DEVELOPMENT AND PRELIMINARY EVALUATION OF THE UNIVERSITY OF CANTERBURY PAEDIATRIC AUDITORY-VISUAL MATRIX SENTENCE TEST

A thesis submitted in partial fulfilment of the requirements for the Degree of

Master of Audiology

in the Department of Communication Disorders

at the University of Canterbury

P. M. Jenkins-Foreman

University of Canterbury

2018

ACKNOWLEDGEMENTS

First and foremost I would like to thank my primary supervisor, Associate Professor Greg O'Beirne, without whom this project never would have come into being, for all of the time and effort that he put into developing the UCAMST and UCAMST-P software, and for his expertise and guidance throughout not only my thesis, but the past two years of the Master of Audiology program. I would also like to express my genuine gratitude to Dr Rebecca Kelly-Campbell, for her guidance on statistical procedures.

I would like to thank Associate Professor Margaret Maclagan for her input and advice regarding lexical difficulty and semantic suitability during the development of the base matrix for the paediatric test.

I would like to acknowledge and thank each and every one of the participants who volunteered their time for this study. Without you none of this would have been possible.

To my classmates – I could not have hoped for a better group of people to go through this crazy, exciting, terrifyingly stressful journey with.

I would like to thank my family, both blood and not, for their unwavering support throughout all of my academic pursuits.

To my best friend and fiancé, Gray, I know it hasn't been easy, thank you for always being there for me in everything I do. To say that I am grateful would be an understatement.

Finally, I would like to dedicate this thesis to my Dad, Marc Foreman, who always inspired me to do everything I do to the best of my abilities, thanks for looking out.

"Wicked" – Marc Foreman

ABSTRACT

Speech recognition measures are a fundamental component of the audiometric test battery, providing valuable information regarding an individual's communication difficulties, extending beyond that conveyed by the audiogram. The University of Canterbury Auditory-Visual Matrix Sentence Test was developed in New Zealand English (O'Beirne, Trounson, McClelland, Jamaluddin, & Maclagan, 2015; Trounson, 2012) with the goal of affording an accurate portrayal of these difficulties encountered in real world scenarios. Owing to the cognitive demands of conventional matrix sentence tests, the current study endeavoured to modify the University of Canterbury Auditory-Visual Matrix Sentence Test to develop an audiometric speech recognition measure suitable for use with the paediatric population in New Zealand. Following this, the current study aimed to evaluate the newly developed paediatric measure, alongside its parent test, in order to establish the equivalence of the sentence lists and the conditions in the auditory-alone and auditory-visual modalities for each test individually. Evaluation of the sentence lists with 43 participants with normal hearing suggested that while the sentence lists were equivalently difficult in the auditory-visual modality, the same was not true of the auditory-alone modality. Further evaluation regarding the equivalence of the conditions within each modality indicated that although the accuracy of estimating a listener's speech recognition threshold was found to be equivalent, the speech recognition threshold values were not. Equivalence is of pivotal importance, allowing speech recognition results to be compared across appointments and clinics; consequently these findings warrant consideration in future research.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

ABSTRACT iii

TABLE OF CONTENTS iv

LIST OF TABLES ix

LIST OF FIGURES x

LIST OF ABBREVIATIONS xii

A NOTE ON NOMENCLATURE xv

CHAPTER ONE 1

1.1 Background 1

1.2 Hearing Impairment 2

1.2.1 Anatomy of Hearing Impairment 2

1.2.2 Detection of Hearing Impairment 4

1.3 Speech Audiometry 4

1.3.1 Speech Recognition Measures Presented in Quiet: Speech Audiometry in New Zealand 5

1.3.2 Limitations of Speech Recognition Measures Presented in Quiet 6

1.4 Speech Recognition Measures Presented in Noise 8

1.4.1 Psychophysical Parameters 9

1.4.2 Selection of an Acoustic Masker 11

1.4.3 SNR Tracking Measures 13

1.5 Stimulus Selection: Sentence or Word Stimuli 16

1.6 The Impact of Working Memory 17

1.7 Paediatric Speech Audiometry 19

1.7.1 Criteria for Paediatric Speech Recognition Measures 20

1.7.2 Stimulus Presentation: Monitored Live Voice or Pre-Recorded Stimuli................21

1.7.3 Existing Paediatric Speech Recognition Measures..21

1.8 Sentence-Based Measures 24

1.9 Development, Normalisation, and Evaluation of the University of Canterbury Auditory-Visual Matrix Sentence Test 26

1.9.1 Overview...26

1.9.2 Rationale Behind the Auditory-Visual Component...27

1.9.3 Recording and Editing 30

1.9.4 Selection of Sentence Stimuli: Visual Considerations...33

1.9.5 Generation of Acoustic Maskers...34

1.9.6 Normalisation of the UCAMST Sentences..34

1.9.7 Word- and Fragment-Specific Normalisation..35

1.9.8 UCAMST: List Equivalence...38

1.9.9 UCAMST: Comparison to International MSTs..38

1.10 Selecting a Response Format 43

1.11 Study Rationale..44

1.12 Development of the University of Canterbury Auditory-Visual Matrix Sentence Test - Paediatric 45

1.12.1 Polish Paediatric MST 45

1.12.2 German Paediatric MST 46

1.13 Evaluation of the UCAMST and the UCAMST-P...47

1.14 Aims and Hypotheses......47

CHAPTER TWO: *METHODS*......52

2.1 Overview......52

2.1 Participants......52

2.1.1 Recruitment......52

2.2 Stimuli......53

2.2.1 Generation of the Paediatric Base Matrix......54

2.2.2 Paediatric Base Matrix Composition......56

2.2.3 Generation of New Sentence Lists......58

2.3 Experimental Instrumentation......60

2.4 Scoring Procedures......61

2.5 Experimental Procedures......61

2.6 Planned Statistical Analyses......67

CHAPTER THREE: *RESULTS*......69

3.1 Overview......69

3.2 Participants......70

3.3 List Equivalence Results......71

3.3.1 UCAMST-P List Equivalence......72

3.3.2 UCAMST List Equivalence......74

3.4 Condition Equivalence Results......77

3.4.1 UCAMST-P Condition Equivalence......77

3.4.2 UCAMST Condition Equivalence......79

3.5 Training Results......81

3.5.1 Effect of Training: UCAMST-P......82

3.5.2 Effect of Training: UCAMST......83

CHAPTER FOUR: *DISCUSSION* 85

4.1 Introduction 85

4.2 List Equivalence 86

4.3 Condition Equivalence 89

4.4 The Impact of Training 90

4.5 Study Limitations and Future Research 92

4.5.1 The Sample 92

4.5.2 Block Testing Structure 94

4.5.3 The Training Effect 95

4.5.4 The Impact of Editing 97

4.5.5 Absence of a Babble Noise Condition 98

4.6 Future research 100

4.6.1 Inclusion of a Picture-Pointing Response Method 100

4.6.2 Piloting with Children 102

4.6.3 Piloting with Individuals with Hearing Impairment 103

4.6.4 Cross-Validation with Other Speech Tests 104

4.6.5 Comparison Between the UCAMST and the UCAMST-P 105

4.7 Exploring the Impact of Working Memory 106

4.7.1 Response Time 106

4.7.2 Confusion Matrices 108

4.8 Concluding Statements 109

REFERENCE LIST 111

APPENDIX A: Ethical Approval 137

APPENDIX B: Recruitment 138

APPENDIX C: Informed Consent 139

APPENDIX D: Sentence Lists 143

APPENDIX E: Data Tables 145

LIST OF TABLES

Table 1. Mean SRT and slope values of international MSTs …………..……………………..**41**

Table 2. Normalisation adjustments and slopes for column one (quantity)……………........**56**

Table 3. Normalisation adjustments and slopes for column two (adjective)..……….……....**57**

Table 4. Normalisation adjustments and slopes for column three (object).…………………**58**

Table 5. Block testing conditions………………………………………………………....**62**

Table 6. Results of the Friedman test for list equivalence. Degrees of Freedom = 15 for all tests. ………………………………………………………….……………….....**72**

Table 7. Results of the Kruskal-Wallis tests for condition equivalence. Degrees of Freedom = 1 for all tests..…………………………………………………………………...**77**

Table 8. Results of the Kruskal-Wallis test for the impact of training. Degrees of Freedom = 1 for all tests. …………………………………………………………………...**82**

LIST OF FIGURES

Figure 1. The characteristic sigmoid shape associated with psychometric functions measuring the relationship between the SNR (dB) and the proportion of correct responses (%). *Image retrieved from McClelland (2015, p. 12).*…………..……10

Figure 2. Comparison of psychometric functions with steep (dashed line) and shallow (solid line) slopes. *Image retrieved from McClelland (2015, p. 13).*……………11

Figure 3. The UCAMST base matrix……………………………………………..……27

Figure 4. Sentence recording method employed in the development of the Danish MST (English translation of the Danish MST; *Wagener et al. (2003, p. 13)*). *Copyright 2016 by Taylor and Francis. Reprinted with permission*………………………31

Figure 5. Illustrates two examples of the way in which sentences are formed from four file fragments. Each example depicts 1) the four file fragments required to generate the sentence, and 2) the precise audio material used from each fragment. *Image retrieved from McClelland (2015, p. 25).*.………………………………………32

Figure 6. Comparison across international MST versions and the UCAMST with regards to slope. *Retrieved from Stone (2016, p. 76).*………………………………………40

Figure 7. Closed-set response panel used for the UCAMST………………………………63

Figure 8. Closed-set response panel use for the UCAMST-P…………………………….64

Figure 9. Written instructions displayed for the UCAMST using the open-set response format..…………………………………………………………………..65

Figure 10. Pop-out scorer used by the researcher to record open-set responses for the UCAMST………………………………………………………………… 66

Figure 11. Pop-out scorer used by the researcher to record open-set responses for the UCAMST-P……………………………………………………………………..66

Figure 12. Average PTA thresholds of participants. Error bars represent the standard deviation of threshold at each frequency..………………………………………71

Figure 13. The speech intelligibility functions for each sentence list in the AA mode of presentation for the UCAMST-P in both the open- (A) and closed-set (B) response formats generated for i) when practice did not immediately precede the condition, ii) when the condition was preceded by practice, and iii) both of these combined...……………………………………………………...…………74

Figure 14. The speech intelligibility functions for each sentence list in the AA mode of presentation for the UCAMST in both the open- (A) and closed-set (B) response formats generated for i) when practice did not immediately precede the condition, ii) when the condition was preceded by practice, and iii) both of these combined... ……………………………………………………...……………...………76

Figure 15. Intelligibility functions of the AA, open-set and closed-set conditions of the UCAMST-P……………………………………………………………………...78

Figure 16. Intelligibility functions of the AV, open-set and closed-set conditions of the UCAMST-P……………………………………………………………………...79

Figure 17. Intelligibility functions of the AA, open-set and closed-set conditions of the UCAMST……………………………………………………………………….80

Figure 18. Intelligibility functions of the AV, open-set and closed-set conditions of the UCAMST-P……..……………………………………………………………….81

Figure 19. Comparisons between the list equivalence findings of Stone (A1 and A2) (2016) and the current study (B1 and B2) for both response formats.……………..………88

LIST OF ABBREVIATIONS

AA	Auditory-Alone
ABG	Air-Bone Gap
Adj Diff	Adjustment Difference
AM	Amplitude Modulation
ANOVA	Analysis of Variance
ASHA	American Speech-Language-Hearing Association
AV	Auditory-Visual
AVE	Auditory-Visual Enhancement
AVI	Auditory-Visual Integration
BKB	Bamford-Kowal-Bench
BKB-SIN	Bamford-Kowal-Bench Speech-in-Noise
BM	Basilar Membrane
DANTALE II	Danish Matrix Sentence Test
dB	Decibels
dB A	A-Weighted Decibels
dB HL	Decibels Hearing Level
dB SNR	Decibels Signal-to-Noise Ratio
dB SPL	Decibels Sound Pressure Level
CVC	Consonant-Vowel-Consonant
HA	Hearing Aid
HI	Hearing Impairment
HINT	Hearing in Noise Test
HINT-C	Hearing in Noise Test for Children

Hz	Hertz
IHCs	Inner Hair Cells
KTT	Kendall Toy Test
MLV	Monitored Live Voice
MST	Matrix Sentence Test
NH	Normal Hearing
NU-CHIPS	Northwestern University Children's Perception of Speech
NZ	New Zealand
NZAS	New Zealand Audiological Society
OHCs	Outer Hair Cells
OlKiSa	Oldenburger Kinder-Satztest
OlSa	Oldenburg Satztest
PB_{max}	Presentation level at which maximal performance is achieved
PI	Performance-Intensity
PPMST	Polish Paediatric Matrix Sentence Test
PTA	Pure-Tone Audiometry
QuickSIN	Quick Speech-in-Noise
RM-ANOVA	Repeated-Measures Analysis of Variance
SD	Standard Deviation
SNHI	Sensorineural Hearing Impairment
SNR	Signal-to-Noise Ratio
SPIN	Speech Perception in Noise
SPSS	Statistical Package for the Social Sciences
SRT	Speech Recognition Threshold
TM	Tympanic Membrane

UCAMST	University of Canterbury Auditory-Visual Matrix Sentence Test
UCAMST-P	University of Canterbury Auditory-Visual Matrix Sentence Test – Paediatric
VA	Visual-Alone
WHO ICF	World Health Organisation International Classification of Functioning, Disability, and Health
WIPI	Word Intelligibility Picture Identification

A NOTE ON NOMENCLATURE

The nomenclature employed throughout this thesis is consistent with that utilised in the model provided by the World Health Organization's International Classification of Functioning, Disability, and Health (WHO ICF; World Health Organisation, 2001).

Accordingly, with a view to recognising the multifaceted essence of hearing impairment, the term “hearing loss” was supplanted by the term “hearing impairment”. Additionally, when referring to individuals with a hearing impairment, wording such as “hearing impaired individuals” or “hearing impaired persons” was avoided in the interest of complying with the WHO ICF principle of universality and preventing the labelling of individuals with hearing impairment as a distinct group.

Furthermore, in an effort to follow the WHO ICF model’s client-centred approach, the term “patient” was replaced by the term “client”. The reasoning behind this approach is that the relationship between the clinician and the client encourages client involvement in the selection of rehabilitation and treatment strategies, consequently improving outcomes.

CHAPTER ONE

1.1 Background

Hearing impairment (HI) is the most common disability impacting mankind (Mathers, Smith & Concha, 2000; Olusanya, Neumann, & Saunders, 2014; WHO, 2008). Although unseen, the effects of HI stretch far beyond oral communication, negatively impacting both wellbeing and quality of life (Bird & O'Beirne, 2015; Dalton, Cruickshanks, Klein, Klein, Wiley, & Nondahl, 2003; Mulrow, Aguilar, Endicott, Tuley, Velez, Charlip, Rhodes, Hill, & DeNino, 1990). The adverse consequences of HI can be severe, with far-reaching impacts surpassing communication difficulties alone and extending to emotional and social isolation, diminished physical health, and negative views concerning quality of life (Kelly-Campbell & Lessoway, 2015; Mulrow et al., 1990; Newman & Sandridge, 2004). Irrespective of whether the listener is a child or an adult, the ability to hear and comprehend speech is an important aspect of daily life. For children, access to auditory input, specifically speech, is of particular importance for the development of oral language skills, educational advancement, and the prevention of stigma (Patel, Moitra, Modi, Contractor, & Kantharia, 2014). For adults, a limited or impaired ability to perceive and comprehend speech can result in increased listening effort and uncertainty concerning the topic of conversation, leading to reduced confidence and even social withdrawal (Arlinger, 2003; Kramer, Kapteyn, & Houtgast, 2006).

Hearing impairment, like all health concerns, is unique for each affected individual. Consequently, the psychosocial impacts of a HI cannot be dictated solely by the audiogram (Mulrow et al., 1990). Therefore, when assessing the impacts of a HI, gaining information pertinent to real world scenarios and expected hearing aid (HA) benefit is imperative.

Speech recognition tests are typically utilised in audiological assessments in order to increase understanding pertaining to such impairments. These tests provide insight into the individual's ability to understand, recognise, and detect speech stimuli (Mendel, 2008). The results of these measures specify the course of auditory rehabilitation and afford an understanding of the communication difficulties encountered in assorted acoustic situations (Dietz, Buschermöhle, Aarnisalo, Vanhagen, Hyyrynen, Aaltonen, & Kollmeier, 2014; Ozimek, Warzybok, & Kutzner, 2010). A vast array of speech recognition measures have been developed, and development continues in a number of areas. It is this continuing development that underpins the premise of this thesis project.

1.2 Hearing Impairment

1.2.1 **Anatomy of Hearing Impairment**

The presence of an abnormality or deformity in either the peripheral auditory system (i.e. the outer, middle, and inner ear) or the central structures (eighth nerve and ascending auditory pathway) is liable produce a HI. The location of the defect dictates the type of HI – sensorineural, conductive, or mixed (encompassing both sensorineural and conductive components) (Patuzzi, 2009; Zeng & Liu, 2006). A conductive HI arises due to a problem within the middle or outer divisions of the ear that physically disrupts the transmission of sound to the cochlea (Donkelaar & Kaga, 2011; Pickles, 2012). Of the multitude of conditions that can result in a conductive HI, a majority can be treated via surgical or medical involvement; consequently, a conductive HI is often considered to be temporary (Bess & Humes, 2008). In contrast, a sensorineural hearing impairment (SNHI) presents when there is a cochlea impairment or damage to the auditory nerve (Bess & Humes, 2008; Donkelaar & Kaga, 2011; Pickles, 2012). When the cochlea is impaired, regions of the basilar membrane (BM) may contain outer and/or inner hair cells (OHCs and IHCs respectively) that are not functioning optimally. Damage to the OHCs impairs the ‘active process’, whereby the vibration of the BM is amplified through the process of electromechanical transduction, which also sharpens the tuning of the BM and therefore its frequency selectivity (Moore, 2013). Impairment of the active process results in the cochlea requiring a higher intensity sound to elicit sufficient vibration of the BM to stimulate the IHCs (Moore, 2013). The resultant reduction in frequency specificity has been found to have an adverse impact on speech intelligibility (Patuzzi, 2009). Damage to, or absence of, IHCs can result in less effective stimulation of the auditory nerve (Moore, 2013). If a region along the BM is completely devoid of functional IHCs, information regarding BM vibration patterns is not transmitted to the brain (Moore, 2013).

As with conductive HI, numerous conditions can cause a SNHI, including ageing, infections, tumours, excessive noise exposure, and ototoxic medications (Donkelaar & Kaga, 2011). The loss of sensory hair cells is the most common cause of SNHI and, as hair cells are unable to regenerate, the HI is generally permanent (Gates & Mills, 2005; Welberg, 2008). There are two distinct subgroups within SNHI that arise based on the specific origin of the abnormality. A cochlear SNHI originates due to interference with the active process (i.e. motor processes) or the ICH function (i.e. sensory processes), whereas a retrocochlear SNHI originates from a deformity beyond the cochlea (Patuzzi, 2009).

The overarching impacts of SNHI are extensive, involving not only the attenuation (i.e. reduced hearing sensitivity) and distortion (i.e. reduced clarity) of sounds, but also reduced speech intelligibility and various psychosocial effects, including impacts on relationships with significant others, spouses, and family members (Kelly-Campbell & Lessoway, 2015; Mulrow et al., 1990; Newman & Sandridge, 2004; Patuzzi, 2009; Plomp, 1978). Thus the ability to determine the extent of the distortion component of a HI plays a pivotal role in an audiological test battery, and speech audiometry is key to accomplishing this (Plomp, 1978).

1.2.2 **Detection of Hearing Impairment**

The degree and origin of the aforementioned forms of HI can be established through audiological assessment. Typically, hearing thresholds are determined by performing pure-tone audiometry (PTA), a subjective test that utilises a behavioural response (often a button press) to hearing a tone. Pure tones are presented at various intensity levels and frequencies to ascertain the quietest level at which a listener can identify a stimulus 50% of the time; this level is taken as the listener's threshold (in dB HL – decibels hearing level) for the frequency tested (Valente, 2009). The frequencies tested generally comprise those most important for speech understanding; conventionally, these are the octave frequencies between 250 and 8000 Hz (Hertz) (Schlauch & Nelson, 2009). Each threshold is recorded on an audiogram, which provides a graphical representation of the listener's hearing sensitivity (dB HL) as a function of frequency (Hz) and allows for the configuration, severity, and type of HI to be determined (Schlauch & Nelson, 2009). Whilst PTA is fundamental to diagnostic audiological assessments, crosschecking the result against corresponding measures, for example speech audiometry, increases its value.

1.3 Speech Audiometry

Speech recognition tests are an integral part of any audiological test battery (Ozimek et al., 2010). Speech audiometry is commonly used as a cross-check against PTA thresholds, and to establish an individual's ability to discriminate and process speech stimuli (Hall, 2008; Hamid & Brookler, 2006; Mendel, 2008). Additionally, speech recognition tests afford a beneficial demonstration of the impact of an individual's HI on everyday auditory communication (Hall, 2008; Hamid & Brookler, 2006). Consequently, the scope and clinical applications of such tests are immense, encompassing a variety of functions from assessing hearing aid candidacy to diagnosing auditory processing disorders (Hall, 2008).

1.3.1 **Speech Recognition Measures Presented in Quiet: Speech Audiometry in New Zealand**

The current practice, with regards to speech recognition testing in New Zealand, is the administration of the meaningful Consonant-Vowel-Consonant (CVC) word lists (Boothroyd, 1968; Boothroyd & Nittrouer, 1988; Purdy, Arlington, & Johnstone, 2000). The test material consists of 10 lists, each consisting of 10 monosyllabic, phonetically balanced words, which are presented auditory-alone in quiet (Boothroyd & Nittrouer, 1988). Each CVC word is presented in isolation, devoid of context, following the carrier phrase “say ____” (e.g. “say light”). After each presentation the listener is required to repeat the recognised word. Phoneme scoring, a method in which a score is awarded based on the number of constituent consonants and vowels correctly identified for each word, is employed (Boothroyd, 2008). Typically, the completion of three word lists at differing levels of intensity (dB HL) is required. The listener’s score on each word list is calculated as a percentage for that intensity level and plotted as a performance-intensity (PI) function (McArdle & Chisolm, 2009). The listener’s speech recognition threshold (SRT; sound pressure level at which 50% of the presented words are correctly identified), and PB_{max} (presentation level at which maximal performance is achieved) can both be estimated based on the PI function (Boothroyd, 2008).

The resultant information obtained from the PI function provides diagnostic value and has several clinical applications. First, the shape of the PI function and the departure from the normative curve can provide insight into the nature of the HI. Second, the SRT, estimated from the PI function, can be employed to provide a valuable cross-check of the reliability of the client’s PTA thresholds for the corresponding ear (Boothroyd, 2008; Mendel, 2008). Lastly, the application of the phoneme method of scoring is advantageous, when compared to simple word scoring methods, as it allows a greater number of items to be tested in a comparatively short period of time and thus generates increased test-retest reliability (Gelfand, 1998). Furthermore, this scoring method reduces the impact of a listener's vocabulary knowledge; subsequently, it is believed to be a more valid method of measuring auditory resolution than whole-word scoring (Boothroyd, 1968b; Olsen, Van Tasell, & Speaks, 1997).

The rationale surrounding the prevalence of word recognition tests (like the meaningful CVC word lists) in the audiological test battery in New Zealand appears to be reasonable due to the efficiency and expanse of information that can be derived from such measures. Nevertheless, previous literature has acknowledged a number of aspects that indicate that the solitary use of word recognition measures in audiological test batteries should be reviewed.

1.3.2 **Limitations of Speech Recognition Measures Presented in Quiet**

The presentation of speech stimuli in the absence of noise fails to consider one of the most common complaints expressed by individuals with a HI – that they struggle to decipher speech in the presence of background noise (Beattie, Barr, & Roup, 1997; Dirks, Morgan & Dubno, 1982; Hochmuth, Brand, Zokoll, Castro, Wardenga, & Kollmeier, 2012; Trounson, 2012). Consequently, measures of speech recognition presented in quiet possess numerous limitations, the foremost of which is the inability to afford a realistic representation of a listener's ability to communicate in a real-world situation. Nevertheless, tests of this format are typically the sole measure of speech recognition in clinical practice, both in New Zealand and globally. Speech stimuli are often presented in the absence of background noise and in isolation (as is the case for the meaningful CVC word lists). The rationale behind this method of testing is that, when compared to alternative designs, it may more accurately capture issues pertaining to audibility, eliminating confounding factors, such as the listener's use of contextual cues or working memory (Wilson, McArdle, & Smith, 2007a).

However, studies have shown that measures of hearing sensitivity and speech recognition in quiet alone are insufficient to establish the communication difficulties encountered by individuals with HI in everyday life, particularly in background noise (Beattie et al., 1997; Carhart & Young, 1976). One consequential shortcoming is that such tests lack the ability to provide information pertaining to a listener's expected real world benefit from amplification (Beattie et al., 1997). Such information is paramount, as research has found that, for some individuals, HAs may intensify the difficulties experienced in background noise (Carhart & Young, 1976; Kelly-Campbell & Lessoway, 2015). Consequently, the need to establish a listener's communication difficulties in complex listening situations is clear.

Also of concern is the sensitivity of diagnostic tests and the degree to which they can distinguish between listeners with normal hearing (NH) and varying degrees of HI. Previous research has demonstrated that the performance of an individual with a mild HI on monosyllabic speech recognition measures presented in quiet may not accurately depict the communication difficulties that the individual encounters (Beattie et al., 1997). Due to the simplicity of such tests, it is believed that they are unable to distinguish between individuals with NH and those with a mild HI (Beattie et al., 1997). This distinction is essential, as previous research has seen benefit from amplification in adults with mild HIs (Kelly-Campbell, Thomas, & McMillan, 2014).

The speech recognition measures presented in quiet that are currently employed in clinical practice in New Zealand are categorised as non-adaptive tests (discussed further in section 1.4.3). Non-adaptive procedures are vulnerable to floor and ceiling effects, where scores of close to 100% or 0% are frequently obtained. This is problematic as it can be challenging to identify meaningful differences in speech recognition abilities (Gifford, Shallop, & Peterson, 2008). For example, subsequent to a score of 100% being obtained, further improvement cannot be recognised.

The essentially exclusive use of word-based speech recognition measures presented in quiet in clinical practice may impinge upon an audiologist's ability to make inferences regarding the client's capabilities in real world communication situations, and thus their suitability for, and potential benefit from, different rehabilitative options. Although efficiency is essential in clinical practice, due to unavoidable time constraints, numerous studies have indicated that the use of speech recognition measures in noise, particularly those utilising sentence stimuli, may be of greater clinical value (Beattie et al., 1997; Carhart & Young, 1976; Dirks et al., 1982).

1.4 Speech Recognition Measures Presented in Noise

Previous literature has suggested that measures of speech recognition that employ background noise, in addition to sentence-based test material, afford a more realistic representation of the ability of a client with a HI to communicate in real-world situations (Grunditz & Magnusson, 2013; Hagerman, 1982; Trounson, 2012). Such measures are advantageous as they provide the audiologist with information pertaining to a client's potential candidacy for different methods of amplification, as well as information that can be applied during the counselling process to outline the shortcomings and benefits of various approaches in order to impart realistic expectations (Humes, 1999; Taylor, 2003; Wilson et al., 2007a). However, despite the longstanding acknowledgement of such tests as an important addition to the audiological test battery, clinical application has only recently commenced (Billings, Penman, Ellis, Baltzell, & McMillan, 2016; Carhart & Tillman, 1970; Dirks et al., 1982). Clinically, a vast array of speech-in-noise measures exist; these measures vary with respect to procedural parameters, in particular the type of stimulus or masking noise presented and the mode of presentation, including various noise or stimulus adaptations (Arlinger, 1998; Taylor, 2003; Wagener & Brand, 2005).

1.4.1 **Psychophysical Parameters**

As with speech recognition tests administered in quiet, performance on speech recognition tests administered in noise is generally specified by a listener's SRT (Brand & Kollmeier, 2002). However, when testing is conducted in noise, the SRT is derived from a psychometric function which represents the listener's performance – number of correct responses, depicted as a percentage intelligibility score (%) – as a function of the signal-to-noise ratio (SNR) (MacPherson & Akeroyd, 2014). Characteristically, psychometric functions of this description are sigmoidal (i.e. 's'-shaped) and are frequently described using two fundamental parameters: the threshold – the stimulus level necessary to achieve a specific performance score (i.e. 50% correct) – and the slope – the proportional rate of change in performance in response to variations in the level of the stimulus (Gilchrist, Jerwood, & Ismaiel, 2005; MacPherson & Akeroyd, 2014). Figure 1 depicts the characteristic sigmoidal shape of a psychometric intelligibility function.

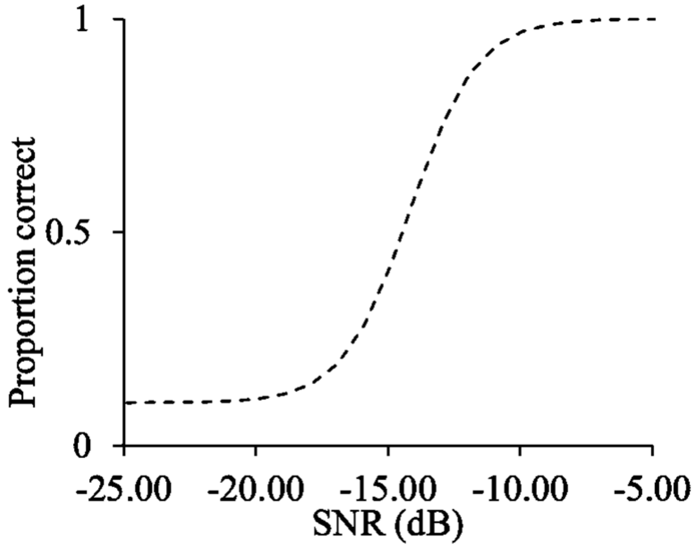

Figure 1. The characteristic sigmoidal shape associated with psychometric functions measuring the relationship between the SNR (dB) and the proportion of correct responses (%). *Image retrieved from McClelland (2015, p. 12).*

With regards to speech recognition tests conducted in noise, the accuracy of the SRT is primarily determined by the slope of the psychometric function at the SRT (Ozimek et al., 2010). This is an orthodox inverse relationship, in which the steeper the slope is at the SRT, the lower the standard deviation (SD) of the SRT (Ozimek et al., 2010). Therefore, the slope of the psychometric function regulates the sensitivity of the test, with a steep slope indicating a more sensitive measure (Ozimek et al., 2010). Accordingly, if a test is highly sensitive, a minor adjustment in the level of the stimulus will result in a sizeable change in the value being measured (Brand & Kollmeier, 2002). This concept is illustrated in Figure 2 by a comparison of the morphology of two psychometric functions with steep and shallow slopes. Tests with a higher degree of sensitivity (i.e. steeper slope) are preferred, as the SRT can be more accurately established in a comparatively fewer number of trials (Francart, van Wieringen, & Wouters, 2011).

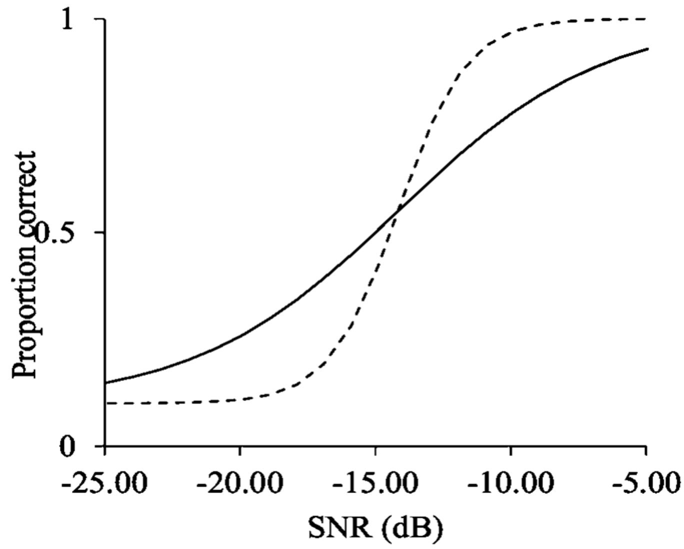

Figure 2. Comparison of psychometric functions with steep (dashed line) and shallow (solid line) slopes. *Image retrieved from McClelland (2015, p. 13).*

The slope of a psychometric function can provide insight into the perceptual benefit a listener is expected to receive from small adjustments in the SNR (MacPherson & Akeroyd, 2014). Consequently, it has been suggested that such information can provide rehabilitation audiologists with the ability to quantify a client's expected gain in perceptual benefit from the improvement in SNR afforded by a HA, and thus assist in establishing the recommendations to be delivered to the client (MacPherson & Akeroyd, 2014). Beyond predicting a client's HA outcomes, this information may also provide an audiologist with valuable insight that can be utilised in the counselling process, particularly with regards to realistic expectations of the HA and expected perceived benefit (Wilson et al., 2007a).

1.4.2 **Selection of an Acoustic Masker**

The specific type of acoustic masker presented is reliant upon both the information required and the test's objective, as one particular type of acoustic masker may be more suitable than another; accordingly, this is an element of speech audiometry that warrants consideration (Francart et al., 2011). Conventionally, multi-talker babble noise and continuous speech-shaped noise are the two types of acoustic masking noise utilised in speech recognition measures (Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004). Previous literature has suggested that, when compared to multi-talker babble noise, continuous speech-shaped noise has less variability. Consequently, the reproducibility of scores obtained using continuous speech-shaped noise is superior (Bacon, Opie, & Montoya, 1998; Killion et al., 2004). The use of continuous speech-shaped noise is therefore liable to be advantageous in a research context as the production of steeper, and thus more sensitive, psychometric functions allows better differentiation between variables (Wagener & Brand, 2005; Francart et al., 2011).

Alternatively, research has indicated that multi-talker babble noise more accurately embodies the speech-in-noise listeners experience in daily listening situations, and thus has greater face validity (Killion et al., 2004). The fluctuating qualities of multi-talker babble noise produce larger amplitude modulations than steady-state background noise (Bacon et al., 1998; Hopkins & Moore, 2009). Amplitude modulation (AM) is the gradual variation in the amplitude of a waveform; these amplitude changes in an acoustic masker create dips in the SNR and provide listeners with a glimpse of the stimuli, an event entitled ‘masking release’ (Füllgrabe, Berthommier, & Lorenzi, 2006; Hopkins & Moore, 2009; Howard-Jones, & Rosen, 1993). Research has shown that listeners with NH typically perform better on speech recognition measures in the presence of fluctuating noise (i.e. multi-talker babble noise) than listeners with a SNHI (Festen & Plomp, 1990; Hopkins & Moore, 2009; Peters, Moore, & Baer, 1998; Wagener & Brand, 2005). This is believed to be due to the fact that listeners with a SNHI have broader auditory filters, and thus the spread of masking is increased, causing the target signal present in the dips of the masker at one frequency to be masked by the acoustic masker present at neighbouring frequencies, which are at a higher level (Glasberg & Moore, 1996; Hopkins & Moore, 2009). As such, research suggests that, for listeners with a SNHI, masking release is usually small or absent (Bacon et al., 1998; Hopkins & Moore, 2009). Therefore, the use of multi-talker babble noise in measures of speech recognition may better distinguish between levels of HI (via SRT) than those measures that use continuous steady-state masking noise (Bacon et al., 1998; Francart et al., 2011; Hopkins & Moore, 2009). Consequently, such measures may be better suited for use in clinical assessment, as they better depict the difficulties listeners with HI face in everyday listening situations (Bacon et al., 1998; Francart et al., 2011; Hopkins & Moore, 2009). The importance of examining the qualities and benefits of different acoustic maskers has been highlighted in the literature, and should be considered when speech measures are employed.

1.4.3 **SNR Tracking Measures**

The reliability and efficiency of the method employed to estimate the SRT in speech intelligibility tests should also be considered. Non-adaptive (i.e. fixed SNR) tests were initially developed in order to more closely replicate the conditions encountered by a listener in a real-world listening environment (Taylor, 2003). Such tests employ different intensity levels, which are established prior to the assessment, and remain constant throughout; this is known as the method of constants (Levitt, 1971). The Speech Perception in Noise (SPIN; Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984) test is an example of a non-adaptive SNR measure. The SPIN test, which is administered in the presence of multi-talker babble noise, involves the listener reciting the last word of each sentence (Bilger et al., 1984). The test is scored as a correct word percentage; individual scores are evaluated based on whether the sentence was considered to be of either high or low predictability with regard to contextual cues (Bilger et al., 1984).

Alternatively, adaptive SNR testing procedures can be used to estimate SRTs. Levitt (1971, p. 467) describes an adaptive procedure as “one in which the stimulus level on any one trial is determined by the preceding stimuli and responses”. The staircase, or up-down, method is the adaptive procedure most commonly employed for measuring sensory thresholds (Brown, 1996; Cornsweet, 1962; Levitt, 1971; Plomp & Mimpen, 1979). Adjustments in the stimulus level (either up or down by an equal and constant 'step size') are dictated by the listener's response (Plomp & Mimpen, 1979). For example, if the listener responds correctly, the presentation level will be reduced by 2 dB (decibels), and vice versa. Examination of the reliability of this adaptive staircase method of measuring SRT has revealed an individual SRT SD of 0.9 dB and a markedly superior slope, 15%/dB as compared to 5%/dB in previous research (Plomp & Mimpen, 1979).

The Hearing In Noise Test (HINT; Nilsson, Soli, & Sullivan, 1994) and the Quick Speech-In-Noise (QuickSIN; Killion et al., 2004) test are both commercially available measures that utilise adaptive SNR procedures. The method employed by the QuickSIN test is a pseudo-adaptive procedure in which the stimulus presentation level remains fixed, whilst the masking noise, four-talker babble, is varied in order to induce 5 dB changes in the SNR (Killion et al., 2004; Taylor, 2003). Scoring is word-based, whereby the listener is granted a correct result for each of the five key words identified for each sentence (Killion et al., 2004; Taylor, 2003). The resultant score is referred to as the "SNR loss", which is defined as the increase in SNR (dB) necessary for a listener with HI to receive speech-in-noise at levels comparable to a listener with NH for a stipulated performance level, usually 50% sentence or word identification (Grant & Walden, 2013; Killion et al., 2004; Tayor, 2003).

Another test capable of providing a measure of SNR loss is the HINT (Nilsson et al., 1994). The HINT presents sentence stimuli, which are adjusted in 2 dB steps, together with speech-shaped masking noise, which is fixed at 65 dB SPL (decibels sound pressure level) (Nilsson et al., 1994). In contrast to the QuickSIN test, scores for the HINT are sentence-based, requiring the listener to correctly recall all of the key words in the sentence to qualify as a correct result (Nilsson et al., 1994). A further distinction between the two tests is that, unlike the QuickSIN test, the HINT utilises a truly adaptive procedure in which the presentation level of the sentence stimuli in each trial is dictated by the listener's response in the preceding trial (Levitt, 1971; Nilsson et al., 1994). Directly measuring a listener's SRT is advantageous as, unlike with scores of percentage correct, floor and ceiling effects can be circumvented. Additionally, the ability of adaptive tests to rapidly and efficiently identify an individual's likely threshold region causes the effectiveness of such tests to exceed that of tests employing the method of constants (Levitt, 1978) whilst reliability and accuracy is maintained (Buss, Hall, Grose, & Dev, 2001; Leek, 2001).

Brand and Kollmeier (2002) proposed an alternative adaptive tracking procedure that simultaneously estimates both the slope and the SRT through the adaptive tracking of two points on the psychometric function (the so-called "pair of compromise"), typically corresponding to the 20% and 80% correct points. This method employs a word scoring system, and has been found, when a minimum of 30 sentences were employed, to attain reliable SRT levels with a SD of 1 dB and slope approximations of 20-30% (Brand & Kollmeier, 2002).

Regardless of the method of adaptive testing employed, the advantage of such measures is that they are highly reliable and efficient, irrespective of any noise or signal alternations (Wagener & Brand, 2005). The information afforded by sentence-based speech-in-noise measures, such as SNR loss, extends beyond that conveyed by the audiogram (Wilson, 2003). Previous literature has suggested that the inconsistencies relating to the perceived deficit among individuals with comparable levels of HI may be accounted for by SNR loss (Killion et al., 2004). Consequently, it has been postulated that acquiring measures pertaining to such deficits may assist clinicians in the formation of recommendations regarding suitable technology levels (Killion et al., 2004). Nevertheless, in audiological rehabilitation, the functional application of SNR loss has not yet been well established, and thus should be applied cautiously.

1.5 Stimulus Selection: Sentence or Word Stimuli

Another aspect of speech recognition measures that requires consideration is the type of speech stimuli employed (Wilson, 2003). As discussed in section 1.3.1, the current practice with regards to speech recognition testing in New Zealand is the administration of the meaningful CVC word lists (Boothroyd, 1968; Boothroyd & Nittrouer, 1988; Purdy et al., 2000). Word-based speech recognition tests are advantageous, as they require less time to administer than sentence-based speech recognition tests and fewer demands are placed on the listener's auditory memory (discussed further in section 1.6) (Wilson et al., 2007a). Nevertheless, despite the frequent employment of such measures, multiple disadvantages are evident (Bosman & Smoorenburg, 1995; Ozimek, Kutzner, Sęk, & Wicher, 2009).

Previous literature has suggested that the utilisation of sentence stimuli, as opposed to word stimuli, in speech audiometry may afford a better estimate of a listener's communication difficulties (Cox, Alexander, & Gilmore, 1999; Hochmuth et al., 2012; Killion et al., 2004). Additionally, the psychometric functions derived from audiometric speech recognition tests that utilise sentence stimuli have been found to be steeper, and thus afford a more accurate measure of SRT, than those of digits and words (Bell & Wilson, 2001; Bosman & Smoorenburg, 1995; McArdle, Wilson, & Burks, 2005; Versfeld, Daalder, Festen, & Houtgast, 2000). Hagerman (1976) documented that by doubling the number of words in a list, the accuracy is improved by $\sqrt{2}$. Therefore, the more words that can be incorporated into a single trial, the greater the steepness of the psychometric functions, and thus the greater the accuracy of the SRT measurements (Hagerman, 1976). Consequently, utilising sentence stimuli, as opposed to single words, allows for the integration of more words, and hence higher accuracy (Hagerman, 1976). Moreover, as sentence stimuli affords the opportunity to examine a listener's capacity to perceive multiple different speech sounds within a solitary trial, the time-efficiency of the assessment can be enhanced (Hochmuth et al., 2012).

Contingent upon the objective of the test, and the cognitive capabilities of the listener, a further drawback of utilising word stimuli is that they are presented individually and, as such, there is no contiguous material affecting the client's answer (Wilson et al., 2007a). Consequently, such tests are not indicative or representative of a realistic listening situation. It has been established that sentence-based test materials increase the validity of the test's capacity to assess a client's ability to hear and understand in a real-world scenario due to the greater dynamic range afforded by sentence stimuli (Dietz et al., 2014; Killion et al., 2004). This is owing to the intonations, fluctuations, pauses, temporal elements, and contextual cues expressed during conversational speech (Nilsson et al., 1994).

Thus, tests that utilise background noise, in addition to sentence-based test material, afford a more realistic representation of the ability of an individual with a HI to communicate in a real-world situation (Hagerman, 1982; Trounson, 2012). Previous research in this field essentially exclusively supports the use of sentence stimuli in speech recognition procedures, due to the rehabilitative value of the information and the extensive understanding of a listener's communication difficulties that can be acquired (Dietz et al., 2014).

1.6 The Impact of Working Memory

The added cognitive load associated with the recognition of sentence stimuli has been extensively examined in the literature (Cervera, Soler, Dasi, & Ruiz, 2009; McArdle et al., 2005; Wilson et al., 2007a). Sentenced based speech recognition measures require the listener to retain the information presented for the length of the sentence, following which the words must be identified, either verbally in the open-set response format, or by selecting the individual words in the closed-set response format. This requires the use of working memory. Working memory is of great importance in auditory speech processing owing to its function in processing and storing information (Cerevera et al., 2009; Daneman & Carpenter, 1980).

Auditory processing has been found to decline with age, and can be explained by peripheral, central, and cognitive factors (Humes, Lister, Wilson, Cacace, Cruickshanks, & Dubno, 2012; Humes, Watson, Christensen, Cokely, Halling, & Lee, 1994; Jerger, Jerger, & Pirozzolo, 1991). Previous research has revealed a robust relationship between age-related deterioration in working memory and decreased auditory performance (Foo, Rudner, Rönnberg, & Lunner, 2007; Hällgren, Larsby, Lyxell, & Arlinger, 2001). Moreover, a significant relationship has been established between the level of HI and cognitive function, with the risk of cognitive deterioration increasing with the level of HI (Lin, Yaffe, Xia, Xue, Harris, & Purchase-Helzner, 2013). Further research regarding this relationship, with respect to speech recognition, has established that SRT estimates can be impacted by memory capacity, with higher estimates of SRT being recorded when memory capacity is reduced (Theunissen, Swanepoel, & Hanekom, 2009; van Rooij and Plomp, 1990).

It has also been acknowledged that the practical application of speech recognition measures should be concise, especially when used with the senior population, in order to account for the higher possibility of results being impacted by age-related cognitive factors (Cervera et al., 2009; van Rooij & Plomp, 1990). Thus, it is apparent that the consideration and investigation of a listener's working memory abilities, prior to employing any sentence-based speech recognition testing, is advisable in order to attempt to minimise the impact of reduced working memory capacity on the test's validity (Craik, 1994; Kramer, Zekveld, & Houtgast, 2009; McArdle et al., 2005; Wilson et al., 2007a). Consequently, due to the cognitive requirements of speech recognition measures that utilise sentence stimuli, the impact of working memory needs to be considered in SRT estimation (Cerevera et al., 2009; McArdle et al., 2005; Wilson et al., 2007a).

These effects of reduced working memory capacity are not only applicable to declines due to aging, but also to development. From birth, working memory continues to develop during childhood and adolescence, increasing the proficiency with which information can be updated (Brocki & Bohlin, 2004; Gathercole, Pickering, Ambridge, & Wearing, 2004; Lendinez, Pelegrina, & Lechuga, 2015; Luna, Garver, Urban, Lazar, & Sweeney, 2004). Accordingly, the necessity for a behavioural speech recognition test that is appropriate for use with the paediatric population and meets several specific criteria has been recognised (Kosky & Boothroyd, 2003).

1.7 Paediatric Speech Audiometry

Speech perception is an essential ability, affording critical information pertaining to general auditory perception capabilities, as well as providing valuable information surrounding the progress of a child's language, speech, cognitive, and reading abilities (Mendel, 2008). Recommendations surrounding amplification by HAs, or cochlear implantation, as well as linguistic learning paradigms are partially established from these speech recognition measures (Mendel, 2008). Consequently, it is evident that the accuracy of these paediatric speech recognition measures in determining a child's capacity to recognise patterns and phonetic segments, in addition to words and sentences, is paramount (Mendel, 2008). In addition to being a valuable tool for monitoring progress, such measures can offer information concerning the formation and employment of supplementary methods of audiological (re)habilitation (e.g. auditory training, speechreading) (Mendel, 2008). Thus, the importance of routine and reliable clinical assessments of the speech recognition abilities of children with HI is imperative.

Analogous to the CVC meaningful word lists used with the adult population in the New Zealand audiological test battery, the Kendall Toy Test (KTT) is presented to the paediatric population. The KTT consists of 10 monosyllabic words presented in quiet, in addition to five practice items. The child is first familiarised with the test items and then asked to point to each item in turn. However, in New Zealand there is no formal manual available for the KTT, and the current presentation method is understood to have been developed from the Australian version of the test, which consisted of five vowel pairs and five distractor items (Antognelli, 1986).

In New Zealand, following familiarisation of the items, it is common for the presenting audiologist to cover their mouth to prevent visual cues. A demonstration of normal hearing is considered to occur when the child gets ≥ 90% of the items correct at 35 dB A (A-weighted decibels). However, there are a wide variety of techniques used to determine the percentage correct score at elevated levels. No normative data exists in New Zealand for the KTT, and the 35 dB A normal hearing 'passing' level was established based on PTA thresholds of ≤ 15 dB HL (Antognelli, 1986). Consequently, given the current administration of the KTT in New Zealand, it is apparent that the validity of the test is not as high as it could be. Taking this into consideration, in addition to the points made above regarding the use of word-based measures of speech recognition, the development of a paediatric sentence-based speech recognition test in New Zealand English appears warranted.

1.7.1 **Criteria for Paediatric Speech Recognition Measures**

Previous literature has outlined multiple criteria and considerations that should be taken into account when developing a clinical test for use with the paediatric population (Kosky & Boothroyd, 2003). These considerations include the test's attentional, cognitive, and motoric demands, in addition to the interest of the task itself, along with the need for motivating factors (Kosky & Boothroyd, 2003). The requirement for performance on the test to be uninhibited by the child's comprehension of vocabulary and higher-level language abilities was also identified to be of importance (Kosky & Boothroyd, 2003; Neumann et al., 2012). Additionally, Kosky and Boothroyd (2003) advised that performance should not be inhibited by a child's speech production skills or lack of phonological knowledge. It was also specified that the child's ability to communicate in real-world listening environments should ultimately be assessed (Kosky & Boothroyd, 2003). Moreover, as the length of paediatric audiometric speech tests is dictated by the child's fatigue, the time efficiency of the test also requires consideration (Neumann et al., 2012).

1.7.2 **Stimulus Presentation: Monitored Live Voice or Pre-Recorded Stimuli**

Another aspect worthy of consideration is whether the test stimuli are presented via a standardised recording or through monitored live voice (MLV). Research has found that the results of speech recognition measures acquired using MLV tend to be better than those obtained using recorded stimuli (Uhler, Biever, & Gifford, 2016). However, overestimation of a child's speech recognition capabilities is possible if such measures are only presented via MLV (Uhler et al., 2016). The practice of employing standardised pre-recorded stimuli is believed to obtain results that better represent the child's abilities, and are not confounded by familiarity effects (Uhler et al., 2016). Owing to the importance of monitoring children with HI, the use of a standardised measure improves the test's accuracy, particularly if the child is being tested across clinics or clinicians (Uhler et al., 2016).

1.7.3 **Existing Paediatric Speech Recognition Measures**

A range of speech recognition measures is available for use with children. The Northwestern University Children's Perception of Speech (NU-CHIPS; Elliott & Katz, 1980) test, and the Word Intelligibility Picture Identification (WIPI; Ross & Lerman, 1970) test employ monosyllabic words that are appropriate, with regards to receptive vocabulary, for use with children aged 3-5 and 4-6 years respectively (Elliott & Katz, 1980; Ross & Lerman, 1970). Both tests are able to be presented in an open-set response format – in which the child is required to respond verbally – or a closed-set response format – wherein the child identifies the word recognised through a picture pointing response (Elliott & Katz, 1980; Ross & Lerman, 1970).

As discussed in section 1.5, although word-based speech recognition measures offer valuable insight regarding everyday receptive communication difficulties, this information is limited (Bell & Wilson, 2001; Neumann et al., 2012). Sentence-based speech recognition tests are believed to be superior for measuring paediatric speech recognition, when compared to equivalent word-based tests, due to a higher degree of sensitivity (i.e. steeper psychometric functions) (Bell & Wilson, 2001; Neumann et al., 2012). Furthermore, measures of speech recognition capable of assessing a greater proportion of words in a limited timeframe attain higher reliability as, for the paediatric population, test length is prominently dictated by fatigue (Neumann et al., 2012). Additionally, the presentation of sentence stimuli is advantageous as it provides information pertaining to the child's capacity to "fill in the blanks" and, subsequently, information on the child's communication abilities in everyday life (Madell, 2008). Thus, it is evident that sentence-based tests are more suitable for use in a clinical paediatric audiometric test battery than word-based tests of the same kind, and the incorporation of such tests has been recommended (Bell & Wilson, 2001; Wagener & Kollmeier, 2005). Nevertheless, word-based sentence tests may be favourable in the diagnosis of central auditory processing disorder, or in the discrimination of minute phonological differences (Neumann et al., 2012).

A variety of paediatric sentence-based speech recognition measures have been developed. One such example is the Bamford-Kowal-Bench (BKB; Bench, Kowal, & Bamford, 1979) test, which is intended for use with children from six years of age with a HI. The BKB employs sentences appropriate for grade one reading level, which are scored based on key word recognition (Bench et al., 1979). The Bamford-Kowal-Bench Speech-In-Noise (BKB-SIN; Etymotic Research, Inc., 2005a; Etymotic Research, Inc., 2005b) test was developed based on the BKB test, and employs BKB sentence lists (Bench et al., 1979). As the BKB-SIN test is administered in the presence of four-talker babble noise, the results are measured by SNR loss (discussed above) (Etymotic Research, Inc., 2005b).

Another sentence-based measure of paediatric speech recognition is the Hearing In Noise Test for Children (HINT-C; Nilsson, Soli, & Gelnett, 1996). The HINT-C was developed in order to evaluate children with profound HIs for cochlear implant candidacy (Nilsson et al., 1996). The HINT-C presents sentences in lists of 10, either in quiet or in combination with speech-shaped masking noise (Nilsson et al., 1996). Scoring for the HINT-C, as with the HINT, is sentence-based, requiring the child to correctly recall all of the key words in the sentence to qualify as a correct result (Nilsson et al., 1994; Nilsson et al., 1996). An adaptive procedure is utilised to determine the threshold at which the child achieves 50% correct (Nilsson et al., 1996).

An alternative sentence-based paradigm is that of matrix sentence tests (MSTs) which, owing to the cognitive demands of conventional Hagerman (1982) MSTs, has been adapted for paediatric use in both the German (Neumann et al., 2012) and Polish (Ozimek et al., 2012) languages by reducing the length of the sentences employed from five words to three, while retaining the integral structure of a MST. Both the Polish Paediatric Matrix Sentence Test (PPMST; Ozimek et al., 2012) and the Oldenburger Kinder-Satztest (OlKiSa; Neumann et al., 2012) were developed through the modification of their parent tests, the Polish Matrix Sentence Test (Ozimek et al., 2010) and the Oldenberge Satztest (OlSa; Oldenburg Sentence Test; Wagener, Kühnel, & Kollmeier, 1999a; Wagener, Brand, & Kollmeier, 1999b; Wagener, Brand, & Kollmeier, 1999c) respectively. These paediatric MSTs are discussed in further detail in section 1.12.

1.8 Sentence-Based Measures

The vast array of sentence-based speech recognition measures that are available can be separated into two main categories. The first category, referred to as 'Plomp-type' sentences (Nilsson et al., 1993; Plomp & Mimpen, 1979), use phonemically balanced sentences, based on meaningful day-to-day speech, yet possess no set grammatical structure (Dietz et al., 2014; Plomp & Mimpen, 1979; Nilsson et al., 1993). Plomp-type sentence tests have been developed for multiple different languages including American English (Nilsson et al., 1994), German, (Kollmeier & Wesselkamp, 1997), Dutch (Plomp & Mimpen, 1979; Versfeld et al., 2000), Swedish (Hällgren, Larsby, & Arlinger, 2006), French (Luts, Boon, Wable, & Wouters, 2008), and, more recently, Polish (Ozimek et al., 2009). Typically, Plomp-type sentence tests are comprised of lists of individual sentences that are both statistically and phonemically equivalent, with statistically insignificant differences across lists with regards to both list-specific SRTs and phonemic distribution (Plomp & Mimpen, 1979). The HINT is an example of a speech test that employs Plomp-type sentences (Nilsson et al., 1994). The test material is comprised of 25 phonemically balanced lists, each comprised of 10 sentences, which are administered in the presence of a spectrally matched masker (King, 2010). The HINT (Nilsson et al., 1994) has since been developed for use in various additional languages and dialects including New Zealand English (Hope, 2010), Swedish (Hällgren et al., 2006), and Cantonese (Wong & Soli, 2005). Notwithstanding the prevalence of Plomp-type tests, studies have shown a high degree of redundancy, which presents complications when repeated retesting is required (Wagener et al., 1999; Dietz et al., 2014).

The second category of sentence tests that may be distinguished is MSTs initially developed by Hagerman (1982) for the Swedish language. Hagerman (1982) intended to develop a standardised speech-in-noise measure that was reliable and efficient, and which offered sufficient speech material for use in HA evaluation. Matrix sentence tests use syntactically fixed but semantically unpredictable sentences, each composed of five words (name, verb, number, adjective, object). Test sentences are generated by choosing one of 10 alternatives for each word in order to form a sentence. For example (English translation; Hagerman, 1982, p. 80), "Karin gave two old buttons." As a result of the fundamental grammatical structure of each sentence, and the alternative word options in each column of the matrix, a total of 10^5 or 100,000 unique sentences can be generated (Hagerman, 1982; Hochmuth et al., 2012). This essentially unrestricted repertoire of sentences is an advantageous aspect of MSTs, as unlike monosyllabic word tests (e.g. the CVC meaningful word lists, which offer 10 lists of 10 words) it enables repeat testing while avoiding the possible implications associated with memorisation (Boothroyd & Nittrouer, 1988). Furthermore, the semantic unpredictability and low redundancy of MSTs eliminates the influence of contextual information on a listener's response (Hochmuth et al., 2012).

Wagener and colleagues (1999a-c) advanced Hagerman's (1982) original concept during development of the German (OlSa; Wagener et al., 1999a-c) and Danish (DANTALE II; Wagener, Josvassen, & Ardenkjær, 2003) versions of the MST, where the importance of co-articulation was considered so as to afford the synthesised sentences with a natural prosody. Other MSTs have since been developed in a number of languages including Dutch (Houben, Koopman, Luts, Wagener, van Wieringen, Verschuure, & Dreschler, 2014), Finnish (Dietz et al., 2014), Spanish (Hochmuth et al., 2012), and Polish (Ozimek et al., 2010). The constant structure of MSTs permits tests of different languages to be compared; similarities in reference intelligibility functions have been identified between the French, Danish, Dutch, and Polish MSTs (Zokoll, Hochmuth, Warzybok, Wagener, Buschermöhle, & Kollmeier, 2013). Despite these similarities, language-specific tests are required as a speaker's dialect and pronunciation can negatively impact a listener's performance

(Hochmuth et al., 2012). Consequently, in order to integrate a MST into the New Zealand context it was necessary to develop a MST using a native speaker of New Zealand English.

1.9 Development, Normalisation, and Evaluation of the University of Canterbury Auditory-Visual Matrix Sentence Test

1.9.1 **Overview**

As discussed previously, current clinical practice for speech audiometry in New Zealand is the presentation of monosyllabic word stimuli in quiet (Orchik, Krygier, & Cutts, 1979). However, the requirement for a New Zealand English MST that could advance the existing audiological test battery, integrate more representative measures of the listening difficulties encountered in everyday listening scenarios, and parallel international progress has been acknowledged. Trounson and O’Beirne (2012) developed the University of Canterbury Auditory-Visual Matrix Sentence Test (UCAMST) with the aim of realising these requirements.

Notwithstanding the availability of the British English MST (Hall, 2006), if used with the New Zealand population its validity would be impacted as a result of the inconsistencies in phonology evident between the two English dialects. When compared to other English dialects, New Zealand English differs in the raised place of production and the formant structure of vowels (Maclagan & Hay, 2007; Wells, 1982). These distinctions provide an explanation for the disparities in pronunciation between dialects, and thus the potential for misinterpretation (Trounson, 2012). Accordingly, certain words from the British English MST were deemed to be unsuitable for use in the UCAMST (Trounson, 2012). The word “tins”, for example, was reasoned to be unsuitable due to the possibility of New Zealand English listeners confusing “tins” with the word “tens” (Trounson, 2012). Previous studies have also shown a significant difference in speech recognition performance between ‘non-native' and native listeners, especially in suboptimal listening conditions, for example in the presence of an acoustic masker (Hochmuth et al., 2012; van Wijngaarden, Steeneken, & Houtgast, 2002; Zokoll et al., 2013).

Due to the evident restrictions surrounding the use of an English MST of a different dialect, development of a New Zealand English MST was necessary. The New Zealand English version was altered from the British English MST (Hall, 2006), to eliminate vowels with the potential to cause confusion for New Zealand listeners in open-set testing (Trounson, 2012). In addition, the word matrix was designed to: i) have an equal distribution of gender specific names across sentence lists; ii) have a fixed number of syllables within each word category, which matched the New Zealand English phoneme distribution; and iii) be grammatically correct and semantically neutral (Hochmuth et al., 2012; Trounson, 2012). Figure 3 illustrates the composition of the UCAMST base matrix.

Name	**Verb**	**Quantity**	**Adjective**	**Object**
Amy	bought	two	big	bikes
David	gives	three	cheap	books
Hannah	got	four	dark	coats
Kathy	has	six	good	hats
Oscar	kept	eight	green	mugs
Peter	likes	nine	large	ships
Rachel	sees	ten	new	shirts
Sophie	sold	twelve	old	shoes
Thomas	wants	some	red	spoons
William	wins	those	small	toys

Figure 3. The UCAMST base matrix.

1.9.2 **Rationale Behind the Auditory-Visual Component**

Unlike previous auditory-alone (AA) MSTs, the UCAMST also incorporated visual-alone (VA) and auditory-visual (AV) modes of presentation, displaying the speaker's face on a computer screen during presentation of the sentence. The inclusion of the AV mode of presentation was intended to increase the validity of the test, as in a real-world scenario listeners can often view the speaker's face during spoken discourse (Mattheyes, Latacaz, & Verhelst, 2009; Tye-Murray, Sommers, Spehar, Myerson, Hale, & Rose, 2008). Utilising cues from both auditory and visual listening modalities is understood to be especially effective in more difficult listening environments, irrespective of whether or not the listener has a HI (Tye-Murray, Hale, Spehar, Myerson, & Sommers, 2014; Tye-Murray et al., 2008; Tye-Murray, Sommers, & Spehar, 2007a). Studies have found that the amalgamation of auditory and visual speech material, when listening in noise, has the ability to significantly increase a listener's speech perception, compared to when speech material is presented in the AA modality (Spehar, Tye-Murray, & Sommers, 2008; Sumby & Pollack, 1954; Tye-Murray, Sommers, & Spehar, 2007b). Additionally, it is understood that as the speech signal weakens, a listener's dependence on visual cues increases considerably (Tye-Murray et al., 2007b). Consequently, previous literature has proposed that each of the three distinct listening modalities (AA, VA, and AV) be evaluated whilst assessing speech recognition capabilities, in order to provide potentially useful diagnostic information (Tye-Murray et al., 2007b).

Based on this premise, all three presentation modalities were incorporated into the design of the UCAMST during its development (Trounson, 2012). Permitting the selection of the modality through which the stimulus is presented allowed the test procedure to be customised based on the objective of the assessment. Consequently, an individual's ability to integrate information from each of the modalities, in turn and in combination, can be examined. It is anticipated that such information will provide an indication of the particular areas contributing to these communication difficulties, and thus be beneficial with regards to forming rehabilitative recommendations (Tye-Murray et al., 2007b).

1.9.2.1 Auditory-Visual Integration

Auditory-visual integration (AVI), commonly referred to as 'speechreading', is a cognitive process in which an individual integrates auditory and visual input information in order to enhance speech perception (Grant & Seitz, 1998). Previous research has differentiated AVI from both visual and auditory speech perception, with higher speech perception scores being found to occur at lower intensities in the auditory-visual condition as compared to the auditory-alone and visual-alone conditions (Tye-Murray et al., 2007a; Most, Rothem, & Luntz, 2009). Furthermore, a listener's ability to integrate auditory and visual inputs has been found to impact speech perception in the mid to high frequency range (i.e. the place of articulation for most consonants), while amplification has been found to provide the most benefit in the low to mid frequencies (i.e. the place articulation and manner of voicing for most vowels) (Walden, Grant, & Cord, 2001). Consequently, AVI and amplification (or NH) may be seen to work in a complementary fashion to enhance speech perception (Dillon, 2012).

The term audio-visual enhancement (AVE) refers to an individual's ability to integrate visual and auditory input information (Tye-Murray et al., 2007a; Tye-Murray & Geer, 2001). Models of audio-visual speech perception suggest that visual enhancement (i.e. the benefit acquired from the presence of both the auditory and visual inputs of a speech signal) is established by an individual's ability to: i) speechread, ii) encode auditory information, and iii) integrate the information acquired from both auditory and visual inputs (Grant, Walden, & Seitz, 1998; Tye-Murray et al., 2007a).

An individual's AVE can be scored by determining the difference between the SNRs required to obtain equivalent percentage correct scores in the AA and AV conditions, when target stimuli are presented in speech-in-noise (MacLeod & Summerfield, 1987). Another method for obtaining and expressing an individual's AVE score involves presenting the target stimuli at the same SNR for the AA and AV conditions, in the presence of speech-in-noise, and examining the difference between the percentage correct scores for each condition (Grant & Seitz, 1998). The latter method was adopted in the current study. However, this method harbours problems, as the percentage correct score for the auditory-alone condition limits the potential AVE. Consequently, normalisation, as depicted in equation (1) (Tye-Murray et al., 2007a) was required.

$$AVE = \frac{(AV - AA)}{(1 - AA)} \tag{1}$$

Note. AVE = auditory-visual enhancement; AV = percentage correct score in auditory-visual condition; AA = percentage correct score in auditory-alone condition. Equation retrieved from Tye-Murray et al., (p. 661, 2007a).

Previous literature has indicated, among other factors, that the age of the individual and the type of stimulus utilised can influence measures of AVE (Rogers, 2012; Sommers, Tye-Murray, & Spehar, 2005; Tye-Murray et al., 2008). The employment of sentence stimuli, over individual words or consonants, has been advised due to the differences observed between the results obtained for each stimulus type, and the greater propensity for sentence-based stimuli to provide a better measure of real-world AVE (Rogers, 2012; Sommers et al., 2005; Tye-Murray et al., 2008). With regards to age-related factors, previous research has found that, for children with NH, older children are able to utilise more visual input information than younger children, as AVI abilities continue to develop throughout childhood (Dick, Solodkin, & Small, 2010; Massaro, Thompson, Barro, & Laren, 1986).

1.9.3 **Recording and Editing**

The methodology used in the development of the UCAMST was identical to that used in the development of its predecessors. Accordingly, the five-word sentences follow the same format as previous MSTs (name, verb, quantity, adjective, object), and are generated from a 5 by 10 word base matrix. The recording method employed in the development of the UCAMST originated, as mentioned previously, during the development of the Danish MST (Wagener et al., 2003), where, unlike the Swedish MST (Hagerman, 1982), the importance of co-articulation was considered so as to afford the synthesised sentences with a natural prosody. In the development of the Danish MST (Wagener et al., 2003) 100 five-word sentences were recorded in a manner that ensured that each word in a particular column was recorded in combination with every word in the neighbouring columns. Consequently, the implementation of the Danish technique in the development of the UCAMST ensured 10 co-articulation specific events for each word in the base matrix. Figure 4 provides an illustration of this method, as exhibited for Index 0 (English translation of the Danish MST; Wagener et al., 2003). Sentences were recorded using this technique for all of the residual indices.

Index	*Name*	*Verb*	*Numeral*	*Adjective*	*Object*
0	Anders	owns	ten	old	jackets
1	Birgit	had	five	red	boxes
2	Ingrid	sees	seven	nice	rings
3	Ulla	bought	three	new	flowers
4	Niels	won	six	fine	cupboards
5	Kirsten	gets	twelve	lovely	masks
6	Henning	sold	eight	beautiful	cars
7	Per	borrows	fourteen	big	houses
8	Linda	chose	nine	white	presents
9	Michael	finds	twenty	funny	plants

Figure 4. Sentence recording method employed in the development of the Danish MST (English translation of the Danish MST; *Wagener et al. (2003, p. 13)*). *Copyright 2016 by Taylor and Francis. Reprinted with permission.*

For the UCAMST, all sentence material was recorded from an actress with a New Zealand English accent (Trounson, 2012). Upon completion of the recording procedure, the recording was edited into 400 file fragments containing distinctive word pairs. These fragments could then be used to generate 100,000 unique five-word sentences. Figure 5 illustrates the alternative editing processes in which complete sentences were generated from the file fragments.

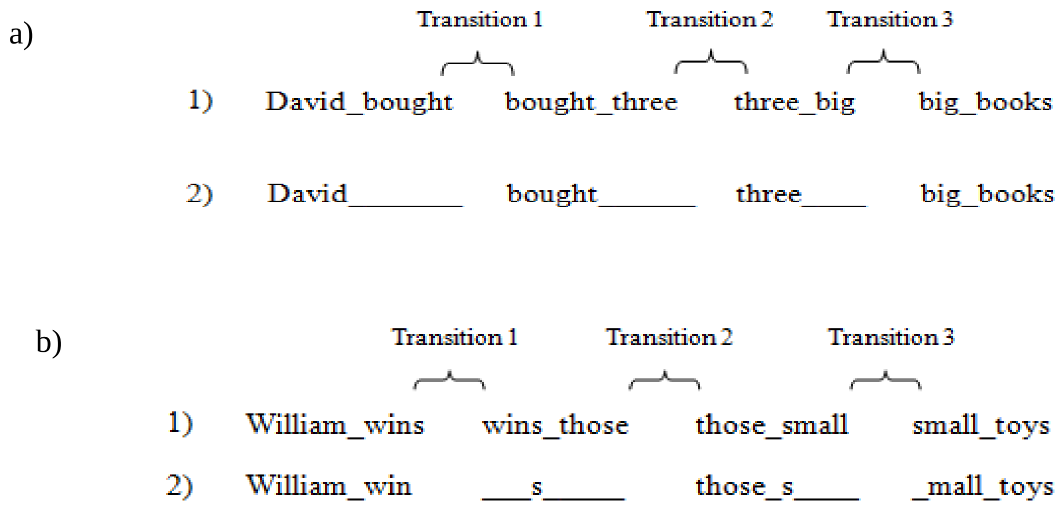

Figure 5. Two examples of the way in which sentences are formed from four file fragments. Each example depicts 1) the four file fragments required to generate the sentence, and 2) the precise audio material used from each fragment. *Image retrieved from McClelland (2015, p. 25).*

As indicated, each sentence generated includes three distinct ‘transitions’ in which one file fragment changes to another. However, the precise location at which these transitions occur is not fixed due to editing revolving around the smoothness and naturalness of the transitions (Trounson, 2012). Although this method has borne significant advances in the quality of the generated sentences, some unnatural sounding sentences have been found to persist (Hochmuth et al., 2012; Houben et al., 2014). Sentences containing these unnatural sounding audio artefacts were not included in the UCAMST final sentence lists. The UCAMST also encountered the added challenge of ensuring that the visual component, in addition to the auditory component, was observed to be natural. Although multiple measures were employed to minimise possible irregularities in the visual recording (Trounson, 2012), a substantial number of the sentences generated showed an evident and unnatural jerk. This jerk artefact was termed 'judder', and referred to discrepancies in the visual component, primarily related to differences in the actress' head position between transitions. Accordingly, the judder artefact necessitated additional inquiry.

1.9.4 **Selection of Sentence Stimuli: Visual Considerations**

As the UCAMST sentences are synthesised from the base matrix, the naturalness of both the auditory and visual components can be impacted during editing or reassembly of the sentences. The naturalness of the visual component of the UCAMST was initially assessed objectively during its development by calculating and comparing the absolute difference in pixel values between the images – the smaller the absolute difference, the smoother the transition between images (Trounson, 2012). Since then, subjective quantifications of naturalness have also been examined, and the absolute pixel difference value has been found to be a significant predictor of subjective rating scores (McClelland, 2015). In order to confirm that the sentence stimuli employed in both the visual and auditory conditions were adequate for presentation as test material, McClelland (2015) evaluated the noticeability of this judder. The judder was subjectively evaluated by listeners with NH based on a 10-point noticeability rating scale, in which a score of 0 represented "no noticeable judder", and a score of 10 represented "highly noticeable judder" (McClelland, 2015). In addition to the "synthesised" test sentences, unedited control sentences, in which no judder occurred, were also presented (McClelland, 2015). The completed sentence material for the UCAMST contained the sentences whose rating scores revealed the least noticeable judder, in addition to the control sentences (McClelland, 2015). The application of this method ensured an adequate final set of sentences appropriate for testing in the visual modality (McClelland, 2015).

1.9.5 **Generation of Acoustic Maskers**

Currently, the UCAMST can be presented in conjunction with two distinct types of masking noise, constant speech-shaped noise (constant noise), and six-talker babble noise (babble noise). The constant noise was produced specifically for the UCAMST by an automated process in which audio recordings were arbitrarily overlaid 10,000 times, generating constant noise with spectral content nearly identical to that of the signal (i.e. the signal and the noise were spectrally matched) (King, 2010). The babble noise was initially produced for an earlier University of Canterbury Master's research project (Spencer, 2011), and was generated though the superimposition of recordings of 20, 6 to 10 word semantically irregular sentences, which were read by six native speakers of New Zealand English (three females and three males).

1.9.6 **Normalisation of the UCAMST Sentences**

The normalisation and naturalness of the synthesized sentences generated by the UCAMST has been the major focus of research pertaining to the test since its development (McClelland, 2015; Stone, 2016). Previous research has endeavoured to optimise measures of speech recognition by attaining high equivalence of the test stimuli (Akeroyd, Arlinger, Bentler, Boothroyd, Dillier, Dreschler, & Kollmeier, 2015; Kollmeier, Warzybok, Hochmuth, Zokoll, Uslar, Brand, & Wagener, 2015; McClelland, 2015). In order to optimise such measures, word-specific intelligibility functions must first be obtained for each of the words recorded; this is typically accomplished by presenting the speech materials at fixed SNRs to listeners with NH (Akeroyd et al., 2015). In this manner, recordings with exceedingly low or high intelligibility can be identified, and level adjustments completed, so as to generate more comparable intelligibility functions (Akeroyd et al., 2015; Kollmeier et al., 2015). Generally, it is recommended that those materials that fail to adequately fit the word-specific intelligibility function be rejected (Kollmeier et al., 2015).

With regards to normalisation of the speech stimuli employed by the UCAMST, McClelland (2015) assessed participants with NH and evaluated the SRTs and slopes of the psychometric functions for each of the 400 recorded sentence fragments. The speech stimuli were administered in babble noise and constant noise at fixed SNRs of -8 dB, - 11.5 dB, -15 dB, and -18.5 dB (McClelland, 2015). However, unlike previous MSTs (i.e. Wagener et al., 2003), the UCAMST presented a unique challenge as each word realisation was not contained within a single file fragment, thus preventing the simultaneous normalisation of both words and fragments (McClelland, 2015). As word realisations were, in some instances, held over two file fragments, level adjustments could be employed to equalise the intelligibility of either the specific word (‘word-specific normalisation’) or the file fragment (‘fragment-specific normalisation’) (McClelland, 2015). Consequently, normalisation was separated into two distinct sections, word-specific normalisation and fragment-specific normalisation. Word-specific normalisation supports the notion that the principal determinant of the intelligibility of a specific word at a specified SNR is determined by the word’s acoustic characteristics, as opposed to the speaker’s presentation of that word (McClelland, 2015). This technique was considered to produce more reliable level adjustments, compared to fragment-specific normalisation, as word-specific normalisation is able to provide access to 10 times more raw psychometric data (McClelland, 2015).

1.9.7 **Word- and Fragment-Specific Normalisation**

In order to normalise the UCAMST stimulus material, fragment-specific intelligibility functions were produced, thus allowing the homogeneity of these functions to be assessed (McClelland, 2015). Psychometric intelligibility functions were generated for the individual file fragments in both constant and babble noise, based on calculations of the mean intelligibility (%) for each fragment across SNRs, and fit to the logistic model depicted in equation (2) (McClelland, 2015). Based on the normalisation procedures of previously published international MSTs (i.e. Dietz, 2014; Hochmuth et al., 2012; Houben et al., 2014; Ozimek et al., 2012), a conservative adjustment limit of ± 3 dB was employed (McClelland, 2015).

Fragment-specific normalisation of the UCAMST was first conducted in the presence

$$SI(L) = \frac{1}{A}\left(\frac{(1 + SI_{max}).(A-1)}{1 + \exp\left(\left[-4.5.\frac{slope}{100}\right].[L - L_{mid}]\right)}\right) \tag{2}$$

Note. SI = speech intelligibility; L = level; L_{mid} *= midpoint;* SI_{max} *= function ceiling; A = number of alternatives;* $\frac{1}{A}$ *= function floor.*

Equation adapted from Kollmeier and Wesselkamp (1997), and Wagener et al. (2003). *Retrieved from Stone (p. 30, 2016).*

of constant noise. Of the 400 file fragments examined, 4% (i.e. 15 file fragments) were identified as unacceptable, and excluded from the final set of stimuli (McClelland, 2015). The persisting fragments generated a mean pre-normalisation midpoint (50% intelligibility) of -10.3 dB SNR (± 2.1 dB SD) (McClelland, 2015; Stone, 2016)[1]. Following integration of the pre-normalisation word-specific intelligibility functions, the data were able to be normalised (McClelland, 2015). In order to obtain increased homogeneity of the post-normalisation functions and improved alignment of the midpoints, the midpoints of the individual word-specific intelligibility functions, pre-normalisation, were modified to match the pre-normalisation mean fragment midpoint of -10.3 dB SNR (McClelland, 2015).

---

[1] Based on the recalibration procedure outlined by Stone (2016) all SNR values quoted from McClelland (2015) have been corrected.

However, level adjustments exceeding the prescribed limit were required, post-normalisation, in order for the intelligibility functions of the words “ships” and “shirts” to achieve acceptable overlap with the remaining post-normalisation functions (McClelland, 2015). Irrespective of this, the mean word-specific midpoint, post-normalisation, in constant noise was expected to be -10.1 dB SNR (± 0.8 dB SD), thus signifying a reduction in the SD of the word-specific midpoint measures intended for use in constant noise of 1.6 dB (McClelland, 2015).

Following the same method as outlined above, the speech stimuli intended for administration in the presence of babble noise were normalised. In the babble noise condition, fragment-specific normalisation resulted in the exclusion of 12% (i.e. 47) of the fragments (McClelland, 2015). The remaining fragments showed a mean midpoint of -11.0 dB SNR (± 2.9 dB SD), revealing that recognition of the UCAMST test stimuli was less challenging in the babble noise condition, as compared to the constant noise condition. As with the normalisation of the word-specific functions for the constant noise condition, the word-specific intelligibility functions were then integrated. Following evaluation of the midpoint for the individual word-specific intelligibility functions, 20 words (i.e. 41% of the original) were found to require level adjustments exceeding the prescribed limit (McClelland, 2015).

Following normalisation, the UCAMST stimuli administered in the presence of constant noise exhibited greater overlap for the predicted post-normalisation intelligibility functions than those administered in babble noise (McClelland, 2015). This difference was affirmed to be a consequence of the lesser proportion of words that required level adjustments exceeding the limit in the constant noise condition (McClelland, 2015). Notwithstanding these disparities in adjustment, the midpoint post-normalisation mean for the test stimuli administered in the presence of babble noise was found to be -11.0 dB SNR

(± 1.9 dB SD), signifying a decrease of 1.7 dB in the SD of the word-specific midpoint measures (McClelland, 2015).

1.9.8 **UCAMST: List Equivalence**

In addition to the normalisation of the individual sentence fragments, the sentence lists themselves have also been examined for equivalence (Stone, 2016). Lists presented in constant speech noise have been found to be equivalent with respect to both SRT and slope, irrespective of response format (Stone, 2016). Based on the consistent method used in the development of MSTs internationally, this result was anticipated, and signifies the capability of the UCAMST sentence list stimuli to be administered in both the open and closed-set response formats interchangeably (Akeroyd et al., 2015; Stone, 2016). However, while lists administered in the presence of babble noise appeared to have comparable SRTs across list stimuli, in both response formats the slope of the psychometric functions were found to vary based on the list presented (Stone, 2016). Consequently, since slope equivalence across sentence lists provides greater confidence in the reliability of SRT estimation, as the UCAMST currently stands, a listener’s SRT may be reliably estimated when the test is administered in the presence of constant noise, but not in the presence of babble noise (Stone, 2016). Concurrent to the current research project, Ripberger (in progress) conducted a study in which the normalisation process for the babble noise condition was continued in order to rectify this issue. Additionally, Ripberger (in progress) endeavoured to evaluate and normalise the sentence list stimuli intended for use in the constant noise condition in the absence of masking noise (i.e. quiet).

1.9.9 **UCAMST: Comparison to International MSTs**

Due to the uniform structure and common methodological standards employed in the development of MSTs internationally, it is possible to compare international MSTs of various languages. Stone (2016) examined the equivalence of the UCAMST stimuli with the previously published international MSTs for the Finnish (Dietz et al., 2014), Dutch (Houben et al., 2014), French (Jansen, Luts, Wagener, Kollmeier, Del Rio, Dauman, James, Fraysse, Vormès, Frachet, Wouters, & van Wieringen, 2012), Norwegian (Øygarden, 2009), Polish (Ozimek et al., 2010), Italian (Puglisi, Warzybok, Hochmuth, Astol, Prodi, Visentin, & Kollmeier, 2014), Danish (Wagener et al., 2003), and Russian (Warzybok, Zokoll, Wardenga, Ozimek, & Boboshko, 2015) languages. Statistically significant differences were found to exist between the previously published international MSTs included in the analysis and the UCAMST stimulus lists with regards to both SRT and slope (Stone, 2016). As illustrated by Figure 6, when compared to the international MSTs analysed, the UCAMST was seen to have a shallower mean slope, which, as the slope of the psychometric function at the SRT primarily determines the accuracy of the SRT, may have implications concerning the accuracy of the estimates of SRT achieved using the UCAMST.

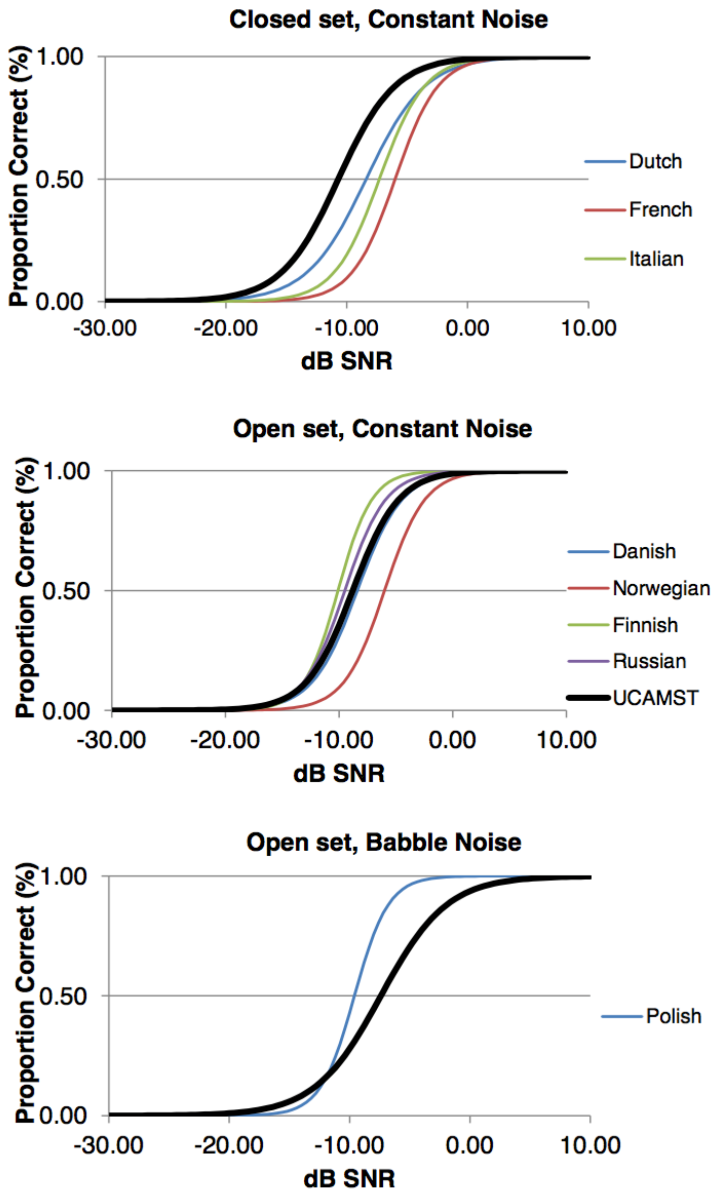

Figure 6. Comparison across international MST versions and the UCAMST with regards to slope. *Retrieved from Stone (2016, p. 76).*

One exception to this finding, however, was the Danish MST (Wagener et al., 2003), which was found to be equivalent to the UCAMST with regards to SRT in the open-set, constant noise condition (Stone, 2016). Nevertheless, the overall findings indicated that the UCAMST speech recognition results are not, as of yet, comparable to those results obtained from international MSTs (Stone, 2016). Table 1 illustrates the distinctions reviewed above, and affords a comparison of the international MSTs and the UCAMST with regards to both SRTs (dB SNR) and slope (%/dB).

Table 1. Mean SRT and slope values of international MSTs.

	MST	***M* SRT (dB SNR)**	***M* Slope (%/dB)**	**Authors**
Constant Noise, Closed Set	Dutch	-8.4 ± 0.2	10.2 ± 0.9	Houben et al. (2014)
	French	-6.0 ± 0.1	14.0 ± 1.6	Jansen et al. (2012)
	Italian	-7.3 ± 0.2	13.3 ± 11.2	Puglisi et al. (2014)
	Malay AV	-10.1 ± 0.2	14.9 ± 1.2	Jamaluddin (2016)
	UCAMST	**-10.7 ± 0.2**	**10.6 ± 0.9**	**Trounson (2012); McClelland (2015); Stone (2016)**
Constant Noise, Open Set	Danish	-8.4 ± 0.2	12. 6 ± 0.8	Wagener et al. (2003)
	Finnish	-10.1 ± 0.1	16.7 ± 1.2	Dietz et al. (2014)
	Norwegian	-6.0 ±0.8	14.0 ± 1.6	Øygarden (2009)
	Russian	-9.5 ±0.2	13.8 ± 1.6	Warzybok et al. (2015)
	UCAMST	**-8.8 ± 0.4**	**12.3 ± 1.5**	**Trounson (2012); McClelland (2015); Stone (2016)**
Babble Noise, Closed Set	Malay AV	-6.4 ± 0.2	12.2 ± 0.7	Jamaluddin (2016)
	UCAMST	**-9.8 ± 0.6**	**7.3 ± 1.4**	**Trounson (2012); McClelland (2015); Stone (2016)**
Babble Noise, Open Set	Polish	-9.6 ± 0.2	17.7 ± 1.6	Ozimek et al. (2010)
	UCAMST	**-7.4 ± 0.5**	**9.1 ± 2.1**	**Trounson (2012); McClelland (2015); Stone (2016)**

Note. M = mean; ± (x) = SD. Adapted from Stone (2016, p. 74).

Prior evaluations across international MSTs have revealed an apparent acceptance range for the SRT and slope values; accordingly the level by which the UCAMST diverges from these versions can be appreciated. Kollmeier and colleagues (2015) compared the reference SRT values across international MSTs exposing acceptable differences between each version. A range of 4.1 dB SNR was observed across the reference SRT values for the international MSTs analysed, with the highest SRT (-6 dB SNR) reported for the French (Jansen et al., 2012) and Norwegian (Øygarden, 2009) MSTs, and the lowest SRT (-10.1 dB SNR) reported for the Finnish (Dietz et al., 2014) and Malay AV (Jamaluddin, 2016) versions (Kollmeier et al., 2015). Although statistically significant differences were found between the stimulus lists for the UCAMST and the international MSTs analysed for each condition, Stone (2016) indicated that, upon inspection of the accompanying intelligibility functions, such disparities appeared to be "marginal", especially in the open-set, constant noise condition. Stone (2016) also concluded that the differences across the mean SRT and slope values reported for the various conditions of the UCAMST were the consequence of rounding error.

Due to the standardised methodology employed in the development of MSTs internationally, several explanations have been considered to account for the differences observed across different versions. It has been postulated that the variation in the SRTs of international MSTs may be associated with the specific characteristics of the speaker, such as gender (Kollmeier et al., 2015). Following the optimisation and evaluation of two distinct versions of the German MST, one employing a male speaker (Wagener et al., 1999) and the other a female speaker (Wagener et al., 2014), a difference in SRT of 2.2 dB SNR was observed (Kollmeier et al., 2015). This disparity in SRT supports the notion that the characteristics of the speaker could potentially impact the homogeneity of MST versions (Wagener et al., 2014). Consequently, with regards to the comparison of the UCAMST to previously published MSTs, it is possible that the degree to which the UCAMST was found to differ may be augmented by the fact that that UCAMST employed a female speaker, while comparative versions, such as the Norwegian (Øygarden, 2009) and Polish (Ozimek et al.,

2010) MSTs, used male speakers. Alternatively, it has been proposed that the attributes of the language itself, such as phoneme frequency, can influence reference SRT values (Kollmeier et al., 2015). Lower SRTs have been observed for the MSTs of the Polish (Ozimek et al., 2010) and Russian (Warzybok et al., 2015) languages; this is hypothesized to be due to high frequency phonemes being more challenging to mask and, consequently, affording a phonetic cue, potentially resulting in lower reference SRT values (Kollmeier et al., 2015).

1.10 Selecting a Response Format

Another aspect of relevance to the current research project was the response format in which the UCAMST test stimuli are administered. The test stimuli employed by MSTs can be presented in either the closed-set response format, in which response options are chosen from a visible word matrix, or the open-set response format, in which verbal responses are made in the absence of such visual cues (Hochmuth et al., 2012). Using a closed-set response format can prove advantageous, as it eliminates the requirement for the presence and participation of a researcher or clinician in the testing process (Hochmuth et al., 2012).

The findings of previous research regarding the impact of the response format on test results have been mixed. Hochmuth et al. (2012) identified a significant difference between the SRTs acquired using open- versus closed-set response formats, as did Stone (2016). Conversely, in the study conducted by Ozimek et al. (2010) no significant differences were identified between the SRTs obtained using the different response formats. However, it is plausible that these differences were based on the extent of the training provided prior to administration of the test materials. For example, in the Ozimek et al. (2010) study, it is possible that, due to the extensive hour-long training session implemented, participants were better acquainted with the test stimuli, resulting in an improvement in global performance. However, during practical clinical presentations of matrix style sentence tests the inclusion of an hour-long training session is not feasible. Thus, in the current study, although training was presented prior to each new test condition, the training was not as extensive as that applied by Ozimek et al. (2010). Consequently, it was anticipated that the SRTs for tests administered utilising the closed-set response format would occur at lower SNRs than those presented using the open-set response format (Hochmuth et al., 2012; Stone, 2016).

With regards to the paediatric population, it is worth noting that the use of an open-set response format is not always appropriate. Speech recognition tests administered using a closed-set response format provide the child with a restricted number of response alternatives for comparison, while an open-set response format requires the child to compare the stimulus material to each of the word possibilities in their lexical memory (Clopper, Pisoni, & Tierney, 2006). Furthermore, if a child is shy, reluctant to respond, or unable to produce discernable speech, employing a test that uses a closed-set response format may lessen the intimidation and expedite the assessment process.

1.11 Study Rationale

Owing to the cognitive demands of conventional Hagerman (1982) MSTs, which comprise a 5 by 10 word matrix, the current research project endeavoured to modify the existing UCAMST in order to produce an audiometric speech recognition test, in New Zealand English, suitable for the paediatric population. As discussed in greater detail in previous sections, modification to produce a speech recognition measure better suited to the paediatric population was paramount in order to minimise the impacts of working memory and fatigue associated with sentence length, as well as to increase the reliability and sensitivity of such measures (Neumann et al., 2012). In addition to the development of a paediatric MST in New Zealand English, the evaluation of this MST was vital, in order to establish the reliability and sensitivity of the MST in estimating SRTs, and, thus, progress the test towards a clinical application.

It is also important to note that previous research pertaining to the evaluation and normalisation of the UCAMST was only conducted for the AA condition, based on preliminary findings regarding the Malay version of the UCAMST (Jamaluddin & O'Beirne, 2015). These findings revealed that when sentences were presented at unfavourable SNRs, the AV condition was equivalent to the VA condition, and thus it was evident that in both conditions listeners were solely reliant on visual cues (Jamaluddin & O'Beirne, 2015). Therefore, the evaluation of the AV component of the UCAMST in the current study was also essential in order to progress the test towards its intended clinical application as a section of the University of Canterbury Adaptive Speech Test (UCAST; O'Beirne, McGaffin, & Rickard, 2012) platform. The UCAST aims to encompass an assortment of audiological tests, including the New Zealand Digit Triplet Test (NZDTT; King, 2011), that are available for both research and clinical applications (O'Beirne et al., 2012).

1.12 Development of the University of Canterbury Auditory-Visual Matrix Sentence Test - Paediatric

The development of the University of Canterbury Auditory-Visual Matrix Sentence Test - Paediatric (UCAMST-P) followed methodology similar to that employed in the development of previously published paediatric MSTs. The following sections will consider such methodology, in addition to the results obtained.

1.12.1 Polish Paediatric MST

Ozimek, Kutzner, and Libisezweski (2012) developed the PPMST by adapting the Polish Matrix Sentence Test (Ozimek et al., 2010) in two main ways (Ozimek et al., 2012). As paediatric sentence tests generally contain 3 words and utilise a simple sentence structure, the matrix was reduced from five columns to three, allowing sentences with a fixed grammatical structure (subject, verb, object) to be synthesised (Ozimek et al., 2012). For example (Ozimek et al., 2012, p. 1123): "*Babacia maluje dom*." (English translation: *"Grandma is painting a house.")*.

In order to prevent synthesis of nonsense sentences, the 16 by 3 matrix was modified to create four separate and independent four by three sub-matrices from which sentences would be synthesised (Ozimek et al., 2012). Normative data, based on verbal responses, uncovered an age effect and established that this version of the PPMST was suitable for children seven years of age and older (Ozimek et al., 2012). A further test, utilising a picture pointing response, was also developed for use with children from three to six years of age, as younger children may find the judgement or execution of a verbal response challenging (Ozimek, et al., 2012). This task incorporated a six-picture array containing the picture corresponding to the sentence presented along with associated alternatives (Ozimek et al., 2012). Significantly higher SRTs were found for children with HIs than those with NH when the sentences were administered in noise (Ozimek et al., 2012).

1.12.2 **German Paediatric MST**

The OlKiSa, a paediatric matrix-style sentence test developed for the German language, utilises pseudo-sentences (number, adjective, object) and has been validated for use with children between the ages of 4 and 10 years (Neumann et al., 2012). The test uses an open-set response format in which children are required to repeat back what was heard. Comparable to the PPMST (Ozimek et al., 2012), an age effect (although not statistically significant) was found between listeners, with younger children in the first year of primary school scoring 1-2 dB higher than children in the second, third, and fourth years of primary school (Wagener & Kollmeier, 2005). The OlKiSa was developed through modification of the OlSa (Wagener et al., 1999a-c), which was devised based on the conventional Hagerman

(1982) matrix-style sentence test for use with both adults and children. However, the results achieved by primary school aged children were found to be less reliable than those achieved by adults (Wagener, Eeenboom, Brand, & Kollmeier, 2005). This discrepancy is believed to be due to the reduced memory span of primary school aged children (Neumann et al., 2012).

1.13 Evaluation of the UCAMST and the UCAMST-P

As with the evaluation of the AA component of the UCAMST (Stone, 2016), the UCAMST-P and the AV component of the UCAMST were evaluated following the guidelines stipulated by Akeroyd et al., (2015) and the methodology employed by previously published international MSTs, as discussed by Stone (2016).

1.14 Aims and Hypotheses

The current thesis project aimed to develop a paediatric version of the UCAMST, the UCAMST-P, and evaluate the difficulty, in each presentation condition, of the sentence lists generated. In order to evaluate list equivalence, the current research project endeavoured to answer the following three research questions:

1) Are the tests lists equivalent in each condition (i.e. AA, open-set; AA, closed-set; AV, open-set; AV, closed-set) with regards to SRT and slope for the:
    a. UCAMST-P
    b. UCAMST
2) Are the open- and closed-set response formats equivalent within each mode of presentation (i.e. AA, open-set vs. AA, closed-set; AV, open-set vs. AV, closed-set) with regards to SRT and slope for the:
    a. UCAMST-P
    b. UCAMST

3) In each of the four test conditions (i.e. AA, open-set; AA, closed-set; AV, open-set; AV, closed-set) is there a significant difference between when the condition is preceded by training and when the condition is not preceded by training for the:
    a. UCAMST-P
    b. UCAMST

Based on previous findings, the ensuing hypotheses were proposed for the current research project:

For research question (1a):

1) That no significant differences would be found between the UCAMST-P sentence lists with regards to <u>SRT</u> in the:
    a. AA, open-set condition
    b. AA, closed-set condition
    c. AV, open-set condition
    d. AV, closed-set condition
2) That no significant differences would be found between the UCAMST-P sentence lists with regards to <u>slope</u> in the:
    a. AA, open-set condition
    b. AA, closed-set condition
    c. AV, open-set condition
    d. AV, closed-set condition

For research question (1b):

3) That no significant differences would be found between the UCAMST sentence lists with regards to <u>SRT</u> in the:
    a. AA, open-set condition

b. AA, closed-set condition

c. AV, open-set condition

d. AV, closed-set condition

4) That no significant differences would be found between the UCAMST sentence lists with regards to slope in the:

    a. AA, open-set condition

    b. AA, closed-set condition

    c. AV, open-set condition

    d. AV, closed-set condition

For research question (2a):

5) That no significant differences would be found between the open-set and closed-set response formats of the UCAMST-P with regards to SRT in the:

    a. AA mode of presentation

    b. AV mode of presentation

6) That no significant differences would be found between the open-set and closed-set response formats of the UCAMST-P with regards to slope in the:

    a. AA mode of presentation

    b. AV mode of presentation

For research question (2b):

7) That no significant differences would be found between the open-set and closed-set response formats of the UCAMST with regards to SRT in the:

    a. AA mode of presentation

    b. AV mode of presentation

8) That no significant differences would be found between the open-set and closed-set response formats of the UCAMST with regards to slope in the:

a. AA mode of presentation

b. AV mode of presentation

For research question (3a):

9) That for the UCAMST-P no significant differences would be found between when the condition is preceded by training and when the condition is not preceded by training with regards to SRT in the:

   a. AA, open-set condition

   b. AA, closed-set condition

   c. AV, open-set condition

   d. AV, closed-set condition

10) That for the UCAMST-P no significant differences would be found between when the condition is preceded by training and when the condition is not preceded by training with regards to slope in the:

   a. AA, open-set condition

   b. AA, closed-set condition

   c. AV, open-set condition

   d. AV, closed-set condition

For research question (3b):

11) That for the UCAMST no significant differences would be found between when the condition is preceded by training and when the condition is not preceded by training with regards to SRT in the:

   a. AA, open-set condition

   b. AA, closed-set condition

   c. AV, open-set condition

   d. AV, closed-set condition

12) That for the UCAMST no significant differences would be found between when the condition is preceded by training and when the condition is not preceded by training with regards to slope in the:

    a. AA, open-set condition
    b. AA, closed-set condition
    c. AV, open-set condition
    d. AV, closed-set condition

CHAPTER TWO:

METHODS

2.1 Overview

As discussed in Chapter one, the purpose of the current research project was to develop a paediatric version of the UCAMST, the UCAMST-P, and evaluate the newly developed test, alongside its parent test, in each presentation condition in order to establish the reliability and sensitivity of the MSTs in estimating SRTs.

Ethical approval for the current research was acquired on 20 March 2017 from the University of Canterbury Human Ethics Committee prior to the research commencing (see Appendix A for a copy of the letter of approval). The procedures employed in the current research were conducted in a manner compliant with those proposed in the ethics application.

2.1 Participants

2.1.1 **Recruitment**

Participants were primarily recruited from the University of Canterbury Department of Communication Disorders via the circulation of an email invitation, outlining the nature of the current research project as well as the inclusion criteria of participants (see Appendix B). Based on G*Power 3.1 calculations the current study required the involvement of 31 participants in order to afford sufficient statistical power. However, due to the testing design employed, outlined in section 2.6 below, 40 participants were needed.

In order to preserve the validity of the findings of the current research project, the following inclusion criteria were employed. First, as previous research has found that when listening to a non-native speaker a listener’s speech intelligibility can be significantly reduced, it was essential that participants be native speakers of New Zealand English so as to ascertain the appropriate application of the UCAMST in the New Zealand context (van Wijingaarden et al., 2002; Zokoll et al., 2013). Second, due to the time and, consequently, attentional requirements of the current study, participants were required to be adults (≥ 18 years of age) as previous literature has found that an individual's capacity to pay constant attention to a task continues to advance through adolescence (Betts, Mckay, Maruff, & Anderson, 2006). The third inclusion criterion, that participants had NH (in accordance with Goodman, 1965), was implemented in order to prevent the data being confounded by HI (Akeroyd et al., 2015). This inclusion criteria specifically stipulated exclusion if an air bone gap (ABG) of ≥ 15 dB HL was identified, due to its propensity to indicate an existing middle ear pathology and, therefore, a possible temporary or permanent threshold shift (Hussain, 2008). Fourth, due to the inclusion of the auditory-visual mode of presentation and the closed-set response format, both of which require visual discernment, participants were required to have good visual acuity (with or without the use of corrective lenses). Finally, as the closed-set response format required the selection of words via a touch screen, participants were required to have no chronic dexterity problems that inhibited these movements.

An honorarium of a $20 Motor Trade Association voucher was presented to all participants in reparation for their time, irrespective of whether a HI was identified during the initial hearing screening.

2.2 Stimuli

As with previous theses conducted on the UCAMST (McClelland, 2015; Stone, 2016; Trounson, 2012), the sentence stimuli for both the UCAMST and the UCAMST-P were presented bilaterally at 65 dB SPL. As per the signal calibrations established by Stone (2016), for each SNR acquired in the current study, 3.85 dB SPL was added. The sentence lists for both the UCAMST and the UCAMST-P were administered in the presence of constant masking noise, at two SNRs, for both the open-set (-11.6 dB SNR and -6.0 dB SNR) and closed-set (-14.0 dB SNR and -7.4 dB SNR) response formats. In order to guarantee that an equal proportion of sentences were administered at both SNRs, each SNR was randomly allocated to half of the sentences in each list for both the UCAMST and the UCAMST-P. These SNRs were employed as a means to approximate the pair of compromise (i.e. the points at which 80% and 20% scores are expected to be obtained) (Brand & Kollmeier, 2002). This method was selected as it allows concurrent estimates of the slope and the psychometric function to be generated, from which the SRT can also be derived (Brand & Kollmeier, 2002). Furthermore, previous literature has suggested that this method affords improved efficiency and accuracy in estimates of SRT (Brand & Kollmeier, 2002; Ozimek et al., 2010).

2.2.1 **Generation of the Paediatric Base Matrix**

In the development of the UCAMST-P, the existing 5 by 10 word base matrix of the UCAMST was modified to generate a new three by six word base matrix. Existing file fragments were edited to create three-word “pseudo-sentences” (consisting of quantity, adjective, object), and the first two columns (name, verb) of the original base matrix were removed. Removal of words from the remaining three columns of the original 5 by 10 base matrix was based on four distinct and largely technical criteria, detailed below.

First, the naturalness of the words following editing had to be considered, as noticeable issues regarding the auditory and visual naturalness of several words in the newly generated pseudo-sentences were identified. As a consequence of the manner in which the file fragments had to be edited, these issues were predominantly associated with the clipped onset of words in the newly generated first column (quantity) – in particular, the words “four”, “six”, “some”, “those”, and “nine”.

The second criterion that was considered in the removal of words from the original UCAMST base matrix was the appropriateness of the lexical content for children, as the intention of this research project was to generate a MST for use with the paediatric population from four years of age. Following discussions with Associate Professor Margaret Maclagan (M. Maclagan, personal communication, December 8, 2016), the following changes were made. Words of lower lexical difficulty were prioritised, while those identified as problematic were removed. In particular, concerns were raised regarding the adjectives "cheap" and "dark", due to potential issues surrounding the semantics of these words (M. Maclagan, personal communication, December 8, 2016). The word "shirts" in the objects column was also identified as potentially problematic (M. Maclagan, personal communication, December 8, 2016). Issues concerning the concept of quantity were also raised, however, M. Maclagan reasoned that a superficial knowledge of the words is anticipated to be sufficient, as the children do not need to pass a comprehension task on the words, they merely have to repeat them. Furthermore, retaining the numerals in the base matrix allowed the UCAMST-P to be comparable to previously published paediatric MSTs (Hagerman & Hermansson, 2015; Neumann et al., 2012).

The third criterion that was considered was the slopes of the psychometric functions generated for each word in both constant noise and babble noise. Words that generated psychometric functions with steeper slopes (i.e. higher slope percentage values) were prioritised, so as to improve the accuracy of the SRT estimates for the UCAMST-P. Following this, any words with a difference between the intended and actual normalisation adjustments (Adjustment Difference; Adj Diff) were considered to be inferior to words that required no adjustment.

The final criterion was for all words retained in the paediatric base matrix to be capable of being presented in the presence of both constant noise and babble noise. Previous research surrounding the UCAMST (McClelland, 2014; Stone, 2016) has suggested that the word “shirts” be removed from the babble noise condition due to the irregularity of the resultant psychometric function and the degree of adjustment required. Consequently, the removal of the word “shirts” from the new three by six base matrix was paramount.

2.2.2 **Paediatric Base Matrix Composition**

As the four distinct removal criteria, outlined above, did not always align, some criteria had to be prioritised over the others. For example, the words “some” and “four” were removed as the fragment editing technique prevented their stand-alone use (see Figure 4). Furthermore, with the removal of the word “dark”, lower lexical difficulty was prioritised over slope in order to maintain a broader age range.

Tables 2 through 4 illustrate the normalisation adjustments and slopes for each of the potential words for inclusion in the new three by six base matrix, in descending order of slope for both the constant and babble noise conditions.

Table 2. Normalisation adjustments and slopes for column one (quantity).

Constant Masking Noise			Babble Masking Noise		
Word	**Slope**	**Adj Diff**	**Word**	**Slope**	**Adj Diff**
some	21.3%	0	four	14.4%	0
those	19.2%	0	some	13.5%	1.22
twelve	17.3%	0	eight	13.1%	0
six	16.8%	0	three	13.1%	0
nine	16.4%	1.19	ten	12.0%	0
four	14.9%	0	nine	11.1%	3.65
eight	14.8%	0	twelve	10.8%	0
three	14.4%	0	two	9.8%	0
ten	13.8%	0	those	9.5%	0
two	11.8%	0	six	9.4%	0

The words “four”, “six”, “some”, and “those” were removed from column one (quantity), as these items would have required additional editing in order to be used, whereas the others did not. The onset of the word “nine”, which sounded unnatural following the initial edit of the file fragment, was able to have a ramped onset integrated to remediate this. Although the adjustment differences for the word “nine” in both the constant and babble noise conditions were larger than optimal, the inability to adequately edit any alternative options resulted in its retention.

Table 3. Normalisation adjustments and slopes for column two (adjective).

Constant Masking Noise			**Babble Masking Noise**		
Word	**Slope**	**Adj Diff**	**Word**	**Slope**	**Adj Diff**
big	16.9%	0	dark	12.9%	0
small	15.8%	0	green	12.3%	0
old	14.6%	0	big	12.0%	0
cheap	14.0%	0	old	11.3%	0
dark	13.8%	0	red	11.1%	0
red	13.4%	0	small	10.3%	0
large	12.9%	0	cheap	10.3%	-1.56
new	12.2%	0	large	9.7%	0.26
green	12.1%	0	new	8.3%	0
good	10.8%	-1.20	good	7.9%	0

In column two (adjective), the words “cheap”, “large”, and “good” were removed based primarily on slope and adjustment difference values, whereas the word “dark” was removed based on lexical difficulty and semantic suitability (M. Maclagan, personal communication, December 8, 2016).

Table 4. Normalisation adjustments and slopes for column three (object).

Constant Masking Noise			Babble Masking Noise		
Word	**Slope**	**Adj Diff**	**Word**	**Slope**	**Adj Diff**
toys	19.9%	0	toys	12.6%	0
books	16.5%	0	books	10.2%	0
shoes	13.7%	0	coats	9.1%	-0.56
coats	13.7%	-0.17	bikes	8.8%	0
spoons	12.7%	0	shoes	8.3%	0
bikes	11.8%	0	spoons	8.2%	0
mugs	11.2%	0	hats	8.2%	-0.31
hats	10.8%	0	mugs	7.2%	0
ships	6.7%	0	ships	7.0%	-0.55
shirts	5.5%	-3.18	shirts	2.4%	-13.67

The words “coats”, “ships”, “shirts”, and “mugs” were removed from column three (object) primarily based on the slope and adjustment difference values. However, the word “hats” was retained over the word “mugs” as it was reasoned that an adjustment difference of -0.31 outweighed the lexical difficulty of the word “mugs”.

2.2.3 Generation of New Sentence Lists

The sentences in the original UCAMST were normalised following arrangement into 16 lists of 10 sentences (see appendix D1). The generation of sentence lists was important because it would have been impossible to evaluate all 100,000 possible sentences. For delivery in constant noise, each of these lists contained exactly one occurrence of each word in each position, and so each list was constrained to have the same average slope value. Effort was taken to homogenise these slope values within each list, so that listeners did not encounter some sentences with a very low slope and others with a very high slope. An iterative procedure (described in Stone, 2016) was employed to generate 16 lists of 10 sentences, each with a very low SD of slope values, and no sentences repeated between lists. For the generation of lists designed for delivery in babble noise, the poor-performing words "wins" and "shirts" were excluded so as to enhance the test's sensitivity and reliability (McClelland, 2015). Consequently, for babble noise, only nine words were used in the verb (column two) and object (column five) columns, resulting in the duplication of one word in each of these columns, for each sub-list in that condition. However, the principle of maintaining homogenous sentence slopes was maintained.

An effort was also made to minimise the judder evident within sentence transitions. Classifications of judder magnitude into "tier groups" were established by the calculated pixel difference value between consecutive video frames for each of the three edited transitions in each five-word sentence (Trounson, 2012). Trounson (2012) classified tier zero and tier one as "no judder", while tiers two through six exhibited increasing judder magnitude. Sentences were rejected if the judder magnitude was tier three or higher for one or more of the transitions, or tier two for all three transitions (Stone, 2016).

The candidate UCAMST-P pseudo-sentences were generated by systematically producing all 216 (6^3) available pseudo-sentences and rejecting those sentences that did not meet certain criteria. For the UCAMST-P, sentences were rejected if the judder magnitude of the transition was classified as tier three or higher. This reduced the number of available pseudo-sentences from 216 down to 162. This small number meant it was feasible to evaluate all available 162 pseudo-sentences in the same timeframe as was allowed for the 16 lists of 10 sentences from the 5 by 10 matrix, and in an almost identical experimental procedure. The 162 pseudo-sentences from the three by six matrix were arranged into 14 lists of 10 and two lists of 11 (see appendix D2), following criteria that ensured that there were: i) no replicate two-word pairs within a single list (e.g. no repeats of "three new" or "red toys") and ii) no identifiable patterns in response positions (e.g. first, third, fifth word in a particular column).

The remaining two sentences, "ten new spoons" and "twelve old bikes", were distributed manually. Although this arrangement was initially for the purposes of gathering the normalisation data, if the lists proved to be equivalent it would be possible to use them in the finished test itself, as it corresponded well with existing paediatric MSTs (Hagerman & Hermansson, 2015; Ozimek et al., 2012).

2.3 Experimental Instrumentation

The preliminary hearing screening was conducted in accordance with the New Zealand Audiological Society best practice guidelines (NZAS; 2016). A calibrated Grason-Stadler GSI 61 clinical audiometer was used to present octave pure-tones, from 250 to 8000 Hz, to participants via Telephonics TDH-50P supra-aural headphones in order obtain audiometric hearing thresholds. Bone conduction thresholds were obtained using a RadioEar B-71 bone transducer at 500, 1000, 2000, and 4000 Hz, again using pure-tones. Participants responded to the pure-tones presented by pressing a response button connected to the GSI 61 clinical audiometer.

Both the preliminary hearing screening and the experimental procedure were conducted in a sound-treated audiological testing booth at the University of Canterbury Speech and Hearing Clinic (Christchurch, New Zealand). Associate Professor Greg O'Beirne developed the software for the UCAMST and UCAMST-P normalisation using LabVIEW. The software was run using an HP EliteDesk 800 G1 and Philips Brilliance 241B monitor, connected to an ēlo touch-sensitive monitor (ēlo ET17115L, Tyco Electronics, CA, USA), which was used to display the visual modality and provide participants with a manner of responding in the closed-set response format for both the UCAMST and the UCAMST-P. Senheiser HD 280 Pro (64 Ω impedance) circumaural headphones were connected to the HP EliteDesk 800 G1 via a Sound Blaster X-Fi Surround 5.1 Pro USB sound card, and used to present the sentence stimuli and masking noise. The resultant data were investigated in Microsoft Excel version 14.7.2, and all statistical analyses were performed on the data using version 24 of the IBM Statistical Package for the Social Sciences (SPSS).

2.4 Scoring Procedures

Based on the findings of McClelland (2015), in which steeper slope scores were obtained for the UCAMST via word scoring, as opposed to fragment scoring, word-based scoring procedures were employed in the current study. Word scoring of sentences calculates the number of words correctly identified for an individual sentence; thus, for the original 5 by 10 version of the UCAMST, scores out of five were given for each sentence, while for the paediatric three by six version of the UCAMST, each sentence was scored out of three.

2.5 Experimental Procedures

Prior to commencing testing, each participant was provided with a consent form and information sheet (refer to Appendix C), and given the opportunity to ask any questions. Each participant was then asked a series of questions relating to their hearing and aspects of their health that have the potential to impact hearing. An otoscopic examination was then performed to ensure that the external ears of each participant were clear of any debris or wax that may affect the audiometric hearing thresholds found. Pure-tone audiometry was performed, as outlined in section 2.4 above, in a sound-treated booth. Each participant was instructed to press the response button when they heard a tone, even if it was very faint. Following testing, the results obtained from this portion of the study were explained to each participant prior to moving forward with the experimental testing section. If a hearing loss was identified, the participant was provided with information concerning appropriate follow-up measures (refer to Appendix C for particulars) and informed that they did not meet the inclusion criteria for the current research.

Each of the participants that met the inclusion criteria was randomly assigned to one of four blocks; each block contained all test conditions, allowing each participant to act as their own control. The assumption was made that the open-set response format must precede the closed-set response format in order to maintain true results for the open-set responses (i.e. eliminating any potential learning effects generated by performing the closed-set test first). Table 5 illustrates the four block testing conditions (presented sequentially from left to right).

Table 5. Block testing conditions.

Block One											
Auditory-alone						**Auditory-visual**					
Open-set			Closed-set			Open-set			Closed-set		
P5	5	3	P5	5	3	P5	5	3	P5	5	3
Block Two											
Auditory-visual						**Auditory-alone**					
Open-set			Closed-set			Open-set			Closed-set		
P5	5	3	P5	5	3	P5	5	3	P5	5	3
Block Three											
Auditory-alone						**Auditory-visual**					
Open-set			Closed-set			Open-set			Closed-set		
P3	3	5	P3	3	5	P3	3	5	P3	3	5
Block Four											
Auditory-visual						**Auditory-alone**					
Open-set			Closed-set			Open-set			Closed-set		
P3	3	5	P3	3	5	P3	3	5	P3	3	5

Note. Auditory-alone = auditory-alone mode of presentation; Auditory-visual = Auditory-visual mode of presentation; Open-set = open-set response format; Closed-set = closed-set response format; P = practice; 5 = five-word sentences from the UCAMST; 3 = three-word sentences from the UCAMST-P.

As described above, all experimental testing was conducted in a sound-treated booth. Participants were seated in front of a touch-sensitive monitor that was used to display the video for the auditory-visual mode of presentation, followed by either the 18-word matrix (UCAMST-P) or 50-word matrix (UCAMST) response panel used during closed-set testing. Figures 7 and 8 illustrate the 50- and 18-word matrix response panels, respectively, used during closed-set testing.

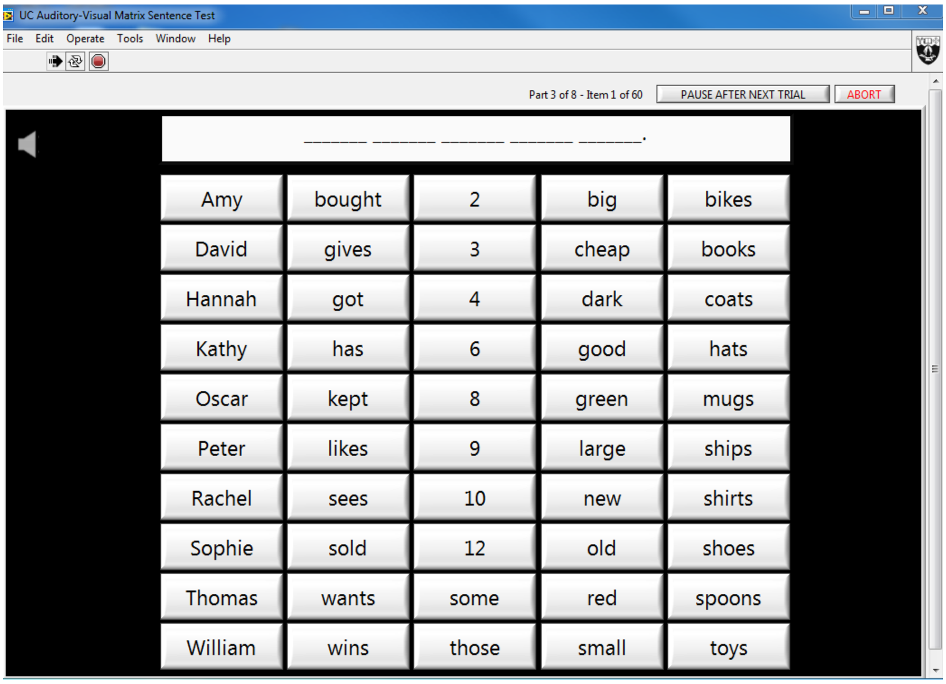

Figure 7. Closed-set response panel used for the UCAMST.

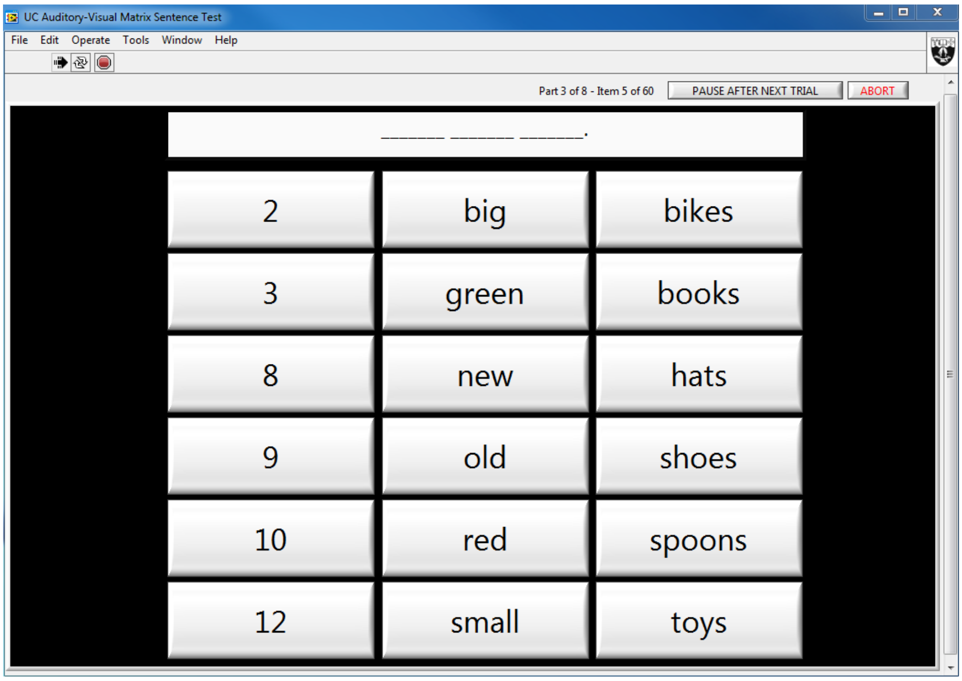

Figure 8. Closed-set response panel used for the UCAMST-P.

Each participant was given verbal instructions explaining that they would hear (and in some instances, see) a sequence of sentences in the presence of background noise of varying intensities. Each participant was then instructed to respond either verbally or by selecting his or her answer on the touch-sensitive monitor. Participants were encouraged to guess when they were uncertain, and informed that in the closed-set response format a full sentence response was required to progress to the next sentence. Finally, for each part of the experiment, additional written instructions appeared on the touch-sensitive monitor outlining the participant's next task. An example of these instructions can be seen in Figure 9.

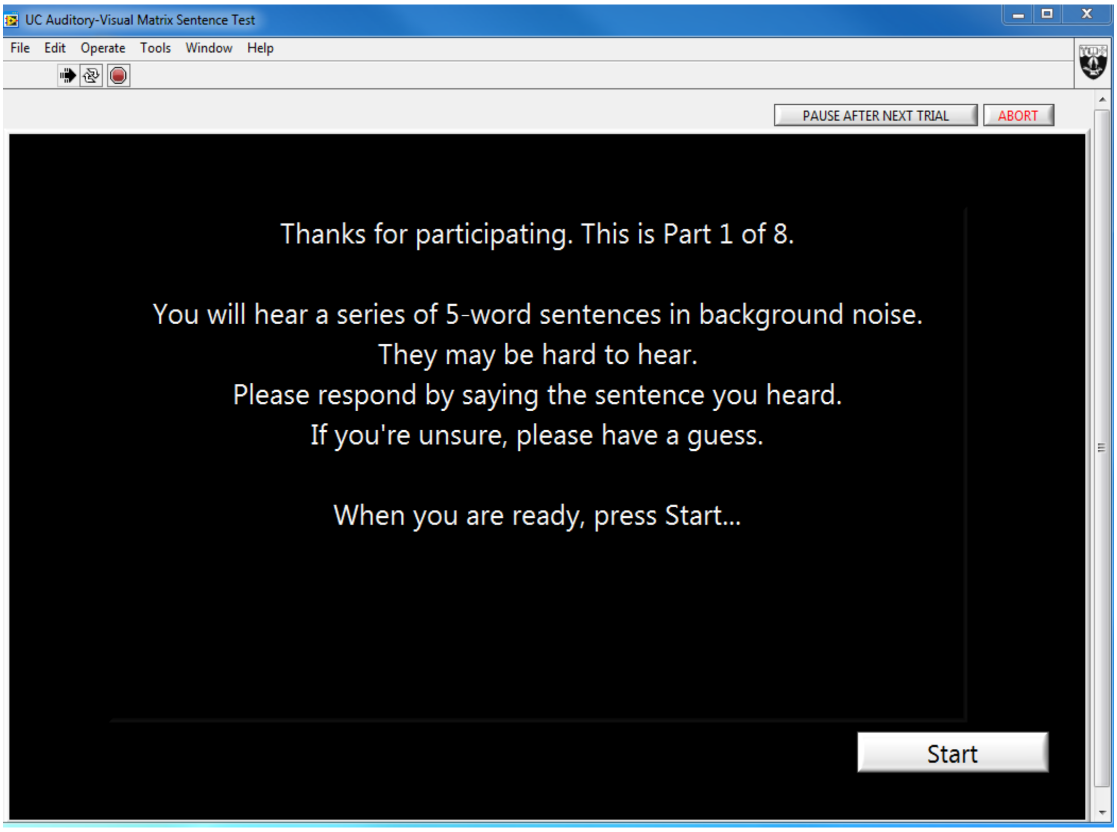

Figure 9. Written instructions displayed for the UCAMST using the open-set response format.

In the open-set response format, the researcher selected the words identified by each participant using an open-set response format pop-out scorer, which was displayed on the Philips Brilliance 241B monitor and not visible to the participant. The layout for the open-set response format pop-out scorer is displayed in Figures 10 and 11 for the UCAMST and UCAMST-P respectively.

Figure 10. Pop-out scorer used by the researcher to record open-set responses for the UCAMST.

Figure 11. Pop-out scorer used by the researcher to record open-set responses for the UCAMST-P.

Irrespective of which block condition participants were assigned to, two practice lists (i.e. 20 sentences) were presented prior to each new mode of presentation and response format in order to ensure understanding of the task and test format and to allow performance to stabilise before commencing testing (Wagener et al., 2003). In addition to the interspersed practice lists, participants were presented with 480 sentences – 40 sentences for each condition for both the UCAMST and the UCAMST-P. The data gathered from these sentences was then utilised in the analyses for the current research project. Due to the length of the experimental testing procedure and the level of concentration required, participants in all four block conditions were encouraged to take rest breaks as needed. The experimental testing procedure took participants approximately 80 minutes to complete, excluding the time taken for such breaks.

2.6 Planned Statistical Analyses

Two separate repeated measures analysis of variance (RM ANOVA) tests were planned to assess hypotheses (1a) to (12d). One RM ANOVA was planned in order to examine the slope and SRT of the UCAMST-P and the other to examine the slope and SRT of the UCAMST. Each RM ANOVA was intended to determine whether differences existed between: (1) each of the 16 sentence lists; (2) the open- and closed-set response formats within each mode of presentation; and (3) when the condition was preceded by training and when the condition was not preceded by training, as well as the interactions between these conditions. However, when the RM ANOVAs were attempted, Box's Test was found to be significant ($p < .001$) for SRT and slope for both the UCAMST-P and the UCAMST data, signifying inequality in the covariances of the variables. Additionally, the assumption of sphericity could not be satisfied for SRT or slope. Based on this, as well as the presence of significant outliers and a lack of normality in the distribution of the SRT and slope values for the UCAMST-P and the UCAMST (discussed further in section 3.1), non-parametric (i.e. assumption free) analyses were employed.

CHAPTER THREE:

RESULTS

3.1 Overview

The following chapter displays the results from the analyses conducted on the data collected in the current study, the implications of which are discussed in Chapter four.

To determine whether the data analyses could be performed using parametric tests, prior to performing the analyses the data were inspected for potential sources of bias that could violate the assumption of normality (i.e. outlying data points or any significant kurtosis or skewness in the distribution). Significant bias was found in the data for each analysis performed; consequently, non-parametric tests were employed. In this study, statistical significance was determined using the Monte Carlo simulation method in SPSS. In the Monte Carlo method, SPSS generates a large number of simulated samples based on the data set. According to North, Curtis, and Sham (2002), the Monte Carlo procedure can be used to calculate p-values when a standard asymptotic distribution cannot be assumed, or if it is not realistic, given the sample size. It is important to note that the Monte Carlo method estimates the p-value using ranked data. Therefore, it is helpful to report confidence intervals along with the estimated p-value. North et al. (2002) also caution that the use of the Monte Carlo method may reduce statistical power, however.

The results pertaining to research question (1), list equivalence, revealed that the sentence lists for both the UCAMST-P and the UCAMST were equivalent with respect to SRT and slope in the AV mode of presentation, irrespective of the response format employed. However, statistically significant differences in SRTs and the slopes of the intelligibility functions were identified between the sentence lists in the AA, open-set and

AA, closed-set conditions for the UCAMST-P and the UCAMST. For research question (2), within each mode of presentation (i.e. AA and AV), the response formats were found to be equivalent with respect to slope for both the UCAMST-P and the UCAMST. However, statistically significant differences in SRT were identified between the open-set and closed-set response formats in both the AA and AV modes of presentation for both tests. In terms of research question (3), no significant impact of training was identified for the UCAMST-P with respect to SRT or slope. For the UCAMST, significant differences in the SRT were identified between when the condition was preceded by training and when the condition was not preceded by training in the AA, open-set and the AA, closed-set conditions. For all of the remaining conditions, no significant differences were found with respect to SRT or slope.

The data for the SRT and slope values for each sentence list in each test condition are shown in Appendices E1 and E2.

3.2 Participants

Participants (n = 43), ranging from 19 to 48 years of age (M = 25 years), were included in the current research project, n = 11 males, and n = 32 females. Upon inspection of the data, no participants were identified as having outlying results; consequently, the data collected from all of the participants included in the study was retained. All participants were native speakers of New Zealand English, with hearing within normal limits bilaterally (in accordance with Goodman, 1965). Figure 12 illustrates the mean audiometric thresholds of the participants.

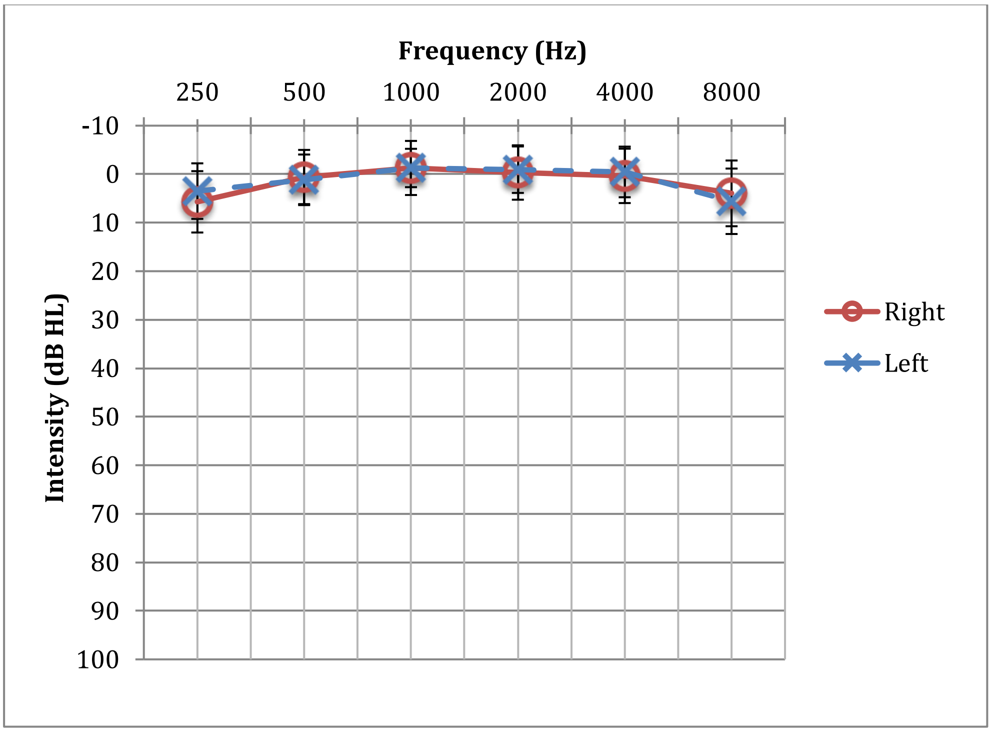

Figure 12. Average PTA thresholds of participants. Error bars represent the standard deviation of the threshold at each frequency.

3.3 List Equivalence Results

The results pertaining to hypotheses (1) and (2) relating to the UCAMST-P, and hypotheses (3) and (4) relating to the UCAMST are outlined in Table 6. In order to examine whether the sentence lists presented within each condition were equivalent, two separate non-parametric related samples Friedman's Two-Way ANOVAs were performed – one conducted on the data relating to the UCAMST-P, and a second conducted on the data relating to the UCAMST.

With regards to the UCAMST-P, the data generally supported the study hypotheses. However, there were three exceptions to this where significant differences were found between the SRT or slope values of the sentence lists. Analogously, for the UCAMST, the data largely supported the study hypotheses, except for in three instances where significant results were obtained. The results shown in Table 6 are described in sections 3.3.1 and 3.3.2 below.

Table 6. Results of the Friedman test for list equivalence. Degrees of Freedom = 15 for all tests.

Test		**UCAMST-P**			**UCAMST**		
Condition	Variable	χ^2	*p*	99% CI	χ^2	*p*	99% CI
AA, Open	SRT	26.471	< .001	< .001 – .001	14.250	.583	.570 – .596
	Slope	27.309	< .001	< .001 – < .001	27.750	< .001	< .001 – < .001
AA, Closed	SRT	22.903	.021	.017 – .024	22.500	.027	.023 – .031
	Slope	19.393	.134	.125 – .143	21.441	.050	.044 – .056
AV, Open	SRT	13.147	.684	.672 – .696	18.265	.213	.203 – .224
	Slope	14.868	.515	.503 – .528	19.809	.114	.106 – .122
AV, Closed	SRT	13.324	.669	.656 – .681	9.923	.902	.894 – .910
	Slope	9.750	.916	.908 – .923	11.578	.805	.795 – .816

Note. AA = auditory-alone; AV = auditory-visual; CI = confidence interval; closed = closed-set response format; open = open-set response format; p = p-value; χ^2 *= chi squared value; SRT = speech recognition threshold.*

3.3.1 **UCAMST-P List Equivalence**

Hypothesis (1) – That no significant differences would be found between the UCAMST-P sentence lists with regards to <u>*SRT*</u> *in the (a) AA, open-set condition; (b) AA, closed-set condition; (c) AV, open-set condition; (d) AV, closed-set condition.*

The Friedman's test conducted on the UCAMST-P data revealed statistically significant differences between the sentence lists when presented in the AA, open-set condition and the AA, closed-set condition with respect to SRT, thus rejecting hypotheses

(1a) and (1b). As reported in Table 6, no statistically significant differences in SRT were found between the UCAMST-P sentence lists when presented in the AV, open-set or closed-set conditions, supporting hypotheses (1c) and (1d).

Hypothesis (2) – That no significant differences would be found between the UCAMST-P sentence lists with regards to slope *in the (a) AA, open-set condition; (b) AA, closed-set condition; (c) AV, open-set condition; (d) AV, closed-set condition.*

As Table 6 shows, no statistically significant differences were revealed between the UCAMST-P sentence lists in the AA, closed-set; AV, open-set; or AV, closed-set conditions with regards to slope, thus supporting hypotheses (2b) through (2d) regarding list equivalence. However, significant differences in the slope of the UCAMST-P sentence lists were identified in the AA, open-set condition, therefore failing to support hypothesis (2a).

Post-hoc analyses were unable to be performed on the data, as discussed in section 4.5.1.1; consequently, plots were generated to aid in the visualisation of the differences and similarities between the sentence lists. Figure 13 shows the speech intelligibility functions for each sentence list of the UCAMST-P in the AA mode of presentation for both response formats. Separate plots were generated for i) when practice did not immediately precede the condition, ii) when the condition was preceded by practice, and iii) both of these combined. The six distinct plots were generated following the removal of the outliers identified by SPSS as “extreme values” (see appendix E). The AV mode of presentation was found to be flawed, as discussed in section 4.1, therefore no plots were generated for this data.

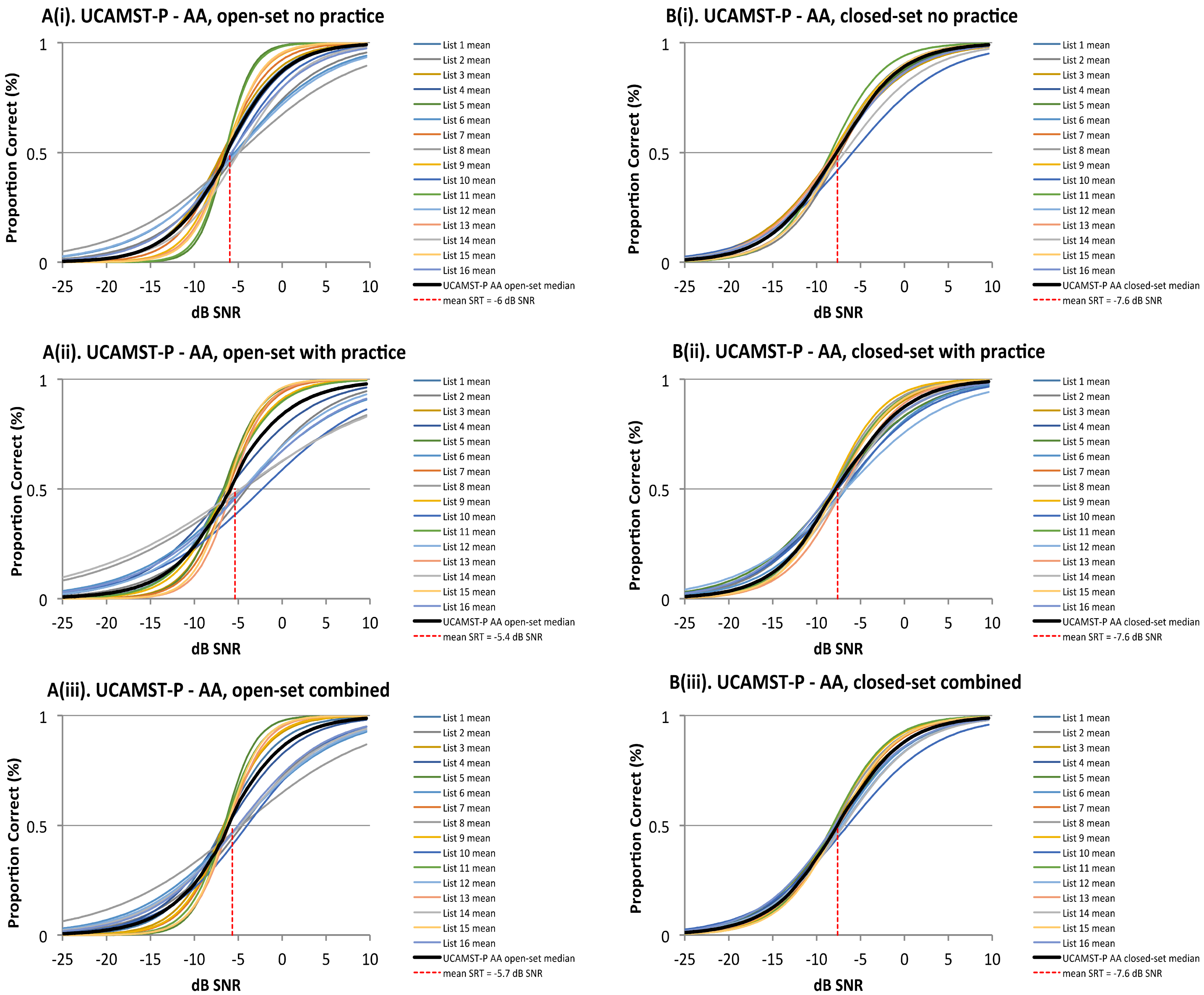

Figure 13. The speech intelligibility functions for each sentence list in the AA mode of presentation for the UCAMST-P in both the open- (A) and closed-set (B) response formats generated for i) when practice did not immediately precede the condition, ii) when the condition was preceded by practice, and iii) both of these combined

3.3.2 **UCAMST List Equivalence**

Hypothesis (3) – That no significant differences would be found between the UCAMST sentence lists with regards to SRT in the (a) AA, open-set condition; (b) AA, closed-set condition; (c) AV, open-set condition; (d) AV, closed-set condition.

The Friedman test conducted on the UCAMST data did not reveal any statistically significant differences between the UCAMST sentence lists when presented in the AA, open-set; AV, open-set; or AV, closed-set conditions with respect to SRT (as shown in Table 6), hence supporting hypotheses (3a), (3c), and (3d) pertaining to UCAMST list equivalence. Conversely, statistically significant differences were identified between the SRT values of the UCAMST sentence lists when presented in the AA, closed-set condition, thus rejecting hypothesis (3b).

Hypothesis (4) – That no significant differences would be found between the UCAMST sentence lists with regards to slope in the (a) AA, open-set condition; (b) AA, closed-set condition; (c) AV, open-set condition; (d) AV, closed-set condition.

As described in Table 6, analysis of the UCAMST data revealed no statistically significant differences between the sentence lists when presented in the AV, open-set or AV, closed-set conditions, supporting hypotheses (4c) and (4d) regarding the equivalence of the UCAMST sentence lists. However, statistically significant differences in the slopes of the speech intelligibility functions were identified for the AA, open-set and AA, closed-set conditions. Consequently, hypotheses (4a) and (4b) were rejected.

As mentioned above, post-hoc analyses were unable to be performed (discussed further in section 4.5.1.1); consequently, plots were generated to afford visualisation of the differences and similarities between the sentence lists. Figure 14 depicts the intelligibility functions for each sentence list of the UCAMST in the AA mode of presentation for both response formats. As before, separate plots were generated for i) when practice did not immediately precede the condition, ii) when the condition was preceded by practice, and iii) both of these combined. Outliers identified by SPSS as “extreme values” (see appendix E) were removed prior to genetrating these plots. Due to the shortcomings of the AV mode of presentation (discussed in section 4.1), no plots were generated for this data.

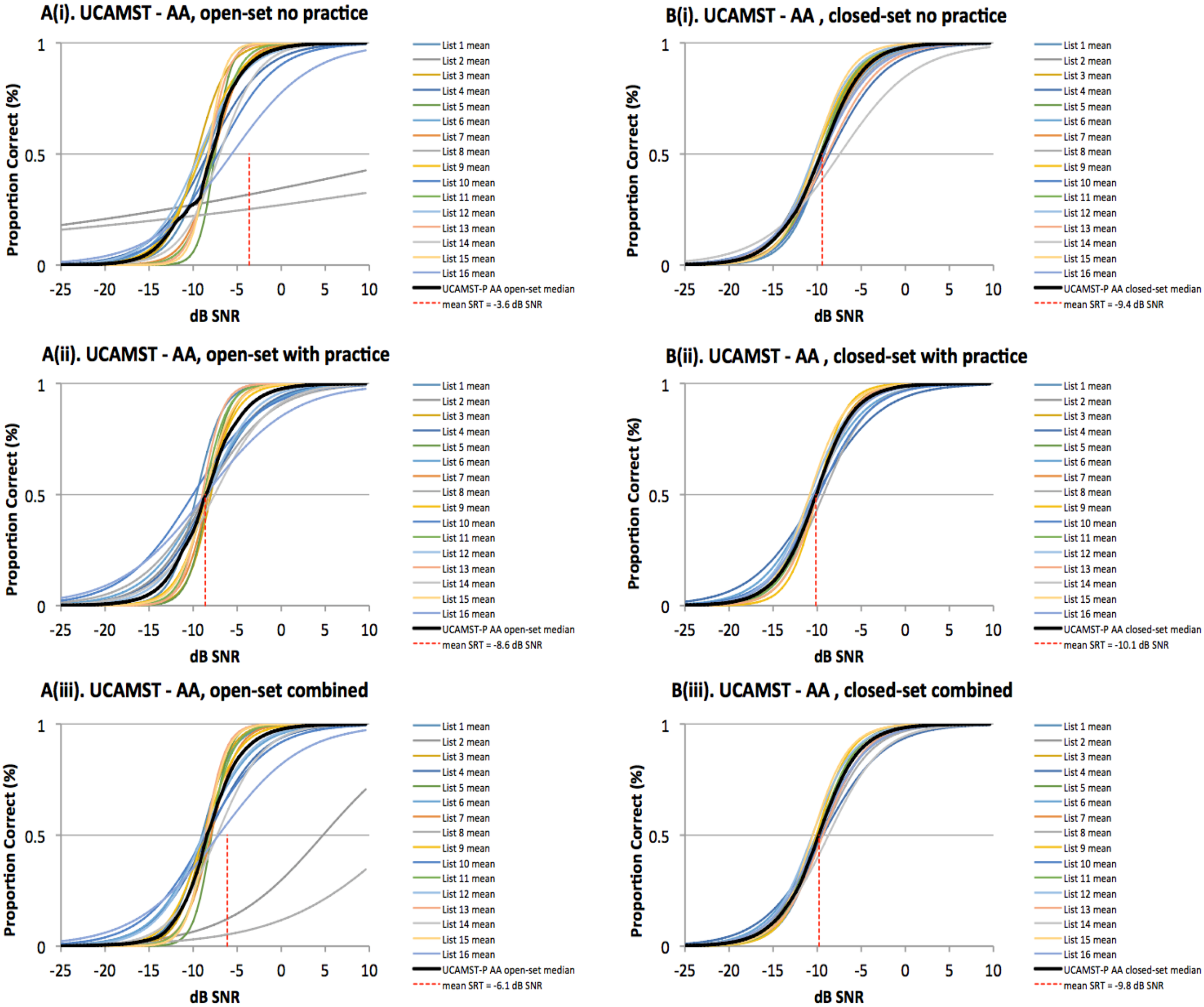

Figure 14. The speech intelligibility functions for each sentence list in the AA mode of presentation for the UCAMST in both the open- (A) and closed-set (B) response formats generated for i) when practice did not immediately precede the condition, ii) when the condition was preceded by practice, and iii) both of these combined.

3.4 Condition Equivalence Results

In order to analyse whether the different conditions employed by the UCAMST-P and the UCAMST were equivalent within each mode of presentation, two separate Kruskal-Wallis one-way ANOVAs were performed on the data. The results pertaining to hypotheses (5) and (6) relating to the UCAMST-P and hypotheses (7) and (8) relating to the UCAMST are reported in Table 7. The results shown in Table 7 are described in sections 3.4.1 and 3.4.2 below.

Table 7. Results of the Kruskal-Wallis tests for condition equivalence. Degrees of Freedom = 1 for all tests.

Test		**UCAMST-P**			**UCAMST**		
Condition	Variable	χ^2	*p*	99% CI	χ^2	*p*	99% CI
AA	SRT	42.936	< .001	< .001 – < .001	34.272	< .001	< .001 – < .001
	Slope	.935	.335	.322 – 0.347	.433	.515	.503 – .528
AV	SRT	7.284	.007	.005 – .009	4.615	.027	.023 – .031
	Slope	1.303	.256	.244 – .267	0141	.713	.701 – .725

Note. AA = auditory-alone; AV = auditory-visual; CI = confidence interval; p = p-value; χ^2 *= chi squared value; SRT = speech recognition threshold.*

3.4.1 **UCAMST-P Condition Equivalence**

Hypothesis (5) – That no significant differences would be found between the open-set and closed-set response formats of the UCAMST-P with regards to <u>*SRT*</u> *in the (a) AA mode of presentation and (b) AV mode of presentation.*

The Kruskal-Wallis ANOVA conducted on the UCAMST-P data revealed significant differences in SRT between the open-set and closed-set response formats for both the AA and AV modes of presentation, therefore failing to support hypotheses (5a) and (5b).

Hypothesis (6) – That no significant differences would be found between the open-set and closed-set response formats of the UCAMST-P with regards to slope in the (a) AA mode of presentation and (b) AV mode of presentation.

The analyses conducted revealed no significant differences in the slopes of the speech intelligibility functions between the open-set and closed-set response formats for the AA or AV modes of presentation, supporting hypotheses (6a) and (6b).

Together, Figures 15 and 16 depict the findings related to hypotheses (5a), (5b), (6a), and (6b), illustrating the variations between the mean SRT for each condition and the similarities between the slopes of the intelligibility functions.

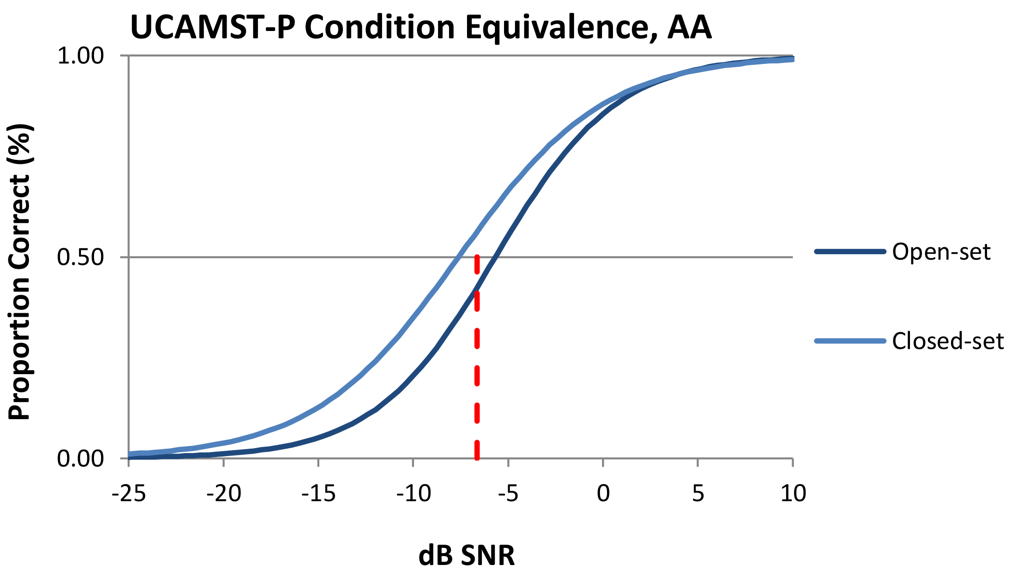

Figure 15. Intelligibility functions of the AA, open-set and closed-set conditions of the UCAMST-P.

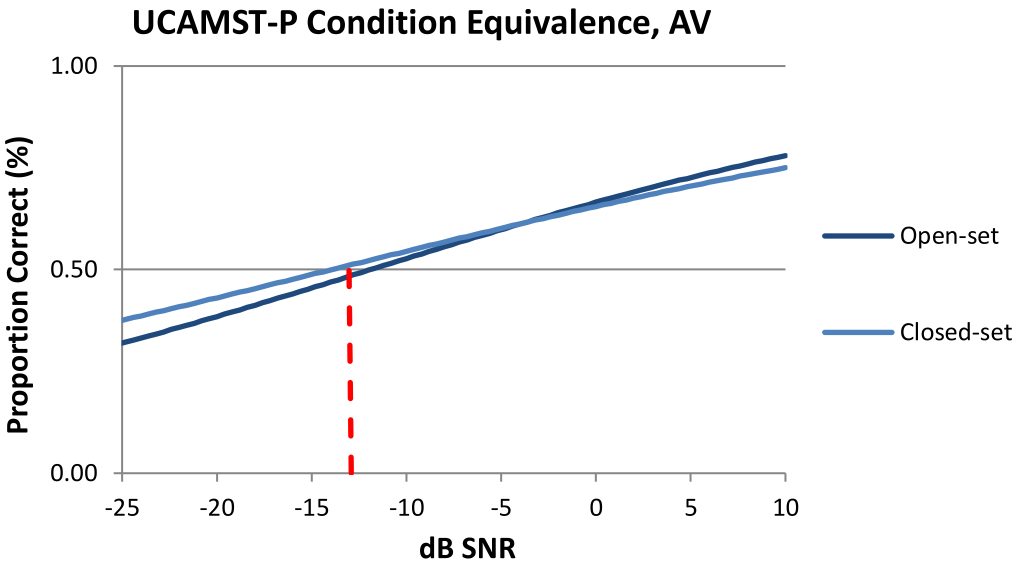

Figure 16. Intelligibility functions of the AV, open-set and closed-set conditions of the UCAMST-P.

3.4.2 **UCAMST Condition Equivalence**

Hypothesis (7) – That no significant differences would be found between the open-set and closed-set response formats of the UCAMST with regards to <u>*SRT*</u> *in the (a) AA mode of presentation and (b) AV mode of presentation.*

The Kruskal-Wallis ANOVA conducted on the UCAMST data revealed significant differences in SRT between the open-set and closed-set response formats for both the AA and AV modes of presentation; thus, hypotheses (7a) and (7b) were rejected.

Hypothesis (8) – That no significant differences would be found between the open-set and closed-set response formats of the UCAMST with regards to <u>*slope*</u> *in the (a) AA mode of presentation and (b) AV mode of presentation.*

As reported in Table 7, no significant differences in the slopes of the speech intelligibility functions were identified between the open-set and closed-set response formats for the AA or AV modes of presentation, supporting hypotheses (8a) and (8b).

Figures 17 and 18 depict the findings related to hypotheses (7a), (7b), (8a), and (8b) displaying the similarities between the slopes of the intelligibility functions and the variations between the mean SRT for each condition.

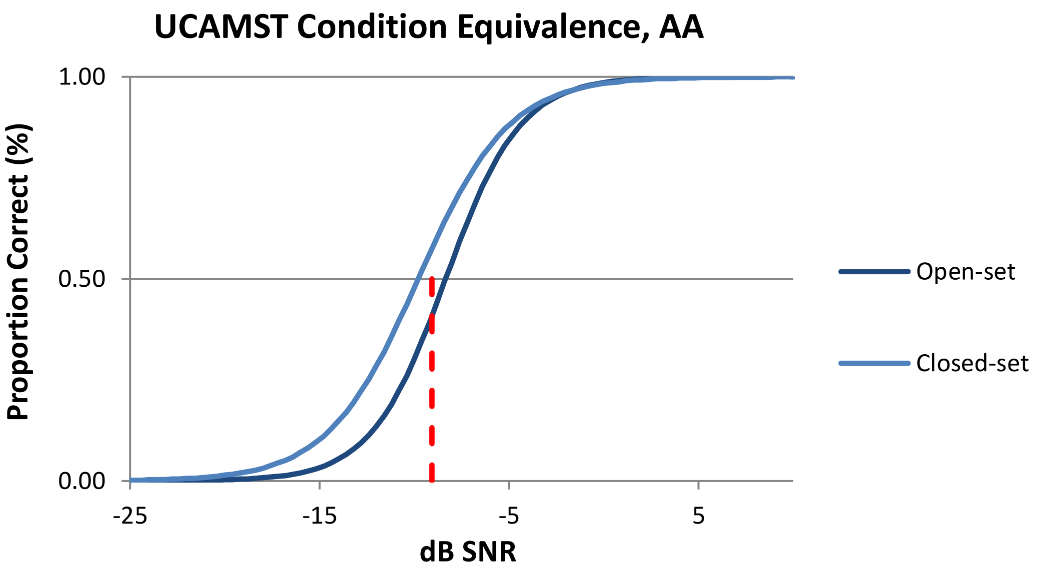

Figure 17. Intelligibility functions of the AA, open-set and closed-set conditions of the UCAMST.

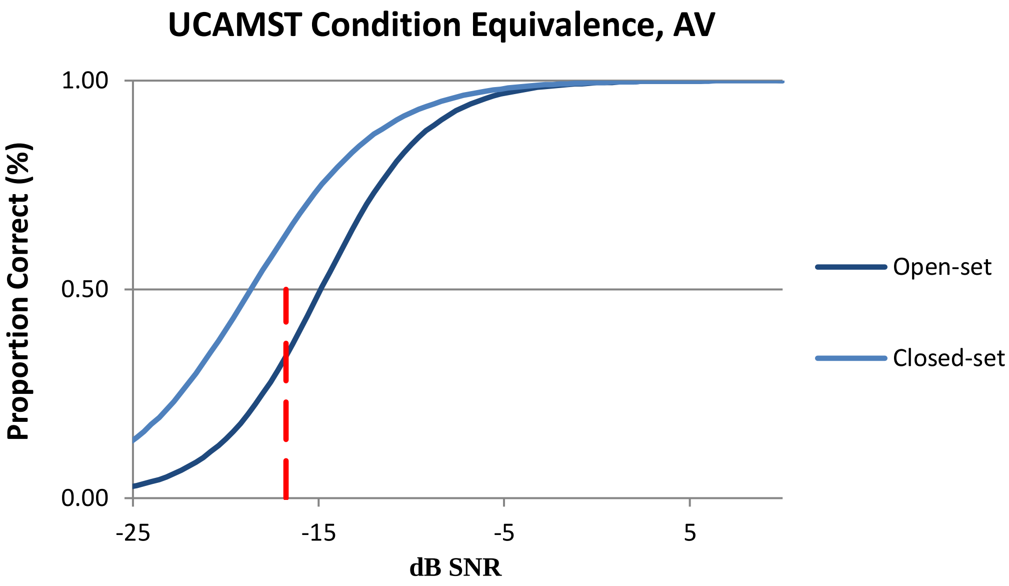

Figure 18. Intelligibility functions of the AV, open-set and closed-set conditions of the UCAMST.

3.5 Training Results

The results pertaining to hypotheses (9) and (10) relating to the UCAMST-P, and hypotheses (11) and (12) relating to the UCAMST are outlined in Table 8. In order to examine whether there was an impact of training, two separate Kruskal-Wallis one-way ANOVAs were performed on the data. The results shown in Table 8 are described in sections 3.5.1 and 3.5.2 below.

Table 8. Results of the Kruskal-Wallis test for the impact of training. Degrees of Freedom = 1 for all tests.

Test		**UCAMST-P**			**UCAMST**		
Condition	Variable	χ^2	*p*	99% CI	χ^2	*p*	99% CI
AA, Open	SRT	1.455	.242	.231 – .253	3.990	.046	.040 – .51
	Slope	.888	.364	.351 – .376	.006	.953	.947 – .958
AA, Closed	SRT	.001	.976	.972 – .980	9.551	.001	< .001 – .002
	Slope	< .001	1.000	1.000 – 1.000	1.642	.205	.194 – .215
AV, Open	SRT	.626	.449	.437 – .462	.091	.778	.767 – .789
	Slope	.960	.344	.332 – .356	.960	.336	.324 – .349
AV, Closed	SRT	1.366	.252	.241 – .263	< .001	1.000	1.000 – 1.000
	Slope	2.627	.107	.099 – .115	2.388	.125	.116 – .133

Note. AA = auditory-alone; AV = auditory-visual; CI = confidence interval; closed = closed-set response format; open = open-set response format; p = p-value; χ^2 *= chi squared value; SRT = speech recognition threshold.*

3.5.1 **Effect of Training: UCAMST-P**

Hypothesis (9) – That for the UCAMST-P no significant differences would be found between when the condition is preceded by training and when the condition is not preceded by training with regards to <u>*SRT*</u> *in the (a) AA, open-set condition; (b) AA, closed-set condition; (c) AV, open-set condition; (d) AV, closed-set condition*.

The analysis performed on the UCAMST-P data revealed no statistically significant differences between when the condition was preceded by training and when the condition was not preceded by training, with regards to SRT, for each of the four test conditions (i.e. AA, open-set condition; AA, closed-set condition; AV, open-set condition; AV, closed-set condition). Consequently, hypotheses (9a) through (9d) were supported.

Hypothesis (10) – That for the UCAMST-P no significant differences would be found between when the condition is preceded by training and when the condition is not preceded by training with regards to slope in the (a) AA, open-set condition; (b) AA, closed-set condition; (c) AV, open-set condition; (d) AV, closed-set condition.

As reported in Table 8, no statistically significant differences were found between when the condition was preceded by training and when the condition was not preceded by training, with regards to the slope, for each of the four UCAMST-P test conditions (i.e. AA, open-set condition; AA, closed-set condition; AV, open-set condition; AV, closed-set condition). Thus, hypotheses (10a) through (10d) were supported.

3.5.2 **Effect of Training: UCAMST**

Hypothesis (11) – That for the UCAMST no significant differences would be found between when the condition is preceded by training and when the condition is not preceded by training with regards to SRT in the (a) AA, open-set condition; (b) AA, closed-set condition; (c) AV, open-set condition; (d) AV, closed-set condition.

The Kruskal-Wallis ANOVA conducted on the UCAMST data revealed significant differences in SRT when the AA, open-set and the AA, closed-set conditions were preceded by training compared to when training did not immediately precede these conditions (as shown in Table 8). Consequently, hypotheses (11a) and (11b) were rejected. Conversely, as no statistically significant differences were found for the AV, open-set or the AV, closed-set conditions, hypotheses (11c) and (11d) were supported.

Hypothesis (12) – That for the UCAMST no significant differences would be found between when the condition is preceded by training and when the condition is not preceded *by training with regards to slope in the (a) AA, open-set condition; (b) AA, closed-set condition; (c) AV, open-set condition; (d) AV, closed-set condition*.

As displayed in Table 8, no statistically significant differences were found between when the condition was preceded by training and when the condition was not preceded by training with regards to the slopes of the intelligibility functions for each of the four test conditions (i.e. AA, open-set condition; AA, closed-set condition; AV, open-set condition; AV, closed-set condition). Thus, hypotheses (12a) through (12d) were supported.

CHAPTER FOUR:

DISCUSSION

4.1 Introduction

The purpose of the current research project was to develop a paediatric MST in New Zealand English (UCAMST-P) by editing the existing 5 by 10 word matrix of the UCAMST into a three by six word matrix, thereby creating three-word pseudo-sentences, better suited to paediatric speech perception testing. Subsequently, it was imperative to evaluate the newly developed UCAMST-P in order to establish the reliability and sensitivity of the MST in estimating SRTs.

In the current study, both the slopes of the speech intelligibility functions and the SRT values were of interest. The homogeneity of SRT values across lists within a single condition is obviously of importance in ensuring the test-retest reliability of the test. Also, as discussed previously, slope values provide information pertaining to the reliability and accuracy of estimates of SRT. Although SRT values are important clinically to provide information on a client's performance relative to the normative values for that particular condition, they are relatively client-specific. Thus, in interpreting the results of the current study, more weight has been afforded to the interpretation of the slope values in order to better establish the reliability and accuracy of these estimates.

It is also worth noting here that the SRT and slope values obtained in the AV mode of presentation were not comparable with those obtained in the AA mode of presentation. The slope values in the AV mode of presentation were found to be drastically lower, irrespective of the test or response format employed. This was reasoned to be due to the additional advantage afforded in the AV mode of presentation being highly correlated with performance in the VA mode of presentation (MacLeod & Summerfield, 1987). Accordingly, the AV mode of presentation was essentially providing information on a participant's lip reading abilities (MacLeod & Summerfield, 1987). Consequently, it is recommended that the AV mode of presentation should only be used at a fixed SNR in conjunction with the AA mode of presentation as a measure of AVE, rather than being used in adaptive-testing mode by itself.

The results of the list equivalence analyses found that while the sentence lists were equivalent when presented in the AV mode of presentation, irrespective of response format, when presented in the AA mode of presentation, the sentence lists were not found to be equivalent. The results of the condition equivalence analyses established differences in performance (i.e. SRT) based on the response format employed, although no significant differences in the slopes of the speech intelligibility functions were identified. Lastly, training was only found to impact SRT in the AA mode of presentation for the UCAMST; aside from this no other impacts of training were identified. Based on these findings, further investigation and adjustment of the composition of the sentence lists employed in both the UCAMST-P and the UCAMST is warranted. This chapter will discuss the implications of these findings with reference to previous literature, outline and consider the limitations of the current study, and propose areas of future research.

4.2 List Equivalence

The first group of hypotheses, relating to research question (1), proposed that the sentence lists employed within each of the four test conditions (i.e. AA, open-set; AA, closed-set; AV, open-set; AV, closed-set) of the UCAMST-P and the UCAMST would be equivalent with regards to the slopes of the speech intelligibility functions and SRTs. While the results of the analyses largely supported these hypotheses, some of the hypotheses were not supported.

Evaluation of the sentence lists presented in the AV mode of presentation, irrespective of response format, revealed non-significant results, suggesting equivalence of the sentence lists with regards to both SRT and slope for the UCAMST-P and the UCAMST. Non-significant results were also obtained for the slope of the UCAMST-P intelligibility functions in the AA, closed-set condition, suggesting equivalence of the sentence lists within this condition with regards to slope. Additionally, the results pertaining to the AA, open-set condition of the UCAMST were found to be non-significant for SRT, indicating equivalence of SRT between the sentence lists within this condition. Due to the standardised methodology utilised in the development of MSTs, these findings are in accordance with what was predicted (Akeroyd et al., 2015) and signify that, in the AV modality, the sentence lists are able to be used interchangeably in the open-set and closed-set response formats for the UCAMST-P and the UCAMST. Furthermore, based on Stone's (2016) findings relating to the equivalence of the UCAMST sentence lists in the AA, constant noise condition, the equivalence of the UCAMST sentence lists in the AA mode of presentation was anticipated.

Evaluation of the UCAMST-P and the UCAMST in the AA mode of presentation revealed a lack of equivalence between the sentence lists. For the UCAMST-P, differences in both SRT and slope were identified in the open-set condition, while in the closed-set condition only differences in slope were evident. For the UCAMST, contrary to Stone's (2016) findings, discrepancies in equivalence were found in the open-set condition with respect to SRT alone, while in the closed-set condition differences in both SRT and slope were apparent. In combination, the lack of equivalence found for the SRT and slope values indicates that the sentence lists were not equally difficult within each condition, and the accuracy with which a listener's SRT could be estimated was not consistent between each of the 16 sentence lists for the UCAMST or the UCAMST-P in either the open- or closed-set condition. The inconsistencies between the findings of Stone (2016) and the current study may be reasoned to be due to differences in the methodology between the two studies; this is discussed further in section 4.5.2. Figure 19 illustrates the comparison between the data collected in the current study, and the comparable data collected by Stone (2016). As all of the conditions tested by Stone (2016) were directly preceded by practice, the comparisons made here employ the results from the current study that were also directly preceded by practice.

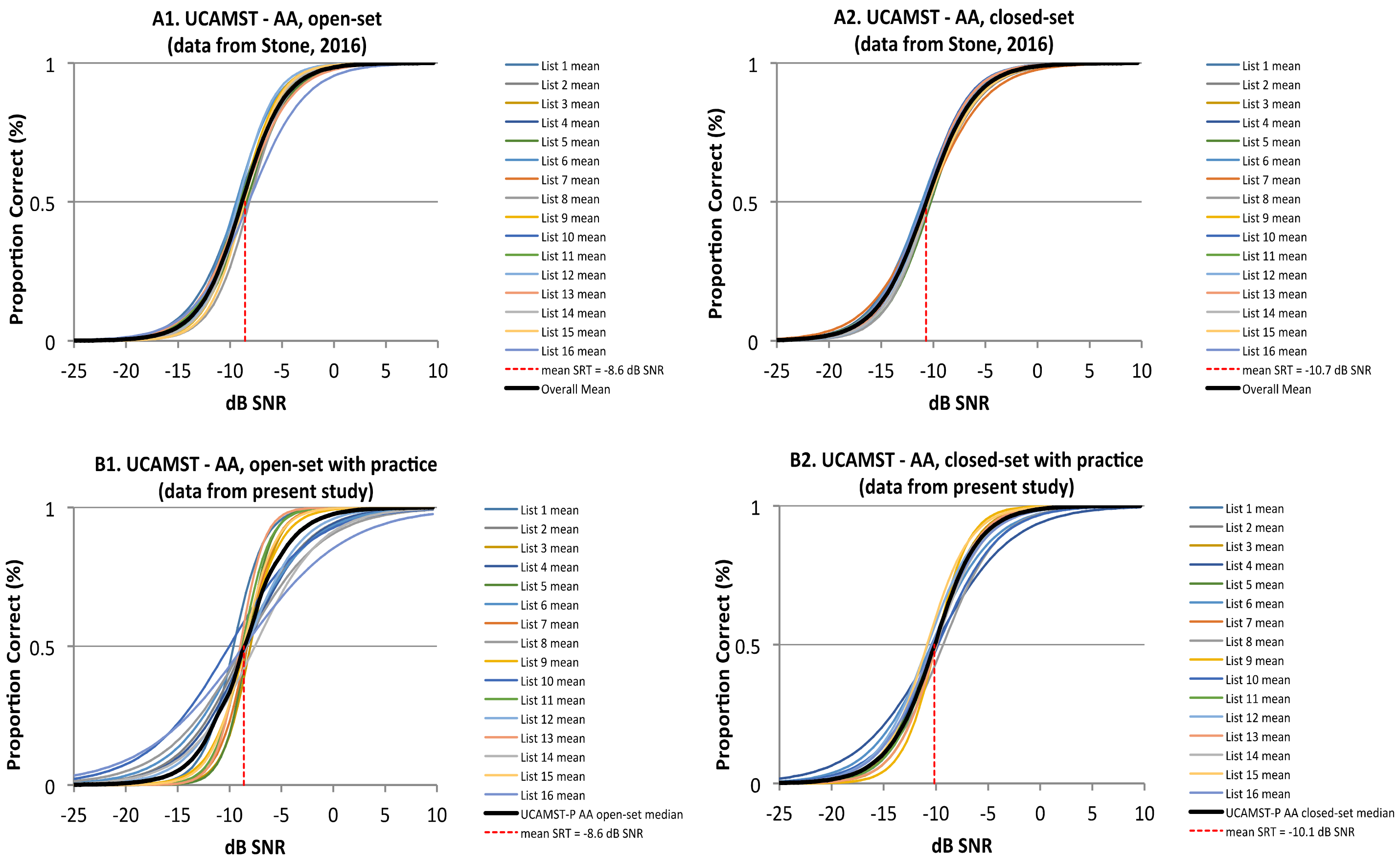

Figure 19. Comparisons between the list equivalence findings of Stone (A1 and A2) (2016) and the current study (B1 and B2) for both response formats.

Due to the manner in which the current study was conducted, additional sentence-specific data was collected. This data provides the opportunity for new sentence lists to be generated that are equivalent with respect to SRT – and more importantly slope – for those conditions where a lack of list equivalence was identified. Achieving equivalence of the sentence lists within each condition is a vital step in progressing the UCAMST-P and the UCAMST towards research and clinical applications as parts of the UCAST test battery (O'Beirne et al., 2012).

4.3 Condition Equivalence

Although estimates of list equivalence afford valuable information regarding the evaluation of newly developed speech recognition measures, the equivalence of each of the test conditions is also a relevant aspect to consider. Thus the reliability of estimates of SRT, and the accuracy of these measures, across test conditions need also be examined in order to determine whether different conditions are able to be used interchangeably. Condition equivalence was investigated for the UCAMST-P and the UCAMST separately. Evaluation of the SRT between the open- and closed-set response formats within each modality (i.e. AA, open-set vs. AA, closed-set; AV, open-set vs. AV, closed-set) revealed significant differences for both the UCAMST-P and the UCAMST, with higher SRTs being obtained in the open-set condition for both the AA and AV modes of presentation. Evaluation of the slopes of the intelligibility functions between the open- and closed-set response formats within each modality found no significant differences for the UCAMST-P or the UCAMST. These findings indicate that the accuracy with which estimates of SRT can be made is equivalent, irrespective of the response format employed, for both tests. However, based on the response format and the modality used, these estimates of SRT are liable to differ.

Previous literature on this matter has identified similar disparities in estimates of SRT between the open- and closed-set response formats, with listeners with HI obtaining higher SRTs on the UCAMST in the open-set condition as compared to the closed-set condition in the AV mode of presentation (Andre, 2016). These findings were suggested to have arisen due to the greater cognitive demands required in the closed-set condition, as listeners are required to retain the test sentence throughout the time in which they select the corresponding words from the base matrix (Andre, 2016). However, it is worth noting that as the study conducted by Andre (2016) employed listeners with HI over 60 years of age, estimates of SRT may have been impacted by cognitive deterioration and/or reduced memory capacity (discussed further in section 1.6) (Lin et al., 2013; Theunissen et al., 2009; Van Rooij and Plomp, 1990). Additionally, as training in noise was not implemented, the estimates of SRT reported by Andre (2016) are liable to have been impacted by the training effect (discussed further in section 4.5.3). Contrary to the findings of the current study, and those obtained by Andre (2016), previous research has found SRTs to be significantly higher in the closed-set condition, indicating that it is the open-set condition that listeners find more difficult, and not the closed-set condition as considered above (Hochmuth et al. 2012; Stone, 2016). Opposing these conflicting results, during the evaluation of the Polish MST, no significant differences in listener performance were identified across response formats (Ozimek et al., 2010). However, as discussed in section 1.10, it is possible that this may be attributed to the extensive hour-long training session implemented in the Ozimek et al. (2010) study to stabilise performance.

Due to the lack of consensus in the literature surrounding the impact of response format on SRT, it is suggested that the response format employed when administering the UCAMST and the UCAMST-P be tailored to the client undergoing testing.

4.4 The Impact of Training

Prior to interpreting the results pertaining to the impact of training, it is important to explain the presence of these analyses. The current study was not designed to examine the training effect – the amount of training required preceding assessment in order to allow SRT measurements to stabilise (Dietz et al., 2014; Hochmuth et al., 2012; Kollmeier et al., 2015; Wagener et al., 2003). Training was incorporated in the current study before each block condition (refer to section 2.6) in order to attempt to allow participants to adjust to the task prior to testing. However, due to the structure of these block conditions, training was not implemented before each test (i.e. the UCAMST or the UCAMST-P) every time a distinct condition was tested. Consequently, in some instances training was presented immediately prior to a condition, and in others it was not. For example, in the AA, open-set condition a participant may have first had training for the UCAMST, then been tested on the UCAMST, and finally been tested on the UCAMST-P without an additional bout of training immediately preceding administration of the UCAMST-P. The limitations and reasoning surrounding this design are discussed further in section 4.5.2. Although the intent of the current study was not to examine the training effect for the UCAMST or the UCAMST-P, investigation into these effects in future research is recommended (refer to section 4.5.3).

Regardless of whether or not investigating the impact of training was the intention of the current study, examining whether or not it had any impact on the data, and therefore the inferences that can be drawn from the analyses, is essential. With respect to the UCAMST-P, no significant differences in SRT or slope were found between when the condition was preceded by training and when the condition was not preceded by training for each of the four test conditions (i.e. AA, open-set condition; AA, closed-set condition; AV, open-set condition; AV, closed-set condition). This suggests that the manner in which practice was implemented in the current study did not significantly impact the results obtained for the UCAMST-P pertaining to list equivalence or condition equivalence. Similarly for the UCAMST, with the exception of two significant findings, no significant differences in SRT or slope were identified between when the condition was preceded by training and when the condition was not preceded by training for each of the four test conditions (i.e. AA, open-set condition; AA, closed-set condition; AV, open-set condition; AV, closed-set condition). The exceptions to this were in the AA open-set and closed-set conditions, where significant differences were found with respect to SRT. Consequently, for the list equivalence and condition equivalence analyses, the SRT values relating to the UCAMST AA open-set and closed-set conditions must be inferred cautiously. However, as discussed above, in the current study more weight has been afforded to the interpretation of the slope of the intelligibility functions than to the SRT, and no significant impact of training on slope was found for the UCAMST or the UCAMST-P, in any of the four test conditions.

4.5 Study Limitations and Future Research

Several limitations arose in the current study that may challenge the results obtained. Thus, in interpreting the results, consideration of the limitations present should be afforded. The following section examines these limitations with reference to how subsequent research in this area may endeavour to avoid such shortcomings.

4.5.1 **The Sample**

4.5.1.1 Lack of Normality

As discussed in section 3.1, the assumption of normality was violated due to significant bias present in the data. As a consequence of this, non-parametric tests had to be employed in place of parametric tests, thus decreasing the statistical power of the study. This posed several limitations. First, as a standard asymptotic distribution could not be assumed, the Monte Carlo method had to be employed to provide p-value estimates, potentially further reducing statistical power (North et al., 2002). Owing to this diminished statistical power, post-hoc analyses were unable to be calculated. This limited both the inferences that can be drawn from the existing results and the analyses themselves, as pairwise comparisons to uncover which of the sentence lists were contributing to the significant differences could not be run. Future research should attempt to resolve this issue of a lack of normality so that parametric statistical analyses can be employed.

4.5.1.2 Generalisability

The generalisability of the sample is also an issue that warrants consideration. The purpose behind evaluating the sentence lists was to provide conformation of equivalence across a sample of participants who are likely to represent the performance typically anticipated for listeners with NH. However, the data collected in the current study may not be truly representative of the actual population's age, linguistic, cultural, and socioeconomic variations in New Zealand. First, due to the manner in which participants were recruited, the current sample consisted largely of individuals, in particular students, from the University of Canterbury (Christchurch, New Zealand). A secondary and related limitation is that, although previous literature affords no recommendations concerning the age distribution of participants, over 85% of the participants included in the current study were aged between 20 and 30 years. Consequently, the current sample represents the performance of a relatively small demographic. Additionally, an observable gender imbalance was present in the current sample with significantly more female listeners volunteering than male. Whether this is a significant factor is unclear, as previous research on this subject is ambiguous with differences in the speech reading abilities of males and females being found to be both significant (Irwin, Whalen, & Fowler, 2006; Ruytjen, Albers, van Dijk, & Willemsen, 2006; Strelnikov, Rouger, Lagleyer, Fraysse, Deguine, & Barone, 2009) and not significant (Auer & Bernstein, 2007; Tye-Murray et al., 2007b). Similar discrepancies regarding gender imbalance have been observed previously in the evaluation procedures of MSTs (Ozimek et al., 2012; Stone, 2016; Wagener et al., 2003). Therefore, although the current data may not have been directly impacted by the factors outlined above, it is recommended that a more representative sample be established in any subsequent research in order to attempt to maintain the generalisability of the findings.

4.5.2 **Block Testing Structure**

Although the presentation and structure of the block testing conditions employed in the current study were beneficial with respect to the amount of information that was able to be collected, the use of these block conditions also presented limitations. As mentioned previously, training was unable to be provided prior to each distinct condition in every block due to the additional time this would have involved during testing, in a regime that was already extensive and cognitively demanding. Another possible limitation associated with the block testing structure was that each participant experienced all four of the conditions in sequence. However, due to the time constrains surrounding the current study, the analyses pertaining to whether or not this had an impact on the data collected are yet to be run; thus the implications of this are currently unclear.

In future research, if the current study were to be re-run in isolation, with the sole focus of evaluating the UCAMST-P and the UCAMST, restructuring the design to examine each of the four conditions (i.e. AA, open-set; AA, closed-set; AV, open-set; AV, closed-set) individually, as recommended by Akeroyd and colleagues (2015), would be beneficial. As in the study conducted by Stone (2016), 16 participants would be required per condition. It is worth noting that, due to the time constraints of the current study, this would have been difficult to achieve as 64 participants would have been required per test (i.e. 128 participants in total). Furthermore, structuring the test in this manner would have forfeited collection of much of the information gathered in the current study that can be used in the future to make direct comparisons between the UCAMST and the UCAMST-P (e.g. with respect to response time – not reported here). Additionally, the use of this block testing structure allowed participants to act as their own control. Although the limitations surrounding the implementation of the block testing structure employed in the current study are evident, the additional information that was able to be obtained due to the use of this design were deemed to outweigh these limitations.

4.5.3 **The Training Effect**

As initially established by Hagerman (1984), significant training, or learning, effects have been reported for MSTs internationally (Dietz et al., 2014; Hochmuth et al., 2012; Kollmeier et al., 2015; Puglisi et al., 2014; Wagener et al., 2003; Warzybok et al., 2015). Wagener and colleagues (2003) defined the training effect as the decrease in an individual's SRT levels with the increasing number of lists administered. This improvement in SRT level is thought to be due to familiarisation with the stimulus materials, response type, and procedure (Wagener et al., 2003). Upon examination of previously published international MSTs, Kollmeier and colleagues (2015) observed training effects for first time users of the MSTs within the first few lists for each language, using both closed- and open-set response formats. A large and significant difference in SRT (1.2 dB) was observed between the first and second lists; however, this change was seen to decrease to barely detectable levels (below 1 dB) following the second measured list (Kollmeier et al., 2015). This phenomenon has been reported for numerous language-specific MSTs (Dietz et al., 2014; Hochmuth et al., 2012; Kollmeier et al., 2015; Wagener et al., 2003). Kollmeier and colleagues (2015) reasoned that this change in an individual's SRT levels, during the initial lists of a MST, is likely language independent and potentially associated with the structure of the task itself. Accordingly, when MSTs are employed to assess speech recognition, it has been recommended that, irrespective of language, two practice lists (each consisting of 20 sentences) be administered prior to the assessment procedure, in order to allow SRT measurements to stabilise (Dietz et al., 2014; Hochmuth et al., 2012; Kollmeier et al., 2015; Wagener et al., 2003). It should also be noted that the training effect phenomenon has been seen to be up to 2 dB higher when responses are obtained via the open-set response format, as opposed to the closed-set response format (Hochmuth et al., 2012; Puglisis et al., 2015; Warzybok et al., 2015). Nevertheless, following completion of the second measured list, differences in training effect between open- and closed-set response formats have been seen to reduce essentially equally (Hochmuth et al., 2012; Puglisis et al., 2015; Warzybok et al., 2015).

As the training effect for the UCAMST is yet to be established, for the current study it was projected based on the results of previously published international MSTs (Akeroyd et al., 2015; Dietz et al., 2014; Hochmuth et al., 2012). Due to the consistency of the methodology employed in the development of MSTs internationally, it was anticipated that in the administration of the UCAMST and the UCAMST-P, the utilisation of two practice lists (i.e. 20 sentences) would be adequate. However, the potential exists for the training period of the UCAMST and the new UCAMST-P to differ from that of previously published international MSTs. If this is the case, then the validity of the results of the current study may be brought into question, as the estimates of slope and SRT obtained may have been influenced by participants continuing to adapt to the task during testing. Examination of the training effect for both the UCAMST and the UCAMST-P in future research is crucial in order to confirm that appropriate practice is afforded prior to testing. Previously, international MSTs have evaluated the training period via the adaptive procedure described by Brand and Kollmeier (2002), in which estimates of SRT are obtained for each list by employing two tracks that are randomly interleaved and converge at the 80% and 20% points. The number of lists utilised in the evaluation of training effects for MSTs internationally is varied, however, typically seven to eight double sentence lists (i.e. 20 sentences per list) were employed (Dietz et al., 2014; Hochmuth et al., 2012; Wagener et al., 2003). Consequently, it is recommended that future research pertaining to the training period for the UCAMST and the UCAMST-P follow the procedure outlined above. Determining the training effect for the UCAMST and the UCAMST-P is a crucial step in the tests' development and advancement towards both research and clinical use.

4.5.4 **The Impact of Editing**

As discussed in section 2.3.3, due to the extent of the editing required to develop the UCAMST-P, only 162 unique pseudo-sentences were able to be generated. In editing the existing 5 by 10 base matrix of the UCAMST to generate the UCAMST-P, the first two columns (name, verb) were removed; this reduced the size of the UCAMST-P base matrix to 3 by 10, allowing 10^3 (i.e. 1000) unique three-word pseudo-sentences to be generated. However, due to the manner in which the file fragments had to be edited, abnormal and unnatural auditory outputs were evident for the newly generated first column (quantity) (discussed in sections 2.3.1 and 2.3.2). A majority of these file fragments could not practically be edited to achieve more natural auditory outputs and, due to the time constraints of the current study, re-editing the file fragments was not feasible. Consequently, out of the 10 words in the quantity column, four had to be removed, forcing the UCAMST-P base matrix to be further reduced to three by six, allowing for 216 possible pseudo-sentences. This reduced number of unique sentences posed limitations surrounding the number of retests an individual would be able to undergo without sentences being repeated.

Accordingly, the question then arose as to whether to retain seven words in columns four and five in order to increase the number of possible sentences to 294 and, subsequently, increase the test-retest capacity of the UCAMST-P. It was proposed that seven buttons be displayed in the closed-set response format for all three columns, increasing the number of possible responses to 343. However, issues emerged surrounding which words from columns four and five of the 5 by 10 base matrix should be introduced to the existing three by six matrix. As illustrated in section 2.3.2, for both of these columns, the words that remained (i.e. those that had not been selected for the three by six base matrix) had lower slope values, higher adjustment difference values, and/or were more lexically difficult than the words included in the three by six base matrix. Consequently, whether or not to include a seventh word from columns four and five became a question of whether to: i) reduce the test's accuracy; ii) increase the test's lexical difficulty, and thus the age of children that the test would be appropriate for; or iii) have a lower number of possible sentences, thereby reducing the test-retest capacity of the UCAMST-P (i.e. decreasing the number of trials that could be run before sentences were reused). It was reasoned that, in this instance, reducing the test-retest capacity of the UCAMST-P outweighed the potential consequences related to increasing the test's lexical difficulty or lowering the test's accuracy. Furthermore, as the presentation of 30 sentences is required to obtain an individual's SRT, it was also reasoned that, even with the reduced number of possible sentences the UCAMST-P offers, the test affords sufficient test-retest capacity, offering a minimum of five tests before the sentences are reused. It is worth noting that while this reduction in the number of unique pseudo-sentences available for the UCAMST-P is not conducive to numerous rounds of repeat testing, the smaller base matrix may reduce search time in the closed-set response format, consequently improving the time efficiency of the test.

4.5.5 **Absence of a Babble Noise Condition**

In the current research project, despite the UCAMST-P being developed for use in the presence of both constant and babble noise, the babble noise condition was not examined. In a previous study conducted by Stone (2016), an undetected software malfunction resulted in non-optimised sentence lists being used in the evaluation of the babble noise condition for the UCAMST. Accordingly, accurate evaluation measures were not available for the UCAMST babble noise condition.

Concurrent to the current research project, Ripberger (in progress) obtained evaluation measures for the UCAMST in babble noise and quiet conditions, using sentence lists optimised for use in each respective condition. Based on Stone's (2016) findings regarding the sentence lists administered in constant noise, it is anticipated that the babble noise test material will be successfully optimised, and equivalence between the sentence lists for the constant and babble noise conditions of the UCAMST will be achieved.

The absence of a babble noise condition in the current study poses limitations with respect to the evaluation and development of the UCAMST and the UCAMST-P in the babble noise condition. Nevertheless, Kollmeier and colleagues (2015) recommended that MSTs use speech-shaped masking noise (i.e. constant noise) with the same spectral content as the speech stimuli. For MSTs internationally, including the UCAMST, constant masking noise has typically been generated through the repeated superimposition of the test's speech materials. Accordingly, constant masking noise can effectively mask the target speech stimuli, as it possesses the same long-term average speech spectrum as the target stimuli (King, 2010). Alternatively, as discussed in section 1.4.2, fluctuating maskers (i.e. babble noise or temporally modified speech noise) produce larger amplitude modulations, yielding depressions in the SNR of the masker envelope (Bacon et al., 1998; Hopkins & Moore, 2009). Individuals with NH are able to take advantage of this temporary release from masking, however, for individuals with HI, masking release is usually small or absent (Bacon et al., 1998; Hopkins & Moore, 2009). Consequently, individuals with NH typically perform better on speech recognition measures in the presence of a fluctuating masking noise than individuals with HI (Festen & Plomp, 1990; Hopkins & Moore, 2009; Peters et al., 1998; Wagener & Brand, 2005).

Wagener and Brand (2005) examined the effect of the type of masking noise employed on the test-retest reliability of the OlSa (Waneger et al., 1999a-c). Greater consistency and predictability in SRT levels was found when the OlSa was administered in the presence of constant noise, while greater variation was observed when fluctuating masking noise was employed, particularly in individuals with HI (Wagener & Brand, 2005). Accordingly, it was recommended that, while using adaptive measurements of SRT in the presence of an acoustic masker to discriminate between individuals with various levels of HI, the masking noise employed possessed spectral properties equivalent to the long-term average speech spectrum of the target stimuli (Wagener & Brand, 2005). Consequently, the administration of the UCAMST and the UCAMST-P in the presence of constant masking noise in the current project appeared justified.

4.6 Future research

4.6.1 **Inclusion of a Picture-Pointing Response Method**

One of the foremost considerations when administering speech recognition measures to the paediatric population is the response method adopted. Previous research has indicated that there are fundamental differences between the distinct informational processing demands that open- and closed-set speech perception tests impose on lexical access with regards to the competition amongst, and activation of, phonetically similar words (Clopper et al., 2006). Existing models of speech perception maintain that word recognition occurs with respect to other phonetically similar words (Clopper et al., 2006). Speech perception tasks employing open-set response formats utilise what is known as “bottom-up” processing, evaluated through lexical memory and acoustic-phonetic activation, whereas closed-set tasks employ “top-down” processing, in which potential responses are evaluated through phonological and lexical competition (Clopper et al., 2006).

Limitations are evident for both open- and closed-set response formats; closed-set formats have been found to be prone to guessing bias and training effects, whereas open-set methods present issues when testing individuals with limited language abilities and/or disordered speech (Ozimek et al., 2012). Consequently, when testing the paediatric population, the verbal responses acquired from open-set response methods may be more difficult to score accurately, especially in cases where the child has atypical speech production (e.g. due to HI) (Calandruccio, Gomez, Buss, & Leibold, 2014). Similarly, closed-set response formats can be problematic when a written response is required, particularly in instances where the child is young, has a developmental delay, or has minimal education (Calandruccio et al., 2014). Nevertheless, a majority of paediatric speech recognition measures utilise closed-set response formats, for example a picture pointing response (Elliott & Katz, 1980; Ross & Lerman, 1970).

Previous research has indicated that picture pointing is a valuable method of evaluating word recognition within the paediatric population (Hall, Grose, Buss, & Dev, 2002; Litovsky, 2005; Ross & Lerman, 1970). Picture pointing has been recommended as a behavioural assessment tool based on the guidelines developed by the American Speech-Language-Hearing Association (ASHA) concerning the audiological assessment of the paediatric population (ASHA, 2004). Ozimek and colleagues (2012) developed a picture pointing response format for the PPMST for use with children from three to six years of age. The PPMST paired the “target word” for each three-word pseudo-sentence with its “visually ambiguous” alternative (Ozimek et al., 2012). The illustrations were incorporated into a six-picture array containing the picture corresponding to the sentence presented, along with associated alternatives (Ozimek et al., 2012). During the development of the UCAMST-P in the current study, pairs of associated alternatives were identified during the generation of the base matrix (i.e. old and new, big and small, red and green). However, as a consequence of the strict time constraints of the current study, it was not feasible to develop a picture pointing response format for the UCAMST-P. The further development and incorporation of a picture pointing response format, as was conducted by Ozimek and colleagues (2012), appears warranted and is recommended. Such development would allow performance on the

UCAMST-P to be uninhibited by speech production skills or lack of phonological knowledge (Kosky & Boothroyd, 2003).

4.6.2 **Piloting with Children**

As the UCAMST-P is intended for use with the paediatric population, piloting the test with children is imperative. Due to the development of working memory throughout childhood and adolescence, the proficiency with which information can be updated progressively increases (discussed further in section 1.6.) (Brocki & Bohlin, 2004; Gathercole et al., 2004; Lendinez et al., 2015; Luna et al., 2004). Accordingly, differences in the reliability of the results obtained by the adult and paediatric populations have been found.

Wagener and Kollmeier (2005) investigated the extent to which the OlSa (Wagener et al., 1999a-c) could be used with primary school aged children with NH. The results of the primary school children were found to be less reliable than the results achieved by adults; this distinction was reasoned to be due to children having a shorter auditory memory span (Wagener and Kollmeier, 2005). In order to circumvent this, the OlKiSa was developed and evaluated for use with younger children (Wagener and Kollmeier, 2005; Neumann et al., 2012).

Based on the discrepancies in the results obtained for the adult and paediatric populations, piloting the UCAMST-P with children appears warranted. Previously, paediatric MSTs have been piloted with children with NH ranging from 4 to 10 years of age; these investigations have utilised an open-set response format and monaural presentation in the presence of either background noise or quiet (Hagermann & Hermansson, 2015; Neumann et al., 2012; Wagener and Kollmeier, 2005). In order to ensure understanding of the task and the test format, and to allow performance to stabilise before commencing testing, practice lists have also been employed prior to the presentation of test items in such investigations (Hagermann & Hermansson, 2015; Neumann et al., 2012; Wagener et al., 2003).

With regards to piloting the UCAMST-P, it would be beneficial to initially undertake evaluation of the test in constant noise, babble noise, and quiet (Wagener and Kollmeier, 2005). Following on from this, it would also be valuable to confirm the age range for which the test is applicable (Stephan & Muigg, 2008) and validate its outputs (Neumann et al., 2012). Additionally, prior to testing, examining the motivating factors of the UCAMST-P may be beneficial. Hagermann and Hermansson (2015) discussed the use of a 67% correct threshold, as opposed to their original 40% threshold, in order to maintain motivation for paediatric testing. Alternatively, auditory, or even visual, reinforcers could be incorporated between presentations in order to increase the child's interest and motivation (discussed further in section 1.7).

4.6.3 **Piloting with Individuals with Hearing Impairment**

To date, the UCAMST has only been investigated with individuals with NH; however, investigation of the UCAMST and the UCAMST-P with individuals with HI is also imperative. Previous literature has established that greater variation exists in the expected SRT between individuals with HI than among individuals with NH (Peters et al., 1998). This has, in part, been attributed to the effect of the spectrum of the masking noise employed (Peters et al., 1998). As discussed previously in section 1.4.2, in instances where the masking noise possesses a spectrum that differs from the spectrum of the target stimuli, a phenomenon known as masking release can arise, wherein dips in the SNR of the acoustic masker can afford the listener with a glimpse of the target stimuli (Füllgrabe et al., 2006; Hopkins & Moore, 2009; Howard-Jones, & Rosen, 1993). Individuals with NH are able to take advantage of this temporary release from masking, however, for individuals with HI, masking release is usually small or absent; accordingly, SRT is affected (Bacon et al., 1998; Hopkins & Moore, 2009). Consequently, for individuals with HI, examination of the anticipated performance for each of the acoustic maskers employed by the UCAMST and the UCAMST-

P is warranted. The purpose of such investigative research would be to ascertain normative data with which to compare the performance of individuals in order to establish the degree of difficulty encountered in the presence of background noise (Akeroyd et al., 2015).

4.6.4 **Cross-Validation with Other Speech Tests**

Further research pertaining to the UCAMST-P should also endeavour to address the cross-validation of these tests with those speech recognition measures commonly used in the audiological test battery in New Zealand. Unlike the research conducted by Stone (2016), in which comparisons were made across MSTs, the aim of such research would be to discern the information afforded by diverse speech recognition measures, in order to ascertain which tests provide information complementary to that offered by the UCAMST-P, for use in clinical practice. Based on the current practice, with regards to speech recognition testing in New Zealand, it is suggested that the UCAMST-P be cross-validated with the KTT (Antognelli, 1986). Additionally, despite inconsistencies in the incorporation of the QuickSIN (Killion et al., 2004) into the audiological test battery in New Zealand, due to the use of sentence stimuli in the presence of an acoustic masker, cross-validation of the UCAMST-P with the QuickSIN is also recommended. Such comparisons of the UCAMST-P with these commercially available speech recognition measures would afford enhanced understanding of the information able to be obtained from the UCAMST-P in relation to its commercially available counterparts. It is expected that valuable insight will be gained with regards to the information obtained from each test, for example, which tests are the most time efficient and suitable for use in a clinical test battery (Wilson et al., 2007a).

Based on the current practice, with regards to speech recognition testing in New Zealand, Ripberger (in progress) evaluated whether a correlation exists between the UCAMST and commonly used word recognition measures in New Zealand, including the meaningful CVC word lists (Boothroyd, 1968; Boothroyd & Nittrouer, 1988; Purdy et al.,

2000) and the QuickSIN (Killion et al., 2004). Previous research conducted by Andre (2016) has established a correlation between the results obtained from the UCAMST and the QuickSIN.

4.6.5 **Comparison Between the UCAMST and the UCAMST-P**

As the UCAMST-P was developed from the UCAMST, evaluating the UCAMST-P with respect to the UCAMST in both the AA and AV modes of presentation, using both open- and closed-set response formats, would be beneficial in identifying any distinction or disparities between the two tests. The data necessary to conduct these evaluations was gathered during the current research study, however it was not analysed or interpreted here, as it did not fall within the scope of the current study.

Nevertheless, preliminary inspection of the data revealed differences between the slopes of the intelligibility functions generated for the UCAMST and the UCAMST-P in the AA and AV modes of presentation for both the open- and closed-set response formats. Overall, in the AA mode of presentation, the slopes of the intelligibility functions generated for the UCAMST-P were found to be shallower – at 6.53 %/dB and 7.93 %/dB for closed- and open-set, respectively – than those generated for the UCAMST -10.41 %/dB and 13.46 %/dB for closed- and open-set, respectively.

As discussed previously, speech intelligibility functions with steeper slopes are able to afford a more accurate and sensitive measure of SRT, as a comparatively smaller change in SNR produces a larger change in SRT (Theunissen et al., 2009). Clinically, the use of more sensitive measures is considered to be valuable due to the limited time available to administer a sizeable battery of tests. Ozimek and colleagues (2010) noted that not only are highly sensitive measures of SRT more accurate, but they also afford a more efficient method of estimating SRT, therefore making such methods especially suitable when efficiency is vital. This rudimentary examination of the data indicates that the UCAMST remains superior to the

UCAMST-P, and thus should still be employed where possible. However, in instances where the UCAMST cannot be used, such as with the paediatric population or with those who have cognitive impairments, use of the UCAMST-P is justified.

4.7 Exploring the Impact of Working Memory

4.7.1 **Response Time**

Research has indicated that a listener's rate of response to an auditory speech stimulus (i.e. response time) is associated with the effort required to interpret the stimulus, as well as the listener's state of fatigue and the mode of presentation (i.e. AA or AV) (Fraser, Gagné, Alepins, & Dubios, 2010). Listeners with NH exert very little effort when listening in background noise, due to what has been termed 'selective gain' (Kerlin, Shahin, Miller, 2010). The brain is able to perform the necessary subconscious procedures that allow selective processing of a specific sound, while simultaneously filtering out irrelevant information (Kerlin et al., 2010). However, for listeners with a HI, listening in background noise has been found to be far more taxing, resulting in greater fatigue due to an increase in the listening effort and concentration needed to understand speech (Kramer et al., 2006).

Previously, listening effort and fatigue have been measured using three main methods: (1) psychophysical measures, (2) self-report, and (3) behavioural measures (Rudner, Lunner, Behrens, Sundewall Thorén, & Rönnberg, 2012). Psychophysical measures of listening effort refer to the recording of those transformations in autonomic and/or central nervous system activity through a task's implementation (McGarrigle, Munro, Dawes, Stewart, Moore, Barry, & Amitay, 2014). These effort related variations in central nervous system activity can be observed by event related potentials, functional magnetic resonance imaging, electroencephalography, and pupillometry (McGarrigle et al., 2014; Wendi, Hietkamp, & Lunner, 2017). Currently, it is understood that only a single attempt at using psychophysical measures to investigate listening related fatigue has been made (McGarrigle et al., 2014). This solitary effort, which examined cortisol levels, revealed no significant differences between school aged children with and without a HI following a full day of schooling (Hicks & Tharpe, 2002). Self-reported measures of listening effort provide insight into the amount of effort an individual invests into speech processing and commonly employ closed-set questionnaires (Gatehouse & Noble, 2004) or rating scales (Rudner et al., 2012). Such measurers of listening effort and fatigue are fast and simple to administer and do not require skilled expertise to deliver or interpret (McGarrigle et al., 2014). However, limitations are evident, as individuals hold different ‘thresholds’ as to what constitutes effort (Hällgren, Larsby, Lyxell, & Arlinger, 2005). Additionally, individuals may interpret the term ‘effort’ differently, relating it to task difficulty or accuracy of performance, as opposed to mental exertion (McGarrigle et al., 2014). Finally, behavioural responses to listening tests may be used as a measure of listening effort. Behavioural measures can be classified as one of two types: (1) multi-tasking paradigm or (2) single-task paradigm. The premise behind multi-task paradigms, such as dual-task methodologies, is that individuals have a ‘limited resource’ of cognitive energy (Kahneman, 1973). Consequently, when an individual is conducting two tasks concurrently and the primary task becomes more demanding, performance on the secondary task is diminished. However, as there is no way to independently measure the resources devoted to each task, it is unclear whether all of the remaining cognitive energy is directed towards the secondary task (Styles, 2006). Alternatively, single-task paradigms employ either a button pressing response (Houben, van Doorn-Bierman, & Dreschler, 2013) or verbal responses to speech stimuli (Gatehouse & Gordon, 1990). Research has suggested that additional information pertaining to an individual’s listening effort, in relation to speech perception, can be ascertained based on the speed of a correct response (Gatehouse & Gordon, 1990; Houben et al., 2013). Moreover, speech processing is thought to correspond with response times (McGarrigle et al., 2014). Consequently, due to the rapid rate at which speech is presented in everyday communication, information concerning slowed speech processing is an essential aspect to consider. Data pertaining to a listener's response time was recorded in the current research project for all of the responses made by each participant in all of the conditions employed. However, this data was not analysed or interpreted here, as it did not fall within the scope of the current study. However, it is anticipated that the response times of listeners will be shorter when responding to the UCAMST-P, as the requirements imposed on the listener's working memory are considerably less.

4.7.2 **Confusion Matrices**

As mentioned prior, a significant amount of additional data was generated and recorded during the current research project; this included the generation of several confusion matrices. For the open-set response format, these confusion matrices depicted the proportion of correct and incorrect responses for each word, in each sentence, in each list for both the UCAMST and the UCAMST-P. For example, in the AA open-set condition of the UCAMST (at -11.62 dB SNR) for sentence 41, "Kathy kept twelve green mugs", the word "Kathy" was recognised correctly 25% of the time and recognised incorrectly 75% of the time. Similarly, for the closed-set response format, the 'confusion' encountered by the participants could be observed to a greater extent. In the closed-set response format we were able identify where the confusion occurred, and what the confusion was. For example, if we look at the same sentence of the UCAMST (i.e. sentence 41) presented in the closed-set condition at -13.97 dB SNR, the word "twelve" was incorrectly recognised as "nine" 33% of the time, and "ten" 33% of the time, while being correctly recognised as "twelve" 33% of the time. Upon examination, this data has the potential to provide extensive information pertaining to the listener's confusion, affording insight into which words are commonly recognised incorrectly. Furthermore, information relating to the order in which words are selected in the closed-set condition, and whether or not the words selected earlier are more often correct than those selected later could offer unique insight into both auditory and working memory.

4.8 Concluding Statements

Speech audiometry is a fundamental component of both the adult and paediatric audiometric test batteries, affording valuable information extending beyond that conveyed by the audiogram. Owing to the advantages associated with MSTs, the implementation of such tests in speech audiometry based research has risen in the last decade. The MST is considered to be valuable both clinically and in a research context due to: i) the capacity to compare results not only between clinics, but also across dialects and languages; ii) the reliability, validity, and efficiency with which estimates of SRT can be produced; and iii) the superior repertoire of sentences that increase test-retest capacity.

Due to the cognitive demands of the existing New Zealand English MST, the UCAMST, the current study aimed to develop an audiometric speech recognition measure suitable for use with the paediatric population in New Zealand, the UCAMST-P. Following this, the current study aimed to evaluate the newly developed UCAMST-P, alongside its parent test, in order to establish the equivalence of the sentence lists for each test individually. Although the AA modality of the UCASMT had previously been evaluated, evaluation of the AV modality was an essential step in progressing the test towards its intended clinical application as a section of the UCAST platform. The results for both the UCAMST and the UCAMST-P suggested that, while the sentence lists were equivalently difficult in the AV modality, the same was not true for the AA mode of presentation. Further evaluation of the equivalence of the conditions within each modality indicated that although the reliability and accuracy with which estimates of SRT could be made were equivalent, the SRTs across each condition were not.

The implications of these findings from a clinical and research standpoint are important, providing information concerning the administration and interchangeablity of the sentence lists for the UCAMST and the UCAMST-P. Due to the differences highlighted by the results, further modifications and evaluation of the sentence lists in the AA modality are warranted for both of these tests. Subsequent research should make use of the additional sentence specific data collected in the current study and endeavour to address the limitations encountered. It is hoped that future research will continue this development and establish equivalence of the sentence lists for both the UCAMST and the UCAMST-P in order to allow these tests to be offered in both a clinical and research setting.

REFERENCE LIST

Akeroyd, M. A., Arlinger, S., Bentler, R. A., Boothroyd, A., Dillier, N., Dreschler, W. A., & Kollmeier, B. (2015). International Collegium of Rehabilitative Audiology (ICRA) recommendations for the construction of multilingual speech tests: ICRA working group on multilingual speech tests. *International Journal of Audiology, 54* (Suppl. 2), 17-22. http://dx.doi.org/ 10.3109/14992027.2015.1030513

Antognelli, T. (1986). The Kendall Toy Test Revisited: The Development of an Australian Version. Conference presentation, VII National Conference of the Audiology Society of Australia.

Arlinger, S. D. (1998). Clinical assessment of modern hearing aids. *Scandinavian Audiology, 27*(4), 50-53. http://dx.doi.org/10.1080/010503998420658

Arlinger, S. (2003). Negative consequences of uncorrected hearing loss – a review. *International Journal of Audiology*, *42*(sup2), 17-20. doi:10.3109/14992020309074639

Auer, E. T., & Bernstein, L. E. (2007). Enhanced Visual Speech Perception in Individuals With Early-Onset Hearing Impairment. *Journal of Speech Language and Hearing Research*, *50*(5), 1157. doi:10.1044/1092-4388(2007/080)

Bacon, S. P., Opie, J. M., & Montoya, D. Y. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *Journal of Speech Language and Hearing Research*, *41*(3), 549. doi:10.1044/jslhr.4103.549

Beattie, R. C., Barr, T., & Roup, C. (1997). Normal and hearing-impaired word recognition scores for monosyllabic words in quiet and noise. *British Journal of Audiology*, *31*, 153-164. http://dx.doi.org/10.3109/03005364000000018

Békésy, G. V. (1947). The Variation of Phase Along the Basilar Membrane with Sinusoidal Vibrations. *The Journal of the Acoustical Society of America*, *19*(3), 452-460. doi:10.1121/1.1916502

Bell, T. S., & Wilson, R. H. (2001). Sentence recognition materials based on frequency of word use and lexical confusability. *Journal of the American Academy of Audiology, 12*(10), 514-522.

Bench, J., Kowal, Å., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *British Journal of Audiology, 13*(3), 108-112. doi:10.3109/03005367909078884

Bess, F. H., & Humes, L. E. (2008). *Audiology: The Fundamentals.* Philadelphia, USA: Lippincott Williams & Wilkins.

Betts, J., Mckay, J., Maruff, P., & Anderson, V. (2006). The development of sustained attention in children: The effect of age and task load. *Child Neuropsychology, 12*(3), 205-221. http://dx.doi.org/10.1080/09297040500488522

Billings, C. J., Penman, T. M., Ellis, E. M., Baltzell, L. S., & McMillan, G. P. (2016). Phoneme and Word Scoring in Speech-in-Noise Audiometry. *American Journal of Audiology*, *25*(1), 75. doi:10.1044/2016_aja-15-0068

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27,* 32-48. http://dx.doi.org/10.1044/jshr.2701.32

Bird, P. A., & O'Beirne, G. A. (2015). Hearing loss in New Zealand – planning for the future. *New Zealand Medical Journal, 128*(1419), 6-8, from: http://www.nzma.org.nz /__data/assets/pdf_file/0003/43725/Ed-Bird-NZMJ1419.pdf

Boothroyd, A. (1968). Developments in speech audiometry. *British Journal of Audiology, 2*(1), 3-10. http://dx.doi.org/10.3109/00381796809075436

Boothroyd, A. (1968b). Statistical Theory of the Speech Discrimination Score. *The Journal of the Acoustical Society of America*, *43*(2), 362-367. doi:10.1121/1.1910787

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America, 84*(1). http://dx.doi.org/101-114. 10.1121/1.396976

Boothroyd, A. (2008). The performance/intensity function: An underused resource. *Ear & Hearing, 29*(4), 479-491. http://dx.doi.org/10.1097/AUD.0b013e318174f067

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment, *Audiology, 34*(5), 260-284. http://dx.doi.org/10.3109/00206099509071918

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America*, *111*(6), 2801-2810. doi:10.1121/1.1479152

Brocki, K. C., & Bohlin, G. (2004). Executive Functions in Children Aged 6 to 13: A Dimensional and Developmental Study. *Developmental Neuropsychology*, *26*(2), 571-593. http://dx.doi.org/10.1207/s15326942dn2602_3

Brown, L. G. (1996). Additional rules for the transformed up-down method in psychophysics. *Perception & Psychophysics*, *58*(6), 959-962.

doi:10.3758/bf03205497

Buss, E., Hall, J. W., Grose, J. H., & Dev, M. B. (2001). A comparison of threshold estimation methods in children 6–11 years of age. *The Journal of the Acoustical Society of America*, *109*(2), 727-731. doi:10.1121/1.1337960

Calandruccio, L., Gomez, B., Buss, E., & Leibold, L. J. (2014). Development and Preliminary Evaluation of a Pediatric Spanish–English Speech Perception Task. *American Journal of Audiology*, *23*(2), 158. doi:10.1044/2014_aja-13-0055

Carhart, R., & Tillman, T. W. (1970). Interaction of Competing Speech Signals With Hearing Losses. *Archives of Otolaryngology - Head and Neck Surgery*, *91*(3), 273-279. doi:10.1001/archotol.1970.00770040379010

Carhart, R. T., & Young, L. (1976). Development of test procedures for evaluation of binaural hearing aids: A final report. *Bulletin of Prosthetics Research, 26*, 9-45. Retrieved from: http://www.rehab.research.va.gov/jour/76/13/2/carhart.pdf

Cervera, T. C., Soler, M. J. Dasi, C., & Ruiz, J. C. (2009). Speech recognition and working memory capacity in young-elderly listeners: Effects of hearing sensitivity, *Canadian Journal of Experimental Psychology, 63*(3), 216-226. http://dx.doi.org/10.1037 /a0014321

Chisolm, T. H., Johnson, C. E., Danhauer, J. L., Portz, L. J., Abrams, H. B., Lesner, S., McCarthy, P. A., & Newman, C. W. (2007). A Systematic Review of Health-Related Quality of Life and Hearing Aids: Final Report of the American Academy of Audiology Task Force on the Health-Related Quality of Life Benefits of Amplification in Adults. *Journal of the American Academy of Audiology*, *18*(2), 151-183. doi:10.3766/jaaa.18.2.7

Clopper, C. G., Pisoni, D. B., & Tierney, A. T. (2006). Effects of Open-Set and Closed-Set Task Demands on Spoken Word Recognition. *Journal of the American Academy of*

Audiology, *17*(5), 331-349. doi:10.3766/jaaa.17.5.4

Cook, J. A., & Hawkins, D. B. (2006). Hearing loss and hearing aid treatment options. *Mayo Clinic Proceedings*, *81*(2), 234-237. http://dx.doi.org/10.4065/81.2.234

Cornsweet, T. N. (1962). The Staircase-Method in Psychophysics. *The American Journal of Psychology*, *75*(3), 485. doi:10.2307/1419876

Cox, R. M., Alexander, G. C., Gilmore, C. (1999). Development of the Connect Speech Test (CST). *Ear and Hearing, 8*(Suppl. 5), 119-126. http://dx.doi.org/10.1097/00003446-198710001-00010

Craik, F. I. M. (1994). Memory changes in normal ageing. *Current Directions in Psychological Science*, *3*(5), 155-158. http://dx.doi.org/10.1111/1467-8721.ep10770653

Dalton, D. S., Cruickshanks, K. J., Klein, B. E., Klein, R., Wiley, T. L., & Nondahl, D. M. (2003). The impact of hearing loss on quality of life in older adults. *The Gerontologist*, *43*(5), 661-668.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19(4),* 450-466. http://dx.doi.org/10.1016/s0022-5371(80)90312-6

Dick, A. S., Solodkin, A., & Small, S. L. (2010). Neural development of networks for audiovisual speech comprehension. *Brain and Language*, *114*(2), 101-114. doi:10.1016/j.bandl.2009.08.005

Dietz, A., Buschermöhle, M., Aarnisalo, A. A., Vanhagen, A., Hyyrynen, T., Aaltonen, O., & Kollmeier, B. (2014). The development and evaluation of the Finnish matrix sentence test for speech intelligibility assessment. *Acta Oto-Laryngologica, 134,* 728- 737. http://dx.doi.org/10.3109/00016489.2014.898185

Dillon, H. (2012). *Hearing aids*. Sydney: Boomerang Press.

Dirks, D. D., Morgan, D. E., & Dubno, J. R. (1982). A procedure for quantifying the effects of noise on speech recognition. *Journal of Speech and Hearing Disorders, 47*, 114-123. http://dx.doi.org/10.1044/jshd.4702.114

Donkelaar, H. J., & Kaga, K. (2011). The Auditory System. In H. J. ten Donkelaar (Ed). *Clinical Neuroanatomy: Brain Circuitry and Its Disorders.* New York: USA, Springer.

Etymotic Research, Inc. (2005a). *BKB-SIN Test. Version 1.03*. Elk Grove Village, IL: Etymotic Research, Inc.

Etymotic Research, Inc. (2005b). *BKB-SIN Test Manual*. Elk Grove Village, IL: Etymotic Research, Inc.

Egeth, H., & Kahneman, D. (1975). Attention and Effort. *The American Journal of Psychology*, *88*(2), 339. doi:10.2307/1421603

Elliott, L.L., & Katz, D. (1980). *Development of a new children's test of speech descrimination (Technical Manual)*. St. Louis, MO: Auditec.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech- reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*(4), 1725-1736. doi:10.1121/1.400247

Fletcher, H. (1940). Auditory Patterns. *Reviews of Modern Physics*, *12*(1), 47-65. doi:10.1103/revmodphys.12.47

Foo, C., Rudner, M., Rönnberg, J., Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *J Am Acad Audiol*, *18*, 553– 566. http://dx.doi.org/10.3766/jaaa.18.7.8

Francart, T., van Wieringen, A., & Wouters, J. (2011). Comparison of fluctuating maskers for speech recognition tests. *International Journal of Audiology, 50*, 2-13. http://dx.doi .org/10.3109/14992027.2010.505582

Fraser, S., Gagne´, J., Alepins, M., & Dubois, P. (2010). Evaluating the Effort Expended to Understand Speech in Noise Using a Dual-Task Paradigm: The Effects of Providing Visual Speech Cues. *Journal of Speech Language and Hearing Research*, *53*(1), 18-33. doi:10.1044/1092-4388(2009/08-0140)

Füllgrabe, C., Berthommier, F., & Lorenzi, C. (2006). Masking release for consonant features in temporally fluctuating background noise. *Hearing Research*, *211*(1-2), 74-84. doi:10.1016/j.heares.2005.09.001

Gatehouse, S., & Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. *British Journal of Audiology*, *24*(1), 63-68. doi:10.3109/03005369009077843

Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology*, *43*(2), 85-99. doi:10.1080/14992020400050014

Gates, G. A., & Mills, J. H. (2005). Presbycusis. *The Lancet*, *366*, 1111-1120. http://dx.doi.org/10.1016/S0140-6736(05)67423-5

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177-190. http://dx.doi.org/10.1037/0012-1649.40.2.177

Gelfand, S. A. (1998). Optimising the reliability of speech recognition scores. *Journal of Speech, Language, and Hearing Research, 41*(5), 1088-1102. Retrieved from: http://search.proquest.com.ezproxy.canterbury.ac.nz/docview/232356121?accountid=14499

Gelfand, S. A. (2010). *Hearing: An Introduction to Psychological and Physiological Acoustics*. London, UK: Informa Healthcare.

Gifford, R. H., Shallop, J. K., & Peterson, A. M. (2008). Speech Recognition Materials and Ceiling Effects: Considerations for Cochlear Implant Programs. *Audiology and Neurotology*, *13*(3), 193-205. doi:10.1159/000113510

Gilchrist, J. M., Jerwood, D., & Ismaiel, H. S. (2005). Comparing and unifying slope estimates across psychometric function models. *Perception and Psychophysics, 67*(7), 1289-1303. http://dx.doi.org/10.3758/BF03193560

Glasberg, B. R., & Moore, B. C. J. (1996). A model of loudness perception applied to cochlear hearing loss. *Auditory Neuroscience, 3,* 289-311.

Goodman, A. (1965). Reference zero levels for pure-tone audiometer. *ASHA, 75*(7), 262-263.

Grant, K. W., & Seitz, P. F. (1998). Measures of auditory–visual integration in nonsense syllables and sentences. *The Journal of the Acoustical Society of America*, *104*(4), 2438-2450. doi:10.1121/1.423751

Grant, K. W., & Walden, T. C. (2013). Understanding Excessive SNR Loss in Hearing-Impaired Listeners. *Journal of the American Academy of Audiology*, *24*(4), 258-273. doi:10.3766/jaaa.24.4.3

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America, 103(5)*, 2677-2690. http://dx.doi.org/10.1121/1.422788

Grunditz, M., & Magnusson, L. (2013). Validation of a speech-in-noise test used for verification of hearing aid fitting. *Hearing Balance and Communication, 18,* 118-125. http://dx.doi.org/10.3109/21695717.2013.782135

Hagerman, B. (1976). Reliability in the determination of speech discrimination. *Scand Audiol*, 5, 219–228. http://dx.doi.org/10.3109/14992027609044991

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian Audiology, 11*(2), 79-87. http://dx.doi.org/10.3109/01050398209076203.

Hagerman, B., & Hermansson, E. (2015). Speech Recognition in Noise in 5-Year-Old Normal Hearing Children. *Canadian Journal of Speech and Language Pathology and Audiology, 39*(1), 52-60.

Hall, J. W. (2008). Diagnostic applications of speech audiometry. *Seminars in Hearing, 4*(3), 179-203. http://dx.doi.org/10.1055/s-0028-1091425

Hall, J. W. (2014). *Introduction to Audiology Today.* New Jersey, USA: Pearson.

Hall, S. J. (2006). *The Development of a New English Sentence in Noise Test and an English Number Recognition test* (Master's Thesis), University of Southampton, United Kingdom.

Hall, J. W., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee Recognition in a Two-Talker Masker and a Speech-Shaped Noise Masker in Adults and Children. *Ear and Hearing*, *23*(2), 159-165. doi:10.1097/00003446-200204000-00008

Hällgren, M., Larsby, B., & Arlinger, S. (2006). A Swedish version of the Hearing in Noise Test (HINT) for measurement of speech recognition. *International Journal of Audiology, 45*(4), 227-237. http://dx.doi.org/10.1080/14992020500429583

Hällgren, M., Larsby, B., Lyxell, B., & Arlinger, S. (2001). Cognitive effects in dichotic speech testing in elderly persons. *Ear Hear, 22*, 120–129. http://dx.doi.org/10.1097/00003446-200104000-00005

Hällgren, M., Larsby, B., Lyxell, B., & Arlinger, S. (2005). Speech understanding in quiet and noise, with and without hearing aids. *International Journal of Audiology*, *44*(10), 574-583. doi:10.1080/14992020500190011

Hamid, M. A., & Brookler, K. H. (2006). Speech audiometry. *Ear, Nose, & Throat Journal, 85*(12), 810-812.

Hicks, C. B., & Tharpe, A. M. (2002). Listening Effort and Fatigue in School-Age Children With and Without Hearing Loss. *Journal of Speech Language and Hearing Research, 45*(3), 573. doi:10.1044/1092-4388(2002/046)

Hochmuth, S., Brand, T., Zokoll, M. A., Castro, F. Z., Wardenga, N., & Kollmeier, B. (2012). A Spanish matrix sentence test for assessing speech reception thresholds in noise. *International Journal of Audiology, 51,* 536-544. http://dx.doi.org/10.3109 /14992027.2012.670731

Hope, R. V. (2010). *Towards the development of the New Zealand hearing in noise test (NZHINT)* (Master's thesis), University of Canterbury, Christchurch, New Zealand. Retrieved from http://ir.canterbury.ac.nz/handle/10092/4815.

Hopkins, K., & Moore, B. C. J. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *Journal of the Acoustical Society of America, 125*(1), 442-446. http://dx.doi.org/10.1121/1.3037233

Howard-Jones, P. A., & Rosen, S. (1993). The perception of speech in fluctuating noise. *Acustica, 78*, 258-272.

Houben, R., Koopman, J., Luts, H., Wagener, K. C., van Wieringen, A., Verschuure, H., & Dreschler, W. A. (2014). Development of a Dutch matrix sentence test to assess speech intelligibility in noise. *International Journal of Audiology, 53*(10), 760-763. http://dx.doi.org/10.3109/14992027.2014.920111

Houben, R., Van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *International Journal of Audiology, 52*(11), 753-761. doi:10.3109/14992027.2013.832415

Humes, L. E. (1999). Dimensions of hearing aid outcome. *Journal of the American Academy* *of Audiology, 10*(1), 26-39.

Humes, L. E., Lister, J., Wilson, R,. Cacace, A., Cruickshanks, K., & Dubno, J. (2012). Central presbycusis: a review and evaluation of the evidence. *Journal of the American Academy of Audiology, 23*(8), 635-666. http://dx.doi.org/10.3766/jaaa.23.8.5

Humes L. E., Watson B. U., Christensen L. A., Cokely C. A., Halling D. A., Lee L. (1994). Factors associated with individual differences in clinical measures of speech recognition among the elderly. *Journal of Speech Language and Hearing Research*, *37*(2), 465. http://dx.doi.org/10.1044/jshr.3702.465

Hussain, S. S. M. (2008). *Conductive Hearing Loss* (Synopsis of Causation). Dundee, Scotland: Ministry of Defence.

Irwin, J. R., Whalen, D. H., & Fowler, C. A. (2006). A sex difference in visual influence on heard speech. *Perception & Psychophysics*, *68*(4), 582-592. doi:10.3758/bf03208760

Jamaluddin, S. A. (2016). *Development and Evaluation of the Digit Triplet and Auditory-Visual Matrix Sentence Tests in Malay* (Master's Thesis). University of Canterbury, Christchurch, New Zealand. Retrieved from: https://ir.canterbury.ac.nz/handle/10092/12288

Jamaluddin, S., & O'Beirne, G. A. (2015). *Development and evaluation of the auditory-visual matrix sentence test in Malay*: Paper presented at the Proceedings of the 7th Biennial Asia Pacific Conference of Speech, Language and Hearing for the Asia Pacific Society of Speech Language and Hearing, Guangzhou, China.

Jansen, S., Luts, H., Wagener, K. C., Kollmeier, B., Del Rio, M., Dauman, R., James, C., Fraysse, B., Vormès., Frachet, B., Wouters, J., & van Wieringen, A. (2012). Comparison of three types of French speech-in-noise tests: A multi-center study.

International Journal of Audiology, 51(3), 164-173. http://dx.doi.org/10.3109/14992027.2011.633568

Jerger J., Jerger S., Pirozzolo F. (1991). Correlational analysis of speech audiometric scores, hearing loss, age and cognitive abilities in the elderly. *Ear Hear, 12 (2),* 103–109. http://dx.doi.org/10.1097/00003446-199104000-00004

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

Kelly-Campbell, R. J., & Lessoway, K. (2015). Hearing aid and hearing assistance technology use in Aotearoa/New Zealand. *International Journal of Audiology, 54,* 308-315. http://dx.doi.org/10.3109/14992027.2014.979952

Kelly-Campbell, R. J., Thomas, K., & McMillan, A. (2014). Audiological outcomes for adults with mild hearing impairment: A pilot study. *Speech, Language and Hearing*, *18*(2), 74-82. doi:10.1179/2050572814y.0000000052

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional Gain Control of Ongoing Cortical Speech Representations in a "Cocktail Party". *Journal of Neuroscience, 30*(2), 620-628. doi:10.1523/jneurosci.3631-09.2010

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 116*(4), 2395-2405. http://dx.doi.org/10.1121/1.1784440

King, S. M. (2010). *Development and Evaluation of a New Zealand Digit Triplet Test for Auditory Screening (Master's Thesis).* University of Canterbury, Christchurch, New Zealand. Retrieved from: http://ir.canterbury.ac.nz/handle/10092/5679

Ko, J. (2010). Presbycusis and its management. *British Journal of Nursing*, *19*(3), 160-165. doi:10.12968/bjon.2010.19.3.46536

Kollmeier, B., Warzybok, A., Hochmuth, H., Zokoll, M.A., Uslar, V., Brand, T., & Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology,* 1-13. http://dx.doi .org/10.3109/14992027.2015.1020971

Kollmeier, B., & Wesselkamp, M. (1997). Development and evaluation of a German sentence test for objective and subjective speech intelligibility. *Journal of the Acoustical Society of America, 102*(4), 2412-2421. http://dx.doi.org/10.1121/1.419624

Kosky, C., & Boothroyd, A. (2003). Validation of an on-line implementation of the imitative test of speech pattern contrast perception (IMSPAC). *Journal of the American Academy of Audiology, 14*(2), 73-83. http://dx.doi.org/10.3766/jaaa.14.2.3

Kramer, S. E., Zekveld, A. A., & Houtgast, T. (2009). Measuring cognitive factors in speech comprehension: The value of using the text reception threshold test as a visual equivalent of the SRT test. *Scandinavian Journal of Psychology*, *50*(5), 507-515. doi:10.1111/j.1467-9450.2009.00747.x

Kramer, S. E., Kapteyn, T. S., & Houtgast, T. (2006). Occupational performance: Comparing normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing and Work. *International Journal of Audiology*, *45*(9), 503-512. doi:10.1080/14992020600754583

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, *63*(8), 1279-1292. doi:10.3758/bf03194543

Lendínez, C., Pelegrina, S., & Teresa Lechuga, M. (2015). Age differences in working memory updating: The role of interference, focus switching and substituting information. *Acta Psychologica*, *157*, 106-113. http://dx.doi.org/10.1016/j.actpsy.2015.02.015

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the*

Acoustical Society of America, 49(2), 467-477. http://dx.doi.org/10.1121/1.1912375

Levitt, H. (1978). Adaptive testing in audiology. *Scandinavian Audiology*(6), 241- 291.

Li, C. M., Zhang, X., Hoffman, H. J., Cotch, M. F., Themann, C. L., & Wilson, M. R. (2014). Hearing Impairment Associated With Depression in US Adults, National Health and Nutrition Examination Survey 2005-2010. *JAMA Otolaryngology, Head, & Neck Surgery, 140*(4), 293.

Li, Y., Healy, E. W., Drane, J. W., & Zhang, J. (2006). Comorbidity between and risk factors for severe hearing and memory impairment in older Americans. *Preventive Medicine*, *43*(5), 416-421. http://dx.doi.org/10.1016/j.ypmed.2006.06.014

Lin, F., Yaffe, K., Xia, J., Xue, Q., Harris, T., & Purchase-Helzner, E. (2013). Hearing loss and cognitive decline in older adults. *JAMA Internal Medicine*, *173*(4), 293. http://dx.doi.org/10.1001/jamainternmed.2013.1868

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of Cognitive Processes From Late Childhood to Adulthood. *Child Development*, *75*(5), 1357-1372. http://dx.doi.org/10.1111/j.1467-8624.2004.00745

Luts, H., Boon, E., Wable, J., & Wouters, J. (2008). FIST: A French sentence test for speech intelligibility in noise. *International Journal of Audiology*, *47*(6), 373-374. doi:10.1080/14992020801887786

Madell, J. R. (2008). Selecting appropriate technology: Hearing aids, FM, and cochlear implants. *The Hearing Journal*, *61*(11), 42. doi:10.1097/01.hj.0000342438.60428.bf

MacPherson, A., & Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends in Hearing, 18*, 1-26. http://doi.dx.org/10.1177/2331216514537722

Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental*

Child Psychology, *41*(1), 93-113. doi:10.1016/0022-0965(86)90053-6

Mathers, C., Smith, A., & Concha, M. (2000). Global burden of hearing loss in the year 2000. *Global Burden of Disease*, *18*, 1-30. Retrieved from: http://www.who.int/healthinfo /statistics/bod_hearingloss.pdf

Mattheyses,W., Latacz, L., & Verhelst, W. (2009). On the importance of audiovisual coherence for the perceived quality of synthesized visual speech. *EURASIP Journal on Audio, Speech, and Music Processing, 1*, 1-12. http://dx.doi.org/10.1155 /2009/169819

McArdle, R.A., & Chisolm, T. H. (2009). Speech Audiometry. In J. Katz., L. Medwetsky., R. Burkard., & L. J. Hood (Eds). *Handbook of Clinical Audiology* (pp. 64-79). Philadelphia, USA: Lippincott Williams & Wilkins.

McArdle, R. A., Wilson, R. H., & Burks, C. A. (2005). Speech recognition in multitalker babble using digits, words and sentences. *Journal of the American Academy of Audiology*, *16*(9), 726-739. http://dx.doi.org/10.3766/jaaa.16.9.9

Maclagan, M., & Hay, J. (2007). Getting fed up with our feet: Contrast maintenance and the New Zealand English “short” front vowel shift. *Language Variation and Change, 19*(1), 1-25. http://dx.doi.org/10.10170S0954394507070020

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*(2), 131-141. doi:10.3109/03005368709077786

McClelland, A. D. (2015). *Refinement and normalisation of the University of Canterbury Auditory-Visual Matrix Sentence Test* (Master’s Thesis). University of Canterbury, Christchurch, New Zealand. Retrieved from: http://ir.canterbury.ac.nz/handle/10092 /10456

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., &

Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal of Audiology*, *53*(7), 433-445. doi:10.3109/14992027.2014.890296

Mendel, L. L. (2008). Current considerations in pediatric speech audiometry. *International Journal of Audiology, 47*(9), 546-553. http://dx.doi.org/10.1080/14992020802252261

Moore, B. C. J. (2013). An introduction to the psychology of hearing (6th ed). Bingley, WYK: Emerald.

Most, T., Rothem, H., & Luntz, M. (2009). Auditory, Visual, and Auditory-Visual Speech Perception by Individuals With Cochlear Implants Versus Individuals With Hearing Aids. *American Annals of the Deaf, 154*(3), 284-292. doi:10.1353/aad.0.0098

Mulrow, C. D., Aguilar, C., Endicott, J. E., Tuley, M. R., Velez, R., Charlip, W. S., Rhodes, M. C., Hill, J. A., & DeNino, L. A. (1990). Quality-of-life changes and hearing impairment: A randomized trial. *Annals of Internal Medicine, 113*(3), 188-194. doi:10.7326/0003-4819-113-3-188

Newman, C. W., & Sandridge, S. A. (2004). Hearing loss is often undiscovered, but screening is easy. *Clevland Clinic Journal of Medicine, 71*(3), 225-232. http:// dx.doi.org/10.3949/ccjm.71.3.225

Neumann, K., Baumeister, N., Baumann, U., Sick, U., Euler, H. A., & Weißgerber, T. (2012). Speech audiometry in quiet with the Oldenburg Sentence Test for Children. *International Journal of Audiology, 51*(3), 157-163. http://dx.doi.org/10.3109/14992027.2011.633935

Nilsson, M. J., Soli, S. D., & Gelnett, D. J., (1996). *Development of the hearing in noise test for children (HINT-C)*. Los Angeles, CA: House Ear Institute.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95*(2), 1085-1099. http://dx.doi.org/10.1121/1.408469

North, B. V., Curtis, D., & Sham, P. C. (2002). A note on the calculation of empirical P values from Monte Carlo procedures. *The American Journal of Human Genetics*, *71*(2), 439-441. doi:10.1086/341527

Øygarden, J. (2009). *Norwegian Speech Audiometry* (Doctoral Thesis). Norwegian University of Science and Technology, Trondheim, Norway. Retrieved from: http://www.diva-portal.org/smash/get/diva2:214670/FULLTEXT01.pdf

Olsen, W. O., Van Tasell, D. J., & Speaks, C. E. (1997). Phoneme and word recognition for words in isolation and in sentences. *Ear & Hearing, 18*(3), 175-188. http://dx.doi.org /10.1097/00003446-199706000-00001

O'Beirne, G. A., McGaffin, A. J., & Rickard, N. A. (2012). Development of an adaptive low-pass filtered speech test for the identification of auditory processing disorders. *International Journal of Pediatric Otorhinolaryngology, 76*(6), 777-782. http:// dx.doi.org/10.1016/j.ijporl.2012.02.039

O'Beirne, G. A., Trouson, R. H., McClelland, A. D., Jamaluddin, S., Maclagan, M. A. (2015). Development of an auditory-visual matrix sentence test in New Zealand English. *12th European Federation of Audiological Societies Congress.* Istanbul, Turkey, 30 May 2015.

Olusanya, B. O., Neumann, K. J., & Saunders, J. E. (2014). The global burden of disabling hearing impairment: A call to action. *Bulletin of the Word Health Organisation, 92*(5), 367-373. http://dx.doi.org/10.2471/BLT.13.128728.

Orchik, D. J., Krygier, K. M., & Cutts, B. P. (1979). A comparison of the NU-6 and W-22 speech discrimination tests for assessing sensorineural hearing loss. *Journal of Speech and Hearing Disorders, 44*(4), 522-527. http://dx.doi.org/10.1044/jshd.4404.522

Ozimek, E., Kutzner, D., & Libiszewski, P. (2012). Speech intelligibility tested by the Pediatric Matrix Sentence test in 3-6 year old children. *Speech Communication, 54*, 1121-1131. http://dx.doi.org/10.1016/j.specom.2012.06.001

Ozimek, E., Kutzner, D., Sęk, A., & Wicher, A. (2009). Polish sentence tests for measuring the intelligibility of speech in interfering noise. *International Journal of Audiology, 48*(7), 433-443. http://dx.doi.org/10.1080/14992020902725521

Ozimek, E., Warzybok, A., & Kutzner, D. (2010). Polish sentence matrix test for speech intelligibility measurement in noise. *International Journal of Audiology, 49,* 444-454. http://dx.doi.org/10.3109/14992021003681030

Patel, H. C., Moitra, M., Modi, A., Contractor, J., & Kantharia, S. L. (2014). Impact of hearing loss on daily life style and schooling among children between 5 and 15 years ace-group. *National Journal of Community Medicine, 5*(1), 73-76.

Patel, M. X., Doku, V., & Tennakoon, L. (2003). Challenges in recruitment of research participants. *Advances in Psychiatric Treatment, 9*(3), 229-238. http://dx.doi.org /10.1192/apt.9.3.229

Patuzzi, R. (2009). Cochlear Mechanics *Encyclopedia of Neuroscience* (pp. 1041-1049). Oxford: Academic Press.

Peters, R. W., Moore, B. C. J., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *Acoustical Society of America, 103*(1), 577-587. http://dx.doi.org/10.1121 /1.421128

Pickles, J. O. (2012). *An Introduction to the Physiology of Hearing* (4th ed.). United Kingdom: Emerald Group Publishing Limited.

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *Journal of the Acoustical Society of America, 63*(2), 533-549. http://dx .doi.org.ezproxy.canterbury.ac.nz/10.1121/1.381753

Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology, 18*, 43-52. http://dx.doi.org/10.3109 /00206097909072618

Puglisi, G. E., Astolfi, A., Prodi, N., Visentin, C., Warzybok, A., Hochmuth, S., & Kollmeier, B. (2014). *Construction and first evaluation of the Italian matrix sentence test for the assessment of speech intelligibility in noise.* Paper presented at the FORUM ACUSTICUM, Kraków: Poland. Retrieved from: http://www.fa2014.agh.edu.pl /fa2014_cd/article/RS/R03C_6.pdf

Purdy, S. C., Arlington, B., & Johnstone, C. (2000). Normative data for the New Zealand recording of the CVC (Revised AB) word lists, *New Zealand Audiological Society Bulletin, 10*(2), 20-29.

Rogers, E. J. (2012). *Development and evaluation of the New Zealand children's-build-a-sentence test (NZ Ch-BAS).* University of Canterbury, Christchurch, New Zealand. Retrieved from: https://ir.canterbury.ac.nz/handle/10092/9433

Ross, M., & Lerman, J. (1970). A Picture Identification Test for Hearing-Impaired Children. *Journal of Speech Language and Hearing Research*, *13*(1), 44. doi:10.1044/jshr.1301.44

Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., & Rönnberg, J. (2012). Working Memory Capacity May Influence Perceived Effort during Aided Speech Recognition in Noise. *Journal of the American Academy of Audiology*, *23*(8), 577-589. doi:10.3766/jaaa.23.7.7

Ruytjens, L., Albers, F., van Dijk, P., Wit, H., & Willemsen, A. (2006). Neural responses to silent lipreading in normal hearing male and female subjects. *European Journal of Neuroscience*, *24*(6), 1835-1844. doi:10.1111/j.1460-9568.2006.05072.x

Scarinci, N., Worrall, L., & Hickson, L. (2008). The effect of hearing impairment in older people on the spouse. *International Journal of Audiology*, *47*(3), 141-151. doi:10.1080/14992020701689696

Schlauch, R. S., & Nelson, P. (2009). Puretone Evaluation. In J. Katz., L. Medwetsky., R. Burkard., & L. J. Hood (Eds). *Handbook of Clinical Audiology* (pp. 30-49). Philadelphia, USA: Lippincott Williams & Wilkins.

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-Visual Speech Perception and Auditory-Visual Enhancement in Normal-Hearing Younger and Older Adults. *Ear and Hearing*, *26*(3), 263-275. doi:10.1097/00003446-200506000-00003

Spehar, B. P., Tye-Murray, N., & Sommers, M. S. (2008). Intra-versus intermodal integration in young and older adults, *Journal of the Acoustical Society of America*, *123*(5), 2858-2866. http://dx.doi.org/10.1121/1.2890748

Spencer, G. A. (2011). *Effects of speaker age on speech understanding and listener effort in older adults* (Master's Thesis). University of Canterbury, Christchurch, New Zealand. Retrieved from: http://ir.canterbury.ac.nz/bitstream/handle/10092/6343/Thesis_Fulltext.pdf?sequence=1&isAllowed=y

Stone, J. M. (2016). *Evaluation of the University of Canterbury Auditory-Visual Matrix Sentence Test* (Master's Thesis). University of Canterbury, Christchurch, New Zealand. Retrieved from: https://ir.canterbury.ac.nz/handle/10092/12047.

Strelnikov, K., Rouger, J., Lagleyre, S., Fraysse, B., Deguine, O., & Barone, P. (2009). Improvement in speech-reading ability by auditory training: Evidence from gender differences in normally hearing, deaf and cochlear implanted subjects. *Neuropsychologia*, *47*(4), 972-979. doi:10.1016/j.neuropsychologia.2008.10.017

Styles, E. A. (2006). *The Psychology of Attention* (2nd ed). Hoboken: Taylor & Francis Ltd.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*(2), 212-215. http://dx.doi.org/10.1121/1.1907309

Taylor, B. (2003). Speech-in-noise tests: How and why to include them in your basic test battery. *The Hearing Journal, 56*(1), 40-43. http://dx.doi.org/10.1097/01.HJ.0000293000.76300.ff

Theunissen, M., Swanepoel, D. W., & Hanekom, J. (2009). Sentence recognition in noise: Variables in compilation and interpretation of tests. *International Journal of Audiology, 48*(11), 743-757. http://dx.doi.org/10.3109/14992020903082088.

Trounson, R. H. (2012). *Development of the UC Auditory-Visual Matrix Sentence Test* (Master's Thesis), University of Canterbury, Christchurch, New Zealand. Retrieved from: http://ir.canterbury.ac.nz/handle/10092/10348

Tye-Murray, N., & Geers, A. (2001). *Children's audio-visual enhancement test.* St. Louis, Mo: Central Institute for the Deaf.

Tye-Murray, N., Hale, S., Spehar, B., Myerson, J., & Sommers, M. S. (2014). Lipreading in school-age children: The roles of age, hearing status, and cognitive ability. *Journal of Speech, Language and Hearing Research, 57*(2), 556-565. http://dx.doi.org/10.1044/2013_JSLHR-H-12-0273

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007a). Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification, 11*(4), 233-241. http://dx.doi.org/10.1177/1084713807307409

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007b). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear & Hearing, 28*(5), 656-668. http://dx.doi.org/10.1097/AUD.0b013e31812f7185

Tye-Murray, N., Sommers, M. S., Spehar, B., Myerson, J., Hale, S., & Rose, N. S. (2008). Auditory-visual discourse comprehension by older and younger adults in favourable and unfavourable conditions. *International Journal of Audiology, 47*(2), 31-37. http://dx.doi.org/10.1080/14992020802301662

Uhler, K., Biever, A., & Gifford, R. H. (2016). Method of Speech Stimulus Presentation Impacts Pediatric Speech Recognition. *Otology & Neurotology*, *37*(2), e70-e74. doi:10.1097/mao.0000000000000911

Wagener, K., & Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal of Audiology, 44*(3), 144-156. doi: 10.1080/14992020500057517

Wagener, K. C., Brand, T., & Kollmeier, B. (1999b). Entwicklung und Evaluation eines Satztests für die Deutsche Sprache II: Optimierung des Oldenburger Satzetests (Development and evaluation of a sentence test for the German language II: Optimization of the Oldenburg sentence test). *Zeitschrift für Audiologie, 38(2),* 44-56.

Wagener, K. C., Brand, T., & Kollmeier, B. (1999c). Entwicklung und Evaluation eines Satztests für die Deutsche Sprache III: Evaluation des Oldenburger Satztests (Development and evaluation of a sentence test for the German language III: Evaluation of the Oldenburg sentence test). *Zeitschrift für Audiologie, 38(2),* 86-95.

Wagener, K. C., Eeenboom, F., Brand,s T., & Kollmeier, B. (2005). Ziffern-Tripel-Test: Sprachverständlichkeitstest über das Telefon. *Proceedings of 8th congress of the German Society of Audiology.*

Wagener, K., Josvassen, J. L., & Ardenkjær, R. (2003). Design, optimization and evaluation of a Danish sentence test in noise. *International Journal of Audiology, 42*(1), 10-17. http://dx.doi.org/10.3109/14992020309056080

Wagener, K., & Kollmeier, B. (2005). Evaluation des Oldenburger Satztests mit Kindern und Oldenburger Kinder-Satztest (Evaluation of the Oldenburg sentence test with children and the Oldenburg children's sentence test). *Zeitschrift für Audiologie, 44*, 134–143.

Wagener, K, C., Kühnel, V., & Kollmeier, B. (1999a). Entwicklung und Evaluation eines Satztests für die Deutsche Sprache I: Design des Oldenburger Satztests (Development and evaluation of a sentence test for the German language I: Design of the Oldenburg sentence test). *Zeitschrift für Audiologie, 38(2),* 4-15.

Walden, B. E., Grant, K. W., & Cord, M. T. (2001). Effects of Amplification and Speechreading on Consonant Recognition by Persons with Impaired Hearing. *Ear and Hearing*, *22*(4), 333-341. doi:10.1097/00003446-200108000-00007

Warzybok, A., Zokoll, M., Wardenga, N., Ozimek, E., Boboshko, M., & Kollmeier, B. (2015). Development of the Russian matrix sentence test. *International Journal of Audiology, 54*(suppl. 2), 35-43. http://dx.doi.org/10.3109/14992027.2015.1020969

Welberg, L. (2008). Auditory system: Making hairs for hearing. *Nature Reviews Neuroscience, 9*(10), 1-2. http://dx.doi.org/10.1038/nrn2496

Wells, J. C. (1982). *Accents of English* (Vol. 1). Cambridge: Cambridge University Press.

Wendt, D., Hietkamp, R. K., & Lunner, T. (2017). Impact of Noise and Noise Reduction on Processing Effort. *Ear and Hearing*, *38*(6), 690-700. doi:10.1097/aud.0000000000000454

Wilson, R. H. (2003). Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *Journal of the American Academy of Audiology, 14*(9), 453-470.

Wilson, R. H., McArdle, R. A., & Smith, S. L. (2007a). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with NH and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research, 50*(4), 844-856. http://dx.doi.org/10.1044/1092-4388(2007/059)

Wong, L. L. N., & Soli, S. D. (2005). Development of the Cantonese hearing in noise test (CHINT). *Ear and Hearing, 26*(3), 276-289. http://dx.doi.org/10.1097/00003446-200506000-00004

World Health Organisation, (2008). *The Global Burden of Disease: 2004 Update.* Heneva: World Health Organisation.

Valente, M. (2009). *Pure-Tone Audiometry and Masking.* Oxfordshire, United Kingdom: Plural Publishing.

van Rooij, J. C. G. M., & Plomp, R. (1990). Auditive and cognitive factors in speech perception by elderly listeners II: Multivariate analyses. *Journal of the Acoustical Society of America, 88*(6). 2611-2642. http://dx.doi.org/ 10.1121/1.399981

van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America, 111*(4), 1906-1916. http://dx.doi.org/10.1121/1.1456928

Versfeld. N. J., Daalder, L., Festen. J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *Journal of the Acoustical Society of America, 107*(3), 1671-1684. http://dx.doi.org/10 .1121/1.428451

Zeng, F. G., & Liu, S. (2006). Speech perception in individuals with auditory neuropathy. *Journal of Speech, Language, and Hearing Research, 49*(2), 367-380. http://dx .doi.org/10.1044/1092-4388(2006/029)

Zokoll, M. A., Hochmuth, S., Warzybok, A., Wagener, K. C., Buschermöhle, M., & Kollmeier, B. (2013). Speech-in-noise tests for multilingual hearing screening and diagnostics. *American Journal of Audiology, 22*(1), 175-178. http://dx.doi.org /10.1044/1059-0889(2013/12-0061).

Appendices

APPENDIX A:

Ethical Approval

A.1 Letter of ethical approval, University of Canterbury Human Ethics Committee.

UC
UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE
Secretary, Rebecca Robinson
Telephone: +64 03 369 4588, Extn 94588
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2017/06/LR

20 March 2017

Pace Jenkins-Foreman
Communication Disorders
UNIVERSITY OF CANTERBURY

Dear Pace

Thank you for submitting your low risk application to the Human Ethics Committee for the research proposal titled "Development of a Paediatric Version of the University of Canterbury Auditory-Visual Matrix Sentence Test".

I am pleased to advise that the application has been reviewed and approved.

Please note that this approval is subject to the incorporation of the amendments you have provided in your email of 10th March 2017.

With best wishes for your project.

Yours sincerely

pp. R. Robinson

Associate Professor Jane Maidment
Chair, Human Ethics Committee

University of Canterbury Private Bag 4800, Christchurch 8140, New Zealand. www.canterbury.ac.nz

F E S

APPENDIX B:

Recruitment

B.1 Email invitation used during recruitment.

Hi everyone,

VOLUNTEERS NEEDED!

We are developing an exciting new speech test call the University of Canterbury Paediatric Auditory-Visual Matrix Sentence Test (UCAMST-P). This test will use both auditory and visual stimuli to diagnose hearing loss in New Zealand children.

If you are:

- are 18 years of age or older
- have normal hearing
- are a native speaker of NZ English
- have no current middle ear pathology (i.e. ear infection)

Then we would like to hear from you!

This study will take place at the University of Canterbury Speech and Hearing Clinic at Creyke Road, Ilam, throughout 2017.

You would be needed for one session of less than two hours, during this time you will:

- receive a free hearing check
- help to develop an exciting new paediatric speech test for use in NZ clinics
- get a first hand look at this new speech test
- receive a $20 fuel voucher

For more information, or to be involved in this project, please contact **Pace Jenkins-Foreman** at **pace.jenkins-foreman@pg.canterbury.ac.nz**, text/call **021 1111 221**, or go to **https://audiovisualstudy.setmore.com/** to book a session time.

Thank you for reading ☺

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee

APPENDIX C:

Informed Consent

C.1 Information sheet provided to participants in the current study (page 1 of 2).

Information Sheet

Full Project Title: Development the University of Canterbury Paediatric Auditory-Visual Matrix Sentence Test

Principal Researcher: Pace Jenkins-Foreman, MAud student (2nd year)
Department of Communication Disorders

Research Supervisor: Associate Professor Greg O'Beirne
Department of Communication Disorders

Associate Supervisor: Dr. Rebecca Kelly-Campbell, Senior Lecturer
Department of Communication Disorders

This study is part of a project to develop a paediatric auditory-visual speech test in New Zealand English, which will supplement the information gathered from other tests typically used in audiology. This project aims to determine if there is a difference in participant's scores between the existing five-word sentence test and the new paediatric three-word sentence test.

The test will take place at the University of Canterbury Speech and Hearing Clinic.

To be eligible to participate, you must:

- be 18 years of age or older
- be a native speaker of New Zealand English
- have normal hearing
- have no current middle ear pathology (i.e. ear infections)

Prior to any testing, you will be asked for a history of your ear health and hearing, which ethnic group you belong to, and your ears will be examined. You will then undergo a hearing check (if you have not provided an audiologist-completed audiogram dated within six months) and I will inform you of the results. If you would like me to, I can write a letter summarising the results if you would like to follow up on this with your GP or an audiologist. If you choose to follow up with your GP, this will be at your own expense. Should an unexpected hearing impairment be discovered, a full audiological assessment will be offered at the University of Canterbury Speech and Hearing Clinic free of charge. If a conductive hearing loss were to be identified during the hearing check, you will receive a $10 fuel voucher for your time.

Following the hearing check the study will take place. The study will consist of eight parts; in each part you will hear, and sometimes see, a series of short sentences in background noise. They may be hard to hear. You will be asked to respond by either saying the sentence you heard or by using the touch screen to select your answers.

The study should take no more than 90 minutes, and will be completed after the hearing check.

This study is being carried out as part of a Masters of Audiology. The information I obtain from you will be used in further development of this test so that it may be used as a diagnostic tool.

I am happy to answer any queries you may have. My phone and email details are provided in case you have any questions at a later date. In recognition of the time and effort involved on your behalf, you will receive an honorarium of $20, as well as a free hearing check.

I have provided a consent form for you to sign prior to participating in this study.

C.1 Information sheet provided to participants in the current study (page 2 of 2).

Signing this indicates your understanding that the data collected in this study will not be anonymous, but it will be confidential, and only viewed by people directly involved in this study (those listed at the top of the first page). Participation is voluntary and you have the right to withdraw without penalty. If you withdraw, I will remove all of the information relating to you.

The project has been reviewed and approved by the University of Canterbury Human Ethics Committee.

For your own reference, please take this form away with you.

With thanks,

Pace Jenkins-Foreman
Second Year MAud Student
Department of Communication Disorders
University of Canterbury
Email: pace.jenkins-foreman@pg.canterbury.ac.nz
Phone: 021 1111 221

Greg O'Beirne, PhD
Primary research supervisor & Associate Professor in Audiology
Department of Communication Disorders
University of Canterbury
Private Bag 4800, Christchurch 8140, New Zealand
Email: gregory.obeirne@canterbury.ac.nz
Phone: +64 3 3642987 ext. 7085

Rebecca Kelly-Campbell
Secondary research supervisor & Senior Lecturer in Audiology
Department of Communication Disorders
University of Canterbury
Private Bag 4800, Christchurch 8140, New Zealand
Email: rebecca.kelly@canterbury.ac.nz
Phone: +74 3 364 2977 ext. 8327

Alternatively, if you have any complaints, please contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz), phone: +64 3 364 2987.

C.2 Consent form signed by all participants in the current study (page 1 of 2).

Consent Form for Persons Participating in Research Studies

Full Project Title: ***Development of a Paediatric Version of the University of Canterbury Auditory-Visual Matrix Sentence Test***

I have read and understand the Information Sheet.

I, ______________________________________ agree to participate in this project according to the conditions in the Information Sheet. I will be given a copy of Information Sheet and Consent Form to keep.

The researcher has agreed not to reveal the participant's identity and personal details if information about this project is published or presented in any public form.

I agree that research data gathered in this study may be published and used in future studies. I provide consent for this publication and the re-use of the data with the understanding that my name or other identifying information will not be used.

I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after five years.

I understand the risks associated with taking part and how they will be managed.

I understand that I can contact the researcher or supervisor for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz)

I would like to receive a report on the findings of the study at the conclusion of the study (please tick one):

Yes □ No □

If yes, please provide a contact email and/or postal address below:

....

C.2 Consent form signed by all participants in the current study (page 2 of 2).

By signing below, I agree to participate in this research project.

Signature ....

Date ....

***Note: All parties signing the Consent Form must date their own signature*. Please return the consent form to the researcher before you actively participate in this research.**

APPENDIX D:

Sentence Lists

D.1 Sentence lists used for the UCAMST, showing the sentences and corresponding sentence numbers.

List 1		**List 2**		**List 3**		**List 4**	
41	Kathy kept twelve green mugs	51	Amy sees twelve cheap spoons	61	Hannah has those red mugs	71	David sold nine red bikes
42	Peter has three good toys	52	Rachel sold nine new books	62	Peter gives twelve cheap toys	72	Peter bought six big ships
43	Oscar sees those red ships	53	Oscar wants three red toys	63	Thomas wants four small bikes	73	Thomas wants some good mugs
44	Thomas got eight cheap books	54	David has six good coats	64	David got ten dark coats	74	Amy likes four new books
45	David bought two big hats	55	Peter got those green shoes	65	Kathy kept six large spoons	75	Oscar got those green spoons
46	Sophie wins ten new shoes	56	Sophie likes two large hats	66	Rachel bought nine big shirts	76	William kept two dark hats
47	Amy sold six small bikes	57	William gives some dark shirts	67	Amy sold eight old shoes	77	Hannah sees twelve large shirts
48	Hannah likes some large shirts	58	Thomas kept ten small ships	68	Oscar likes some new ships	78	Rachel wins eight old coats
49	Rachel gives nine dark spoons	59	Kathy bought four big mugs	69	William wins three good books	79	Kathy has three small toys
50	William wants four old coats	60	Hannah wins eight old bikes	70	Sophie sees two green hats	80	Sophie gives ten cheap shoes
List 5		**List 6**		**List 7**		**List 8**	
81	Oscar gives six dark coats	91	Hannah gives those green hats	101	Hannah got those large shoes	111	Oscar wants twelve dark shoes
82	Hannah sees ten small ships	92	Sophie has two dark spoons	102	Thomas wants three small books	112	David kept six red ships
83	William wins two red hats	93	Thomas sees some old shirts	103	Oscar sold some dark shirts	113	Rachel got nine cheap hats
84	Sophie has nine cheap spoons	94	Peter sold six small coats	104	William sees twelve new ships	114	Thomas gives some green spoons
85	Thomas wants some large shoes	95	William likes three good shoes	105	Amy bought eight big bikes	115	Hannah wins two small bikes
86	Amy got eight good toys	96	David bought nine big ships	106	Peter gives ten cheap toys	116	Amy sees ten old coats
87	Rachel bought four big mugs	97	Amy kept twelve new bikes	107	Rachel wins four old coats	117	Peter has those large toys
88	David likes those green shirts	98	Rachel wins ten large mugs	108	Sophie has six good spoons	118	Sophie bought three big shirts
89	Peter sold three old books	99	Kathy wants four red toys	109	David likes two red mugs	119	William sold four good mugs
90	Kathy kept twelve new bikes	100	Oscar got eight cheap books	110	Kathy kept nine green hats	120	Kathy likes eight new books
List 9		**List 10**		**List 11**		**List 12**	
121	Amy gives twelve dark coats	131	Thomas likes two small spoons	141	David sold three large coats	151	Rachel gives those cheap spoons
122	David wins those cheap shirts	132	Kathy got some cheap shoes	142	Rachel has twelve red shoes	152	Kathy likes three good books
123	Kathy sold nine red books	133	Rachel wins three red mugs	143	Hannah gives six dark mugs	153	Oscar has twelve old coats
124	William has some new spoons	134	Oscar kept six green ships	144	Thomas sees eight small ships	154	Hannah sees nine new bikes
125	Thomas sees eight small hats	135	Sophie bought ten big shirts	145	Oscar likes some new shirts	155	Peter got some green mugs
126	Rachel got two good toys	136	Peter gives eight good toys	146	Sophie got nine cheap hats	156	Amy wants four red toys
127	Oscar kept six green mugs	137	Hannah sold those large bikes	147	Amy wants those green toys	157	William wins two dark shoes
128	Hannah likes three large shoes	138	William has nine old books	148	Kathy wins four old books	158	Sophie kept eight large shirts
129	Peter bought four big ships	139	Amy sees four new coats	149	Peter bought ten big spoons	159	Thomas bought ten big ships
130	Sophie wants ten old bikes	140	David wants twelve dark hats	150	William kept two good bikes	160	David sold six small hats
List 13		**List 14**		**List 15**		**List 16**	
161	William sold eight old mugs	171	Hannah gives some old spoons	181	Peter wins nine green spoons	191	Peter got three dark toys
162	Rachel got six dark coats	172	Thomas sees those green ships	182	Oscar has twelve large shoes	192	Rachel sold four red shoes
163	Kathy kept three small bikes	173	David wants twelve red mugs	183	Amy gives ten dark toys	193	Amy sees ten new bikes
164	Peter wins ten green toys	174	Rachel has eight dark books	184	David kept six good hats	194	Kathy likes some good mugs
165	Hannah gives those red shirts	175	William kept six good shirts	185	Thomas likes three new books	195	William kept those large shirts
166	David wants four good books	176	Sophie wins two small bikes	186	Sophie wants those red shirts	196	Thomas wants twelve small coats
167	Oscar sees twelve new shoes	177	Kathy sold three new toys	187	Kathy sold some small bikes	197	Sophie has two green books
168	Amy bought nine big ships	178	Peter got nine cheap shoes	188	Rachel got four cheap coats	198	Hannah gives nine cheap hats
169	Thomas has some cheap hats	179	Amy bought four big hats	189	William sees eight old ships	199	David wins eight old spoons
170	Sophie likes two large spoons	180	Oscar likes ten large coats	190	Hannah bought two big mugs	200	Oscar bought six big ships

D.2 Sentence lists used for the UCAMST-P, showing the sentences and corresponding sentence numbers.

List 1		**List 2**		**List 3**		**List 4**	
2	eight big books	21	eight old hats	13	eight new bikes	10	eight green shoes
47	nine green spoons	28	eight red shoes	69	nine small hats	23	eight old spoons
54	nine new toys	61	nine red bikes	94	ten old shoes	52	nine new shoes
88	ten new shoes	81	ten green hats	101	ten red spoons	65	nine red spoons
95	ten old spoons	132	three old toys	110	three big books	89	ten new spoons
138	three red toys	139	three small bikes	117	three green hats	93	ten old hats
156	twelve green toys	150	twelve big toys	158	twelve new books	109	three big bikes
177	twelve small hats	161	twelve new spoons	172	twelve red shoes	128	three old books
202	two old shoes	194	two new books	192	two green toys	174	twelve red toys
207	two red hats	214	two small shoes	203	two old spoons	180	twelve small toys
						213	two small hats
List 5		**List 6**		**List 7**		**List 8**	
29	eight red spoons	14	eight new books	9	eight green hats	17	eight new spoons
33	eight small hats	96	ten old toys	16	eight new shoes	27	eight red hats
48	nine green toys	100	ten red shoes	60	nine old toys	38	nine big books
50	nine new books	114	three big toys	64	nine red shoes	97	ten red bikes
58	nine old shoes	127	three old bikes	108	ten small toys	115	three green bikes
90	ten new toys	159	twelve new hats	137	three red spoons	129	three old hats
142	three small shoes	173	twelve red spoons	154	twelve green shoes	134	three red books
151	twelve green bikes	176	twelve small books	163	twelve old bikes	157	twelve new bikes
167	twelve old spoons	183	two big hats	181	two big bikes	178	twelve small shoes
182	two big books	191	two green spoons	197	two new spoons	190	two green shoes
				200	two old books		
List 9		**List 10**		**List 11**		**List 12**	
1	eight big bikes	3	eight big hats	12	eight green toys	11	eight green spoons
18	eight new toys	59	nine old spoons	25	eight red bikes	26	eight red books
51	nine new hats	66	nine red toys	53	nine new spoons	63	nine red hats
92	ten old books	76	ten big shoes	98	ten red books	78	ten big toys
102	ten red toys	120	three green toys	105	ten small hats	91	ten old bikes
130	three old shoes	125	three new spoons	143	three small spoons	131	three old spoons
135	three red hats	166	twelve old shoes	146	twelve big books	144	three small toys
152	twelve green books	175	twelve small bikes	160	twelve new shoes	148	twelve big shoes
179	twelve small spoons	193	two new bikes	168	twelve old toys	195	two new hats
187	two green bikes	206	two red books	199	two old bikes	205	two red bikes
List 13		**List 14**		**List 15**		**List 16**	
19	eight old bikes	20	eight old books	24	eight old toys	22	eight old shoes
30	eight red toys	36	eight small toys	37	nine big bikes	39	nine big hats
46	nine green shoes	45	nine green hats	57	nine old hats	62	nine red books
49	nine new bikes	86	ten new books	84	ten green toys	77	ten big spoons
74	ten big books	121	three new bikes	87	ten new hats	82	ten green shoes
99	ten red hats	136	three red shoes	116	three green books	133	three red bikes
141	three small hats	149	twelve big spoons	124	three new shoes	140	three small books
155	twelve green spoons	165	twelve old hats	169	twelve red bikes	153	twelve green hats
196	two new shoes	209	two red spoons	210	two red toys	162	twelve new toys
215	two small spoons	211	two small bikes	212	two small books	216	two small toys

APPENDIX E:

Data Tables

E.1 The SRT values for each sentence list in each condition for the UCAMST-P and the UCAMST.

1

				SRT (dB SNR)																
				List 1	**List 2**	**List 3**	**List 4**	**List 5**	**List 6**	**List 7**	**List 8**	**List 9**	**List 10**	**List 11**	**List 12**	**List 13**	**List 14**	**List 15**	**List 16**	**Mean ± SD (no outliers)**
UCAMST-P	Auditory alone	Open-set	With practice	-6.71	-4.04	-6.54	-6.16	-6.68	-4.49	-6.14	-4.38	-6.51	-2.27	-6.59	-4.46	-5.74	-4.77	-6.25	-4.30	-5.38 ± 1.3
			No practice	-6.60	-5.03	-6.81	-6.23	-6.46	-5.40	-6.67	-4.90	-6.53	-5.69	-6.50	-5.20	-6.37	-4.94	-6.31	-5.75	-5.96 ± 0.68
		Closed-set	With practice	-8.00	-7.74	-7.97	-7.56	-7.91	-6.66	-7.89	-8.18	-8.24	-6.99	-8.00	-6.67	-7.00	-6.97	-7.66	-8.12	-7.6 ± 0.55
			No practice	-8.13	-7.31	-7.98	-7.54	-7.71	-8.00	-7.40	-7.92	-8.14	-5.90	-8.38	-7.49	-7.96	-6.83	-7.53	-7.73	-7.62 ± 0.6
	Auditory-visual	Open-set	With practice	-301.84	-10.50	-13.96	-10.89	-10.22	-9.99	-19.26	-8.12	-28.28	-10.18	-20.06	-10.26	-22.81	-9.67	-8.99	-10.09	-11.71 ± 3.77
			No practice	-9.55	-12.91	-30.98	-15.35	-10.22	-10.22	-11.06	-11.96	-13.01	-12.64	-12.21	-9.06	-13.21	-13.29	-14.02	-50.46	-12.05 ± 1.81
		Closed-set	With practice	-11.86	-14.68	-38.80	-19.61	-13.78	-15.53	-26.32	-13.74	-20.56	-14.35	-12.22	-14.52	-18.11	-11.67	-15.88	-13.41	-14.15 ± 1.83
			No practice	-15.13	-14.79	-10.71	-13.79	-13.80	-22.11	-14.03	-14.79	-14.52	-12.17	-13.14	-17.55	-10.81	-13.41	-14.59	-12.73	-13.73 ± 1.73
UCAMST	Auditory alone	Open-set	With practice	-9.61	-8.57	-7.98	-8.32	-8.29	-8.80	-8.30	-8.64	-8.46	-10.02	-8.78	-8.36	-8.96	-7.55	-8.65	-8.59	-8.62 ± 0.58
			No practice	-8.40	18.07	-9.57	-8.34	-7.53	-8.98	-7.64	36.71	-9.10	-7.21	-7.68	-9.32	-8.22	-6.99	-8.31	-5.62	-8.06 ± 1.06
		Closed-set	With practice	-10.19	-10.04	-10.00	-10.10	-10.28	-10.42	-10.20	-9.31	-10.00	-9.78	-10.12	-10.77	-9.93	-10.00	-10.85	-10.33	-10.15 ± 0.36
			No practice	-9.25	-10.15	-9.62	-8.61	-9.68	-9.65	-9.23	-9.48	-10.16	-9.22	-9.68	-10.33	-8.80	-7.43	-10.16	-9.39	-9.43 ± 0.72
	Auditory-visual	Open-set	With practice	-42.87	-15.46	-39.24	-11.83	-13.18	-12.48	-23.31	-16.15	-16.05	-12.40	-13.72	-15.32	-24.84	-11.85	-14.14	-14.42	-15.37 ± 3.98
			No practice	-40.06	-12.48	-187.78	-12.56	-25.65	-11.80	-12.00	-12.68	-167.61	-17.54	-32.10	-13.44	-11.84	-12.34	-17.22	-11.77	-14.28 ± 4.11
		Closed-set	With practice	-16.30	-15.24	-17.12	-25.88	-15.62	-16.88	-15.41	-18.48	-17.87	-15.13	-16.00	-16.29	-16.88	-24.20	-20.63	-17.21	-17.82 ± 3.15
			No practice	-24.74	-15.19	-14.24	-15.70	-33.22	-17.06	-20.58	-15.35	-14.33	-15.98	-41.45	-21.76	-22.43	-14.64	-31.64	-15.98	-19.52 ± 6.19

E.2 The slope values for each sentence list in each condition for the UCAMST-P and the UCAMST.

2

				Slope (%/dB)																
				List 1	**List 2**	**List 3**	**List 4**	**List 5**	**List 6**	**List 7**	**List 8**	**List 9**	**List 10**	**List 11**	**List 12**	**List 13**	**List 14**	**List 15**	**List 16**	**Mean ± SD (no outliers)**
UCAMST-P	Auditory alone	Open-set	With practice	8.20	5.23	10.81	5.14	11.51	4.04	11.04	2.93	8.88	3.88	8.24	4.62	12.52	2.75	12.82	4.19	7.3 ± 3.6
			No practice	7.55	5.26	7.88	7.36	16.84	4.64	9.32	3.70	11.06	6.83	15.52	4.49	11.63	6.80	12.18	5.91	8.56 ± 3.91
		Closed -set	With practice	7.51	5.66	7.03	5.28	5.07	5.52	7.42	7.80	8.45	5.05	7.51	4.27	7.53	6.30	7.68	5.50	6.47 ± 1.28
			No practice	6.90	7.63	5.63	6.45	6.44	6.61	7.15	6.67	6.81	4.78	8.27	7.30	5.87	5.40	7.41	5.99	6.58 ± 0.9
	Auditory-visual	Open-set	With practice	0.02	1.79	1.14	1.87	2.15	1.84	0.52	3.29	0.36	2.08	0.24	2.21	0.36	2.75	2.68	1.96	1.58 ± 1.01
			No practice	1.69	1.16	0.36	0.73	1.90	2.39	1.63	1.46	1.44	0.98	1.13	2.62	1.26	1.20	0.83	0.18	1.31 ± 0.65
		Closed -set	With practice	1.54	1.09	0.23	0.66	1.24	1.14	0.43	1.03	0.68	1.14	1.86	1.40	0.67	1.85	0.93	1.32	1.08 ± 0.47
			No practice	0.96	1.35	2.09	1.27	1.34	0.47	1.29	1.35	1.40	1.81	1.50	0.78	2.64	1.32	1.24	1.00	1.22 ± 0.33
UCAMST	Auditory alone	Open-set	With practice	16.83	8.16	18.92	8.31	21.58	7.51	17.88	6.52	14.38	6.32	18.69	9.41	21.72	7.78	16.00	5.10	12.82 ± 5.95
			No practice	12.30	0.88	14.54	7.98	28.06	9.91	15.35	0.67	11.09	7.68	17.47	9.20	22.21	11.10	26.24	5.52	12.51 ± 7.96
		Closed -set	With practice	12.00	10.54	12.87	6.76	10.75	8.52	11.68	9.44	15.40	8.66	11.78	10.36	12.64	10.68	12.14	9.95	10.89 ± 2.04
			No practice	12.27	9.93	11.09	7.73	9.95	9.18	10.50	8.83	10.56	11.18	11.61	10.90	8.39	5.80	12.68	8.46	9.94 ± 1.8
	Auditory-visual	Open-set	With practice	1.27	7.50	1.69	12.58	12.07	12.94	2.46	6.36	7.31	12.38	11.22	6.73	3.33	14.82	10.66	6.74	8.13 ± 4.39
			No practice	1.70	10.58	0.16	9.63	2.25	11.11	7.89	12.57	0.02	2.61	1.91	11.65	8.91	14.15	5.15	4.52	6.55 ± 4.78
		Closed -set	With practice	12.95	9.66	7.10	3.99	8.40	7.28	9.22	5.23	10.02	14.92	9.27	8.11	7.28	4.24	5.05	7.03	8.11 ± 2.96
			No practice	3.74	8.86	7.31	10.87	1.43	4.30	4.64	7.71	8.38	8.83	1.46	2.42	3.86	10.46	1.87	8.83	5.94 ± 3.3

Note: The outliers highlighted in Appendices E.1 and E.2 (defined as "Extreme values" by SPSS), were removed for graphing purposes only and were retained in all non-parametric analyses.