the point at which listeners would most accurately recognize its identity (Hillenbrand et al., 1995). In continuous speech, the production of this target is limited by the speed and accuracy of articulatory movement. When a talker's rate of speech does not allow enough time for the articulators to move from the production of surrounding consonants to a vowel's target position, a less distinct perceptual token is produced. On a spectrogram, this is seen when formants undershoot the speaker's articulatory target, falling short of the vowel's steady state realization (Moon & Lindblom, 1994). When significant undershoot occurs in a set of vowels, they will present with a more centralized pattern of F1 and F2 values.

Increased segment durations and vowel centralization have commonly been reported as prominent features of the dysarthrias. However, these features have also been inconsistently reported to occur in healthy ageing, and it is unclear to what degree these changes co-occur in older speakers. To understand how these speech features differ in NZ speakers with dysarthria, we must first recognize the level of natural variation that occurs amongst healthy speakers—especially within older age groups. The current chapter provides an introduction to the acoustic measurement of vowels through an investigation of vowel production in healthy older speakers of NZE. This study aimed to determine whether measures of vowel duration and centralization: (1) were correlated across older speakers, and (2) changed as a function of age. The chapter also serves to provide normative data concerning aspects of temporal variation and vowel articulation in NZE.

2.3 VOWEL CHANGES IN OLDER SPEAKERS

Within the speech-development literature, numerous studies have examined how the acoustic signal of vowels changes with age (S. Lee, Potamianos, & Narayanan, 1999; Vorperian & Kent, 2007). From infancy to adulthood, these studies have used acoustic measures to provide insight into both anatomical changes of the vocal tract and the development of neuromuscular control. In contrast, there has been only limited interest in the vowel articulation of older adults. But information about older talkers' speech production remains important—both to gain insight into how speech changes with age and to delineate normal acoustic variation from changes associated with acquired neurological disease.

Vowel centralization has been associated with reduced intelligibility in both normal speakers and those with motor speech disorders (Ferguson & Kewley-Port, 2007; Liu, Tsao, & Kuhl, 2005; McRae, Tjaden, & Schoonings, 2002; Neel, 2008; Tjaden & Wilding, 2004). Generally speaking, vowel centralization has been said to occur because of a reduction in the speed and amplitude of articulatory movements (Forrest, Weismer, & Turner, 1989; Moon & Lindblom, 1994). Some researchers have also suggested that this might cause vowels to centralize with advancing age, as a result of overall decreases in the speed and accuracy of motor control, reduced auditory feedback, and diminished cognitive-linguistic functioning in older speakers (Liss, Weismer, & Rosenbek, 1990).

Although it is generally accepted that ageing results in some changes to the speech production mechanism (Kahane, 1981), studies comparing groups of younger and older participants have produced conflicting evidence on how vowel production is affected (Benjamin, 1982; Liss et al., 1990; Rastatter, McGuire, Kalinowski, & Stuart, 1997; Torre III & Barlow, 2009; Xue & Hao, 2003). Based on studies thus far, two patterns of spectral vowel change have been theorized to occur as part of the ageing process: (i) centralization and (ii) generalized decreases in F1 and F2 across all frequencies. However, acoustic evidence in support of either theory is still far from clear.

Given the variability in results, it is worth considering whether other speech-related factors may be affecting the relationship between age and vowel formant values. There is an abundance of evidence to suggest that speech rate slows as people get older (e.g. Harnsberger, Shrivastav, Brown Jr, Rothman, & Hollien, 2008; Jacewicz, Fox, O'Neill, & Salmons, 2009; Ramig, 1983; Shewan & Henderson, 1988; Smith, Wasowicz, & Preston, 1987); yet few studies have considered how naturally occurring differences in speech rate might contribute to changes in vowel formant values. The next section introduces the issue of rate variability and explores its relationship with vowel articulation.

2.4 EFFECT OF SPEAKING RATE ON VOWEL CENTRALIZATION

When people intentionally alter their speech rate they produce corresponding changes to their vowel formants. For example, both Fourakis (1991) and Moon and Lindblom (1994) found that when speakers increased their vowel duration, they produced formants that were closer to their "idealized" steady-state frequencies. Hence, when a person is prompted to produce longer speech segments, they are likely to make changes to the spectral quality of their

vowels (for more detailed review, see Tsao et al., 2006). However, these effects are observed when speakers are compelled to change their speech—to talk clearer or slower. They do not tell us whether natural rate variations between speakers are also capable of influencing vowel production.

It is unclear whether people with naturally slower articulatory rates produce more spectrally distinct vowels. It could be that faster speakers habitually produce formants with steeper slopes than slower speakers, enabling them to reach the same acoustic targets in a shorter time frame. If this were the case, one could hypothesize that all speakers use similar articulatory gestures to produce vowels regardless of their natural speaking rate. However, there is reason to believe that differences may exist in the gestures fast and slow speakers use. Tsao and Weismer (1997) found that faster and slower speakers demonstrated considerable differences in the maximum articulatory rate they reached when prompted to read "as fast as you can." Habitually slow speakers exhibited reduced maximum articulatory rates relative to fast talkers. Tsao and Weismer (1997) suggested that this was due to differences in speech motor control between the two groups. If this is true, these speakers might use different articulatory strategies to produce vowels.

If speakers use different articulatory strategies based on their habitual speaking rate, it is likely that a relationship would exist between speakers' vowel durations and the acoustic distance between their vowel formants. Tsao et al. (2006) examined this hypothesis using two groups: 15 "slow" and 15 "fast" speakers. Tsao et al. (2006) measured individuals' vowel durations as well as formant frequencies taken from the temporal midpoint of their /i/, /æ/, /u/, and /a/ vowels. They found that the average vowel space area (VSA) of the two groups were virtually identical. There was no evidence of a systematic, across speaker relationship between average vowel duration and either F1 or F2 values in any of the vowels examined. The results from Tsao et al. (2006) suggest that while people's speech rates may differ, they employ similar configurations of the vocal tract to produce speech segments. But these results may not provide the full picture. Tsao et al. (2006) extracted formant measurements from one static time-point, at the temporal centre of the vowel. It could be that faster and slower speakers reach a vowel's target position—or an approximation of this target—at different stages of the vowel's duration. Tsao et al. (2006) speculated that taking only one measurement of formant frequency, from the midpoint of speakers' vowels, could have obscured possible differences in formant movement between the fast and slow speakers studied. Hence, in order to compare VSA across speakers, a measurement point reflecting a vowel's target, steady state formant position may be required—irrespective of the time-point

at which this target is reached. It is hypothesized that the use of different formant measurement points might alter speakers' VSAs. Specifically, that measurements taken from vowels' steady state articulatory targets may produce larger VSAs than those extracted from vowels' temporal midpoints.

Older speakers present an interesting group in which to evaluate the relationship between vowel duration and measurements of VSA. As people age, they have tendency to speak slower (Harnsberger et al., 2008; Jacewicz et al., 2009; Ramig, 1983; Smith et al., 1987) and produce longer vowel segments (Benjamin, 1982; Harnsberger et al., 2008; Liss et al., 1990). Thus far, none of the studies that have examined vowel production changes in older speakers controlled for differences in speech rate. Furthermore, there has been little examination of whether changes in vowel spectral quality continue to progress past the age of 65.

2.5 SUMMARY AND AIMS OF THE CURRENT STUDY

This study explores the relationship between average vowel duration and VSA in a cohort of older speakers of NZE. The relationship is analysed using formants extracted from the vowels' temporal midpoints and an additional measurement point reflecting their target, steady state formant position—as it was theorized that this might capture anticipated differences in formant movement between fast and slow speakers. As a secondary aim, this study also examined changes in average vowel duration and spectral vowel quality in older speakers. Specifically, this investigation explored (a) whether vowel durations lengthened between the ages of 65 and 90, (b) whether vowel formants lowered between the ages of 65 and 90, and (c) whether VSA decreased between the ages of 65 and 90.

2.6 METHODS

2.6.1 Participants

The study included 149 speakers of New Zealand English (NZE) (55 males and 94 females), aged between 65 and 90 years. The average age of the participants was 72.7 years (SD = 5.3), with 42 participants aged 65 to 69 years, 60 participants aged 70 to 74 years, 28 participants aged 75 to 79 years, 14 participants aged 80 to 84 years, four participants aged 85 to 89 years, and one participant aged 90. Speakers reported no previous history of neurological

impairment or speech and language disorders, and all scored within the normal range (i.e., >26) on the Montreal Cognitive Assessment, a screening tool that identifies individuals with mild cognitive impairment (Nasreddine et al., 2005). Due to dialectal variations in NZE, the study excluded any speakers who had lived outside New Zealand or in the province of Southland— where there is residual use of post-vocalic /r/—between the ages of 0 and 18 (Trudgill, Maclagan, & Lewis, 2003). At the time of recording, all participants were free of colds or other respiratory issues that may have affected their speech.

2.6.2 Speech Stimuli

Each speaker attended a single recording session. Recordings took place in a quiet room, with an investigator present. Participants were asked to read "the Grandfather passage" (see Appendix A) in their normal speaking voice after familiarizing themselves with the content of the passage. During recording, participants were seated at a table and their speech was recorded using a Zoom H4n recorder placed on the table in front of them (at an approximate distance of 30 cm). Digital audio recordings of the speakers were made at 22.05 kHz with 16 bits of quantization.

2.6.3 Extraction of Acoustic Data

2.6.3.1 Segmentation of the data set

Prior to conducting the acoustic analysis, all recorded passages were transcribed, and then automatically segmented at the phoneme level using the hidden Markov model toolkit (HTK) (Young et al., 2002). Phoneme segments were labelled in Praat (Boersma & Weenink, 2012) based on the "Origins of New Zealand English Miner" orthographic-phonemic dictionary (Fromont & Hay, 2008), constructed from Celex (Baayen, Piepenbrock, & Gulikers, 1996) and additional hand labelled entries. Three trained analysers manually checked the accuracy of all phoneme boundaries according to standard segmentation criteria (Peterson & Lehiste, 1960). This was conducted using visual examination of the waveform and wide-band spectrogram, and through the use of auditory cues. The primary indicators for the onset and offset of vowels were changes to formant structures, voicing, and waveform amplitude. Vowel onset boundaries were identified at the start of the pitch period coinciding with the onset of regular formant structure. Vowel offset boundaries were identified by changes in formant structure at the point where the pitch period ended, and where there was a corresponding drop in waveform amplitude. The amplitude, shape, and lack of frication of

successive pitch periods were also used in determining boundaries. Since the HTK segmentation was completed at the individual phoneme level, if the person manually checking phoneme boundaries was uncertain of a boundary for consecutive phonemes, the boundary derived from automatic segmentation was retained.

2.6.3.2 Measurement of mean vowel duration

Vowel onset and offset times were extracted from the checked data set using a custom Praat script. Each speaker's average vowel duration (in milliseconds) was calculated from all vowels produced in the passage reading. The durations of these vowels were summed and divided by the number of vowels produced across the passage. This approach ensured that if a speaker repeated or missed a word in the passage reading, the missing vowel was accounted for. For example, if the total duration of a speaker's vowel productions was 13.93 s, but they only produced 148 vowels across the passage, their average vowel duration would be 94 ms. On average, participants produced 154 vowels when reading the passage.

2.6.3.3 Measurements of formant frequencies

Three tokens of the NZE START [ɐː], FLEECE [iː], and THOUGHT [oː] vowels were selected from the passage for the measurement of VSA. These tokens tend to elicit the most extreme front [iː], open [ɐː], and back [oː] vowel positions in NZE (Watson et al., 1998). The [ɐː] vowel was extracted from two occurrences of the word "grandfather"[3] and one occurrence of the word "answers."[4] The [iː] vowel was extracted from two occurrences of the word "each" and once occurrence of the word "three." The [oː] vowel was extracted from one occurrence of the words "organ," "short," and "more." Due to reading errors, speakers occasionally missed one of the selected tokens. In this instance, the remaining two tokens were used. If a speaker repeated a sentence containing one of the vowel tokens, the second repetition was always selected for analysis.

F1 and F2 frequencies of these vowels were extracted from the checked data set using custom Praat scripts. To begin with, the formant tracks of the first five formant frequencies were obtained via Praat using the Burg linear predictive coding (LPC) algorithm, with a Gaussian window length of 25 ms, a time step of 2.5 ms between the centres of consecutive

[3] The primary stress in 'grandfather' usually occurs on the first syllable; however, there was always adequate stress on the second syllable to produce a distinctive [ɐː] token.
[4] In NZE, 'answers' always contains an START vowel rather than a TRAP vowel.

windows, a maximum formant value of 5.5 kHz for females and 5 kHz for males, and a pre-emphasis from 50 Hz (Boersma & Weenink, 2012). F1 and F2 were then extracted at two measurement points. The first, a "midpoint" measurement of F1 and F2, was taken at the temporal midpoint of each vowel. The second, an "articulatory point" measurement, was taken at a single time point between the vowel's "onset" (defined at 20% of the vowel's duration) and "offset" (defined as occurring at 80% of the vowel's duration). Articulatory point measurement criteria were designed with the aim of extracting values at a time where there was minimal movement in formant tracks—for the best approximation of the vowels' steady-state target. For the front vowel, [i:], this point was set at peak F2 frequency; for the open [ɐ:] vowel the target was extracted when F1 was at its maximum; and for the back [o:] vowel the target point was taken when the lowest value of F2 was reached (Watson & Harrington, 1999; Watson et al., 1998). While constraining these measurement points to between 20% and 80% of the vowels' duration did somewhat limit the scope of F1 and F2 deviation, without this constraint, the automated script would select measurement points that reflected transitional movements towards neighbouring phonemes.

Criteria for visual checking of formant values were devised to identify potential errors in automatic formant tracking. These were as follows: (1) if F2 was less than 100 Hz above F1, it was manually checked; (2) if values were three standard deviations (SD) from the F1 or F2 mean of all tokens, they were manually checked; (3) in the case of the [o:] vowel, further checks were made if the frequency of F2 was below 500 Hz. Corrections to tracking errors were made by adjusting the time-step settings or visually adjusting the measurement point time (in keeping with the midpoint and articulatory point criteria).

2.6.3.4 Vowel space area

VSA was calculated using F1 and F2 of the [ɐ:], [i:] and [o:] vowels. Two VSA values were calculated—one using midpoint formant values and another using articulatory point formant values. F1 and F2 values for the three [ɐ:], [i:] and [o:] word tokens were averaged for each speaker. Triangular VSA was constructed by plotting these values as coordinates in a F1/F2 plane, and calculating the resulting triangular area using the formula: $Hz^2 = 0.5 \times ABS[F1[i:] \times (F2[ɐ:] - F2[o:]) + F1[o:] \times (F2[i:] - F2[ɐ:]) + F1[ɐ:] \times (F2[o:] - F2[i:])]$, where ABS = absolute value, F1[i:] = first formant frequency of the [i:] vowel, and so on. Given the NZE dialect, a measure of triangular VSA provides a more accurate representation of vowel dispersion than quadrilateral VSA (Maclagan, 2009).

2.6.3.5 Reliability of acoustic measures

To determine inter- and intra-rater reliability of the measures, 10% of textgrids were manually re-segmented for reliability by the original three analysers. The scripts to determine vowel duration and formant values were then re-administered. As an additional measure, a further 10% of the hand-checked formant values were manually re-measured by the first author and a trained research assistant. The average inter-rater difference in speakers' vowel duration scores was within 3.0 ms of the original values. Intra-rater vowel duration scores were also, on average, within 3.0 ms of these values. The reanalysis found F1 intra-rater reliability scores averaged within 13.2 Hz of original values, while F2 scores were within 35.0 Hz. Average inter-rater variation was 16.6 Hz for F1 values and 46.1 Hz for F2.

2.7 RESULTS

2.7.1 Comparison of VSA from Articulatory and Midpoint Values

Graphs depicting speakers' articulatory and midpoint formant values are presented in Figure 2.1. The use of these different measurement points resulted in statistically significant differences in VSA [$t(54) = 15.5$, $p < 0.001$]. For males, the average midpoint VSA was significantly smaller ($M = 142381$ Hz^2, $SD = 42281$ Hz^2) than the average articulatory point VSA [$M = 206798$ Hz^2, $SD = 54481$ Hz^2, $t(54) = 15.5$, $p < 0.001$]. This was also the case for females, with the midpoint formants producing significantly smaller midpoint VSA values ($M = 244023$ Hz^2, $SD = 78907$ Hz^2) than those extracted from the articulatory points [$M = 375744$ Hz^2, $SD = 115924$ Hz^2, $t(93) = 16.8$, $p< 0.001$]. However, despite these differences, the two measures of VSA were highly correlated [$r(147) = 0.85$, $p < 0.001$] (see Figure. 2.2).

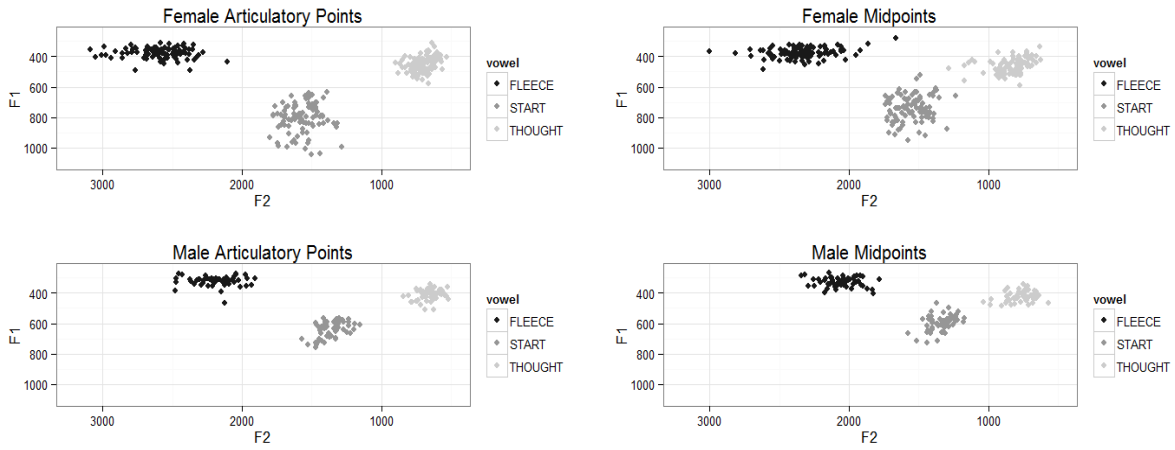

Figure 2.1. A comparison of midpoint and articulatory point VSA values for male and female speakers.

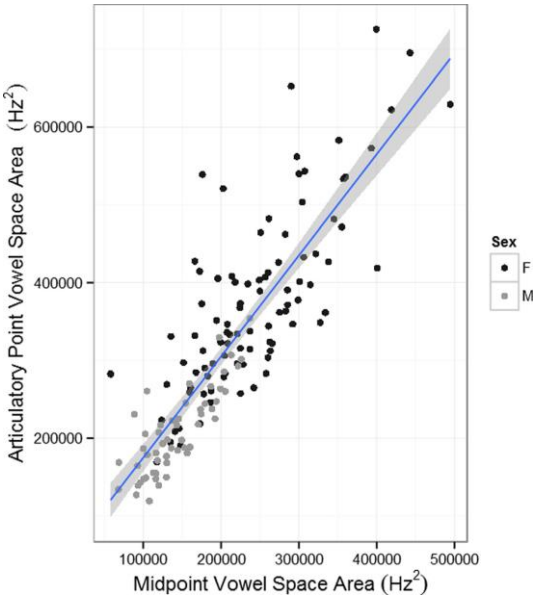

Figure 2.2. A comparison of midpoint and articulatory point VSA values for male and female speakers. Shaded area indicates 95% confidence interval of regression estimate.

2.7.2 Relationship between Average Vowel Duration and VSA

To understand how vowel duration might affect spectral vowel quality, models of speakers' midpoint VSA and articulatory point VSA measures were created. In addition to calculating

VSA in Hz^2, speakers' formants were measured using the Bark frequency scale. Modelling VSA in $Bark^2$ did not make any significant changes to the overall findings of this study. For this reason, in the models presented below, all measurements of VSA were calculated in Hz^2.

2.7.2.1 Predictors of midpoint VSA

A series of linear regression models were used to analyse the effect of speakers' age, sex, and average vowel duration on their midpoint VSA. The analysis began with a full model that examined the interaction between age, sex, and average vowel duration. It proceeded in a backward-stepwise iterative fashion seeking to reduce the full model to a model containing only significant effects (with alpha set at 0.05). The final model included a main effect of sex [b = 102997 (11349), $p < 0.001$], with males producing significantly smaller VSA values than females. There was also a significant effect of average vowel duration [b = 958 (406), $p = 0.02$], with speakers who exhibited longer vowel durations producing larger vowel spaces. No interactions were significant. Overall, this model accounted for 37% of the variance in speakers' VSA values.

2.7.2.2 Predictors of articulatory point VSA

The same statistical procedure as detailed above was then conducted using VSA derived from articulatory points as the dependent variable. The results of this model fitting procedure were similar to the results for midpoint VSA. There were no significant interactions between the variables, and speaker age did not significantly affect VSA ($p > 0.05$). The final model included a main effect of sex [b = 171973 (15954), $p < 0.001$], with males producing significantly smaller vowel space areas than females. Again, there was a significant main effect of speakers' average vowel durations [b = 2141 (571), $p < 0.001$]. Speakers with longer vowel durations exhibited larger VSAs. As hypothesized, vowel duration had a greater effect on speakers' VSA measurements when compared to the model of VSA derived from midpoint values. Overall, the articulatory point model of VSA accounted for 46% of the variance in speakers' VSA values.

2.7.3 Vowel Duration as a Function of Age

There were no significant differences in average vowel duration between male (M = 107 ms, SD = 13.6) and female speakers [M = 106 ms, SD = 13.5, $t(147) = 0.614$, $p = 0.54$]. For this reason, the relationship between average vowel duration and age was examined after

collapsing the data across sex. The analysis revealed a weak but significant correlation between age and average vowel duration [r(147) = 0.25, p = 0.003], shown in Figure 2.3.

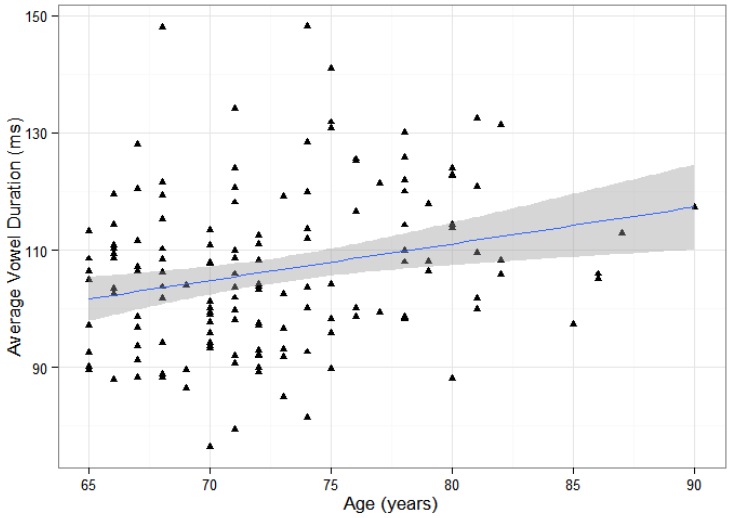

Figure 2.3. Average vowel duration as a function of age. Shaded area indicates 95% confidence interval of regression estimate.

2.7.4 Relationship between Formant Values and Age

As reported in section 2.7.2, no relationship between speakers' ages and VSAs was found. However, as speakers with lower formant frequencies would not necessarily exhibit smaller VSAs, it remained possible that reductions in the absolute frequency of speakers' vowel formants might still be observed with advancing age. To address this question, an aggregate formant frequency was created for each speaker by summing their [ɐ:], [i:], and [o:] F1 and F2 values. The vowels were examined together to reduce the chance of type 1 errors associated with multiple statistical comparisons. In female speakers, there was no significant relationship between aggregate formant frequencies and age [midpoints: r(92) = 0.20, p > 0.05; articulatory points: r(92) = 0.15, p > 0.05]. Similarly, in males, there was no relationship between speakers' ages and aggregate formant frequencies [midpoints: r(53) = 0.20, p > 0.05; articulatory points: r(53) = 0.24, p > 0.05]. These data suggest that reductions in F1 and F2 were not contributing to the relationship (or lack of relationship) between VSA

and age. Hence, Figure 2.1 also provides a normative representation of vowel production in healthy older speakers of NZE.

2.8 DISCUSSION

This study examined the relationship between average vowel duration and spectral vowel quality across a group of 149 NZE speakers aged 65 to 90 years. The purpose of this study was threefold. Firstly, the study aimed to acoustically measure aspects of vowel dispersion and speech duration in healthy, older speakers of NZE, to determine the normal range of values we should expect for these speakers. The distribution of speakers' vowel space and duration measurements is discussed in the next section.

Secondly, to better understand the relationship between these variables, the study aimed to determine whether participants who had a natural tendency to speak more slowly (i.e., longer vowel durations) would also exhibit more acoustically distinct vowel segments. When people intentionally slow their speech rate, they tend to produce vowel formants that are closer to their "idealized" steady-state frequencies. However, previous research has found no relationship between inter-talker speech rate and spectral vowel quality (Tsao et al., 2006). It was hypothesized that this relationship existed, but that it might only be apparent when formant measurements were extracted from an articulatory-based measurement point. As hypothesized, a significant relationship between speakers' average vowel durations and their VSA was identified. This indicated that speakers who had slower speech rates did, on average, produce more spectrally distinct vowels. However, in contrast to expectations, this relationship was present both when formants were extracted from a defined articulatory point, and when measurements were taken at the vowels' temporal midpoints.

The third aim was to investigate whether the age of speakers included in the study had any effect on these measurements. Results showed that there were measurable increases in average vowel durations amongst speakers between the ages of 65 and 90—though the magnitude of the effect was subtle. In contrast, vowel formants did not change. It was suggested that the use of a habitually slower speaking rate may assist some speakers in maintaining acoustically distinct speech segments.

2.8.1 Measurement of Vowel Articulation in NZE

The main motivation for examining vowel articulation in this population of NZ speakers, was to better understand how these measurements (that are investigated in later chapters of this thesis) might be affected by the NZE dialect. This study included some measurements that were adapted from those typically used in the motor speech literature. For example, in measuring VSA, tokens of the NZE START [ɐ:], FLEECE [i:], and THOUGHT [o:] vowels were used to calculate a triangular space. As stated earlier, these tokens tend to elicit the most extreme front [i:], open [ɐ:], and back [o:] vowel positions in NZE (Watson et al., 1998). This selection of phonemes differs from that which is normally used in studies of dysarthria (see Kent & Kim, 2003; Lansford & Liss, 2014b), as studies typically select the /u/ phoneme (the vowel found words such as "stew" or "goose") to represent a back vowel. In NZE, this phoneme is produced much further forward as a [ʉ:] vowel sound. The effect that changing the /u/ phoneme had on the size of the vowel spaces reported in this study is hard to quantify. In comparing the midpoint VSA measurements, it is clear that the values for the current group older speakers are broadly consistent with (i.e. within one standard deviation of) the values reported for younger, healthy US speakers in Lansford and Liss (2014b). The values for the articulatory VSA are also within one standard deviation of VSA values reported in Sapir et al. (2010) based on formant frequencies extracted by Hillenbrand et al. (1995) from healthy US speakers. Hence it seems the substitution of the /u/ phoneme for NZE speakers had no clear effect on the overall size and variation present in the triangular vowel space measure.

In contrast, the average vowel durations reported in this study appeared to be different from previous reports of US English. Although it is difficult to draw any robust conclusions between studies which have used different speech stimuli and recording conditions, NZE vowel durations appeared shorter than those typically reported in the US speech literature. For example, Benjamin (1982) reported average vowel durations of older speakers of between 156-169 ms (when measuring vowels in words which were not given special emphasis). Liss et al. (1990) also reported consistently longer vowel durations across their groups of older speakers, although these measures came from only a small set of content words. Liss et al. (1990) indicated vowel durations of between 164 and 219 ms were typical in a stressed vowel position. Both these measures are much higher than the 106-107ms averages found in the male and female groups. The average vowel durations reported in this study were also considerably shorter than each of the vowel sounds reported by Hillenbrand et al. (1995) for healthy US adults – although this is likely related to the fact that the Hillenbrand et al. (1995) data were taken from a stressed, /h_d/ context.

Data suggesting that NZE speakers have shorter vowel durations is consistent with previous literature which has examined speech rate in NZ and US speakers (Robb et al., 2004). However, it is important to note that the current recordings contained many function words and unstressed vowels, which are known to have significantly shorter durations (Umeda, 1975). Hence the number of function words in the passage reading may have contributed to the generally short vowel durations for healthy speakers reported in this thesis.

2.8.2 Effect of Average Vowel Duration on Vowel Space Area

In general, there was considerable variation in the average vowel durations of older adults in this cohort (ranging between 77 and 148 ms across the reading passage). These vowel durations had a significant effect on VSA. The regression model, which included significant effects for sex and average vowel duration, indicated that for every 1 ms increase in speakers' average vowel duration, there was an average increase of 958 Hz^2 in their midpoint VSA. When articulatory point VSA was examined, the average increase was 2141 Hz^2 (again, in a model where sex was held constant). Speaker age did not influence these findings, a result which is further discussed in the next section. From the current analysis, it appeared that people who spoke more slowly did, on average, produce more acoustically distinct vowels. These findings contrast to the results presented in Tsao et al. (2006)—though, as elaborated upon below, it is possible that these differences reflect the differing degrees of freedom used to test the relationship across studies.

While significant, the effects of speakers' average vowel duration on their VSA were subtle in this study. Increases of 958 Hz^2 and 2141 Hz^2 may seem like large changes in VSA, but these differences can occur with only small shifts in F1 and F2 values. For example, in the case of midpoint VSA, average F1 and F2 values (for both male and female speakers) would have to change by less than 1 Hz across the three vowels to accommodate a 958 Hz^2 VSA increase. To produce a 2141 Hz^2 increase, articulatory point formant frequencies would have to change by less than 1 Hz across vowels for the average female speaker, and 1 to 2 Hz for the average male.

When average vowel duration was controlled for, females' midpoint VSA were estimated to be 102997 Hz^2 larger than males. When articulatory point values were used to calculate VSA, speaker sex produced an even larger effect—with an estimated difference of 171973 Hz^2 between males and females in the model. Vocal tract size clearly plays a role in determining a person's VSA, and we expected that variations in vocal tract size would

produce significant differences between the sexes. While the effect of average vowel duration is considerably smaller than the differences in VSA between males and females, it is compelling because it accounts for variance over and above what we can attribute to simple anatomical differences. It is hypothesized that the effect of average vowel duration must be related to the articulatory gestures that speakers use to produce vowel sounds—with slower speakers producing less formant undershoot (as suggested in previous studies of intra-speaker rate variation, see Moon & Lindblom, 1994).

While the two VSA measurements used to index spectral vowel quality were highly correlated, there were significant differences between the size of speakers' articulatory and midpoint vowel spaces. The differences between the two VSA measures suggest that formant values extracted from vowels' temporal midpoints do not always occur at the phonemes' idealized steady state frequencies. Formant values extracted from the articulatory point appeared to come closer to these idealized frequencies, producing larger vowel spaces across speakers. The current regression models indicated that speakers' sex and average vowel duration had a greater effect on their articulatory point, as opposed to midpoint, VSA values. This is consistent with the hypothesis that midpoint formant values—which are likely to reflect different points of articulation amongst different speakers—may obscure differences in people's vowel production. In this investigation, the use of a standard articulatory measurement point accounted for around ten percent more variation in speakers' VSA values. These data indicate that the use of an articulatory measurement point may reduce the effect that different formant trajectories have on the measurement of speakers' F1 and F2 values. For this reason, this method of formant measurement may be valuable in future investigations of vowel space differences between speakers.

While there were benefits to using the articulatory measurement point, speakers' average vowel durations had a significant effect on their VSA regardless of the measurement point used. Given this, the differences between the findings of the current study and those of Tsao et al. (2006) may be reflective of the degrees of freedom used to test the relationship ($df = 13$ vs 146), as opposed to the formant measurement points used to construct VSA. Across speakers, changes in average vowel duration had only a subtle effect on VSA, and it is likely that a high number of participants was necessary to detect this relationship.

2.8.3 Changes in Vowel Articulation with Age

A secondary aim of this paper was to investigate changes in average vowel duration and spectral vowel quality in older speakers. Specifically, we were interested in whether (a) vowel duration increased between the ages of 65 and 90, (b) whether vowel formants continued to lower between the ages of 65 and 90, and (c) whether VSA decreased between the ages of 65 and 90.

2.8.3.1 Changes in vowel duration with age

Across speakers, statistically significant increases in average vowel duration were found with advancing age. These results were consistent with previous research showing that speech segment durations tend to be longer in older speakers (Benjamin, 1982; Harnsberger et al., 2008; Liss et al., 1990). This study goes a step further, providing evidence that speech segment duration also shows measurable change amongst speakers over 65. As previous studies have suggested, age-related neuromuscular degeneration, in conjunction with slower processing times and reduced auditory feedback (Ramig, 1983; Zraick, Gregg, & Whitehouse, 2006), may limit the speed at which older speakers are able to produce speech segments. However, results from the current study appear to indicate that this slower speech rate may be an adaptive speech strategy enacted by older adults—given those participants with longer average vowel durations also tended to produce more acoustically distinct vowels. For this reason, it is not clear that increased segment durations in older speakers are a direct result of neuromuscular declines.

This hypothesis is consistent with recent work by Mefferd and Corder (2014). Mefferd and Corder (2014) examined speakers' lip and jaw movements in a syllable repetition task. Females of four different age groups (ranging between 22 and 95 years) were prompted, via metronome cueing, to strike two fixed targets placed below their lower lip or jaw. Mefferd and Corder (2014) found the ability to increase the speed of lip and jaw movements, in response to cues, did not reduce with age. However, during faster metronome paces, only the older participants maintained their jaw displacements. The older adults tended to produce larger, more accurate, movements across the speed conditions. Naturally, maintaining these larger jaw displacements meant older adults' movements tended to take longer, despite having a faster rate of movement than their younger counterparts.
The authors concluded that the speed of older speakers' articulatory movements was not physiologically limited in any way. While older adults showed an ability to increase their lip and jaw speeds, they seemed less able to regulate their speed relative to the distance of their

movement. From these data, Mefferd and Corder (2014) hypothesized that a slow speech rate may primarily be a compensatory strategy to maintain speech accuracy in the presence of diminished articulatory control. In the present vowel data, the increased acoustic space between vowel targets in speakers with longer vowel durations is consistent with this idea. Vowel precision appears to be increased in older people who adopt a speech pattern with longer vowel durations.

2.8.3.2 Changes in vowel formants with age

Between the ages 65 and 90, there was no evidence that vowel formant frequencies were lowering. Furthermore, there was no evidence of changes to VSA as age increased. In general, previous studies have provided evidence for theories of vowel change through group comparisons of older speakers and relatively young adults (Benjamin, 1982; Liss et al., 1990; Rastatter et al., 1997; Torre III & Barlow, 2009; Xue & Hao, 2003). It could be that the changes observed within speakers over 65 are subtler, and perhaps less consistent, than these group differences. Given the level of natural variation in formant frequencies amongst different speakers, changes due to ageing might only be apparent through longitudinal investigations that make comparisons within the same speakers across time (e.g. Endres, Bambach, & Flösser, 1971). Lowering of formant frequencies has been theorized to occur due to progressive anatomical changes associated with ageing—such as craniofacial growth and lowering of the larynx—which create larger resonating spaces within the vocal tract (Xue & Hao, 2003). However, the rate and degree of these anatomical changes is unclear. Therefore, it is also possible that expansions of the vocal tract usually occur before the age of 65 and may exhibit less influence on formant frequencies beyond this age.

In the case of vowel centralization, there is another possibility why a relationship between age and VSA was not apparent in the present study. Prior evidence of an association between ageing and vowel centralization involved speakers who were 18 years older, on average, than the participants in this study (Liss et al., 1990). These older speakers may have demonstrated a greater level of declining physiology than this study's participants. Furthermore, this data set included only older subjects in good general health, with no evidence of mild cognitive decline.

2.9 CONCLUSIONS

The current study demonstrated that, when applied to healthy control speakers, VSA measures adapted for NZE produce a similar range of values to those reported in previous US studies. The data presented in this chapter also suggest that there may be differences in the speech segment durations and articulatory rates of older NZ speakers, and that these differences may change as speakers age.

Notably, the current study found that people who habitually spoke more slowly, with longer vowel durations, produced larger vowel spaces on average. As discussed, there are several well-known factors that influence peoples' speech rate—including speakers' age, dialect, and physiological condition. This study provides new evidence demonstrating that these natural suprasegmental differences are also related to the spectral quality of phonemes. This suggests that speakers use different articulatory strategies to achieve acoustic targets based on their habitual speaking rate.

This study was not without its limitations. The use of a reading passage to elicit vowel production likely produced different vowel formant values to those that would be observed in conversational speech. In more conversational contexts, speakers tend to show decreases in average vowel duration coupled with a higher degree of vowel centralization. For this reason, vowels produced in sentence reading tasks exhibit less centralization than vowels produced in conversation, but more centralization than vowels produced in single words (Laan, 1997; van Bergem, 1995). It is possible that, in order to see differences in vowel centralization with age, a citation form vowel production task might be required—a task which demands greater movement of speakers' vocal tracts.

This investigation also raised questions about the relationship between age and speech rate. Specifically, in older people, it is unclear whether limitations in neuromuscular speed are directly responsible for speakers' reduced articulatory rate—or whether a slower rate of speech and longer vowel durations are a compensatory mechanism for maintaining articulatory precision. This study showed that speakers with longer vowel segments tended to display larger vowel spaces than others in their cohort. This indicated that a slower rate of speech was associated with more acoustically distinct vowels. It therefore seems unlikely that longer vowel durations are a direct result of limited neuromuscular control or impairment. Instead, it appears that the lengthening of vowel durations acts as a successful behavioural strategy for maintaining articulatory precision in older speakers.

In conclusion, this study goes some way towards elucidating changes to vowel articulation that occur as speakers age. The fact that speakers appear able to maintain vowel space area between the ages of 65 to 90 suggests that they make gradual adjustments to their

articulation to compensate for inherent anatomical and physiological changes that occur to the aging speech system. A slower rate of speech appears to be one of these adjustments.

2.9.1 Summary

This chapter provided information about how speech prosody and vowel articulation are likely to vary in NZE. This knowledge is important before attempting to interpret measures of prosody and vowel space in speakers with dysarthria. The next chapter builds on the information presented in this chapter and begins to explore acoustic measures of vowel articulation in dysarthric speech. As discussed in chapter one, this thesis was particularly interested in gathering objective and reliable measurements of speakers' overall speech severity. Measures of vowel articulation are very sensitive to changes in speech—and regularly show correlations with perceptual measures of intelligibility (Ferguson & Kewley-Port, 2007; Liu et al., 2005; Neel, 2008; Tjaden & Wilding, 2004). However, the high degree of naturally occurring variation observed in older NZE speakers' vowel measurements raises questions about the validity of these measures in indexing our perceptions of dysarthria severity. The next chapter explores this issue in further detail.

Chapter Three

Measuring Vowel Centralization and Dysarthria Severity: A Comparison of Techniques

Chapter three is an adaptation of the article titled "Assessing vowel centralization in dysarthria: A comparison of methods", which has been recently accepted at the Journal of Speech, Language and Hearing Research. In some sections the text has been modified and additional information has been provided to ensure consistency and relevance to the current chapter and thesis.

3.1 PREFACE

In the clinical management of communication disorders, the severity of dysarthria is determined through perceptual measurements of speech. As discussed in chapter one, listeners' perceptions of speech disorder provide important information about the functional impact of dysarthria. This is because perceptual measures allow us to make inferences about a client's level of disability in everyday communicative situations. For example, if listeners are not able to distinguish a person's speech impairment from healthy ageing (i.e. their speech disorder is not noticeable to listeners) then the dysarthria is unlikely to have much effect on the person's ability to communicate in everyday life. For these reasons, improvement in perceptual measures (e.g. intelligibility, naturalness) is usually a goal of speech therapy and a common way to mark progress in treatment (Duffy, 2013).

However, there are limitations to relying on perceptual ratings as a benchmark in dysarthria assessment. One issue is that listeners vary in their ability to parse dysarthric speech (Choe, Liss, Azuma, & Mathy, 2012). For this reason, perceptual measures may not provide meaningful units of comparison between different studies. Another concern, which presents in clinical practice, is that speech therapists can learn to understand dysarthric speech over time. Indeed, even a very short period of exposure to the speech of a person with dysarthria can improve a listener's ability to parse their speech signal a week later (Borrie et al., 2012). Therefore, additional techniques are needed to provide objective measurements of dysarthria which can be accurately replicated and compared over time.

Acoustic measures provide greater consistency and objectivity when measuring the features of dysarthric speech. But to validate these measures for clinical use, we need to know how well they index perceptual measures that are used in clinical practice. The previous chapter examined measurements of VSA as an index of articulatory precision. Results demonstrated that the size of speakers' VSA was significantly affected by speakers' sex and exhibited considerable variation amongst healthy speakers. Thus, it appears that the size of speakers' VSA does not provide a precise index of motor impairment. It was hypothesized that different methods of measuring VSA— such as changes in the formant extraction point—may affect its ability to capture speakers' articulatory gestures. If this is true, changes to the way vowel dispersion is measured may influence its relationship with perceptual ratings of dysarthria. The goal of this chapter is to examine the relationship between acoustic measures of vowel centralization and perceptual ratings of dysarthria by comparing different methods of acoustic and perceptual analyses. This chapter presents the

first measurements of the baseline speech signal in speakers with dysarthria—measures which provide an important component of the analyses in chapter four.

3.2 INTRODUCTION

Acoustic analysis of vowel sounds offers an objective assessment tool for measuring speech production in those with dysarthria. However, there are significant limitations in using acoustic metrics to infer information about listeners' perceptions of the disorder. Although studies have consistently reported an association between acoustic vowel centralization and perceptual measures, the strength of these relationships remains highly variable (Lansford & Liss, 2014a). Linking measurements of the speech signal to perceptual outcomes is an important component of validating acoustic metrics for clinical use. Understanding causes of variation in the relationship between acoustic and perceptual data is a first step towards establishing stronger links between these variables.

Centralization of vowel formants has been associated with reduced intelligibility in both healthy speakers and those with motor speech disorders (e.g. Ferguson & Kewley-Port, 2007; Liu et al., 2005; Neel, 2008; Tjaden & Wilding, 2004). In the motor speech literature, the most common way of measuring vowel centralization is through the calculation of vowel space area (VSA) – using the first and second formants of a dialect's corner vowels. Static vowel formant values can be extracted across a range of word tokens, enabling measurements to be taken from a variety of speech stimuli. Unfortunately, VSA measurements have high inter-speaker variability and have traditionally demonstrated variable success in distinguishing healthy and disordered speech (Sapir et al., 2010). Indeed, VSA has been reported to account for both between 6-8% (Tjaden & Wilding, 2004) and 69% (H. Kim, Hasegawa-Johnson, & Perlman, 2011) of the variance in perceptual ratings of dysarthria (for a more detailed review see Lansford & Liss, 2014a).

As Lansford and Liss (2014a) speculate, much of this inconsistency may be due to differences in the underlying nature of participants' dysarthria from one study to another (for example, there have been large differences in the severity of participants' dysarthria across studies). However, we hypothesize that this is not the only cause. Across studies, there are many differences in methods that are overlooked when results are summarized. This chapter will compare procedures used to measure vowel centralization and listeners' perceptions of dysarthria. The aim of this study is twofold. Firstly, we will determine whether (and to what degree) changes in analysis procedures affect the relationship between vowel centralization

measurements and perceptual ratings. Secondly, in order to make recommendations for future studies, we will determine which set of procedures produce the strongest relationship between the acoustic and perceptual measurements. To accomplish this, we will evaluate the following techniques: 1) the time-point of formant extraction, 2) the calculation of vowel centralization, and 3) the perceptual measurement of speech disorder.

3.2.1 Time-point of Formant Extraction

In the motor speech literature, formant measurements are almost universally taken from vowels' temporal midpoints. The rationale being that this provides a consistent measurement point that is as temporally removed from adjacent consonants as possible. However, it is well recognized that neighbouring consonants can affect formant values across the entire vowel segment (Hillenbrand, Clark, & Nearey, 2001) and, for this reason, the temporal midpoint may not necessarily provide the best representation of a vowel's steady-state formant frequency. Weismer and Berry (2003) also demonstrated that the shape of formant movements can vary from speaker to speaker. This suggests that speakers might reach a vowel's steady-state target—or an approximation of this position—at different stages of the vowel's duration. If this is the case, midpoint vowel measurements may obscure differences in formant movement between speakers.

To address this issue, more flexible measurement point criteria have been suggested (see section 2.6.3.3 or Fletcher, McAuliffe, Lansford, & Liss, 2015). These criteria would enable us to extract an approximation of the vowel's steady-state target, irrespective of the time-point it is reached. However, in the study of dysarthria, this approach has not been explored. Thus, while we suspect that a flexible formant measurement point may be more successful in indexing speakers' articulatory impairment, there are currently no data to support this hypothesis.

3.2.2 Calculation of Vowel Centralization

Our acoustic metrics should index speech motor impairment while limiting the degree of inter-speaker variation that is unrelated to speech disorder. However, no matter where they are extracted from, static vowel formants will always be affected by inherent differences in the size of speakers' vocal tracts. This variation obscures differences in vowel production that are due to changes in articulatory movement.

A number of methods aim to normalize anatomical and physiological differences between speakers' vowel formants in order to reduce the effect of age and sex on these measures (Clopper, 2009). However, in the study of motor speech disorders, many of these techniques can introduce problems. For example, normalizing the distances between speakers' vowels has the potential to remove information about the degree of articulatory movement they make (i.e. as the speaker moves from the production of one vowel to another). In fact, even some of the VSA differences between healthy male and female speakers may reflect differences in articulatory movement—with females seeking to expand the acoustic distance between their vowels (Cox, 2006; Diehl, Lindblom, Hoemeke, & Fahey, 1996).

One method of vowel normalization—which reduces variance caused by the size of the vocal tract—is to transform the frequency scale used to measure formants (Clopper, 2009). Transformations of frequency measurements are classed as 'vowel intrinsic' methods of normalization, and use only acoustic information contained within a single vowel to alter its formants. The aim of these methods, broadly speaking, is to model human vowel perception—not to eliminate physiological differences between speakers. Measuring formants in Bark reduces the absolute variance between speakers' VSAs. However, it is unclear whether this also reduces inter-speaker differences in articulatory impairment. H. Kim et al. (2011) found a strong relationship (i.e. $R^2 = 0.69$) between measurements of VSA in Bark and intelligibility scores for speakers with cerebral palsy. Although the study did not directly compare different units of measurement, the strong relationship between triangular VSA and their intelligibility measurements suggests that there may advantages to using Bark frequency units.

There have been investigations into other methods of normalizing inter-speaker variations in vocal tract size with the aim of maintaining differences in articulatory movement. In a recent paper, Sapir et al. (2010) suggested that using a ratio of each person's formant values would normalize inter-speaker variance in the magnitude of formant values, while preserving information about vowel centralization. They advocated use of the Formant Centralization Ratio (FCR) – which weighs formants that are likely to increase as a result of vowel centralization against formants which are expected to lower (Sapir et al., 2010). Lansford and Liss (2014a) found that FCR produced a stronger correlation between vowel centralization and listeners' perception of dysarthric speech than measures of VSA. In this case, the FCR measure was able to account for 15% more of the variance in speaker's intelligibility than a triangular VSA (despite using the same formant measurements).

Although these results were promising, data on this new measurement tool is lacking. It is not yet clear whether the FCR is able to consistently index our perceptions of dysarthria severity—or how this new measurement compares to vowel intrinsic methods of vocal tract normalization.

3.2.3 Perceptual Measurement of Speech Disorder

VSA is commonly indexed against some form of speech intelligibility measurement—for example, listener transcriptions of words and phrases (H. Kim et al., 2011; Lansford & Liss, 2014a; Liu et al., 2005). Although orthographic intelligibility measures are consistently linked to VSA, there are limitations to their use. In particular, orthographic transcription of dysarthric speech is limited in its ability to detect mild articulatory impairment (Sussman & Tjaden, 2012). That is, a listener may exhibit a perceptible dysarthria, but be given similar scores to healthy speakers on transcription intelligibility tests. Rating scales offer a useful alternative—allowing listeners to indicate that they detect speech impairment, even if they can still understand the words spoken. Many studies of vowel centralization have found a relationship between acoustic measures and scaled ratings of intelligibility (Y. Kim et al., 2011; McRae et al., 2002; Tjaden & Wilding, 2004; Turner et al., 1995; Weismer, Jeng, Laures, Kent, & Kent, 2001)—though the strength of these relationships remains highly variable.

When rating scales are used to measure speech impairment in dysarthria, listeners are usually asked to rate intelligibility or "how easy" the speaker is to understand (Y. Kim et al., 2011; Tjaden & Wilding, 2004; Turner et al., 1995; Weismer et al., 2001). However, it is possible that these ratings of intelligibility might be prone to the same issues as transcription based intelligibility scores. For example, even when listeners detect mild articulatory impairment, they may still rate a speaker as very easy to understand. To combat this issue, more global ratings of speech severity have been proposed (Sussman & Tjaden, 2012). Sussman and Tjaden (2012) found that scaled estimates of speech severity were able to distinguish speakers with mild dysarthria more successfully than transcription based intelligibility scores. But while they suggested that the instructions we give listeners are important in measuring dysarthria, the study did not directly compare different listener prompts (i.e. prompts to rate "intelligibility" vs. prompts to rate "speech severity"). Hence, it is not clear whether the instruction to rate "speech severity"—as opposed to intelligibility—made any difference to the sensitivity of their rating scale.

There are limited data to evaluate how listener instructions affect the measurement of dysarthric speech. Previously, Weismer et al. (2001) compared ratings of "intelligibility" with "speech severity" and found little difference in the amount that each rating predicted acoustic changes in VSA. There may, however, be a reason why ratings of 'intelligibility' performed particularly well in this study. When Weismer et al. (2001) gave instructions to rate speech intelligibility, listeners were also told to focus on articulatory precision. It is possible that by focusing on articulatory precision listeners produced ratings that were more sensitive to mild dysarthria. In contrast, instructions to rate "speech severity" required the listener to focus on all aspects of possible speech disorder: including parameters of nasality, prosody, vocal quality and respiration. Although these parameters are likely to be affected by dysarthria, they do not directly influence vowel centralization. For this reason, to best index changes in acoustic vowel production, it may be beneficial to have listeners rate a speaker's speech precision irrespective of other speech subsystem impairment.

In summary, there are a number of methodological factors that might affect the relationship between vowel centralization and listeners' perceptions of dysarthric speech. However, it is unclear to what degree these factors are capable of changing this relationship—and therefore contributing to the variable results reported in previous studies. This study will evaluate methods of acoustic and perceptual analysis to determine what effect they have on the relationship between measurements of vowel centralization and listeners' perceptions of dysarthria. In doing so, this study aims to determine which measures produce the strongest relationship between these variables—to provide the clearest acoustic index of dysarthria severity. Specifically, this investigation will compare the use of different 1) formant extraction time-points 2) methods of vocal tract normalization and 3) listener ratings of dysarthria. The results address several questions: (1) Do these changes in method produce significantly different perceptual ratings and measurements of vowel dispersion? (2) Are the resultant measurements able to distinguish individuals with dysarthria from healthy older speakers? (3) Do these methodological changes strengthen the relationship between the acoustic and perceptual measures?

3.3 METHOD

3.3.1 Speakers

Sixty-one speakers of New Zealand English (NZE) (42 males and 19 females), aged between 43 and 89 years, participated in this study. Of these speakers, 44 were diagnosed with dysarthria. The dysarthria varied in severity, with speakers classed as mild (n=16), mild-moderate (n=9), moderate (n=8), moderate-severe (n=4) and severe (n=7). Dysarthria subtypes and perceptual classification of severity were provided by three experienced speech-language pathologists via a consensus rating procedure, based on speakers' recordings of the Grandfather Passage. Biographical details are supplied in Table 3.1. The remaining 17 speakers, who reported no history neurological impairment or speech and language disorders, acted as healthy controls. The group diagnosed with dysarthria had a mean age of 65 years, while the control group had a mean age of 66 years.

Table 3.1

Demographic Information for Speakers with Dysarthria

Participant Number	Sex	Age	Medical Aetiology	Severity of Disorder
1	F	48	traumatic brain injury	mild-moderate
2	M	60	traumatic brain injury	moderate
3	M	55	traumatic brain injury	mild-moderate
4	F	67	Progressive supranuclear palsy	mild
5	F	68	Freidreich's ataxia	mild
6	F	70	Parkinson's disease	mild-moderate
7	M	75	Parkinson's disease	moderate
8	F	79	Parkinson's disease	mild
9	M	56	cerebellar ataxia	mild
10	F	45	Wilson's disease	mild
11	M	53	Undetermined neurological disease	moderate
12	M	55	Undetermined neurological disease	moderate
13	M	58	brainstem stroke	moderate
14	M	76	Parkinson's disease	mild
15	M	67	Parkinson's disease	mild-moderate
16	M	77	Parkinson's disease	mild
17	M	67	Parkinson's disease	mild
18	M	79	Parkinson's disease	moderate

19	M	71	Parkinson’s disease	moderate
20	M	71	Parkinson’s disease	mild-moderate
21	F	83	Parkinson’s disease	mild
22	M	68	Parkinson’s disease	mild
23	F	73	Parkinson’s disease	mild-moderate
24	M	89	Parkinson’s disease	mild
25	M	58	Parkinson’s disease	mild
26	M	81	Parkinson’s disease	moderate-severe
27	M	73	Parkinson’s disease	mild
28	M	79	Parkinson’s disease	mild
29	M	77	Parkinson’s disease	moderate-severe
30	M	69	Parkinson’s disease	moderate
31	M	69	Parkinson’s disease	mild
32	M	65	Parkinson’s disease	mild-moderate
33	M	68	Parkinson’s disease	mild
34	M	47	traumatic brain injury	severe
35	M	64	spinocerebellar ataxia	severe
36	F	69	cerebral palsy	severe
37	F	60	multiple sclerosis	moderate-severe
38	M	55	Huntington’s disease	severe
39	F	53	multiple sclerosis	mild-moderate
40	F	47	Huntington’s disease	moderate-severe
41	M	43	hydrocephalus	severe
42	M	60	cerebral palsy	severe
43	M	72	stroke	severe
44	F	46	brain tumor	mild-moderate

Note. F = female; M = male

3.3.2 Speech Stimuli

Each speaker attended a single recording session. Recordings took place in a quiet room, with an investigator present. Participants were asked to read the Grandfather Passage (see Appendix A) in their normal speaking voice after familiarizing themselves with passage. Two

participants with dysarthria required assistance reading the passage. In these instances, the first author would read full sentences from the passage, with the speaker repeating the sentences immediately afterwards. For 58 participants, digital audio recordings were made via an Audix HT2 headset condenser microphone, positioned approximately five centimetres from the mouth. Digital audio recordings of these speakers were made at 48 kHz with 16 bits of quantization. The remaining three participants were female control speakers who were recorded as part of the study presented in the previous chapter. These participants were recorded using a Zoom H4n recorder placed on the table in front of them (at an approximate distance of 30 centimetres). Their audio recordings were made at 22.05 kHz with 16 bits of quantization. As part of formant extraction procedure, all sound files were later resampled to a lower frequency as per the Burg LPC algorithm described in the next section.

3.3.3 Extraction of Acoustic Data

3.3.3.1 Segmentation of the data set

The recordings were transcribed, automatically segmented to the phoneme level, and labelled in Praat using the same methods described in section 2.6.3.1. The accuracy of all phoneme boundaries was checked by a team of four trained analysers who visually examined of the waveform and wide-band spectrogram, and listened for auditory cues. As described in the previous chapter, the primary indicators for the onset and offset of vowels were changes to formant structures, voicing and waveform amplitude. Vowel onset boundaries were identified at the start of the pitch period coinciding with the onset of regular formant structure. Vowel offset boundaries were distinguished by changes in formant structure at the end of the pitch period, where there was a corresponding drop in waveform amplitude. The amplitude, shape, and lack of frication of successive pitch periods were also used to determine boundaries. Since the HTK segmentation was completed at the phoneme level, if the person checking phoneme boundaries was uncertain in discriminating boundaries for consecutive phonemes, the boundary derived from automatic segmentation was kept in place.

3.3.3.2 Extraction of formant values

For consistency, this study used the same three tokens of the NZE START [ɐː], FLEECE [iː], and THOUGHT [oː] vowels that were examined in the previous chapter. Due to reading errors, speakers occasionally missed one of the selected tokens. In this instance, the remaining two tokens were used. In instances of dysfluency, where speakers repeated certain

word tokens, the average formant value across word repetitions was used. The formant tracks of the first five formant frequencies were obtained via Praat using the Burg LPC algorithm, with a Gaussian window length of 25 ms, a time step of 6.25 ms between the centres of consecutive windows, a maximum formant value of 5.5 kHz for females and 5 kHz for males, and a pre-emphasis from 50 Hz (Boersma & Weenink, 2012). Formant one (F1) and formant two (F2) measurements were extracted from two measurements points in each vowel. Criteria for the formant measurement points are outlined below. Each set of vowel formants was measured in Hz and also transformed into the Bark frequency scale (Traunmüller, 1990).

3.3.3.3 Midpoint formant values

Midpoint values were automatically extracted using a custom Praat script. All formant tracks were then visually checked. If the midpoint values selected by the script did not accurately represent the formant that was being measured (i.e. the formant track was not centred on the correct formant band) the measurement point was adjusted by hand.

3.3.3.4 Articulatory point formant values

The articulatory point criteria were designed with the aim of extracting values at the time where there was the least movement in the formant tracks—for the best approximation of the vowels' steady-state target. For the front [iː] vowel, this point was set at peak F2 frequency; for the open [ɐː], formants were extracted when F1 was at its maximum; and for back [oː] vowel, when the lowest value of F2 was reached. Articulatory point formant values were all automatically extracted using a custom praat script. All vowel measurement points were then visually checked. As described above, if the values selected by the script did not accurately represent the formant that was being measured, the measurement point was adjusted by hand.

It should be noted that the criteria used in this chapter differed slightly from those described in section 2.6.3.3. When examining speakers with dysarthria, it was observed that the use of Praat scripts resulted in a much larger number of formant tracking errors as compared with healthy speakers. It appeared that this was due to wider and less distinct formant regions for those with dysarthria. Given these errors, all the midpoint and articulatory point formant values extracted by Praat were hand checked. Because all values were visually checked in this study, we did not place the same restrictions on the scripts to avoid capturing information about the neighbouring consonant. Instead, we allowed the articulatory point formant values

to be selected from across the entire length of the vowel. An example of how the midpoint and articulatory extraction points might differ in a speaker with dysarthria is shown visually in Figure 3.1.

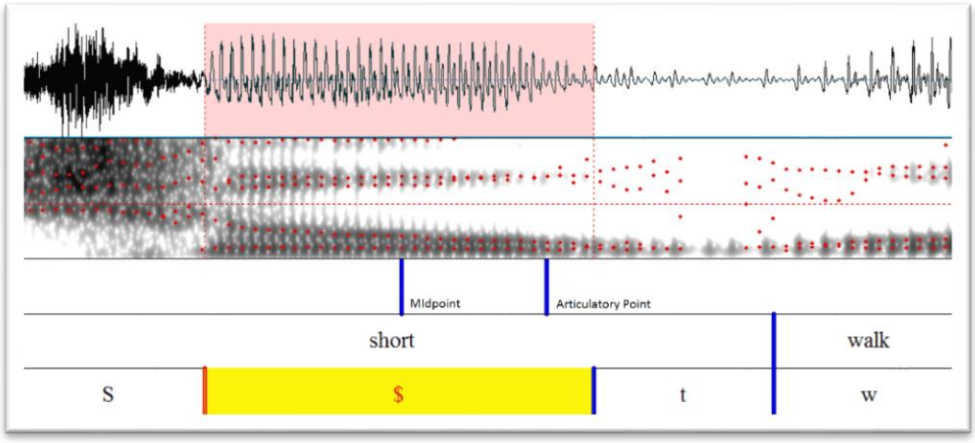

Figure 3.1. Example of the two extraction points within a speaker's [oː] vowel.

3.3.4 Description of the Acoustic Metrics

Vowel centralization was calculated with two metrics of vowel articulation. The measures employed are described below.

3.3.4.1 Vowel space area

Vowel space area (VSA) was calculated using F1 and F2 of the [ɐː], [iː] and [oː] vowels. Given the NZE dialect, a measure of triangular VSA (using the THOUGHT vowel as opposed to GOOSE) provides a more accurate representation of vowel dispersion than quadrilateral VSA (Maclagan, 2009). F1 and F2 values for the three [ɐː], [iː] and [oː] word tokens were averaged for each speaker. Triangular vowel space area was constructed by plotting these values as coordinates in a F1/F2 plane, and calculating the resulting triangular area using the formula: $Hz^2 = 0.5 \times ABS[F1[iː] \times (F2[ɐː] - F2[oː]) + F1[oː] \times (F2[iː] - F2[ɐː]) + F1[ɐː] \times (F2[oː] - F2[iː])]$, where ABS = absolute value, F1[iː] = first formant frequency of the [iː] vowel, and so on.

3.3.4.2 Formant centralization ratio

Given the dispersion of vowels in NZE, the Formant Centralization Ratio (FCR) metric was adapted from Sapir et al. (2010)[5]. It was calculated with the same average F1 and F2 values for each speaker as detailed above, again, using the THOUGHT vowel as opposed to GOOSE. Therefore, FCR was realized as: (F2[oː] + F2[ɐː] + F1[iː] + F1 [ɐː]) ÷ (F2[iː] + F1[oː]).

The procedures in sections 3.3.3 and 3.3.4 resulted in eight different formant centralization measurements for each speaker, outlined in Table 3.2.

Table 3.2

Combinations of Acoustic Vowel Metrics

Formant Measurement Point	Unit of Measurement	Vowel Centralization Metric
Temporal Midpoint	Hz	VSA
Temporal Midpoint	Hz	FCR
Temporal Midpoint	Bark	VSA
Temporal Midpoint	Bark	FCR
Articulatory Target	Hz	VSA
Articulatory Target	Hz	FCR
Articulatory Target	Bark	VSA
Articulatory Target	Bark	FCR

Note. VSA = Vowel Space Area, FCR = Formant Centralization Ratio

3.3.5 Reliability of Acoustic Measures

To determine inter- and intra-rater reliability of the measures, 10% of textgrids were manually re-examined for reliability. Phoneme boundaries were manually rechecked and scripts to obtain vowel formant values were re-administered. The newly generated vowel formants values were visually checked in the same manner described in section 3.3.3.3. In the

[5] The formula provided by Sapir et al. (2010) is given as: (F2/u/ + F2/ɑ/ + F1/i/ + F1/u/) / (F2/i/ + F1/ɑ/). In NZE, the vowel in THOUGHT is produced much further back than the vowel in GOOSE. For this reason, its inclusion better represents the overall vowel dispersion of the NZE speakers.

case of midpoint formant values, the reanalysis found F1 intra-rater reliability scores averaged within 12 Hz of original values, and F2 scores were within 22 Hz. The average inter-rater difference was 26 Hz for F1 values and 46 Hz for F2. The reanalysis of the articulatory points found F1 intra-rater reliability scores within 16 Hz of original values, and F2 scores within 23 Hz. Average inter-rater differences were 35 Hz for F1 values and 29 Hz for F2.

3.3.6 Perceptual Task

3.3.6.1 Listeners

Listeners consisted of two randomly assigned groups of 14 adults (aged 18 to 47). The listeners were native speakers of NZE, who did not have training in the assessment of dysarthria. All listeners passed a pure tone hearing screening at 20dB HL for 500, 1000, 2000, and 4000Hz in both ears.

3.3.6.2 Listening stimuli

Due to the large amount of speech data collected in this study, only a small portion of the reading passage was used to gather perceptual ratings. The phrase: “he slowly takes a short walk in the open air each day” was selected for this purpose. Across the speaker group, this phrase was free from reading errors. For all recordings, the average intensity of the phrase was scaled to 70dB to provide a similar perceived loudness.

3.3.6.3 Procedure

All listeners completed the rating task in one session. The listening task was programmed on E-prime and speech stimuli were played through Panasonic RP-HT 161 stereo headphones. In group one, listeners were asked to rate how easy the speaker was to understand; while in group two, listeners were asked to rate the speakers’ speech precision. An example of the instructions participants received is provided in Appendix B.

Although each group was given different prompts, all other rating procedures were identical. Before beginning the experiment, listeners completed a short practice task, to familiarize them with the rating procedure, and allow them the opportunity to adjust the volume of the computer to a comfortable level. In the practice task, listeners were exposed to three recordings. These included a speaker with severe dysarthria, a speaker with mild-

moderate dysarthria, and one healthy older speaker. These speakers were not included in the main experiment.

The main experiment consisted of 61 phrases—one from each speaker listed in Table 3.1. The phrases were randomly presented twice, giving a total of 122 trials for every listener. In every trial the listeners were presented with a prompt to either rate "the speaker's speech precision" or "how easy is the speaker to understand?". Listeners pressed a button to hear the recording play and clicked on a visual analogue scale to place a copy of the button onto the scale. For listeners in group one, the scale ranged from "easy" at one end to "difficult" at the other. For the second group, the scale ranged from "precise" to "imprecise". Listeners were able to adjust their rating as often as they wished before selecting to move to the next trial.

The raw output of these judgments was an integer between 0 and 100 for each stimulus phrase. For each listener, the average and standard deviation of all ratings was calculated. This information was used to compute a z score for every speaker that was rated by the listener. For example, a numeric rating given by 'listener one' would be converted in the following manner:

$$\frac{\text{rating of speaker by listener one} - \text{average rating given by listener one}}{\text{standard deviation of listener one's ratings}}$$

This z score procedure ensured that listeners who tended to give speakers higher ratings (while placing bigger spaces between different speaker ratings on the VAS) would not have a larger influence the peoples' average ratings. After applying this z score procedure, the scores of all listeners were averaged to determine the final rating for that speaker.

3.3.6.4 Reliability of the perceptual task

To assess intra-rater reliability, Pearson's product-moment correlations (across ratings of the same speech samples) were calculated based on listeners' raw ratings (i.e. scores between 0-100). For intelligibility ratings, the average intra-rater correlation between the first and second presentation of the phrases ranged from .70 to .95, with a mean of .88. For ratings of speech precision, the intra-rater correlations were between .86 and .96, with a mean of .90. To assess inter-rater reliability, intraclass correlations (ICC) were calculated (as described in Sheard et al., 1991). The obtained ICC (2,1) coefficients were .677 for intelligibility ratings and .835 for speech precision ratings.

3.4 RESULTS

The results of this study are discussed in three parts, in order to address: (1) whether measurements of vowel dispersion and perceptual ratings were affected by changes in methods, (2) whether these measurements were able to distinguish individuals with dysarthria from healthy older speakers, and (3) whether methodological changes strengthened the relationship between the acoustic and perceptual measures.

3.4.1 Differences in Measurements

3.4.1.1 Method of formant extraction

The use of the two measurement points resulted in statistically significant differences in the size of speakers' VSAs. For example, in speakers with dysarthria, the average midpoint VSA was significantly smaller (M = 147315 Hz^2, SD = 74337 Hz^2) than the average articulatory point VSA (M = 207575 Hz^2, SD = 86283 Hz^2, $t(43) = 11.2$, $p = < .001$). This was also the case for control speakers, with the midpoint formants producing significantly smaller VSA values (M = 217220 Hz^2, SD = 81373 Hz^2) than those extracted from the articulatory point (M = 293539 Hz^2, SD = 106344 Hz^2, $t(16) = 7.7$, $p < .001$). However, despite these differences, the two measures were highly correlated ($r(59) = .93$, $p < .001$).

3.4.1.2 Unit of measurement and vowel centralization metric

Table 3.3 provides mean FCR and VSA values of male and female speakers calculated using Hertz and Bark, and across the two measurement points. The results indicate that, as expected, formant vales for males and females are more similar when measured in Bark. These data suggest that differences caused by the size of the vocal tract are indeed reduced when the Bark scale is used. Table 3.3 also demonstrates that the mean difference between males and females is reduced when vowel centralization is measured using the FCR, as opposed to VSA. Together, the combined use of the Bark scale and the FCR eliminated any significant differences in vowel centralization measurements between male and female speaker groups—both when midpoint ($t(59) = 1.05$, $p > .05$) and articulatory point formant values were used ($t(59) = 0.57$, $p > .05$).

Table 3.3

Measurement Differences Between Males and Females

	Articulatory Target Measurement			
	VSA - Hz^2	VSA - $Bark^2$	FCR - Hz	FCR - Bark
Male	195295 *(80218)*	8.493 *(3.25)*	1.048 *(0.12)*	1.265 *(0.11)*
Female	311635 *(91527)*	11.192 *(2.89)*	0.987 *(0.08)*	1.249 *(0.07)*
	Temporal Midpoint Measurement			
	VSA - Hz^2	VSA - $Bark^2$	FCR - Hz	FCR - Bark
Male	137331 *(65552)*	5.974 *(2.60)*	1.132 *(0.12)*	1.348 *(0.11)*
Female	231932 *(78255)*	8.394 *(2.60)*	1.061 *(0.10)*	1.319 *(0.08)*

Note. VSA = Vowel Space Area; FCR = Formant Centralization Ratio

3.4.1.3 Perceptual correlates of dysarthric speech

The relationship between the two perceptual measurements –"ease of understanding" and "speech precision"– is shown in Figure 3.2. Overall, listeners' perceptions, as measured by the two different rating instructions, were highly correlated ($r(59) = .98$, $p < .001$). However, while closely related, the data points in Figure 3.2 appeared to have a curvilinear relationship. This observation was confirmed by comparing a simple linear regression against a second-degree polynomial model of the two variables. A comparison of the models revealed that the curvilinear, polynomial model accounted for significantly more variance in the data ($F(1, 58) = 21.03$, $p < .001$). The existence of a curvilinear relationship demonstrates that there are differences in the way the two perceptual measures index mild, moderate and severe dysarthria. For example, speakers with a mild dysarthria tended to exhibit higher *z* scores (i.e. scores that were further away from the mean) for speech precision than for intelligibility. In contrast, speakers with more severe dysarthria tended to be slightly further from the mean in their intelligibility scores. This suggested that ratings of intelligibility and speech precision were distributed differently, with ratings of speech precision producing a larger range of scores for speakers above the mean (i.e. those with less impairment).

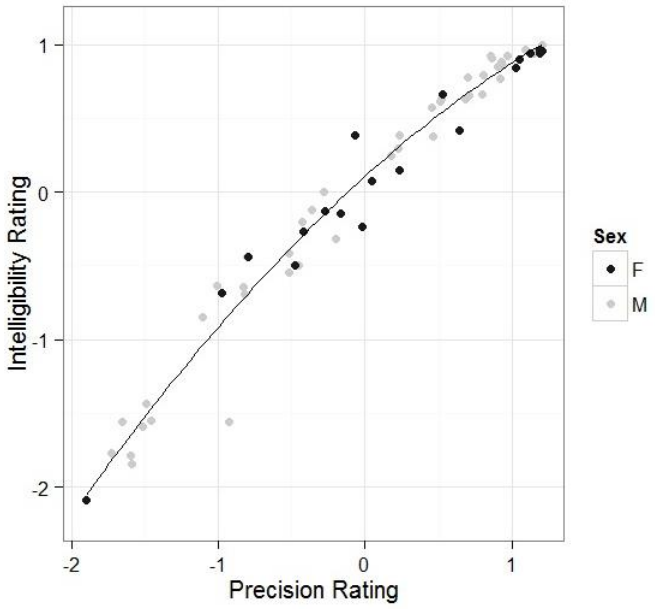

Figure 3.2. Relationship between listeners' ratings of intelligibility and speech precision. M= male, F = female

3.4.2 Measurement Differences between Speakers with and without Dysarthria

Differences in perceptual and acoustic measurements between speakers with dysarthria and healthy controls and summarized in Table 3.4. The perceptual measurements were combined for male and female speakers after determining that there were no significant differences between the sexes for ratings of intelligibility ($t(59) = 0.85$, $p > .05$) or speech precision ($t(59) = 0.95$, $p > .05$). Perceptual ratings of speech precision produced a greater mean difference between speakers with and without dysarthria than ratings of intelligibility (after both measures had been z scored).

The acoustic measures were compared separately in groups of male and female speakers. All measurements produced statistically significant differences between the speakers with and without dysarthria (at $p < .05$). However, it was apparent that some measures were able to separate the two groups more clearly than others (i.e. there was less overlap in the distribution of measurements across the two groups, as indicated by higher *t* values). Firstly, formants taken with a flexible extraction point were able to consistently produce higher *t* values in comparisons between the speakers with and without dysarthria.

Measuring formant values in Bark units also produced consistently higher *t* values. In contrast, the FCR did not perform consistently better than measures of VSA in distinguishing speakers with dysarthria—as measures of VSA in $Bark^2$ produced particularly high *t* values in these group comparisons.

Table 3.4

Differences in Perceptual Ratings and Vowel Dispersion Metrics in Participants with and without Dysarthria

Measurement	Average in Speakers with Dysarthria	Average in Control Speakers	*t* value	*p* value
Rating of Speech Precision	-0.382 *(0.805)*	0.987 *(0.198)*	6.895	<.001
Rating of Intelligibility	-0.339 *(0.834)*	0.877 *(0.110)*	5.960	<.001
Male Speakers				
VSA in $Bark^2$ using a flexible formant extraction point	7.635 *(3.034)*	10.912 *(2.635)*	3.178	.003
FCR using formants measured in Bark from a flexible extraction point	1.293 *(0.113)*	1.186 *(0.066)*	2.967	.005
VSA in $Bark^2$ using formants from the temporal midpoint	5.341 *(2.479)*	7.755 *(2.156)*	2.864	.007
FCR using formants measured in Hz from a flexible extraction point	1.077 *(0.124)*	0.968 *(0.070)*	2.746	.009
FCR using formants measured in Bark from the temporal midpoint	1.372 *(0.111)*	1.280 *(0.064)*	2.594	.01
VSA in Hz^2 using a flexible formant extraction point	178294 *(77630)*	243206 *(69879)*	2.441	.02
FCR using formants measured in	1.157	1.061	2.325	.03

Hz from the temporal midpoint	*(0.128)*	*(0.078)*		
VSA in Hz^2 using formants from the vowels' temporal midpoint	124062 (*64460)*	174724 *(55396)*	2.317	.03
Female Speakers				
FCR using formants measured in Bark from a flexible extraction point	1.279 *(0.060)*	1.185 *(0.028)*	3.615	.002
VSA in $Bark^2$ using a flexible formant extraction point	9.958 *(2.088)*	13.866 *(2.641)*	3.496	.003
VSA in $Bark^2$ using formants from the temporal midpoint	7.287 *(2.187)*	10.794 *(1.682)*	3.464	.003
FCR using formants measured in Hz from a flexible extraction point	1.018 *(0.071)*	0.920 *(0.031)*	3.224	.005
FCR using formants measured in Bark from the temporal midpoint	1.351 *(0.080)*	1.248 *(0.029)*	3.024	.008
VSA in Hz^2 using a flexible formant extraction point	277397 *(64148)*	385816 *(103108)*	2.829	.01
VSA in Hz^2 using formants from the vowels' temporal midpoint	202764 *(68454)*	295128 *(61605)*	2.814	.01
FCR using formants measured in Hz from the temporal midpoint	1.097 *(0.104)*	0.984 *(0.029)*	2.571	.02

Note: Standard deviations across groups are shown in italics. All t values were derived from two sample, independent t tests. Equal variance between groups was assumed after applying Levene's test. Absolute *p* values are reported, with no corrections made for multiple comparisons.

3.4.3 Relationship between Acoustic and Perceptual Measures

Figure 3.3 plots the strongest and weakest relationships found between the acoustic and perceptual measures in male and female speakers. There were no signs of non-linearity in these relationships that would indicate changes in the acoustic-perceptual relationship for speakers with differing levels of articulatory impairment. For this reason, speakers with and

without dysarthria were assessed together in order to capture a complete range of speakers' articulatory capabilities.

To evaluate the relationship between the vowel centralization and perceptual ratings of dysarthria, a series of Pearson correlation analyses were performed. These are summarized in Table 3.5. Table 3.5 shows that, in both male and female speakers, there were common methodological approaches that improved the association between perceptual and acoustic measurements. In combination, changes to the formant extraction point, unit of measurement, metric of vowel centralization, and listener instructions resulted in 17% more variance being accounted for in males (i.e. an increase from 17 to 34%), and 27% more variance accounted for in females (from 49 to 76%). Overall, the strongest was relationship between acoustic and perceptual measures—in both male and female speakers—was achieved by using a flexible formant extraction point, Bark units, the FCR metric, in combination with listener ratings of speech precision.

Table 3.5

Relationships Between Acoustic Vowel Metrics and Perceptual Measures

	Females			
	Speech Precision Rating		Intelligibility Rating	
Vowel Space Metric	Correlation Coefficient	Explained Variance	Correlation Coefficient	Explained Variance
FCR using formants measured in Bark from a flexible extraction point	-.873	76%	-.839	70%
FCR using formants measured in Bark from the temporal midpoint	-.854	73%	-.810	66%
FCR using formants measured in Hz from a flexible extraction point	-.852	73%	-.836	70%

VSA in $Bark^2$ using formants from the temporal midpoint	.832	69%	.774	60%
VSA in $Bark^2$ using a flexible formant extraction point	.808	65%	.750	56%
FCR using formants measured in Hz from the temporal midpoint	-.798	64%	-.764	58%
VSA in Hz^2 using formants from the vowels' temporal midpoint	.751	56%	.710	50%
VSA in Hz^2 using a flexible formant extraction point	.739	55%	.698	49%

Males

	Speech Precision		Intelligibility	
Vowel Space Metric	Correlation Coefficient	Explained Variance	Correlation Coefficient	Explained Variance
FCR using formants measured in Bark from a flexible extraction point	-.584	34%	-.565	32%
FCR using formants measured in Hz from a flexible extraction point	-.581	34%	-.568	32%
VSA in $Bark^2$ using a flexible formant extraction point	.576	33%	.556	31%
FCR using formants measured in Bark	-.550	30%	-.523	27%

from the temporal midpoint				
FCR using formants measured in Hz from the temporal midpoint	-.548	30%	-.531	28%
VSA in Hz^2 using a flexible formant measurement point	.516	27%	.500	25%
VSA in $Bark^2$ using formants from the temporal midpoint	.505	26%	.475	23%
VSA in Hz^2 using formants from the temporal midpoint	.435	19%	.407	17%

Note. All correlations have *p* values of less than .01.

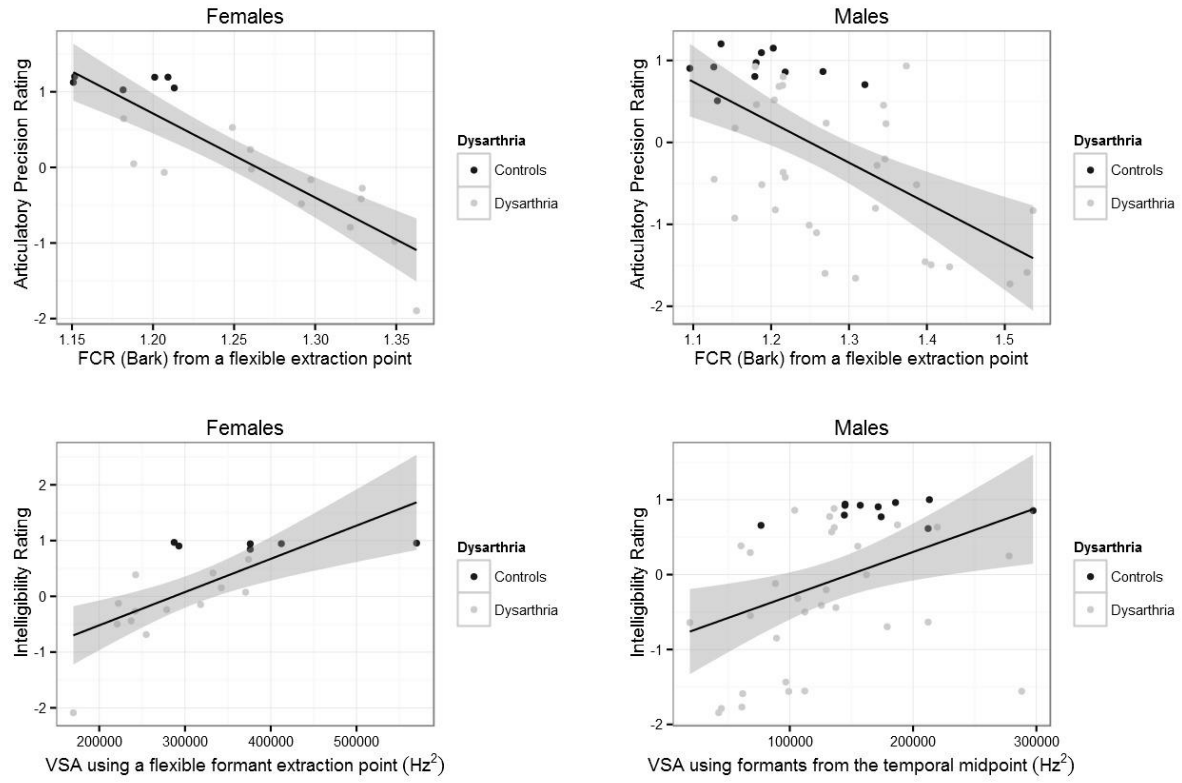

Figure 3.3. A comparison of the relationships between acoustic vowel metrics and perceptual

measures, plotted by sex. Shaded areas indicate 95% confidence interval of regression estimates.

3.5 DISCUSSION

This study aimed to determine which procedures would result in the strongest relationship between measurements of vowel centralization and listeners' perceptions of dysarthria. Previous literature has reported considerable variability in the correlations between these measurements, and it is unclear to what degree different procedures might be contributing to this inconsistency. It was hypothesized that there were several methodological changes that might affect the relationship between our acoustic and perceptual measurements.

3.5.1 Method of Formant Extraction

Two different formant extraction points were explored: a static temporal midpoint and a flexible articulatory point. It was hypothesized that the articulatory point might better index speakers' articulatory impairment by indexing a larger degree of vocal tract movement between vowels. The first question was whether the use of the articulatory point extraction criteria made any notable difference to the magnitude of formant values and the resultant VSAs. Consistent with data presented in section 2.7.1, study findings indicated that VSA values were considerably larger when formants were extracted from the articulatory point. Overall, these results provide further support for the hypothesis that the articulatory point produces more acoustically distinct formant values.

The aim of extracting formants from the articulatory point was to determine whether this method would strengthen the relationship between vowel centralization metrics and perceptual ratings. The results indicated that formant measurements taken from the articulatory point tended to explain more of the variation in speakers' perceptions of dysarthria. The articulatory point formants were used in four vowel centralization metrics—applied to both male and female data. On average, they resulted in an increase of six percent in the variance accounted for by perceptual ratings in the male data, but only three percent in the female data.

The reason for the decreased performance in the female data appears to be due to a single speaker. Figure 3.3 demonstrated that the relationship between the articulatory point

VSA and listeners' perceptions of dysarthria was skewed by one high VSA value in the female subset. This outlier meant females' midpoint VSAs were more closely related to perceptual measures than the corresponding articulatory point VSAs. When the FCR was applied, the outlier was no longer apparent. These data suggest that articulatory point values are generally able to capture more information about speakers' perceived severity—but they may also be more sensitive to changes in vocal tract size. Hence, to achieve a strong relationship with perceptual ratings of dysarthria, it may be advisable to use articulatory point criteria in conjunction with the FCR procedure recommended by Sapir et al. (2010), to help normalize differences in vocal tract size.

3.5.2 Unit of Measurement and Vowel Centralization Metric

Outside of the motor speech disorder literature, it is common to convert formant data to Bark units before calculating VSA, in order to provide an appropriate auditory scaling of frequency (Ferguson & Kewley-Port, 2007; Neel, 2008). It was hypothesized that measuring vowel centralization using Bark units would increase the relationship between acoustic and perceptual measures by reducing the effect that differences in the size of the vocal tract had on speakers' VSAs. There is evidence that the use of Bark units did reduce these anatomical differences. In the current study, transformation of F1 and F2 to Bark units resulted in a reduction in the difference in VSA between male and female speakers (i.e. the measurements became less than one standard deviation apart). As expected, this also occurred when FCR was used in place of VSA. However, it was only when Bark units were applied to the FCR that the differences between the sexes became insignificant. This finding indicates that Bark units have the potential to reduce inter-speaker variations over and above what the FCR alone is able to accomplish, and that the use of the FCR does not necessarily render Bark units redundant.

Within this dataset, the combined use of Bark units and the FCR provided a scale of values that could be interpreted together, regardless of a person's sex. The ability to plot and interpret these data as one group enhances sample size and increases the power to detect a relationship between our acoustic and perceptual measures. It is interesting that, in isolation, the application of the FCR was not able to completely eliminate group differences between males and females—in contrast to findings reported by Sapir et al. (2010) and Lansford and Liss (2014a). Differences may have persisted because of sociophonetic differences between

the sexes (Cox, 2006; Diehl et al., 1996) or, simply, because complex differences in vocal tract size could not be easily normalized in this population.

While it was evident that the use of FCRs and Bark units reduced variance in the acoustic measurements, the question remained: do these techniques also eliminate important information regarding articulatory movement? The results presented in Table 3.5 suggest that this is not the case. In comparison to triangular VSA, the use of an adapted FCR consistently improved the acoustic-perceptual relationship amongst speakers (accounting for an average of 6% more variance in males and 11% in females). This result was consistent with findings from previous studies which have compared FCR to triangular VSA (Lansford & Liss, 2014a; Sapir et al., 2010), and demonstrates the utility of using formant ratios as a measurement tool in other dialects. As hypothesized, the use of the Bark unit of frequency tended to increase the relationship between VSA measures and perceptual ratings (with a 6% increase in variance accounted for in males and 10% in females). When the FCR was used, there was much smaller effect of using Bark units (an average of 5% increase in the female data, with no increase observed in the male data).

It is worth noting that both the original FCR and our adapted formula utilize measurements of only three corner vowels. In dialects with more corner vowels, the inclusion of additional formants in this ratio may benefit the measurement's validity. For example, Lansford and Liss (2014a) found that the FCR was able to account for over 15% more variance in intelligibility measurements than triangular VSA (when utilizing the same three vowels). However, this difference reduced to just over 3% when the quadrilateral vowel space was used—indicating that a set of four vowels may more adequately index the vowel dispersion of their US speakers. In the case of NZE, the triangular shape of the dialect's vowel dispersion lends itself well to the three-vowel FCR and, for this reason, may have boosted the success of the measurement tool (Maclagan, 2009).

Overall, when making these inter-speaker comparisons, it appears that formant ratios have the capacity to map more strongly to our perceptual impressions than vowel space measurements. However, in dialects with a more quadrilateral dispersion of vowels, the effect of Bark units on quadrilateral VSA (and the inclusion of more vowels in formant ratios), warrants further examination.

3.5.3 Perceptual Correlates

Findings of the current study suggested that small changes to listener prompts may affect the way ratings of dysarthria are distributed (as demonstrated in Figure 3.3). While it is apparent that the two measurements used in this study were highly correlated, the distribution of ratings meant that there was less variation amongst the 'above average' scores of intelligibility. This was consistent with the hypothesis that ratings of 'ease of understanding' may not be as sensitive to mild speech disorder as ratings of 'speech precision'. Indeed, data presented in Table 3.4 indicated that ratings of speech precision tended to better separate the speakers with dysarthria from healthy controls. Across all metrics of vowel centralization, ratings of speech precision explained the most acoustic variance between speakers (an average of 7% in females and 2% in males). Although the improvement was subtle, it is suggested that ratings of speech precision may capture changes in vowel centralization more successfully than ratings of intelligibility.

There are many ways to perceptually scale dysarthria severity that were not investigated in this study. For example, several previous studies have focused on comparisons of equal interval scales and direct magnitude estimates (DME) of speech disorder (e.g. Eadie & Doyle, 2002; Schiavetti, Metz, & Sitler, 1981; Zraick & Liss, 2000). These studies have suggested that listeners will not necessarily divide speech stimuli into intervals with an equal magnitude of change between them (i.e. the magnitude of change between a rating of 1 and 2 may be different to the change between 5 and 6).

The current investigation used visual analogue scales (VAS) to rate dysarthria. Unlike equal interval scales, VAS do not force listeners to partition speech samples into categories—and may have allowed the listeners to better index differences in the magnitude of speakers' intelligibility and speech precision. In the current study, VAS enabled listeners to record their judgments quickly, with high reliability. The resultant ratings were able to account of up to 76% of the variance in vowel centralization measures—providing good evidence of their utility in measuring the speech signal. It is possible that DME may also have produced results sensitive to acoustic change. Comparisons of DME and VAS ratings should be explored in future work, particularly when large numbers of speech stimuli are being assessed. These comparisons should focus on the ability of the scales to index objective changes in the speech signal—rather than simply comparing the distributions of listener scores.

3.5.4 Limitations

3.5.4.1 Use of peripheral vowel space to index articulatory disorder

It is well recognised that isolated measures of F1 and F2 do not provide the complete range of information necessary for accurate phoneme identification (Hillenbrand et al., 2001). Even in vowels that are traditionally classed as "monophthongs", differences in formant movement and duration have been demonstrated to significantly influence listeners' abilities to correctly classify vowel phonemes (Nearey & Assmann, 1986; Neel, 2008). Ideally, when assessing vowel quality, we should not only collect information about discrete F1 and F2 values, but also consider the degree and velocity of formant movement (Nearey & Assmann, 1986; Neel, 2008), vowel duration (Neel, 2008), and information from the third formant (Hillenbrand et al., 2001). The articulatory point vowel space measures used in this study take into account only the most extreme articulatory positions used to produce peripheral vowels. Previous work has suggested that these measurements are particularly well related to dysarthria severity—as opposed to acoustic information from less peripheral vowels in the F1/F2 space (Lansford & Liss, 2014a). This may be because the production of peripheral vowels provides a greater physiological challenge for speakers with dysarthria. For example, speakers with dysarthria are prone to articulatory undershoot, and this may be most evident when comparing articulatory positions that require a large range of vocal tract movement. In contrast, healthy speakers (and perhaps speakers with a milder dysarthria) may not show any restrictions in their range of articulatory movement. For these speakers, it is likely that subtler differences in articulatory accuracy and coordination are better correlated with speech precision ratings. For example, Neel (2008) suggested that vowel space measures provide little insight into vowel identification errors that occur when listening to healthy speech. Instead, a more detailed analysis of the distinctiveness of neighbouring vowels in the F1/F2 was required to account for listener errors. Indeed, as NZE has developed, some vowel phonemes have become completely superimposed in the F1/F2 space (Maclagan & Hay, 2007). In these cases, it appears that speakers' ability to produce vowels that are well distinguished by formant movement and duration could be particularly important (Maclagan & Hay, 2007). Future work may benefit from examining these measurements in neighbouring vowels amongst speakers with mild forms of dysarthria.

3.5.4.2 Generalization to other dialects

In recommending changes in measurement procedures, careful consideration must be given to the generalizability of this study's results. Despite substituting the NZE [oː] vowel in place of the /u/ phoneme (which has been traditionally used in the dysarthria literature), the average

midpoint VSA and FCR values in this study did not significantly differ from results presented in other large scale studies of US speakers with dysarthria (Lansford & Liss, 2014b; Sapir et al., 2010). For example, to compare the values obtained from the adapted FCR measure to values collected using the formula presented in Sapir et al. (2010), the average FCRs (generated from midpoint vowel formants) were examined. In the current study, speakers with dysarthria had an average FCR of 1.14 (*SD* = 0.12) while healthy speakers had an average FCR of 1.03 (*SD* = 0.07). These results lie directly between those reported by Lansford and Liss (2014b) and Sapir et al. (2010)—who both recruited speakers from a similar geographic region of the United States (US) and used the original FCR formula. The adapted FCR produced averages for NZE speakers (both with and without dysarthria) that were within one standard deviation of the values reported in these studies.

In comparison to previous literature, the only formant which consistently showed differences within our formulas was the F2 [oː]—which was lower than values reported for the US /u/ phoneme (Sapir et al., 2010; Turner et al., 1995). Evidently, there are considerable variations reported in VSA and FCR values, as well as differences in midpoint vowel formant values of the /a/, /i/ and /u/ phonemes, amongst healthy speakers of US English (Lansford & Liss, 2014b; Sapir et al., 2010; Turner et al., 1995). For this reason, it is difficult to determine how differences in the raw VSA and FCR values were influenced by the NZE dialect.

3.5.4.3 Effects of speaker sex

This study found large differences in the acoustic perceptual relationships demonstrated by male and female speakers. Considerably stronger correlations were produced amongst female speakers—indicating that changes in their acoustic measurements were more closely related to listeners' perceptions of dysarthria. Given the smaller number of female participants, random sample variation may have played a role in producing this result. However, this study is not the first to find a stronger link between perceptual and acoustic vowel measures amongst female speakers (Lansford & Liss, 2014a). It is possible that current methods of indexing vowel centralization might influence differences between the sexes. Both metrics amplify F1/F2 changes differently depending on the magnitude of speakers' formants, and this may account for some of the differences in the acoustic-perceptual relationship across the sexes. Examining the same set of speakers over time may help elucidate how the relationship between vowel centralization metrics and perceptual measurements is affected by differences in the magnitude of baseline formant values.

3.5.5 Summary

This chapter explored the ability of different acoustic and perceptual measures to index speakers' baseline dysarthria severity. The study found that perceptual ratings of speech precision were highly successful at distinguishing speakers with dysarthria from healthy controls—and were linked closely with objective measures of the speech signal. Vowel centralization measures taken from the articulatory point, and normalized using the Bark units and the FCR, produced strong correlations with these perceptual ratings. This acoustic metric had the added advantage of providing a scale of values that could be interpreted together, regardless of a person's sex. Taken together, the techniques suggested in this study were able to double the amount of variance accounted for in the acoustic-perceptual relationship amongst male speakers. In females, the amount of variance accounted for increased from 49 to 76%. This demonstrates that the procedures chosen when taking these measurements can have an important influence on a study's results. In future vowel space studies, it is recommended that researchers consider more flexible formant extraction points and different normalization procedures. Furthermore, it should be noted that perceptual measurements of dysarthria are not an inflexible standard, and the procedures we use to rate dysarthria should be carefully considered when any acoustic metrics are being assessed. Perceptual rating tasks that allow listeners to indicate when they hear impaired speech—even if the signal is intelligible—appear to be advantageous when indexing changes in mild speech disorder.

Having valid, objective, and reliable measurements of speakers' baseline severity is important in order to examine and interpret interspeaker variability in assessment, and in treatment programs. Specific to this dissertation, differences in baseline severity may explain some of the inter-speaker variations in intelligibility gains that occur in response to speech cueing strategies. Rate control techniques, in particular, have been posited to influence speakers differently depending on their dysarthria severity (Hammen et al., 1994; Pilon et al., 1998). For example, speakers with more severe dysarthria have been said to have more "potential" to increase their intelligibility in response to treatment cues (Hammen et al., 1994). It has also been posited that slow cued speech negatively impacts a speakers' "naturalness" and, for this reason, may be less effective in speakers who already have high intelligibility (Pilon et al., 1998).

Based on this investigation, perceptual ratings of speech precision and acoustic measurements of FCR (derived from a flexible extraction point and calculated in Bark) were selected to be used in the next chapter, to model speakers' responses to common treatment cues. The next chapter explores these baseline speech measures—in combination with established acoustic indices of voice quality, intonation, articulation rate and rhythm—to predict the degree to which speakers are able to improve their intelligibility following cues to speak louder and reduce their rate of speech.

Chapter Four

Predictors of Intelligibility Gain: A Treatment Simulation Study

Chapter four is an adaptation of the article titled "Predicting Intelligibility Gains in Individuals with Dysarthria from Baseline Speech Features", which is currently under review at the Journal of Speech, Language and Hearing Research. In some sections the text has been modified and additional information has been provided to ensure consistency and relevance to the current chapter and thesis.

4.1 PREFACE

Chapters two and three offered insight into the acoustics of NZE and the effect of slower speech rates on measures of vowel articulation. In addition, through the analyses completed in these chapters, we refined the techniques used to assess speakers' dysarthria severity in this thesis. As a result of the previous chapter's study, reliable measurements of participants' dysarthria severity were obtained—and these measurements proved to be sensitive to the presence of dysarthria (see section 3.3.4.2, section 3.3.6.4 and section 3.4.4).

The current chapter uses this information to address the central goal of the thesis: to predict treatment outcomes based on speech feature analyses. In order to model the responses of a large sample of speakers with dysarthria, this thesis focuses on treatment simulations—by examining the effects of two common therapist cueing strategies on listener ratings of intelligibility. As mentioned in chapter one, loud and slow cueing strategies have shown promising improvements in blinded listeners' understanding of dysarthric phrases (Hammen et al., 1994; Neel, 2009; Patel, 2002; Patel & Campellone, 2009; Pilon et al., 1998; Tjaden & Wilding, 2004; Turner et al., 1995; Van Nuffelen et al., 2010; Van Nuffelen et al., 2009; Yorkston et al., 1990). However, the level of intelligibility gain achieved is highly variable across participants, with many speakers unable to achieve significant improvements.

There are two central aims of this chapter. Firstly, cues to speak louder and reduce speech rate will be compared to examine their effect on speakers' intelligibility. Secondly, this study will assess whether measurements of speakers' baseline speech features can account for the variation observed in their intelligibility gains. In examining these intelligibility outcomes, we seek to model whether baseline perceptual and acoustic speech measures can help to distinguish speakers who achieved better outcomes with one treatment strategy over the other.

4.2 INTRODUCTION

As discussed in chapter one, for speakers with dysarthria, increasing intelligibility is central to improving communicative participation and is therefore a common goal of speech therapy. Several treatment studies have reported improved intelligibility by training speakers to use behavioural speech modifications (e.g. Cannito et al., 2012; Wenke et al., 2011). The two primary forms of behavioural speech modifications enacted in these studies—and, reportedly, in the clinical remediation of dysarthria—are changes to speakers' vocal loudness and speech

rate (Miller, Deane, Jones, Noble, & Gibb, 2011). As reviewed in section 1.3, loud and slow cued speech strategies are the foundation of many well-established treatment programs (i.e. LSVT) and approaches (e.g. pacing boards, delayed auditory feedback). However, the success of these techniques is unclear. Not all speakers with dysarthria have benefited from programs that use these strategies (Cannito et al., 2012). Partly as a result of this variation in individual response, treatment studies often fail to demonstrate intelligibility improvements across speaker groups (e.g. Lowit et al., 2010; Mahler & Ramig, 2012).

To predict whether a client will benefit from a larger program of treatment, it is important to first determine that they are able to change their speech in response to cues, and that these changes confer greater intelligibility. As reviewed in the first chapter, cueing strategies applied directly to dysarthric speech often produce significant improvements in listeners' ratings and understanding of dysarthric phrases (e.g. Hammen et al., 1994; Neel, 2009; Yorkston et al., 1990) However, as discussed in section 1.3.2, even amongst participants with the same aetiology and dysarthria subtype, there has been considerable variation in treatment effects observed across studies—and not all individuals have responded positively to loud and slow treatment cues. As with more complex programs of treatment, this variation across participants can prevent group outcomes from reaching statistical significance (Pilon et al., 1998; Van Nuffelen et al., 2010).

Section 1.3.2 reviewed a number of studies which have reported intelligibility gains following cues to speak louder and control rate of speech. Two issues were identified: (1) there has been limited investigation of the effect of loud and slow speech cues on intelligibility in speakers with non-Parkinsonian/hypokinetic dysarthria and (2) the evidence for the use of particular treatment strategies based on dysarthria subtype is limited (i.e., if a person has hypokinetic dysarthria, cues to increase loudness do not always result in improved intelligibility). Hence, more detailed profiles of the characteristics of speakers who benefit from different treatment strategies are required.

4.2.1 Issues with Categorizing Speakers who have Dysarthria

Based on the evidence presented, there is no 'one-size-fits-all' strategy for improving intelligibility in dysarthria. Thus, to determine the suitability of treatment strategies on an individual, speaker-specific level, it is important to attempt to understand reasons for speakers' variable outcome measures. At present, treatment studies commonly use neurologic aetiology and the Mayo classification system to group participants with dysarthria prior to

intervention (e.g. Cannito et al., 2012; Lowit et al., 2010). As homogeneity of speech characteristics is assumed within each dysarthria classification, detailed examination of the perceptual or acoustic features of participants' baseline speech patterns is not usually completed. But this approach may be problematic. As discussed in the first chapter, it has long been presumed that certain patterns of muscle disorder will contribute to similar speech symptoms, making it easier to generalize speech treatments to a group with "the same" type of neurological impairment. However, the rationale behind inferring different speech symptoms based on changes in the strength or steadiness of muscles in non-speech tasks has been heavily challenged (see Weismer, 2006 for discussion). At present, there is little speech production data to suggest that people who share a subtype have closely aligned speech features (Y. Kim et al., 2011). There has also been poor reliability evidenced between the clinical diagnosis of dysarthria and perceptual classification of professionals blinded to its aetiology (Fonville et al., 2008; Van der Graaff et al., 2009). These findings suggest that people with the same dysarthria subtype may not necessarily share distinct and recognizable speech symptoms. If knowing a person's dysarthria subtype does not offer reliable information about the type—let alone the severity—of various disordered speech features, we are limited in what we can infer from studies which group together speakers based on their dysarthria subtype or neurological diagnosis. This leaves us with limited evidence of exactly which diagnostic speech features are most important for selecting treatments (e.g., does the presence of speech feature A mean that a loud speech strategy may facilitate greatest perceptual improvements?). In order to determine the appropriateness of a treatment technique for any given individual, it seems possible that we should utilize more information about the unique combinations of features that occur in speech output when making treatment decisions. In approaching treatment decisions in this way, we disregard any assumptions about the speech signal that are based purely on neurologic aetiology.

4.2.2 The Relationship between Speech Features and Treatment Outcomes

In discussions of dysarthria treatment strategies, it is commonly presumed that certain strategies are more appropriate for some dysarthria subtypes than others (see Duffy, 2013, pages 421-423). For example, speech treatments which focus on getting speakers to increase loudness have been suggested to be beneficial for people with hypokinetic dysarthria (e.g. Duffy, 2013; Fox et al., 2002; Fox et al., 2006). However, the reasons that researchers choose

to focus on loud cued speech in this population usually relate to the speech features that are thought to be associated with the subtype—rather than the underlying neurologic aetiology. For example, the LSVT literature suggests that loud speech is most beneficial for speakers exhibiting reduced respiratory and laryngeal drive, with speech that presents as quiet and breathy, with less variation in pitch and amplitude (Fox et al., 2002). Because these speech features are considered hallmark features of hypokinetic dysarthria, we consider loud cued speech to be a good treatment option for this population. However, to this author's knowledge, correlations between measurements of baseline speech symptoms and treatment gains following LSVT have never been reported. Therefore, we do not have clear evidence to suggest which speech characteristics are indicative of speakers who will benefit from loud cued speech—and, importantly, which symptoms may act as contraindications of this treatment.

There is additional evidence for this call to look beyond broad subtype categories—and devise ways of providing more individually tailored recommendations for speech therapy. Using loudness as a strategy to improve intelligibility has not traditionally been recommended for speakers with spastic dysarthria as it is thought to exacerbate hyperadduction of the vocal folds (see Duffy, 2013, pages 421-423). However, there is case study evidence to show that strategies focused on increasing loudness can increase the intelligibility of speakers diagnosed with dysarthrias containing some spastic components (e.g. D'Innocenzo, Tjaden, & Greenman, 2006).

In addition, it seems that loud cued speech is not necessarily superior to rate reduction techniques for *all* speakers with PD. For example, while Tjaden and Wilding (2004) put forth results suggesting loud speech exacts greater intelligibility than slow cued speech, McAuliffe et al. (2014) present contrasting results. Indeed, McAuliffe et al. (2014) found that when speakers with PD reduced speech rate, they produced a significantly higher proportion of correct listener responses, as compared with increased vocal loudness. These data suggest that different groups of speakers with PD—and associated hypokinetic dysarthria—may benefit from different treatment strategies. To help understand why these differences occur, we need more information about the unique characteristics of participants in these studies.

In contrast to loud cued speech, rate control techniques have traditionally been recommended for speakers across dysarthria subtypes (Duffy, 2013, pages 421-423) Indeed, in the case of slow cued speech, there is evidence to suggest that a person's dysarthria subtype will provide limited information about whether a rate modification technique will be beneficial. For example, Van Nuffelen et al. (2010), examined rate control techniques across

six dysarthria subtypes (including unspecified mixed dysarthrias). Many of the speakers they examined did not increase intelligibility in response to the slow speech cues—and there was no indication that speakers of any one subtype were more likely to increase their intelligibility than speakers of another.

As mentioned in the summary of chapter three, it has been speculated that speakers' baseline severity (as opposed to their dysarthria subtype) may affect their response to rate control techniques (Hammen et al., 1994; Pilon et al., 1998). Speakers with more severe dysarthria have been said to have more "potential" to increase their intelligibility in response to treatment cues. That is to say, if speakers exhibit severely reduced intelligibility in their baseline speech, there will be no ceiling effect to the treatment gains that can be made (Hammen et al., 1994). In addition, it has been posited that slow cued speech might negatively impact a speaker's "naturalness" and, for this reason, may be less effective in speakers who already have high intelligibility (Pilon et al., 1998). However, evidence supporting improved treatment outcomes in speakers with more severe dysarthria has been inconsistent (Van Nuffelen et al., 2009). It seems likely that the effect of increased intelligibility gains as dysarthria severity increases will depend somewhat on the treatment strategy being tested. For example, there is evidence that certain treatment strategies are more appropriate for speakers with severe dysarthria—while others promote greater gains in speakers with moderate dysarthria (Hunter et al., 1991). Thus, while we suspect baseline severity will affect speakers' responses to treatment, it is unclear in exactly what direction these effects might occur.

4.2.3 Summary

In summary, there is little evidence that grouping participants by the Mayo System or their dysarthria aetiology provides any real insight into whether individuals will benefit from speech modification strategies, or why certain participants achieve best results in one treatment condition over another. It is hypothesized that, in order to identify the types of speakers who will achieve success with certain behavioural strategies (in addition to the types of speakers who will not), we need a deeper understanding of these participants' baseline speech features. Ideally, these features should be measurable across participants with dysarthria, so that we can make predictions about whether a treatment strategy is appropriate for any speaker—regardless of their dysarthria aetiology and subtype.

The current study explores whether various acoustic and perceptual features of participants' baseline (i.e., recorded in a 'habitual' speech condition) speech are able to predict their responses to treatment techniques. There were three purposes of this investigation. The first was to compare how cues to speak louder and reduce speech rate affected speakers' intelligibility. The second was to determine whether features of speakers' baseline speech were able to account for the variation observed in their intelligibility gains. The third aim was to investigate the speakers who improved their intelligibility in response to treatment cues, and to model which of the two treatment strategies was most appropriate for each person.

In examining speakers' baseline speech, particular attention was given to perceptual ratings of speakers' speech severity, as this was hypothesized to be an important determinant of treatment outcome. Acoustic analyses were conducted to gather discrete measurements of vowel articulation, rhythm, speech rate, intonation, and voice quality. This information was used determine which features of speech, or clusters of features, best served to predict whether an individual speaker would benefit from reduced rate or increased loudness as intervention strategies.

4.3 METHOD

4.3.1 Speakers

Fifty speakers of New Zealand English (NZE) (35 males and 15 females), aged between 43 and 89 years, contributed speech recordings for this study. Of these speakers, 43 were diagnosed with dysarthria ranging from mild through to severe. Dysarthria subtypes and severity were determined via consensus judgements of the first three authors. The remaining seven speakers, who reported no history neurological impairment, acted as healthy controls. The group diagnosed with dysarthria had a mean age of 65 years, while the control group also had a mean age of 67 years. Table 3.1 in chapter three provides biographical information about the speakers with dysarthria who participated. However, it should be noted that this table—associated with the previous study—contains a total of 44 speakers with dysarthria (rather than 43). The discrepancy results from the removal of one speaker from the current study. The speaker was a 55 year old male with a moderate dysarthria resulting from an undetermined neurological disease. The speaker was removed because the passage readings they produced in the loud and slow treatment conditions contained too many reading errors and speech interruptions to provide a set of comparable speech samples across conditions.

4.3.2 Speech Stimuli

Speakers attended a single recording session. Recordings took place in a quiet room, with a single investigator present. Digital audio recordings were made via an Audix HT2 headset condenser microphone, positioned approximately five centimetres from the mouth, at a sampling rate of 48 kHz with 16 bits of quantization. These procedures are the same as outlined in the previous chapter, however, in this study, the three female control speakers who were recorded with a different microphone were not included.

The Grandfather Passage was used to elicit a sample of participants' baseline speech, as well as samples simulating two common treatment strategies (see Appendix A). For the baseline condition, speakers were asked to read the passage in their everyday speaking voice, after they had familiarized themselves with passage. As mentioned in the previous chapter, two participants with dysarthria required assistance reading the passage. In these instances, the first author would read full sentences from the passage, with the speaker repeating the sentences immediately afterwards. To create the treatment simulations, a magnitude scaling procedure was used to elicit louder and slower speech. For the slow condition, speakers were asked to say each phrase at "what feels like half your normal speed". While, for the loud condition, speakers were asked to read each phrase "at a level that feels like twice as loud as normal" (Tjaden & Wilding, 2004). The order of the simulations was randomized across speakers.

4.3.3 Procedures

This study was conducted in two parts: (1) a perception experiment where listeners rated the intelligibility of speakers' baseline, loud and slow speech recordings, (2) a perceptual and acoustic analysis of the speakers' baseline speech features. These are described below.

4.3.4 Step One: Determining Intelligibility Gains

Intelligibility gains in response to loud and slow speaking cues were measured using a listener-rating task. Twenty-five listeners made judgments of intelligibility along a visual analogue scale. All listeners were between the ages of 18 and 35 and all passed a hearing screening before beginning the experiment. In each trial, the listeners were presented with three linguistically-matched phrases from one of the study's speakers (one baseline, one loud, one slow). All phrases were between 11-14 syllables, and the first and last sentences of the reading passage were not included. Each of the phrases was represented onscreen graphically

with an icon (e.g., a loud phrase was represented by a triangle, a slow phrase by a circle). The recordings were all scaled to the same average dB SPL, and presented at the same volume throughout the experiment. Listeners were prompted to "rate how easy you find the phrase to understand" and then place the specific phrase's icon at a point along the scale (i.e., if the slow condition was easier to understand than the baseline condition, the slow icon would be deposited further along the scale). An example of the visual presentation is shown in Figure 4.1.

To measure intelligibility gain, the average distance that the loud and slow tokens were placed from the baseline speech token was calculated for each listener. This average was used to compute a z score for the amount of change (in either the negative or positive direction) that occurred as a result of the two treatment conditions. For example, the z score for listener one's rating of speaker one in the loud condition would be derived in the following manner: Firstly, the absolute value of the difference between the listener's rating of speaker one's baseline speech and loud cued speech would be calculated. Then, the average difference in ratings that the listener gave to all speakers' baseline speech and their loud and slow cued speech would be subtracted from this value. The resultant number would then be divided by the standard deviation of the difference in ratings that the listener gave to speakers' baseline speech and cued speech. This creates a z score value.

By definition, the z score is lowest when the listener gives the baseline and cued speech the same rating. Therefore, when we extract the difference between the lowest possible z score and all other z score ratings given by a listener, we end up with a series of positive scores (when the baseline and treatment tokens are rated differently), and scores of 0 (when they are rated the same). The final ratings given to each speaker can then be made positive or negative depending on the direction of the change. After completing these processes, ratings from the 18 listeners were averaged to give one 'group average' rating per speaker.

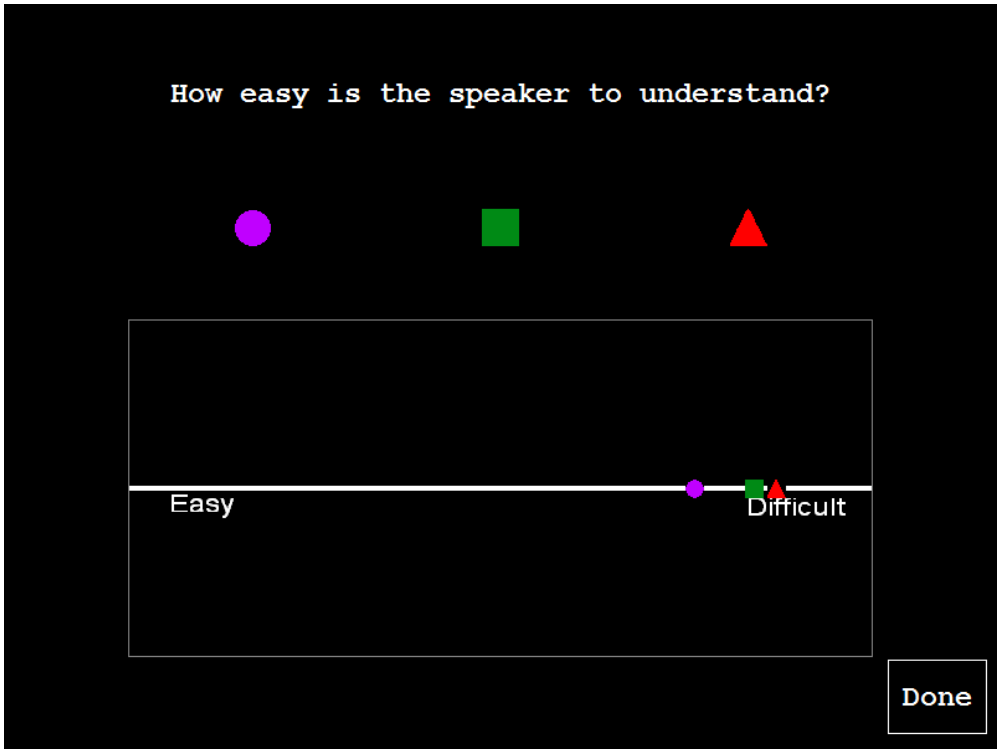

Figure 4.1. Visual analogue scale presented to listeners in the experiment to determine intelligibility gains.

4.3.4.1 Reliability of measurements of intelligibility gain

Because the perceptual experiment required sustained focus on the part of the listener, checks were included to ensure that participants were consistent in their ratings throughout the experiment. Trials from each speaker were presented twice for this purpose, at random stages throughout the experiment. Only listeners who placed the baseline and treatment tokens in the same order in over 65% of the repeated trials were included in the final results. After removing the less reliable participants, eighteen listeners remained. On average, the eighteen listeners were consistent in the order they placed the tokens in 77% of repeated trials. Inter-rater reliability was assessed in a similar way by examining the proportion of listeners who placed the baseline and treatment tokens in the same order within trials (cases where listeners gave two samples identical ratings were excluded). Naturally, the proportion of listeners who agreed on the "most intelligible" sample varied considerably across speakers. For example, when a speaker did not produce noticeable changes in the treatment conditions, listeners tended to be less consistent in their preferences. For this reason, listeners' agreement ranged from between 50% (indicating no consensus) to 100% (indicating total agreement about

which sample was most intelligible). On average, across trials, 76% of listeners' agreed in their preferences for the treatment tokens vs. baseline speech.

4.3.5 Step Two: Analyses of Baseline Speech Features

4.3.5.1 Perception task: Speech severity.

To measure speakers' baseline speech severity, a second perceptual experiment was completed. In this task, a separate group of 14 listeners were asked to make judgments of speakers' baseline speech precision along a visual analogue scale. These listeners and speech stimuli are described in section 3.3.6.1 and 3.3.6.2. The decision to use the 'speech precision' ratings to index baseline severity, as opposed to the 'ease of understanding' ratings was made based on the results of the previous chapter (which indicated that these values were more sensitive to mild dysarthria). As described in the previous chapter, listeners were exposed to an identical phrase for each speaker. They rated one speaker at a time, with each trial repeated once for reliability purposes. The raw ratings were z-scored for each listener before being averaged. Further details of this protocol are outlined in section 3.3.6.3. The ICC (2,1) coefficient used to assess the inter-rater reliability of the raw ratings was .835, and the average intra-rater correlation (across ratings of the same speech samples) was $r(852) = .89$, $p < .001$.

4.3.5.2 Acoustic analysis

To complete the acoustic analysis, each speakers' baseline recording of the grandfather passage was transcribed and then automatically segmented at the phoneme level using the Hidden Markov Model Toolkit (Young et al., 2002). All automatically derived phoneme boundaries were then visually checked for accuracy by a team of trained analyzers, using standard criteria (Peterson & Lehiste, 1960). If any uncertainty arose in discriminating boundaries for consecutive phonemes (e.g., /t/ and /s/) the boundary selected through the automatic segmentation was retained. For further information about this process see (chapter one or Fletcher et al., 2015).

After manual checking, seven acoustic metrics were extracted from across each speaker's baseline speech recordings. To index articulation and speech prosody, measurements of articulation rates, the pairwise variability index of vowels (vPVI), formant centralization ratios, and the standard deviations of speakers' fundamental frequency (F0) and intensity were extracted. These measurements are similar to those examined in Kim et

al. (2011), and were selected because they were reported to be useful either for predicting intelligibility deficits or for distinguishing amongst different types of dysarthria. For example, certain dysarthria subtypes have been reported to have quantifiable differences in their articulation rates and vPVI scores (Liss et al., 2009)[6], while variations in speakers' F0 and intensity are thought to differ based on their dysarthria severity (Bunton et al., 2000; Metter & Hanson, 1986; Schlenck, Bettrich, & Willmes, 1993). In addition to these measures, two indices of voice quality were also included: smoothed cepstral peak prominence (Hillenbrand & Houde, 1996) and the amplitude of the first harmonic. These measurements were included because aspects of voice quality are thought to differ considerably amongst speakers with dysarthria (Darley, Aronson, & Brown, 1975), and preliminary research indicates that these measurements may be able to index differences in listeners' perceptions of breathiness and strain (Cannito, Buder, & Chorna, 2005; Hillenbrand & Houde, 1996). In summary, the seven acoustic measures were chosen in order to gain an objective account of differences in the speech signal across speakers. The following list details how these measurements were extracted and calculated:

1) Mean articulation rate was calculated in syllables per second using a custom praat script. Pauses over 50 milliseconds were excluded from the calculation (Robb et al., 2004).

2) Normalized Pairwise Variability Index (PVI) for vowel duration, (as described in Liss et al., 2009), was extracted using a custom praat script.

3) Pitch variation (fundamental frequency standard deviation in Hz taken from across the passage) was extracted using a custom praat script

4) Intensity variation (intensity standard deviation in dB taken from across the passage) was extracted using a custom praat script

5) Vowel centralization was calculated using formants from three tokens of each vowel that were manually measured in bark. A formant centralization ratio (FCR) was calculated for each speaker, adapted from Sapir et al. (2010) and calculated for NZE using the formula: (F2[oː] + F2[ɐː] + F1[iː] + F1 [oː]) ÷ (F2[iː] + F1[ɐː]) (see the previous chapter or Fletcher, McAuliffe, Lansford, & Liss, in press for further details of this protocol). Formant values were extracted from the articulatory points described in section 3.3.3.4.

[6] Although see Lowit (2014) for evidence of variability in vowel PVI measures amongst speakers with hypokinetic dysarthria.

6) Smoothed cepstral peak prominence (CPPS). The procedure for attaining this measure is described in Hillenbrand and Houde (1996). However, because the reading passage had been phonemically segmented, only vowel sounds were selected for the voicing analyses. A Hanning-window was applied to the 60 ms segment at the temporal center of each vowel in the passage, with shorter vowel segments excluded from the analyses.

7) First harmonic amplitude (H1A). Again, the procedure for attaining this measure is described in Hillenbrand and Houde (1996). The amplitude of the first harmonic is considered relative to the second (i.e. the amplitude of the second harmonic is subtracted from the first). In contrast to Hillenbrand and Houde (1996), the analyses only included vowel sounds, as described in the CPPS measure.

To determine the reliability of these measures, 10% of the phonemically segmented textgrids were manually re-examined. Phoneme boundaries were manually re-segmented by a different rater and scripts to obtain the acoustic metrics were then re-administered. Overall, there was an average correlation of .94 with the original measurements.

4.4 RESULTS

4.4.1 Effect of Loud and Slow Cued Speech on Intelligibility

Speakers' intelligibility changes in response to different treatment simulations are summarized in Figure 4.2. Analysis revealed considerable variation in the magnitude of speakers' intelligibility gains in the loud and slow conditions. However, speakers who made intelligibility gains in the one condition also tended to demonstrate intelligibility gains in the other. Hence, improvement in the two conditions was linked ($r(48) = .55$, $p < .001$). Overall, 35 speakers showed some degree of improvement using one or more of the treatment strategies. In contrast, the baseline speech of the remaining fifteen speakers was rated as more intelligible than either treatment condition. Amongst the group that did not benefit from treatment cues, healthy control speakers were disproportionally represented. Five out of seven healthy controls had baseline speech samples that were rated as the most intelligible. In contrast, only 10/43 of the speakers with dysarthria were rated higher in the baseline condition.

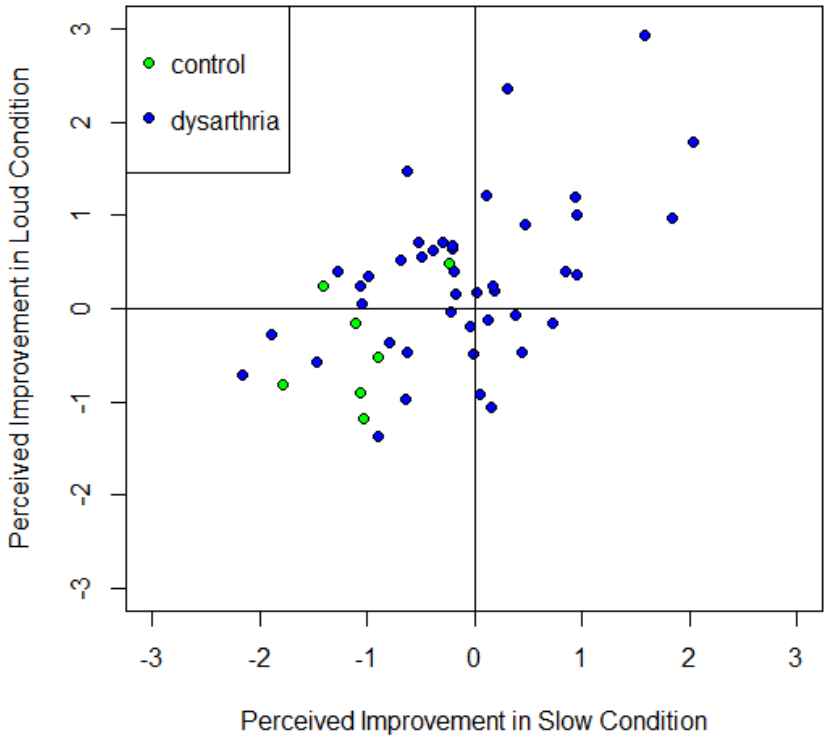

Figure 4.2. The relationship between speakers' intelligibility gain indices in the loud and slow cueing conditions. The units on the X and Y axes depict the distance, in standard deviations, between listeners' ratings of the cued speech and their ratings of the baseline speech samples.

4.4.2 Acoustic Measures of Dysarthric Speech

Table 4.1 reports the average acoustic measurements for both speakers with dysarthria and the healthy, older speakers included in this study. Average values for the FCR metric and the perceptual ratings of baseline speech precision were previously reported in Table 3.4, so are not repeated in this table. It is interesting that the measurements reported in Table 4.1 show much less distinction between speakers with dysarthria and healthy control speakers than the vowel dispersion measurements explored in the previous chapter. Indeed, there were no statistically significant differences between the two groups in any of the six measures reported (with all two sample *t* tests demonstrating $p>.05$).

Table 4.1

Average Values of Acoustic Measures Across Speakers with Dysarthria and Healthy Controls

Acoustic Measure	Healthy Controls	Dysarthria
Articulation Rate (syllables per second)	4.094 *(0.53)*	3.685 *(1.01)*
Vocalic Pairwise Variability Index	62.113 *(6.23)*	59.578 *(7.37)*
Standard Deviation of Fundamental Frequency (Hz)	27.634 *(11.03)*	26.280 *(12.76)*
Standard Deviation of Speech Intensity (dB)	16.190 *(1.15)*	15.905 *(2.56)*
First harmonic amplitude - second (dB)	-0.232 *(4.24)*	0.555 *(4.68)*
Cepstral Peak Prominence (dB)	16.657 *(1.01)*	17.154 *(2.79)*

Note. Standard deviations across groups are provided in parenthesis.

4.4.3 Predicting Intelligibility Improvement

A series of linear regression models were used to analyse the effect of speakers' baseline speech features on their intelligibility improvement in the two treatment conditions. The aim of these models was to better characterize both the types of speakers who made large gains in response to treatment cues, as well as those who did not. For this reason, models one and two included all speakers in the dataset. Before the models were run, speakers' acoustic and perceptual features were investigated for sources of multicollinearity. Correlations between different speech features ranged from between .06 to .64. While there were many statistically significant correlations, none were high enough to raise concern (see Table 4.2 for more detail). Because of the large set of acoustic variables, backward stepwise regression was conducted to identify a subset of speech features that were predictive of speakers' intelligibility improvement. Model selection proceeded in a backward-stepwise iterative fashion seeking to create a predictive model which contained only significant effects (with alpha set at 0.05). This process resulted in the creation of two models: one to predict improvement in the loud condition, and one to predict improvement in the slow condition.

Table 4.2

Across Speaker Correlations between Targeted Acoustic Measurements

Feature	1	2	3	4	5	6	7	8
1) Perceptual Ratings (higher rating = less severe)	-							
2) Formant Centralization Ratio	**-.61****	-						
3) Articulation Rate	**.44****	-.21	-					
4) First Harmonic Amplitude	.18	-.06	**.64****	-				
5) Smoothed Cepstral Peak Prominence	-.23	.15	**-.56****	**-.47****	-			
6) Pitch Variation	-.10	-.02	**-.43****	**-.40****	.19	-		
7) Intensity Variation	-.15	.22	**-.49****	**-.38****	**.51****	**.36***	-	
8) Normalized Vocalic Pairwise Variability Index	**.31***	-.19	**.35***	.07	.08	-.22	-.06	-

Note. Bolded correlations are statistically significant, *p<.05, **p<.01

4.4.3.1 Model one: Intelligibility gains in response to cues to speak slower

Model one examined the degree that speakers changed their intelligibility in the slow condition relative to their baseline speech sample. Model one found that the level of intelligibility improvement made in the slow condition was best predicted by both speakers' baseline speech severity and their vocalic pairwise variability index. To compare the relative effect of these features, all regression coefficients were standardized. The final model included a main effect for listener ratings of speech precision [β = -0.59 (0.13), $p < .001$], indicating that speakers with more severe dysarthria produced significantly greater intelligibility improvements when cued to slow down. There was also a significant effect for measurements of vowel PVI [β = 0.33 (0.11), $p = .006$]. This suggests that speakers with a greater degree of temporal variability in their speech segments were able to better utilize slow speech as a strategy for increasing intelligibility. However, this relationship was only apparent once baseline speech severity was held constant. Overall, this model accounted for 34% of the variance in speakers' responses to treatment cues. Interactions between the variables were not significant.

4.4.3.2 Model two: Intelligibility gains in response to cues to speak louder

Model two examined the degree that speakers changed their intelligibility when cued to speak louder. Model two revealed that improvements in the loud condition were best predicted by speakers' articulation rate and their baseline speech severity. The final model had a main effect for articulation rate [β = 0.52 (0.12), $p < .001$], suggesting that speakers with a faster rate of speech produced significantly greater intelligibility improvements when cued to speak louder. There was also a significant effect for ratings of speech precision [β = -0.41 (0.13), p = .003], indicating that speakers with more severe dysarthria made greater intelligibility gains (though this relationship was only apparent once speakers' articulation rates were held constant). Again, all regression coefficients were standardized and interactions between variables were not significant. Overall, this model accounted for 31% of the variance in speakers' responses to treatment cues.

4.4.4 Choosing between Treatment Cues

Models one and two give us insight into speech features which can be used to determine the appropriateness of loud or slow treatment cues for any given person. However, Figure 4.2 shows that some speakers benefited considerably more from one speech modification as opposed to the other. Hence, the question remains: which characteristics of dysarthria can we use to identify speakers who will perform better with one strategy over another? To answer this, a subset of 35 speakers who demonstrated a positive change in intelligibility in response to either the loud or slow treatment strategy were examined. These participants were divided into two groups: 1) those who produced greater intelligibility gains in the loud condition (n=24), and 2) those who produced greater gains in the slow condition (n=11).

As previously discussed, intelligibility improvement was measured as an average of listeners' z scored ratings (with each unit representing one standard deviation of change between the baseline and treatment conditions). Participants in group 1 demonstrated an average improvement in the loud condition of 0.76 standard deviations. Participants in group 2 demonstrated an average improvement in the slow condition of 0.70 standard deviations. The two groups were coded separately and group membership was used as dependent variable for a series of binomial regression models. This modelling aimed to determine whether there were differences in the speech features of participants who did better in one treatment conditions as opposed to another. Again, model selection proceeded in a backward-

stepwise iterative fashion seeking to create a predictive model which contained only significant effects (with alpha set at 0.05).

4.4.4.1 Model three: Speakers' most successful strategy

Model three examined whether speakers made greater treatment gains in the loud or slow treatment condition. The model revealed that speakers' highest rated treatment strategy was best predicted by the baseline measurement of their first harmonic amplitude (H1A). The final model contained only one main effect for H1A [β = -1.36 (0.53), p = .01], suggesting that a speaker's baseline voice quality was the best determinant of whether one treatment condition would be more successful than another. As in previous models, the regression coefficient was standardized and there were no significant interactions. The odds ratio revealed that for each standard deviation increase in a speaker's H1A scores, they were 3.9 times more likely to perform better in the loud condition—as opposed to slow. In comparison to the null model, the inclusion of H1A significantly improved the fit of the model (χ^2 (1) = 9.13, p = .003).

4.5 Discussion

This chapter explored whether it was possible to predict speakers' responses to common treatment cues given their baseline speech characteristics. Previous studies that have explored the effect of these cues on speech intelligibility have generally grouped participants together based on their dysarthria aetiology or subtype. However, there is evidence to suggest that people who share a dysarthria aetiology or subtype may not necessarily have similar speech features (Y. Kim et al., 2011). Furthermore, it seems that a person's aetiology and subtype does not provide particularly useful information about the degree to which they will improve their speech intelligibility in response to treatment—or information to decide whether one strategy is more appropriate than another (McAuliffe et al., in press).

The current study investigated measurements of speakers' baseline speech features to see if there were other, more objective speech assessment data that could be used to make predictions about individuals' responses to treatment. The study included a range of speakers with dysarthria and analysed each of their baseline speech features in the same way—regardless of their underlying aetiology. There were three main aims of this study. The first was to compare how cues to speak louder and reduce speech rate changed speakers' intelligibility. The second was to determine whether features of speakers' baseline speech

were able to account for the variation observed in their intelligibility gains. The third aim was to model which of the two treatment strategies was most appropriate for each person. The results of each of these investigations will be discussed in turn.

4.5.1 Intelligibility Gains Following Cues to Speak Louder and Slow Rate of Speech

Across speakers, variable intelligibility gains were observed following cues to speak louder and reduce rate of speech. This variation in treatment response was consistent with findings from previous investigations of loud and slow cueing strategies, as reviewed in the first chapter (Hammen et al., 1994; Neel, 2009; Pilon et al., 1998; Tjaden & Wilding, 2004; Turner et al., 1995; Van Nuffelen et al., 2010; Van Nuffelen et al., 2009; Yorkston et al., 1990).

The level of variation in ratings of loud cued speech and slow cued speech was reasonably similar (as demonstrated in Figure 4.2). However, the slow cued speech samples were more likely to be perceived as causing a reduction in listeners' 'ease of understanding'. This finding may, in part, reflect the manner in which intelligibility gains were measured in the perceptual experiment. To be specific, the perceptual experiment asked listeners to rate how easy each speaker was to understand in order to index their 'intelligibility gains'. Consequently, the measurements taken were not an objective tally of the words each listener was able to correctly transcribe. Instead, this protocol allowed the listener to express their own preferences and biases for certain speech samples. It has been reported that listeners have a natural preference for speech samples that are of a similar —or slightly faster—rate than their own speech (Street, Brady, & Putman, 1983). Street et al. (1983) found that when a healthy person speaks in this range, they are rated as significantly more socially attractive and competent to listeners than when speaking at a slower rate. For this reason, it is likely that listeners are naturally predisposed to prefer faster speech samples in cases where the objective intelligibility of the two samples is similar.

It is interesting to note that while the majority of speakers with dysarthria were perceived to be somewhat easier to understand in at least one of the cued speech conditions, this was not the case for the healthy older speakers. This may indicate that when a speaker is relatively healthy (i.e., a control group member) and intelligible (as indicated by high ratings of baseline speech precision) listeners are less likely to view behavioural changes to the speech signal positively. The tendency for negative ratings of control speakers in the slow

condition indicate that this cueing strategy may be perceived as being less natural than loud cued speech or may require more listener effort in cases when the speech signal is relatively unimpaired.

It was also interesting to note that there was a significant correlation between the intelligibility gains made by speakers in each of the treatment conditions. This result was not particularly surprising—it appears that speakers who are able to use one strategy with great success are also more likely to be able to successfully employ other speech modification strategies. However, as most dysarthria treatment studies explore the effects of only one program at a time, we often do not consider whether a particular group would achieve similar results with a different speech therapy approach. The correlation between speakers' intelligibility gains following cues to speak louder and reduce speech rate indicate the importance of remaining open-minded in our approaches to treating dysarthria. Just because one strategy might be successful for a particular participant does not necessarily mean it is the only speech modification that will produce positive outcomes.

4.5.2 Predictors of Intelligibility Gain in the Slow Condition

In combination, baseline measurements of speech precision and temporal vowel variation were significant predictors of speakers' intelligibility improvement when prompted to speak slower. Our model demonstrated that speakers with more severe dysarthria (i.e. lower ratings of baseline speech precision) tended to make larger intelligibility gains. This result was not unexpected. Baseline severity has previously been hypothesized to affect speakers' intelligibility improvement in exactly this manner (Hammen et al., 1994; Pilon et al., 1998). Indeed, there have been several ideas posited as to why speakers with more severe dysarthria might benefit more from rate control strategies. For example, it has been suggested that intelligibility gains would exhibit a ceiling effect in speakers with highly intelligible baseline speech (Hammen et al., 1994). This would mean that—amongst speakers who improved their intelligibility—those with more severe dysarthria would have the ability to make larger differences to their ratings. It has also been suggested that slow cued speech could negatively impact the "naturalness" of a person's speech (Pilon et al., 1998). As discussed in the previous section, listeners have a tendency to prefer speech that is of a similar rate to their own. People with highly intelligible speech who significantly slow down their natural speaking rate may, for this reason, also be perceived as requiring more effort to understand.

When perceptual ratings of speech severity were held constant, speakers who had a larger PVI in the duration of their vocalic segments tended to make greater intelligibility improvements. A high normalized vocalic PVI occurs as a result of increased durational differences from one syllable to the next. Hence, it is thought to be associated with speech that has more variation in stress (Liss et al., 2009). Syllabic stress helps listeners' in their segmentation of the dysarthric speech signal (Liss, Spitzer, Caviness, Adler, & Edwards, 1998). Our model suggests that speakers who have very poor temporal differentiation of stressed and unstressed syllables may not be able to employ rate control strategies as effectively. It could be that when these speakers try to extend the duration of their speech sounds, they do not differentiate the length of their vowels. This may cause listeners to experience even more difficulty correctly detecting their stressed syllables than they would when the speech was faster. For this reason, a cue to slow down might further exacerbate the perception of 'unnatural' or 'robotic' speech.

4.5.3 Predictors of Intelligibility Gain in the Loud Condition

Speakers' intelligibility gains in the loud condition were best predicted by a model that included information about their articulation rate as well as a perceptual measure of their baseline speech severity. This model suggests that, when baseline severity is controlled for, people who speak at a faster articulation rate tend to exhibit larger intelligibility gains in the loud condition. Conversely, when baseline articulatory rates are held constant, speakers with more severe dysarthria tend to make greater intelligibility gains. Baseline speech severity predicted intelligibility improvement in a similar manner for both loud and slow cued speech (although the effect was stronger for slow cued speech). The reasons for this effect of severity are likely to be similar to those discussed in the previous section. For example, cues to speak loud may have a negative impact on perceived naturalness to some degree.

It is interesting that a faster articulatory rate was predictive of intelligibility gains in the loud condition but not the slow condition. One explanation is that speakers with a faster articulatory rate may exhibit a range of related characteristics that make loud speech an appropriate treatment strategy. For example, cues to speak loud have been theorised to specifically address breathiness by improving vocal fold adduction in speakers with dysarthria (Baumgartner et al., 2001). In the current study, measures of articulatory rate were strongly correlated with measures of acoustic voice quality (see Table 4.2). Specifically, H1A demonstrated a strong, statistically significant relationship with articulatory rate. Previous

studies have reported that H1A is positively correlated with the perception of breathiness (Hillenbrand & Houde, 1996), but is hypothesized to be negatively correlated with measures of creakiness, vocal strain and the perception of roughness (when breathiness is controlled for) (Cannito et al., 2005). The relationship between H1A and articulatory rate suggests that speakers with a slow rate of speech may exhibit a more strained speech quality, while speakers with a normal or increased rate might be more likely to exhibit breathiness. Indeed, it is possible that articulatory rate is more sensitive than our measurements of CPP and H1A to differences between 'breathy' vs 'strained' and 'effortful' speech. Hence, it is possible that positive relationship between speakers' articulatory rate and improvement in the loud condition may be related to a number covarying factors.

4.5.4 Comparing Treatment Cues in the Same Speakers

This final question investigated in this chapter was whether there were characteristics of dysarthria that could be used to identify speakers who performed better with one strategy over another. In this case, measures of H1A were a significant predictor of whether speakers would demonstrate greater success with cues to speak louder as opposed to cues to speak slower. There were no further variables that accounted for significant variation in this model. It is worth noting that this model attempted to predict a broad, binary outcome, and—in contrast to models one and two—had less speaker data available to train on. This may account, to some degree, for its simplicity. As discussed in the previous section, measures of H1A may be affected differently in speakers with a breathier voice quality and speakers who are perceived to sound tense or strained. From a theoretical point of view, it seems likely that loud cued speech would be more effective in breathy speakers with hypoadduction of the vocal folds. In contrast, data from the previous section suggests it may be contraindicated in speakers who have a very slow speech—who perhaps also exhibit an effortful or strained voice quality. This may account for the significance of H1A measure in distinguishing the speakers who are more likely to be successful with cues to speak loud.

4.5.5 Summary

In summary, this study found that features of speakers' baseline speech—including information about their dysarthria severity and acoustic measures of the dysarthric speech signal—were able to predict their level of success in response to different treatment strategies. As expected, features of the dysarthric speech signal were not able to account for

all the variation observed in speakers' intelligibility gains. It remains likely that factors related to participants' cognitive abilities, motivation and fatigue significantly affect their responses to the speech modification strategies, as discussed in section 1.4. However, the ability to account for around 1/3 of the variance in listeners' perceptions of their intelligibility gains has considerable clinical importance. Furthermore, the binomial model of speakers' most successful strategy revealed that changes to H1A made participants almost four times more likely to be more successful with one strategy as opposed to another. These preliminary data demonstrate new assessment methods that could be used to select and group participants for future treatment studies. Being able to more specifically target the types of speakers who are likely to make large intelligibility gains has the potential to promote much stronger group outcomes in these studies.

Data from this study also provide the beginnings of an evidence base for clinical decision making that can account for a wider variety of presenting dysarthrias. The assessment protocol used in this chapter can be applied to any speaker, regardless of their underlying aetiology. Indeed, if more speakers were added to this analysis, it is likely that the models would be able to incorporate even more baseline speech variables—and hence be able to account for increasingly individualised presentations of dysarthric speech. This could provide a pathway to more individually targeted and adaptive approaches to speech modification and motor learning in dysarthria therapy.

There are, however, several factors that currently limit the clinical utility of the assessment techniques we used to predict speakers' responses to speech modification strategies. For example, this study relied heavily on a perceptual measurement of baseline severity gathered from a large group of listeners. While similar perceptual measures are common in research studies, they are difficult to replicate from one study to another due to their subjective nature. They also require a large group of listeners to ensure their reliability and validity. This presents a challenge in a clinical setting. In addition, the acoustic features employed in this study were labour intensive, requiring detailed manual segmentation of the speech signal into its component phonemes. In an effort to address these issues, chapter five of this thesis explores the use of automated speech data extraction procedures. However, the targeted analysis of different perceptual speech characteristics remains very important. The features examined in this chapter have a clear perceptual interpretation. This 'interpretability' helps to provide a theoretical understanding of the strategies we use in clinical practice. As assessment protocols continue to change and develop, the ability to interpret our assessment data remains incredibly important. For this reason, targeted measurements of speech

features—as time-consuming as they may be—are pivotal in advancing our understanding of exactly what makes dysarthric speech sound disordered.

Chapter 5

A Follow-up Investigation using Automated Acoustic Analyses

5.1 Preface

The previous chapter found that speakers' responses to loud and slow cueing strategies could be predicted to some degree by features of their baseline speech. However, as discussed in the summary, the results of this study remain somewhat limited. If assessments of speech features are to be more commonly used in the selection of participants for future treatment studies, they will need to become more time efficient. In addition, evidence that these models are able to generalize to speakers that they have not been trained on is required. To explore whether these requirements are possible, this chapter introduces data gathered from automated acoustic analyses techniques. Unlike the acoustic measures used in chapter four, these techniques do not require prior segmentation of phonemes and can be applied to relatively short speech samples.

This study examines features generated from Long Term Average Spectra (LTAS), Envelope Modulation Spectra (EMS) and Mel-Frequency Cepstral Coefficients (MFCCs). The aim of the chapter is not only to provide alternate models of speaker's intelligibility gains—but also to demonstrate that the models presented in this thesis are generalizable to new groups of speakers. For this reason, the baseline speech features presented in the previous chapter are also compared within a cross-validation analyses. This analysis allows us to compare the performance of the more targeted features of severity, rate, rhythm, voice quality and vowel articulation that were used in chapter four against those generated through automatic speech analyses.

5.2 Introduction

As discussed throughout this thesis, there has been considerable variation reported concerning the degree that speakers with dysarthria benefit from cues to speak louder and reduce speech rate (Hammen et al., 1994; Neel, 2009; Patel, 2002; Patel & Campellone, 2009; Pilon et al., 1998; Tjaden & Wilding, 2004; Turner et al., 1995; Van Nuffelen et al., 2010; Van Nuffelen et al., 2009; Yorkston et al., 1990). The results presented in the previous chapter are consistent these reports (see section 4.1). However, this thesis demonstrates that information about participants' baseline speech can be used to make predictions about their response to different treatments cues (see sections 4.4.2.1 and 4.4.2.2). Indeed, in the case of both loud and slow speech, selected baseline speech features were able to account for around 1/3 of the variation in listeners' ratings of intelligibility gain. Given these findings, it is

probable that assessment data from other acoustic analyses may also be able to model aspects of this variation.

This chapter assesses the ability of statistical models to predict the intelligibility gains of speakers that they have not been trained on. One of the main barriers in applying models from the previous chapter to new groups speakers is the time needed to calculate reliable measures of articulatory rate, vowel PVI and overall severity. For example, there were considerable resources required to collect listener ratings of baseline speech precision in order to form a valid and reliable index of dysarthria severity (as discussed in chapter three). Furthermore, hundreds of hours were spent hand-checking the Praat phoneme segmentation prior to application of scripts to generate articulatory rate and vowel PVI measures. Automated acoustic analyses offer a promising alternative. The automated features examined in this chapter are able to be computed for a large numbers of speakers in a matter of minutes. They can also be replicated more easily, as they do not require researchers to make any judgements concerning phoneme or syllable boundaries. For these reasons, automated acoustic assessment presents a promising method of selecting participants for future treatment studies.

5.2.1 Automated Measurements of the Dysarthric Speech Signal

Recently, there has been considerable interest in the application of automated speech measurements as a method of gathering faster, less invasive assessments of speakers' disease progression (see Bayestehtashk, Asgari, Shafran, & McNames, 2015 for review). For example, Bayestehtashk et al. (2015) found automated acoustic assessments to be reasonably effective in modelling the overall severity of speakers' PD. They were able to produce a model that could account for 61% of the variance in speakers' scores on the motor subscale of the Unified Parkinson's Disease Rating Scale on cross-validation. Indeed, automated speech analysis is considered particularly advantageous in marking the progression of neurological diseases because it is able to be completed remotely without the presence of a clinician.

One of the most common sets of features used in automated speech analyses are measurements extracted from MFCCs. Broadly speaking, MFCCs are used to capture information about the spectral structure of speech over time—in a manner which approximates the way we perceive speech sounds. MFCCs provide the most widely used representations of the speech signal in automated speech recognition programs (Han, Chan,

Choy, & Pun, 2006). In the study of dysarthria, they are used with the aim of measuring subtle changes in the movement of articulators (Khan, Westin, & Dougherty, 2014). These measures have been suggested to be particularly effective in detecting the presence of Parkinson's disease, with statistical models showing an ability to correctly distinguish over 80% of speakers with Parkinson's disease from healthy controls (Bocklet, Nöth, Stemmer, Ruzickova, & Rusz, 2011; Bocklet, Steidl, Nöth, & Skodda, 2013).

Another promising tool in the automatic evaluation of dysarthria is the measurement of EMS. EMS represent modulations that occur in the amplitude of the speech signal. Slow rate modulations in amplitude can provide information about individual's articulatory rate as well as any sudden changes in loudness or interruptions to the speech signal. For example, vowels are usually marked by sections of the speech signal with a high amplitude, indicating the presence of a syllable nucleus. The speed at which changes in amplitude occur can therefore depict the rate that speakers produce syllables. Liss et al. (2010) explored measures derived from EMS that were taken from a range of frequency bands within the speech signal. They found that these features were 95% accurate in classifying speakers with dysarthria from healthy controls on cross validation. Furthermore, they demonstrated 67% accuracy in their ability to classify individual into five speaker groups. These groups included four different dysarthria subtypes, in addition to a group of healthy control speakers.

Indeed, the ability to separate speakers of different dysarthria subtypes was particularly notable, as this study is one of the few reports of quantifiable differences in the speech signal between Mayo system groups (Y. Kim et al., 2011). However, it should be noted that Liss et al. (2010) selected these speakers because they exhibited the "cardinal" perceptual features thought to be associated with their subtype. Hence, it is unclear whether these differences between subtypes would remain if a wider range of speakers with hypokinetic, hyperkinetic, ataxic and mixed dysarthrias were included. Nevertheless, findings from Liss et al. (2010) suggest that EMS measures may be particularly sensitive to perceptual differences between speakers with similar dysarthria severities.

Measurements of LTAS have also been used to analyse differences in individuals with dysarthria. LTAS provide a representation of the average spectral information contained in the speech signal across a relatively long period (i.e. they provide information about the spectral content across whole phrases, rather than within specific phonemes). Previous studies have found statistically significant differences in LTAS measures between people with PD and healthy controls (Dromey, 2003). Furthermore, Tjaden, Sussman, Liu, and

Wilding (2010) found some significant correlations between LTAS measures and perceptual ratings of severity—although these were variable across different dysarthria groups.

LTAS is also often used to analyse changes in voice quality. Improvements in voice quality can be demonstrated through a strengthening of lower frequency components of the LTAS and a weakening of upper frequency components (Cannito et al., 2005). For example, Tanner, Roy, Ash, and Buder (2005) observed that speakers with functional dysphonia had lower spectral means and standard deviations following behavioural therapy. The reduction in spectral mean accounted for 14% of the variance in ratings of perceived voice improvement. There is also evidence that LTAS measures can detect changes in nasality, with amplitudes around 250 Hz showing significant changes when speakers simulate hypernasality (de Boer & Bressmann, 2015).

5.2.2 Summary and Aims of the Current Study

In summary, methods of automated speech analyses demonstrate an ability to describe speech differences in dysarthria. The tools hold particular promise for clinical applications because of their ability to generate rapid, objective measurements. Unlike the acoustic features examined in chapters two, three and four, these automated analyses are unaffected by differences in the way researchers segment speech sounds. Indeed, these measures are imminently replicable across copies of the same sound file. Therefore, the current study seeks to utilize these methods in order to reproduce models of intelligibility gain similar to those established in the previous chapter.

This follow-up study had two primary aims. The first was to predict the intelligibility gain index scores of participants with dysarthria and healthy controls based on automatically generated acoustic measures. The models produced were evaluated based on their ability to generalize to speakers they were not trained on. The second aim of this chapter was to compare the performance of the models created using automated feature sets against models generated using the targeted acoustic variables examined in chapter four.

5.3 Method

5.3.1 Speakers and Speech Stimuli

Fifty speakers read a standard passage (7 healthy older individuals; 43 with dysarthria) in habitual, loud and slow speaking modes. The speakers investigated in this chapter were the

same as in the previous chapter. For further biographical details about these speakers see Table 3.1. Treatment simulations were elicited via magnitude scaling. All procedures were the same as reported in the previous chapter. For further details about the recording protocol see sections 4.3.1 and 4.3.2.

5.3.2 Perceptual Experiment to Determine Intelligibility Gain

Eighteen listeners rated intelligibility on a visual analogue scale. The listeners used in this chapter were the same group described in section 4.3.4.1 of the previous chapter. In each trial, they were presented with three phrases from one speaker (one baseline, one loud, one slow) and prompted to place corresponding icons along the scale. Two "intelligibility gain" indices were calculated for each speaker, one for change in the slow condition, and one for change in the loud condition. Again, the procedures for collecting listener ratings and calculating intelligibility gain were identical to those described in the previous chapter. For further details about the procedures used in the perceptual experiment, and the method for calculating of intelligibility gain, see section 4.3.4.

5.3.3 Baseline Speech Analysis

The following features were obtained via MATLAB scripts, using standard procedures (as previously reported in Berisha, Sandoval, Utianski, Liss, & Spanias, 2013; Jiao, Berisha, Tu, & Liss, 2015; Liss et al., 2010). The same phrases and recordings described in section 3.3.6.2 were analysed in order to obtain the EMS, LTAS and MFCC feature sets for each speaker. However, this chapter only examines the 50 speakers who participated in the previous study—rather than all 61 participants from the investigation in chapter three.

5.3.3.1 Mel-frequency cepstral coefficients

MFCCs are coefficients that collectively describe the shape of a Mel frequency cepstrum. Unlike the cepstrum used in the previous chapter (to calculate cepstral peak prominence), the Mel frequency cepstrum was created from speech frequency bands equally spaced along the Mel scale. The Mel scale is used in order to better approximate the way humans hear sound (with similarities to the Bark scale discussed in chapter three). In this study, the MCFFs were calculated using a filter bank approach, where the speech signal was filtered into 39 frequency bands distributed evenly along the Mel scale. Within the 39 MFCCs calculated

from these bands, six different statistics were computed: 1) mean, 2) standard deviation, 3) range, 4) pairwise variability, 5) skew, and 6) kurtosis. This resulted in 234 MFCC features.

5.3.3.2 Envelope modulation

EMS depicts the variations that occur in the amplitude of individual's speech signals—as well as in selected frequency bands. Before obtaining the EMS, recordings of the 50 speakers were filtered into nine octave bands with centre frequencies of 30, 60, 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. Amplitude envelopes were taken from the nine bands as well as the full speech signal. Low-pass filters with a cut-off of 30 Hz were then applied to the amplitude envelopes, to capture the slower changes in amplitude that occur across words and phrases. Fourier analyses were used to quantify the temporal modulations in the signal. Six EMS metrics were computed for each of the 9 bands and the full signal: 1) Peak frequency, 2) Peak amplitude, 3) Energy in the spectrum from 3-6 Hz, 4) Energy in spectrum from 0-4 Hz, 5) Energy in spectrum from 4-10 Hz, and 6) Energy ratio between 0-4 Hz band and 4-10 Hz band. This resulted in 60 EMS features.

5.3.3.3 Long-term average spectra

LTAS provide a representation of the average spectral information within speakers' phrases. Similarly to EMS, the speech signal was passed through an octave filter, breaking it into nine separate bands. The centre frequencies of these bands was 1.6, 63.1, 125.9, 251, 501, 1000, 1995, 3981, and 7943 Hz. For each of the nine octave bands, and the full signal, we extracted: 1) the normalized average RMS energy, 2) the RMS energy range, 3) the normalized RMS energy range, 4) skew, 5) kurtosis, as well as the standard deviation of RMS energy normalized relative to both 6) the total RMS energy, and 7) the RMS energy in each band (not applicable in the analysis of the full signal). At this point, the data was framed using a 20ms rectangular window with no overlap in order to calculate: 8) the pairwise variability of RMS energy between successive frames, 9) the mean of the framed RMS energies, and 10) the normalized mean of the framed RMS energies. This produced 99 LTAS features.

5.3.4 Statistical Analyses

To predict speakers' intelligibility gains using their automated baseline speech features, two regression models were developed. Model one predicted the degree that speakers changed

their intelligibility in the slow condition relative to their baseline speech sample. Model two predicted the degree that speakers changed their intelligibility in the loud condition relative to their baseline speech sample. The models' predictive power was assessed by determining the correlations between the automated acoustic features chosen by the model and the two intelligibility gain measures.

The large number of features extracted from speakers' baseline speech (a combined total of 393 features per speaker) meant that standard stepwise regression methods needed to be applied with caution. For example, a forward stepwise regression (with alpha set to $p = .05$) would likely continue adding variables until it overfit the perceptual data. Hence, a cross validation procedure was applied to determine the total number of features to be included in the model. Cross validation was achieved by training models on 90% of the speakers, and testing their predictive power on the remaining 10% (test speakers). Model training was completed ten times, using a different set of test speakers each time. The predictive power was averaged across the ten repetitions. Cross validation was conducted each time a new variable was added to the forward stepwise regression. The results of this procedure were used to determine at which point in the forward regression the predictive power of the cross validated model was highest.

We also compared the cross-validated accuracy of models built using the automated acoustic metrics against models built using the more targeted dysarthria measurements reported in the previous chapter. As described in section 4.3.5, these targeted measures included: a perceptual rating of speech severity, as well as acoustic measures of speakers' articulation rates, formant centralization ratios, the amplitudes of their first harmonics relative to second, cepstral peak prominences, the standard deviation of their pitch and amplitude, and the pairwise variability index of their vowels and consonants.

5.4 Results

5.4.1 Intelligibility Gain in Response to Cueing

As previously reported in section 4.4.1, there was considerable variation in speakers' intelligibility gains. Thirty-five speakers showed some degree of improvement using one or more of the treatment strategies. The baseline speech of the remaining 15 speakers was rated as more intelligible than either treatment condition. Figure 4.2 depicts speakers' intelligibility gain scores following cues to speak louder and reduce rate of speech.

5.4.2 Stepwise Regression and Cross-Validation on Untrained Speakers

Figures 5.1 and 5.2 detail the relationship between participants' baseline speech features and their intelligibility gains in the loud and slow speaking conditions. This relationship is evaluated at each step in the forward regression model building process. As described in section 5.3.4, the relationships in these figures demonstrate the performance of the models on speakers that they have not been exposed to or trained on.

Both the automated assessment data and the targeted acoustic metrics were able to predict the intelligibility gains of speakers they had not been trained on. As demonstrated in figures 5.1 and 5.2, at their peak, the automated assessment data were able to account for 25% of the variance in speakers' intelligibility gains in response to the loud cue (i.e. a .5 correlation between the independent variables in the regression equation and the measurement of intelligibility gain gives an R^2 value of .25), and over 10% of the variance in their responses to the slow cue. The targeted baseline measures presented in chapter four were able to account for over 19% of the variance in speakers' responses to the loud cue and almost 17% of their variance in response to the slow cue. Figures 5.1 and 5.2 demonstrate that, in both speech conditions, models constructed from the automated feature set decrease in their predictive performance once more than two variables are added. This indicated that it was appropriate to stop the forward regression after two steps, to avoid overfitting the model. The targeted baseline measures did not demonstrate the same reduction in performance. No matter which set of training speakers was used, these models always stopped adding targeted measures to the regression after three steps. Interestingly, when models of intelligibility gains in the slow condition were constructed with these measures, baseline severity was always added to the regression first, regardless of the speaker group it was trained on.

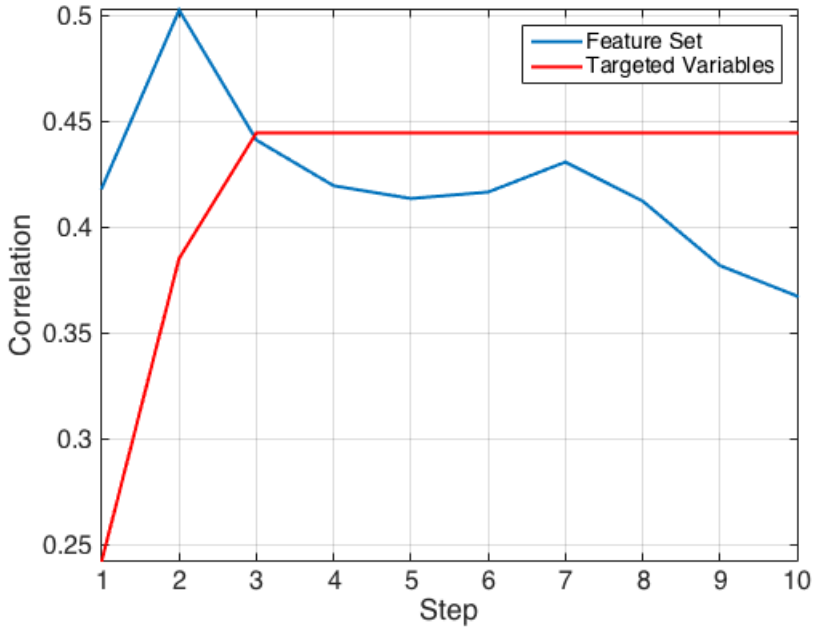

Figure 5.1. Relationship between test speakers' baseline speech features and their intelligibility gain in the loud condition as given by forward stepwise regression models. The blue line represents the automated features set and the red line represents the targeted baseline measures. At each step on the x axis, one additional feature was added to the forward regressions (if $p < .05$).

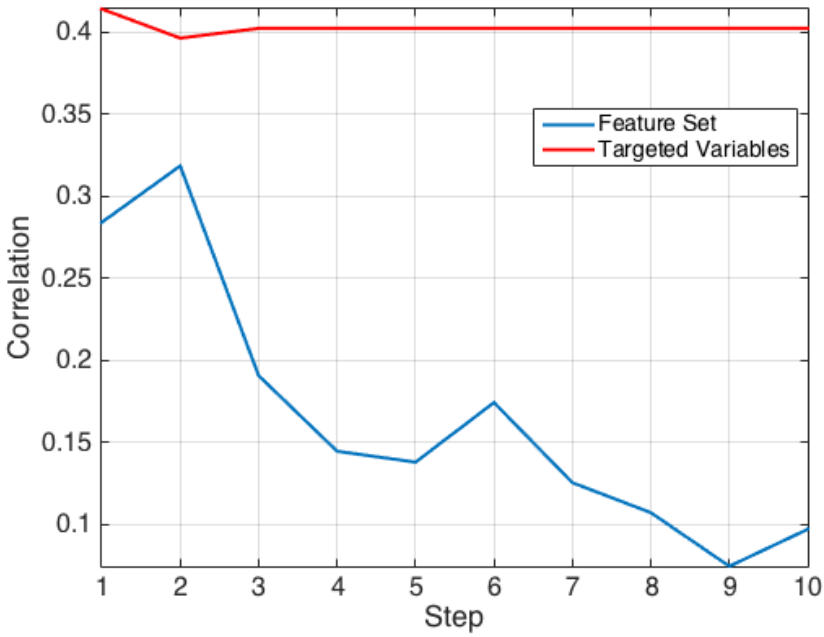

Figure 5.2. Relationship between test speakers' baseline speech features and their intelligibility gain in the slow condition as given by forward stepwise regression models. As with Figure 5.1, the blue line represents the automated features set and the red line represents

the targeted baseline measures. At each step on the x axis, one additional feature was added to the forward regressions (if $p < .05$).

5.4.3 Final Models of Intelligibility Gain

5.4.3.1 Predicting intelligibility gains in the loud condition

Based on the results from the cross validation, a two-step forward regression was conducted to retrain the model on all the speakers' automatically derived data points. The first two variables to emerge from the regression were as follows: 1) From EMS filtered around 250 Hz, energy in the region of 3–6 Hz (divided by overall amplitude of spectrum) and 2) From the EMS filtered around 500 Hz, the frequency in the spectrum (from 0-10 Hz) that had the greatest amplitude.

5.4.3.2 Predicting gains in the slow condition

Again, a two-step forward regression was conducted on all the speakers' automatically derived data points. The first two variables to emerge from this regression were as follows: 1) From the EMS filtered around 500 Hz, the frequency in the spectrum (from 0-10 Hz) that had the greatest amplitude, and 2) from the LTAS filtered around 1995 Hz, kurtosis in root mean squared energy.

5.5 Discussion

This current study presented a follow-up to chapter four by exploring whether automated acoustic analyses could be used to predict speakers' responses to common treatment strategies. The results presented in chapter four suggested that around 1/3 of the variance in speakers' perceived intelligibility gains could be attributed to features of their baseline speech. However, as noted in section 4.5.5, these models had several limitations, most notably the considerable resources required to extract the data.

This chapter focused on the performance of models on cross validation. Cross validation procedures can be used to assess how accurately statistical models will predict the intelligibility gains of new groups of speakers. Overall, this study found that both the automated feature set and the targeted measurements from chapter four could predict the intelligibility gains of speakers they had not been trained on. However, the amount of

variation they were able to account for in the loud and slow speaking conditions was quite different. The outcome of the cross validation process was used to determine an appropriate method of model selection using the automated acoustic feature sets. The cross validation procedure and the final models that resulted from these analyses are discussed in turn.

5.5.1 Cross Validation of Models

Cross validation revealed that the models built using targeted baseline measures, and those built using the automated feature set, varied in their ability to predict the intelligibility gains of speakers they had not been trained on. The automated measures showed a clear reduction in their cross-validated performance after two features had been added to the regression. This problem is unsurprising given the large number of features examined—and demonstrates evidence of model overfitting. Overfitting occurs when dependent variables try to model random noise in the dataset—for example, fluctuations in speakers' intelligibility gains that are completely unrelated their baseline speech. Allowing a model to choose from 393 features in order to describe the variation in 50 data-points (i.e. the differences in intelligibility gains between 50 people), increases the likelihood that overfitting will occur (as there is likely to be a variable available that can account for more random—but statistically significant—levels of variance). Overfit models will generally have poor accuracy in predicting the outcomes of new groups of speakers. Figures 5.1 and 5.2 demonstrate clear evidence of this occurring as the forward stepwise regression progresses. In contrast, when the targeted acoustic measures were trained on 90% of the data-points, they tended to stop adding features after two or three steps of the forward regression were complete—and did not display the same tendency towards overfitting.

Overall, both groups of features demonstrated a stronger ability to predict speakers' improvements in response to cues to speak louder (as opposed to slower). When participants were cued to speak loud, the automated feature set had a greater ability to capture the variation in intelligibility gains amongst speakers that had not been trained on, provided the models were kept simple. This suggests that there is additional acoustic information—beyond that captured by the measures in chapter four—that is important in determining the effectiveness of cues to speak louder. These data demonstrate a limitation of the chapter four study. The eight measurements used in the previous chapter were clearly not able to adequately index all the differences in speakers' baseline speech that were predictive of their intelligibility gains.

In contrast, in the slow condition, the targeted measures from chapter four were more accurately able to predict speakers' intelligibility gains. In this case, the first feature to be added to the forward regression was always baseline speech severity. This measurement was not an acoustic feature. Figure 5.2 suggests that none of the automated speech features were able to account for speakers' variations in intelligibility gain as accurately as these perceptual ratings. This is not surprising given that attempts to acoustically index intelligibility have typically relied on composite measures of a large number of acoustic variables in order to achieve high accuracy (Falk, Chan, & Shein, 2012). As discussed in chapter three, more targeted measurements of vowel articulation have sometimes demonstrated strong correlations with subjective ratings of intelligibility—but there is no single measure which can accurately account for all the information in listeners' perceptual impressions.

5.5.2 Final Model Selected using Automated Baseline Speech Analyses

One of the difficulties in presenting models that contain automated acoustic features is how to interpret their meaning. Currently, the automated analyses used in this study do not translate to readily interpretable speech features. For example, while envelope modulation spectra are known to correlate with speech rhythm metrics, it is difficult to say exactly what different amplitude fluctuations in different frequency bands might represent perceptually (Liss et al., 2010). Because we do not have clear evidence of the perceptual correlates of these speech features, it also is difficult to interpret how these features might relate to different pathophysiologies. Recent work by Berisha, Liss, Sandoval, Utianski, and Spanias (2014) is beginning to elucidate this, by modelling different perceptual qualities using similar automated features to the current study.

With that caveat, the final models constructed with the automated features to predict intelligibility gain incorporated measurements taken from the EMS and LTAS. Intelligibility gains in the loud condition were predicted by two EMS measures, filtered around 250 Hz and 500 Hz respectively. It is likely that these two lower filters capture information about vowels, while filtering out most information about speakers' consonant production (Fry, 1979). The first significant feature was a measurement of the proportion of energy in the 3-6 Hz region of the EMS. Amplitude envelopes with high energy around 3-6 Hz are representative of a normal rate of speech. Hence, speakers who have less energy inside that region may have abnormal fluctuations in speech rate. The second significant EMS feature

examined the frequency of the EMS that had the greatest amplitude. This feature indexes speech rate in a more direct manner—with high energy at lower frequencies of the EMS indicating a speaker with a slower speech rate. These results are congruous with data presented in chapter four, which suggested that measurements of articulation rate were a strong predictor of speakers' intelligibility gain in the loud condition.

Intelligibility gain in the slow condition was predicted by one EMS measure and one LTAS measure. The EMS feature, which was filtered around 500 Hz, examined the frequency of the EMS that had the greatest amplitude. As described in the previous paragraph, this feature is likely to be indicative of speakers' overall speech rate. The second predictive feature in this model was an LTAS measurement filtered around 1995 Hz. It is difficult to interpret why energy in this region would be a significant predictor of intelligibility, but it is possible that this frequency band provides information about speakers' second formants. The LTAS feature selected by the model provided a measurement of kurtosis in root mean squared energy. Kurtosis describes the "peakedness" of a set of data—indicating differences in its distribution. It may be that variations in the energy around 1995 Hz provide information about the degree to which speakers' change the shape of vocal tract (thereby inducing changes in F2).

In summary, the features chosen by forward regression may not have clear perceptual correlates—but their selection suggests they can be used as an indicator of how speakers respond to common treatment strategies. Furthermore, this information has the considerable benefit of being gathered without any manual checking or subjective judgments from researchers. As discussed in the previous chapter, objective diagnostic information is important for researchers wanting to develop more specific inclusion criteria for treatment studies. Furthermore, as automated measures continue to be incorporated into more user-friendly applications, these data may also be used to help provide recommendations for clinicians when choosing between different treatment programs. However, further development of the models presented in this chapter is required. Ideally, these models would benefit from training on a much larger groups of speakers. The inclusion of more data-points in model training is likely to improve the cross-validated accuracy of models generated with three, four or five variables. This would allow more confidence in their application to new groups.

Chapter 6

Summary, Limitations and Future Directions

6.1 Summary

This thesis investigated the speech features of healthy ageing individuals and participants with a range of dysarthria aetiologies. The primary purpose was to determine whether features of the baseline speech signal could predict speakers' responses to common speech modification strategies used in speech therapy. Chapter one discussed the effect of dysarthria and the role of behavioural therapy techniques in speech remediation. This chapter also highlighted the variable patterns of speech intelligibility gains observed in studies of dysarthria treatment. A review of the literature suggested that traditional systems of dysarthria classification may result in participants who exhibit significant differences in their baseline speech features being grouped together for treatment studies. It was posited that variations in speakers' baseline speech features could have considerable influence on the success or otherwise of common speech modification techniques. Therefore, a closer examination of the effect of differences in the speech signal on speakers' intelligibility gains was proposed.

Before exploring the speech features of speakers with dysarthria, consideration was given to the methods used to measure the speech signal in NZE datasets. It was acknowledged that each person has a unique acoustic speech signal that can be systematically affected by many factors unconnected to the presence and severity of dysarthria. Chapter two explored several of these factors including age, sex, and speech dialect. Overall, it was found that age had a significant effect on the duration of speakers' speech segments, but sex did not. In contrast, sex had a large effect on speakers' VSA measures, but their age did not account for significant additional variance. The impact of speakers' dialect could not be directly quantified in this study. However, the NZE dialect appeared to have some influence on the duration of speech segments, with the older NZE speakers producing speech segments relatively quickly (as compared to data presented in Liss et al. (1990) and Benjamin (1982)). Although adaptations were made to the methods used to measure VSA in NZ speakers, there was no clear evidence of quantitative differences in size and variation of VSA measurements as compared to previous studies of US speakers (Lansford & Liss, 2014b; Sapir, Ramig, Spielman, & Fox, 2010).

Chapter two also investigated whether there were naturally occurring relationships between acoustic measurements of prosody and articulation. It was found that measures of vowel duration and spectral vowel dispersion were significantly correlated in healthy older speakers of NZE. The duration of speech segments naturally decreased with age—but VSA

measures did not change. It was posited that reduced speech in older speakers may act as a natural compensatory mechanism, helping them to preserve the accuracy of their articulatory movements with increasing age.

Taken together, data from chapter two highlight several points to consider when measuring the speech of people with dysarthria. Firstly, they provide evidence that an adapted VSA metric is appropriate for measuring vowel dispersion in NZ speakers with dysarthria, with similar levels of overall variation observed compared to previous studies of US speakers. However, they also reveal that there may be important differences to consider in the techniques used to index vowel dispersion. For example, data from this chapter suggested that speakers might reach a vowel's steady-state target—or an approximation of this position—at different stages of the vowel's duration. VSA measurements were also significantly affected by speakers' sex. A difference of this magnitude is problematic when combining speakers' acoustic features within a single model of 'intelligibility gain'. In terms of speech prosody, it was demonstrated that vowel durations could be affected by healthy ageing—perhaps as a result of age-related neuromuscular changes. The NZE dialect may also be a factor in these measures, as it is possible that cultural differences cause NZE speakers with dysarthria to produce faster rates of speech and a lower PVI. Hence, the raw data presented in tables 3.4 and 4.1 needs to be interpreted with these factors in mind.

Chapter three refined the methods used to collect acoustic measures of vowel articulation and perceptual ratings of speech severity by investigating the link between vowel centralization and listener ratings of dysarthria. Perceptual ratings of 'speech precision' were found to be particularly sensitive to the presence of dysarthria—and provided the best index of acoustic measures of vowel dispersion. These ratings also had high reliability. Acoustic indices of vowel articulation exhibited the strongest relationship with perceptual measures when vowel formants were extracted from a flexible measurement point, measured in Bark, and applied to a ratio to calculate their centralization (i.e. the FCR). These methods were able to eliminate significant differences in the size of acoustic measures between male and female speakers. Overall, this investigation produced more valid, objective, and reliable indices of speakers' baseline severity—by utilising both acoustic measures of vowel articulation and perceptual rating scales.

Chapter four revealed that these measurements of baseline severity accounted for significant variability in speakers' intelligibility gains. To assess changes in intelligibility, listeners rated how easy they found the participants' speech to understand following cues to increase loudness and reduce speech rate. In addition to the perceptual ratings and acoustic

measures of vowel production, this investigation measured targeted features of speakers' prosody, voice quality and intonation (including articulatory rate, PVI, CPP, H1A, and the standard deviation of their pitch and speech intensity). Statistical models revealed that intelligibility gains in the loud condition were best predicted by speakers' baseline articulatory rate and listener ratings of baseline speech precision. Intelligibility gains in the slow condition could be modelled by ratings of speech precision and baseline PVI. Both models were able to account for around 1/3 of the variance in intelligibility gains. For speakers who increased their intelligibility in one or more of the treatment conditions, the H1A measure provided the strongest indication of which speech modification strategy would result in optimal improvement. Changes to H1A scores made participants almost four times more likely to be more successful with one strategy as opposed to another. Given these findings, it was hypothesized that assessment data from other forms of acoustic analysis may be able to model aspects of variation in speakers' responses to treatment.

Chapter five followed up on this line of reasoning, investigating whether time-efficient, automated measurements of the baseline speech signal could similarly account for differences in speakers' intelligibility gains. The performance of features derived from MFCCs, LTAS and EMS was compared against the targeted measurements extracted in chapters three and four. Cross-validation techniques were used to determine how well the models could perform on speakers they had not been trained on. When the optimal number of speech features were included in a forward regression model, 17% of the variance in speakers' responses to cues to speak slower could be accounted for by the targeted speech features. The automated measurements were only able to account for around 10%. In contrast, in the loud condition, both feature sets exhibited a stronger performance. The automated features were able to account for up to 25% of the variance in speakers' intelligibility gains—while the targeted measures accounted for 19%. Thus, this study offered evidence that automated feature sets—which are time efficient and require no subjective judgments of researchers—could be used diagnostically to guide treatment decisions.

Overall, the research described in this thesis offers preliminary evidence for the development of a new framework which could be used to identify speakers likely to achieve positive outcomes with common speech therapy strategies. Investigations in chapters three, four and five incorporated a wide variety of presenting dysarthrias. The possibility that researchers could classify speakers likely to achieve positive treatment outcomes based on their presenting speech features has the potential to facilitate clinical decision making within an evidence-based framework, and ultimately, promote stronger group treatment outcomes.

In addition, a theoretical understanding of the features that predict positive treatment outcomes may make it easier for clinicians to provide evidence-based recommendations and justify the funding of different treatment programs for their clients.

6.2 Limitations

Although the studies contained in this thesis have many research and clinical implications, there are a number of factors which limit the application of these findings. Some of these limitations have already been discussed alongside the findings presented in each chapter. However, a more comprehensive review of the overall thesis is required. The following section presents limitations of the current work with regards to the following three methodological variables: (1) sample size, (2) therapist cueing strategies, and (3) measurements of intelligibility gain.

6.2.1 Sample Size

This thesis developed models to predict intelligibility gains across speakers with dysarthria, in order to identify people who were likely to achieve successful outcomes in response to different treatment strategies. The limited sample of speakers with dysarthria included within this study represents the largest limitation to model development and the generalizability of the models' findings. For example, it is unlikely that the group of speakers represented the full range of speech characteristics and severity of dysarthria present in the larger population. Therefore, it is unlikely that the models generated in this thesis are able to account for all combinations of speech features found in speakers with dysarthria. It is also possible that an overrepresentation of certain dysarthria aetiologies (e.g., those with PD) skewed the patterns of speech features observed in this thesis—and therefore may have influenced the size of the effects reported in the models. As discussed in chapter one, speakers with the same aetiology and subtype do not necessarily share more similar acoustic speech features (Y. Kim et al., 2011). However, there are certain cardinal features of dysarthria which have been commonly documented in this group, including breathiness, reduced loudness and a faster rate of speech (Duffy, 2013). In the loud condition, articulatory rate was a significant predictor of intelligibility gains. H1A was a significant predictor of speakers' most successful cueing strategy. It is possible that the speakers with PD drove the strength of these effects. For example, speakers with PD and unusually fast rates or atypically breathy voices may have

had more influence on the results because they represented a relatively large proportion of the sample.

The speaker sample in this thesis was also limited in its overall size. The studies described in chapters four and five had a total of fifty speakers, 43 of whom exhibited dysarthria. Chapter four revealed that only a small number of baseline speech variables were able to predict statistically significant levels of variance in speakers' intelligibility gains. Larger studies may be necessary to determine whether subtler effects exist. With a larger group of participants, smaller effects may reach statistical significance in a multiple regression. Issues with the small sample size were particularly apparent in chapter five. This study found there were many automated features that were able to account for significant variance in models of intelligibility gain. However, the size of these models had to be constrained because of their tendency to overfit the small dataset. The inclusion of more data-points in model training would be likely to improve the cross-validated accuracy of models generated with three, four or five variables—allowing for more complex models to be produced.

Research involving a wider range of languages and speech dialects is also important to better understand the acoustics of dysarthric speech. All participants in this thesis were speakers of NZE. Chapter one discussed differences in the articulation rates of speakers with NZE. Articulation rates and other factors related to speech prosody were significant predictors of intelligibility gain in the models presented in chapters four and five. Therefore, it is possible that differences in the speech prosody of NZE speakers with dysarthria may have produced slightly different effect sizes in these models. For this reason, follow-up studies are needed to test whether the findings in this thesis generalise to other dialects, or other languages.

6.2.2 Therapist Cueing Strategies

This thesis investigated two types of cueing strategies: increased loudness and reduced speech rate. These cues were chosen because they are the basis of many well-established treatment programs (i.e. the Lee Silverman Voice Treatment program (LSVT)) and strategies (e.g. pacing boards, delayed auditory feedback). However, it is acknowledged that there are many methods that can be used to elicit changes in loudness and rate (e.g. Van Nuffelen et al., 2010). This thesis used direct magnitude scaling to elicit changes in loudness and rate (Tjaden & Wilding, 2004). However, speakers might be expected to exhibit slightly different

intelligibility gains depending on the exact instructions given by the clinicians (Lam & Tjaden, 2013). For this reason, the models may not be as accurate in predicting intelligibility gains when different instructions to speak loud and reduce speech rate are used.

In addition, there was no attempt in this thesis to measure to what degree the participants accurately followed the cues to speak louder and reduce speech rate. For example, it is likely that some speakers made considerably more effort to speak louder than others. It is also possible that some participants were unable to produce noticeably louder speech—or were unable to maintain this speech pattern throughout the reading passage. In the case of slow cued speech, there was no controlling for the reduction in rate that speakers made—or the manner in which they reduced their rate. For example, it was observed that some speakers were more inclined to insert pauses between words while others extended the duration of each syllable. It is likely that the manner and degree to which each speaker enacted production changes influenced the individual's resultant intelligibility gain. However, the models in chapter four and five do not account for these differences. Hence, when a speaker made very little change to intelligibility between the baseline and treatment conditions, the reasons for this outcome are not entirely clear. Intelligibility may have remained static because the speaker did not significantly change their speech or because listeners did not judge the speech changes to have improved their intelligibility. In future studies, a closer examination of changes that participants made to their loudness and rate of speech would be beneficial in order to better understand these data.

6.2.3 Measurements of Intelligibility Gain

The measurements of intelligibility change in the loud and slow conditions form the cornerstone of the models presented in chapters four and five. The methods used in this thesis to measure intelligibility have already received some discussion in chapters three and four. Measurements were obtained from a listener rating task using a VAS. As discussed in chapter three, rating scales—especially VAS—are considered to be more sensitive to subtle changes in speech production than measurements of listener transcription accuracy (Sussman & Tjaden, 2012). This is especially true in cases where the speech signal remains highly intelligible, as rating scales allow listeners to indicate that they detect changes, even if they can still understand the words spoken. However, as discussed in chapter four, this protocol also allows listeners to express their own preferences and biases for certain speech samples—and this may have adversely affected ratings of the slow condition in some cases.

Another limitation of the procedures used to index 'intelligibility gain' was the short samples of speech that listeners were exposed to. The phrases that listeners rated were between 11-14 syllables long. These speech samples were taken from the middle of a reading passage and may not have been representative of the speakers' overall performance throughout the task. Indeed, Yunusova, Weismer, Kent, and Rusche (2005) demonstrated that speakers' intelligibility can naturally fluctuate across breath groups. For this reason, it is likely that speakers' intelligibility gains would have differed slightly if a different phrase had been randomly selected for comparison. These fluctuations introduce an additional source of random noise to the models presented in chapters four and five.

6.3 Future Directions

In summary, the final chapter of this thesis demonstrated the potential of using fast, automated assessments to provide recommendations for treatment strategies based on characteristics of individuals' speech patterns. This research is an important first step in the development of personalized medicine approaches in speech therapy. However, before these data can be used to make personal recommendations for speech therapy, further research is needed to address whether these intelligibility gains can be maintained across time, and to examine whether other speech modification strategies may be beneficial for these speakers. To address these questions, future studies will require an expansion: 1) the speaker sample used to build statistical models, 2) the outcome measurements used to train the models, and 3) the number of treatment strategies examined.

The development of new personalized medicine approaches has applications which extend far beyond the recommendation of different therapy strategies. Automated acoustic assessments have the potential to provide immediate feedback on a speaker's response to activities within a speech therapy session. Hence, these assessments could be used to make online adaptations to therapy protocols, by altering the intensity and complexity of therapy activities. Clearly these applications will require a better understanding of how different acoustic measures map to our perceptual impressions of the speech signal. The development of automatic assessments of intelligibility for speakers with dysarthria is a growing area of research (Berisha, Utianski, & Liss, 2013) and the inclusion of these measures within a personalised treatment programme represents an obvious next step for this research theme.

6.4 Conclusions

Understanding differences in the way speakers respond to various behavioural cueing strategies is important in the development of individualised approaches to speech treatment. This thesis demonstrated that it is possible to model individual differences in response to behavioural cues using objective measurements of the speech signal. These data contribute to a stronger theoretical understanding of the mechanisms by which cues to speak louder and reduce speech rate improve speech intelligibility. Furthermore, this thesis provides a platform for the development of more individualised approaches to speech therapy—by demonstrating that it is possible to use rapid automated assessments to make recommendations regarding certain treatment strategies. Ultimately, it is hoped these approaches will contribute to the development of a stronger evidence base to support speech therapy treatment for speakers with dysarthria.

REFERENCES

Awan, S. N., Roy, N., Zhang, D., & Cohen, S. M. (2015). Validation of the Cepstral Spectral Index of Dysphonia (CSID) as a Screening Tool for Voice Disorders: Development of Clinical Cutoff Scores. *Journal of Voice*.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). Celex2. *LDC96L14, Linguistic Data Consortium, Philadelphia*.

Baumgartner, C. A., Sapir, S., & Ramig, L. O. (2001). Voice quality changes following phonatory-respiratory effort treatment (LSVT®) versus respiratory effort treatment for individuals with Parkinson disease. *Journal of Voice*, *15*(1), 105-114.

Bayestehtashk, A., Asgari, M., Shafran, I., & McNames, J. (2015). Fully automated assessment of the severity of Parkinson's disease from speech. *Computer Speech & Language*, *29*(1), 172-185.

Benjamin, B. J. (1982). Phonological performance in gerontological speech. *Journal of Psycholinguistic Research*, *11*(2), 159-167.

Berisha, V., Liss, J., Sandoval, S., Utianski, R., & Spanias, A. (2014). *Modeling pathological speech perception from data with similarity labels.* Paper presented at the Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference.

Berisha, V., Sandoval, S., Utianski, R., Liss, J., & Spanias, A. (2013). *Selecting disorder-specific features for speech pathology fingerprinting.* Paper presented at the Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference.

Berisha, V., Utianski, R., & Liss, J. (2013). *Towards a clinical tool for automatic intelligibility assessment.* Paper presented at the Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference.

Bocklet, T., Nöth, E., Stemmer, G., Ruzickova, H., & Rusz, J. (2011). *Detection of persons with Parkinson's disease by acoustic, vocal, and prosodic analysis.* Paper presented at the Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop.

Bocklet, T., Steidl, S., Nöth, E., & Skodda, S. (2013). *Automatic evaluation of parkinson's speech-acoustic, prosodic and voice related cues.* Paper presented at the Interspeech.

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program], Version 5.1.04, retrieved from http://www.praat.org/.

Borrie, S. A., McAuliffe, M. J., Liss, J., Kirk, C., O'Beirne, G. A., & Anderson, T. (2012). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes, 27*(7-8), 1039-1055.

Bunton, K., Kent, R. D., Kent, J. F., & Rosenbek, J. C. (2000). Perceptuo-acoustic assessment of prosodic impairment in dysarthria. *Clinical linguistics & phonetics, 14*(1), 13-24.

Cannito, M. P., Buder, E. H., & Chorna, L. B. (2005). Spectral amplitude measures of adductor spasmodic dysphonic speech. *Journal of Voice, 19*(3), 391-410.

Cannito, M. P., Suiter, D. M., Beverly, D., Chorna, L., Wolf, T., & Pfeiffer, R. M. (2012). Sentence intelligibility before and after voice treatment in speakers with idiopathic Parkinson's disease. *Journal of Voice, 26*(2), 214-219.

Choe, Y., Liss, J. M., Azuma, T., & Mathy, P. (2012). Evidence of cue use and performance differences in deciphering dysarthric speech. *Journal of the Acoustical Society of America, 131*(2), EL112-EL118. doi: http://dx.doi.org/10.1121/1.3674990

Clopper, C. G. (2009). Computational methods for normalizing acoustic vowel data for talker differences. *Language and Linguistics Compass, 3*(6), 1430-1442.

Cox, F. (2006). The acoustic characteristics of/hVd/vowels in the speech of some Australian teenagers. *Australian journal of linguistics, 26*(2), 147-179.

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.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech, Language and Hearing Research, 12*(3), 462.

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

de Boer, G., & Bressmann, T. (2015). Application of Linear Discriminant Analysis to the Long-term Averaged Spectra of Simulated Disorders of Oral-Nasal Balance. *The Cleft Palate-Craniofacial Journal*.

de Lau, L. M. L., & Breteler, M. (2006). Epidemiology of Parkinson's disease. *The Lancet Neurology, 5*(6), 525-535.

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.

Diehl, R. L., Lindblom, B., Hoemeke, K. A., & Fahey, R. P. (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics, 24*(2), 187-208.

Dromey, C. (2003). Spectral measures and perceptual ratings of hypokinetic dysarthria. *Journal of Medical Speech-Language Pathology, 11*(2), 85-95.

Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management*: Elsevier Health Sciences.

Dykstra, A. D., Hakel, M. E., & Adams, S. G. (2007). *Application of the ICF in reduced speech intelligibility in dysarthria.* Paper presented at the Seminars in speech and language.

Eadie, T. L., & Doyle, P. C. (2002). Direct magnitude estimation and interval scaling of pleasantness and severity in dysphonic and normal speakers. *The Journal of the Acoustical Society of America, 112*(6), 3014-3021.

Easton, A., & Bauer, L. (2000). An acoustic study of the vowels of New Zealand English. *Australian journal of linguistics, 20*(2), 93-117.

Endres, W., Bambach, W., & Flösser, G. (1971). Voice spectrograms as a function of age, voice disguise, and voice imitation. *The Journal of the Acoustical Society of America, 49*, 1842.

Falk, T. H., Chan, W.-Y., & Shein, F. (2012). Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility. *Speech Communication, 54*(5), 622-631.

Feigin, V. L., Lawes, C. M. M., Bennett, D. A., & Anderson, C. S. (2003). Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *The Lancet Neurology, 2*(1), 43-53.

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research, 50*(5), 1241-1255.

Fletcher, A. R., McAuliffe, M. J., Lansford, K., & Liss, J. M. (in press). Assessing vowel centralization in dysarthria: An analysis of methods. *Journal of Speech, Language and Hearing Research*.

Fletcher, A. R., McAuliffe, M. J., Lansford, K. L., & Liss, J. M. (2015). The relationship between speech segment duration and vowel centralization in a group of older speakers. *The Journal of the Acoustical Society of America, 138*(4), 2132-2139.

Fonville, S., van der Worp, H. B., Maat, P., Aldenhoven, M., Algra, A., & van Gijn, J. (2008). Accuracy and inter-observer variation in the classification of dysarthria from speech recordings. *Journal of Neurology, 255*, 1545-1548.

Forrest, K., Weismer, G., & Turner, G. S. (1989). Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. *The Journal of the Acoustical Society of America, 85*(6), 2608-2622.

Fourakis, M. (1991). Tempo, stress, and vowel reduction in American English. *The Journal of the Acoustical Society of America, 90*(4), 1816-1827.

Fox, C. M., & Boliek, C. A. (2012). Intensive voice treatment (LSVT LOUD) for children with spastic cerebral palsy and dysarthria. *Journal of Speech, Language, and Hearing Research, 55*(3), 930-945.

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.

Fox, C. M., Ramig, L. O., Ciucci, M. R., Sapir, S., McFarland, D. H., & Farley, B. G. (2006). *The science and practice of LSVT/LOUD: neural plasticity-principled approach to treating individuals with Parkinson disease and other neurological disorders.* Paper presented at the Seminars in speech and language.

Fromont, R., & Hay, J. (2008). ONZE Miner: the development of a browser-based research tool. *Corpora, 3*(2).

Fry, D. B. (1979). *The physics of speech*. Cambridge, UK: Cambridge University Press.

Hammen, V. L., Yorkston, K. M., & Minifie, F. D. (1994). Effects of temporal alterations on speech intelligibility in parkinsonian dysarthria. *Journal of Speech, Language and Hearing Research, 37*(2), 244.

Han, W., Chan, C.-F., Choy, C.-S., & Pun, K.-P. (2006). *An efficient MFCC extraction method in speech recognition.* Paper presented at the Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on.

Harnsberger, J. D., Shrivastav, R., Brown Jr, W. S., Rothman, H., & Hollien, H. (2008). Speaking Rate and Fundamental Frequency as Speech Cues to Perceived Age. *Journal of Voice, 22*(1), 58-69. doi: http://dx.doi.org/10.1016/j.jvoice.2006.07.004

Herd, C. P., Tomlinson, C. L., Deane, K. H. O., Brady, M. C., Smith, C. H., Sackley, C. M., & Clarke, C. E. (2012). Comparison of speech and language therapy techniques for speech problems in Parkinson's disease. *Cochrane Database of Systematic Reviews,*

(8). http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD002814.pub2/abstract doi:10.1002/14651858.CD002814.pub2

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97*, 3099.

Hillenbrand, J., & Houde, R. A. (1996). Acoustic Correlates of Breathy Vocal QualityDysphonic Voices and Continuous Speech. *Journal of Speech, Language, and Hearing Research, 39*(2), 311-321.

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America, 109*(2), 748-763.

Howe, T. J. (2008). The ICF Contextual Factors related to speech-language pathology. *International Journal of Speech-language Pathology, 10*(1-2), 27-37.

Hunter, L., Pring, T., & Martin, S. (1991). The use of strategies to increase speech intelligibility in cerebral palsy: An experimental evaluation. *International Journal of Language & Communication Disorders, 26*(2), 163-174.

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change, 21*(2), 233-256.

Jiao, Y., Berisha, V., Tu, M., & Liss, J. (2015). Convex weighting criteria for speaking rate estimation. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on, 23*(9), 1421-1430.

Kahane, J. C. (1981). Anatomic and physiologic changes in the aging peripheral speech mechanism. *Aging: Communication Processes and Disorders*, 21-45.

Kent, R. D., & Kim, Y.-J. (2003). Toward an acoustic typology of motor speech disorders*. *Clinical Linguistics & Phonetics, 17*(6), 427-445.

Kent, R. D., Weismer, G., Kent, J. F., Vorperian, H. K., & Duffy, J. R. (1999). Acoustic studies of dysarthric speech: Methods, progress, and potential. *Journal of Communication Disorders, 32*(3), 141-186.

Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics, 26*(2), 117-128.

Khan, T., Westin, J., & Dougherty, M. (2014). Cepstral separation difference: A novel approach for speech impairment quantification in Parkinson's disease. *Biocybernetics and Biomedical Engineering, 34*(1), 25-34.

Kim, H., Hasegawa-Johnson, M., & Perlman, A. (2011). Vowel contrast and speech intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica*, *63*(4), 187-194.

Kim, Y., Kent, R. D., & Weismer, G. (2011). An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. *Journal of Speech, Language and Hearing Research*, *54*(2), 417.

Laan, G. P. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication, 22*(1), 43-65.

Lam, J., & Tjaden, K. (2013). Intelligibility of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1429-1440.

Langstrof, C. (2006). Acoustic evidence for a push-chain shift in the intermediate period of New Zealand English. *Language variation and change*, *18*(02), 141-164.

Lansford, K. L., & Liss, J. M. (2014a). Vowel acoustics in dysarthria: Mapping to perception. *Journal of Speech, Language, and Hearing Research, 57*(1), 68-80.

Lansford, K. L., & Liss, J. M. (2014b). Vowel acoustics in dysarthria: Speech disorder diagnosis and classification. *Journal of Speech, Language, and Hearing Research, 57*(1), 57-67.

Lee, C. M., Yildirim, S., Bulut, M., Busso, C., Kazemzadeh, A., Lee, S., & Narayanan, S. (2004). Effects of emotion on different phoneme classes. *The Journal of the Acoustical Society of America, 116*(4), 2481-2481.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America, 105*(3), 1455-1468.

Liss, J. M., LeGendre, S., & Lotto, A. J. (2010). Discriminating dysarthria type from envelope modulation spectra. *Journal of Speech, Language and Hearing Research, 53*(5), 1246.

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *The Journal of the Acoustical Society of America, 104*(4), 2457-2466.

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

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., & Caviness, J. N. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of Speech, Language and Hearing Research, 52*(5), 1334.

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America, 117*, 3879.

Lowit, A., Dobinson, C., Timmins, C., Howell, P., & Kröger, B. (2010). The effectiveness of traditional methods and altered auditory feedback in improving speech rate and intelligibility in speakers with Parkinson's disease. *International journal of speech-language pathology, 12*(5), 426-436.

Maas, E., Robin, D. A., Austermann Hula, S. N., 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.

Mackenzie, C. (2011). Dysarthria in stroke: A narrative review of its description and the outcome of intervention. *International journal of speech-language pathology, 13*(2), 125-136.

Maclagan, M. (2009). Reflecting connections with the local language: New Zealand English*. *International Journal of Speech-Language Pathology, 11*(2), 113-121.

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*(01), 1-25.

Mahler, L. A., & Ramig, L. O. (2012). Intensive treatment of dysarthria secondary to stroke. *Clinical linguistics & phonetics, 26*(8), 681-694.

Marchant, J., McAuliffe, M. J., & Huckabee, M.-L. (2008). Treatment of articulatory impairment in a child with spastic dysarthria associated with cerebral palsy. *Developmental neurorehabilitation, 11*(1), 81-90.

Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., & De Bodt, M. (2010). Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *Journal of Voice, 24*(5), 540-555.

McAuliffe, M. J., Fletcher, A. R., Kerr, S. E., Anderson, T., & O’Beirne, G. (in press). Effect of dysarthria type, speaking condition and listener age on speech intelligibility. *Journal of Speech, Language and Hearing Research*.

McAuliffe, M. J., Kerr, S. E., Gibson, E. M., Anderson, T., & LaShell, P. J. (2014). Cognitive–Perceptual Examination of Remediation Approaches to Hypokinetic Dysarthria. *Journal of Speech, Language, and Hearing Research, 57*(4), 1268-1283.

McRae, P. A., Tjaden, K., & Schoonings, B. (2002). Acoustic and perceptual consequences of articulatory rate change in Parkinson disease. *Journal of Speech, Language and Hearing Research, 45*(1), 35.

Mefferd, A. S., & Corder, E. E. (2014). Assessing articulatory speed performance as a potential factor of slowed speech in older adults. *Journal of Speech, Language, and Hearing Research, 57*(2), 347-360.

Metter, E. J., & Hanson, W. R. (1986). Clinical and acoustical variability in hypokinetic dysarthria. *Journal of Communication Disorders, 19*(5), 347-366.

Miller, N., Deane, K. H. O., Jones, D., Noble, E., & Gibb, C. (2011). National survey of speech and language therapy provision for people with Parkinson's disease in the United Kingdom: therapists' practices. *International Journal of Language & Communication Disorders, 46*(2), 189-201.

Miller, N., Noble, E., Jones, D., & Burn, D. (2006). Life with communication changes in Parkinson's disease. *Age and Ageing, 35*(3), 235-239.

Moon, S. J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America, 96*, 40.

Nasreddine, Z. S., Phillips, N. A., Bédirian, 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 Geriatrics Society, 53*(4), 695-699.

Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America, 80*(5), 1297-1308.

Neel, A. T. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language and Hearing Research, 51*(3), 574.

Neel, A. T. (2009). Effects of loud and amplified speech on sentence and word intelligibility in Parkinson disease. *Journal of Speech, Language and Hearing Research, 52*(4), 1021.

Niimi, M., & Nishio, S. (2001). Speaking rate and its components in dysarthric speakers. *Clinical Linguistics & Phonetics, 15*(4), 309-317.

Nokes, J., & Hay, J. (2012). Acoustic correlates of rhythm in New Zealand English: A diachronic study. *Language Variation and Change, 24*(01), 1-31.

Palmer, R., Enderby, P., & Hawley, M. (2007). Addressing the needs of speakers with longstanding dysarthria: computerized and traditional therapy compared. *International Journal of Language & Communication Disorders, 42*(sup1), 61-79.

Patel, R. (2002). Prosodic control in severe dysarthria: Preserved ability to mark the question-statement contrast. *Journal of Speech, Language and Hearing Research, 45*(5), 858.

Patel, R., & Campellone, P. (2009). Acoustic and perceptual cues to contrastive stress in dysarthria. *Journal of Speech, Language and Hearing Research, 52*(1), 206.

Pennington, L., Smallman, C., & Farrier, F. (2006). Intensive dysarthria therapy for older children with cerebral palsy: findings from six cases. *Child Language Teaching and Therapy, 22*(3), 255-273.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32*, 693.

Pilon, M. A., McIntosh, K. W., & Thaut, M. H. (1998). Auditory vs visual speech timing cues as external rate control to enhance verbal intelligibility in mixed spastic ataxic dysarthric speakers: a pilot study. *Brain Injury, 12*(9), 793-803.

Ramig, L. A. (1983). Effects of physiological aging on speaking and reading rates. *Journal of Communication Disorders, 16*(3), 217-226.

Ramig, L. O., Countryman, S., Thompson, L. L., & Horii, Y. (1995). Comparison of two forms of intensive speech treatment for Parkinson disease. *Journal of Speech, Language, and Hearing Research, 38*(6), 1232-1251.

Rastatter, M. P., McGuire, R. A., Kalinowski, J., & Stuart, A. (1997). Formant frequency characteristics of elderly speakers in contextual speech. *Folia Phoniatrica et Logopaedica, 49*(1), 1-8.

Robb, M. P., Maclagan, M. A., & Chen, Y. (2004). Speaking rates of American and New Zealand varieties of English. *Clinical Linguistics & Phonetics, 18*(1), 1-15.

Robertson, S. (2001). The efficacy of oro-facial and articulation exercises in dysarthria following stroke. *International Journal of Language & Communication Disorders, 36*(sup1), 292-297.

Rosen, K. M., Kent, R. D., Delaney, A. L., & Duffy, J. R. (2006). Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers. *Journal of Speech, Language and Hearing Research, 49*(2), 395.

Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language and Hearing Research, 53*(1), 114.

Sapir, S., Spielman, J., Ramig, L. O., Hinds, S. L., Countryman, S., Fox, C., & Story, B. (2003). Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on ataxic dysarthria: A case study. *American Journal of Speech Language Pathology, 12*(4), 387-399.

Schiavetti, N., Metz, D. E., & Sitler, R. W. (1981). Construct validity of direct magnitude estimation and interval scaling of speech intelligibility: Evidence from a study of the hearing impaired. *Journal of Speech, Language and Hearing Research, 24*(3), 441.

Schlenck, K.-J., Bettrich, R., & Willmes, K. (1993). Aspects of disturbed prosody in dysarthria. *Clinical linguistics & phonetics, 7*(2), 119-128.

Sellars, C., Hughes, T., & Langhorne, P. (2005). Speech and language therapy for dysarthria due to non-progressive brain damage. *Cochrane Database of Systematic Reviews*, (3). http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD002088.pub2/abstract doi:10.1002/14651858.CD002088.pub2

Sheard, C., Adams, R. D., & Davis, P. J. (1991). Reliability and agreement of ratings of ataxic dysarthric speech samples with varying intelligibility. *Journal of Speech, Language, and Hearing Research, 34*(2), 285-293.

Shewan, C. M., & Henderson, V. L. (1988). Analysis of spontaneous language in the older normal population. *Journal of Communication Disorders, 21*(2), 139-154.

Simmons, K. C., & Mayo, R. (1997). The use of the Mayo Clinic system for differential diagnosis of dysarthria. *Journal of Communication Disorders, 30*(2), 117-132.

Smith, B. L., Wasowicz, J., & Preston, J. (1987). Temporal characteristics of the speech of normal elderly adults. *Journal of Speech, Language and Hearing Research, 30*(4), 522.

Spencer, K., & Beukelman, D. (2001). Evidence-based practice guidelines for dysarthria: Management of velopharyngeal function. *Journal ofMedical Speech Language Pathology, 9*(4), 257-274.

Statistics New Zealand. (2007). *New Zealand's 65+ Population: A statistical volume*. Wellington.

Street, R. L., Brady, R. M., & Putman, W. B. (1983). The influence of speech rate stereotypes and rate similarity or listeners' evaluations of speakers. *Journal of Language and Social Psychology, 2*(1), 37-56.

Sussman, J. E., & Tjaden, K. (2012). Perceptual measures of speech from individuals with Parkinson's disease and multiple sclerosis: Intelligibility and beyond. *Journal of Speech, Language, and Hearing Research, 55*(4), 1208-1219.

Tanner, K., Roy, N., Ash, A., & Buder, E. H. (2005). Spectral moments of the long-term average spectrum: Sensitive indices of voice change after therapy? *Journal of Voice, 19*(2), 211-222.

Tjaden, K., Lam, J., & Wilding, G. (2013). Vowel Acoustics in Parkinson's Disease and Multiple Sclerosis: Comparison of Clear, Loud, and Slow Speaking Conditions. *Journal of Speech, Language, and Hearing Research, 56*(5), 1485-1502.

Tjaden, K., Rivera, D., Wilding, G., & Turner, G. S. (2005). Characteristics of the lax vowel space in dysarthria. *Journal of Speech, Language, and Hearing Research, 48*(3), 554-566.

Tjaden, K., Sussman, J. E., Liu, G., & Wilding, G. (2010). Long-term average spectral (LTAS) measures of dysarthria and their relationship to perceived severity. *Journal of Medical Speech-Language Pathology, 18*(4), 125-133.

Tjaden, K., & Wilding, G. E. (2004). Rate and loudness manipulations in dysarthria: acoustic and perceptual findings. *Journal of Speech, Language and Hearing Research, 47*(4), 766.

Torre III, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of communication disorders, 42*(5), 324-333.

Trudgill, P., Maclagan, M., & Lewis, G. (2003). Linguistic Archaeology The Scottish Input to New Zealand English Phonology. *Journal of english linguistics, 31*(2), 103-124.

Tsao, Y.-C., & Weismer, G. (1997). Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component. *Journal of Speech, Language, and Hearing Research, 40*(4), 858.

Tsao, Y.-C., Weismer, G., & Iqbal, K. (2006). The effect of intertalker speech rate variation on acoustic vowel space. *The Journal of the Acoustical Society of America, 119*, 1074.

Turner, G. S., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech, Language and Hearing Research, 38*(5), 1001.

Umeda, N. (1975). Vowel duration in American English. *The Journal of the Acoustical Society of America, 58*(2), 434-445.

van Bergem, D. R. (1995). Perceptual and acoustic aspects of lexical vowel reduction, a sound change in progress. *Speech Communication, 16*(4), 329-358.

Van der Graaff, M., Kuiper, T., Zwinderman, A., Van de Warrenburg, B., Poels, P., Offeringa, A., . . . De Visser, M. (2009). Clinical identification of dysarthria types

among neurologists, residents in neurology and speech therapists. *European neurology*, *61*(5), 295-300.

Van Nuffelen, G., De Bodt, M., Vanderwegen, J., Van de Heyning, P., & Wuyts, F. (2010). Effect of rate control on speech production and intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica*, *62*(3), 110-119.

Van Nuffelen, G., De Bodt, M., Wuyts, F., & Van de Heyning, P. (2009). The effect of rate control on speech rate and intelligibility of dysarthric speech. *Folia Phoniatrica et Logopaedica*, *61*(2), 69-75.

Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language and Hearing Research*, *50*(6), 1510.

Watson, C. I., & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America*, *106*(1), 458-468.

Watson, C. I., Harrington, J., & Evans, Z. (1998). An acoustic comparison between New Zealand and Australian English vowels∗. *Australian journal of linguistics*, *18*(2), 185-207.

Watson, C. I., Maclagan, M., & Harrington, J. (2000). Acoustic evidence for vowel change in New Zealand English. *Language variation and change*, *12*(01), 51-68.

Weismer, G. (2006). Philosophy of research in motor speech disorders. *Clinical linguistics & phonetics*, *20*(5), 315-349.

Weismer, G., & Berry, J. (2003). Effects of speaking rate on second formant trajectories of selected vocalic nuclei. *The Journal of the Acoustical Society of America*, *113*(6), 3362-3378.

Weismer, G., Jeng, J.-Y., Laures, J. S., Kent, R. D., & Kent, J. F. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatrica et Logopaedica*, *53*(1), 1-18.

Weismer, G., & Kim, Y.-J. (2010). Classification and taxonomy of motor speech disorders: What are the issues? In B. Maassen & P. H. H. M. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. pp. 229-241). Cambridge, UK: Oxford University Press.

Wenke, R. J., Theodoros, D., & Cornwell, P. (2011). A Comparison of the Effects of Lee Silverman Voice Treatment and Traditional Therapy on Intelligibility, Perceptual

Speech Features, and Everyday Communication in Nonprogressive Dysarthria. *Journal of Medical Speech-Language Pathology, 19*(4), 1.

Xue, S. A., & Hao, G. J. (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study. *Journal of Speech, Language and Hearing Research, 46*(3), 689.

Yorkston, K. M. (1996). Treatment efficacy: dysarthria. *Journal of Speech, Language and Hearing Research, 39*(5), S46.

Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of communication disorders, 11*(6), 499-512.

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., Hammen, V. L., Beukelman, D. R., & Traynor, C. D. (1990). The effect of rate control on the intelligibility and naturalness of dysarthric speech. *Journal of Speech and Hearing Disorders, 55*(3), 550.

Yorkston, K. M., Strand, E. A., & Kennedy, M. R. (1996). Comprehensibility of Dysarthric SpeechImplications for Assessment and Treatment Planning. *American Journal of Speech-Language Pathology, 5*(1), 55-66.

Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, J., Ollason, D., . . . Woodland, P. (2002). The HTK-Book 3.2. *Cambridge University, Cambridge, England*.

Yunusova, Y., Weismer, G., Kent, R. D., & Rusche, N. M. (2005). Breath-Group Intelligibility in Dysarthria Characteristics and Underlying Correlates. *Journal of Speech, Language, and Hearing Research, 48*(6), 1294-1310.

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.

Zraick, R. I., & Liss, J. M. (2000). A comparison of equal-appearing interval scaling and direct magnitude estimation of nasal voice quality. *Journal of Speech, Language, and Hearing Research, 43*(4), 979-988.

Appendix A: The Grandfather Passage

You wish to know all about my grand**fath**er. Well, he is nearly 9**3** years old, yet he still thinks as swiftly as ever. He dresses himself in an old, black frock coat, usually with several buttons missing. A long beard clings to his chin, giving those who observe him a pronounced feeling of the utmost respect. Twice **each** day he plays skillfully and with zest upon a small **org**an. Except in the winter when the snow or ice prevents, he slowly takes a **short** walk in the open air **each** day. We have often urged him to walk **more** and smoke less but he always **ans**wers, "Banana oil!" Grand**fath**er likes to be modern in his language.

Note: *Syllables in bold indicate where the New Zealand point vowels, used in the analyses of VSA and FCR, were extracted from the reading passage.*

Appendix B: Listener Rating Instructions

For the articulatory precision ratings, the following instructions were given: “In this experiment, you will rate people's speech precision. Precise speech sounds crisp, with clear and accurate enunciation. Some of the people you will hear have speech disorders which affect the precision of their speech. Your job is to judge each person's speech precision.”
For the ratings of ease of understanding, the following instructions were given: “In this experiment, you will rate how easy it is to understand different speakers. Some of the people you will hear have speech disorders which affect how easy they are to understand. You will make your rating by placing a mark on a scale.”