FURTHER EVALUATION AND VALIDATION OF THE UNIVERSITY OF CANTERBURY 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

A. R. Ripberger

University of Canterbury

2018

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation for my primary supervisor, Dr Greg O'Beirne, whose dedication and enthusiasm for this project is inspiring. I am very grateful for all the work you have done to ensure the completion of this thesis despite the (exceptionally stringent) time constraints we encountered at different points throughout the research process.

To my secondary supervisor, Dr Rebecca Kelly-Campbell, I am so genuinely thankful for your assistance and guidance for the statistical aspects of this project and for always making the time to meet with me and reply to my desperate emails!

I would also like to acknowledge the MAuD staff. Your support and accommodating attitude was incredibly valued especially during the data collection phase of this project.

To the Palmerston North Bay Audiology team, thank you for letting me use your clinic to test participants and doing your utmost to assist me in any way possible.

Thank you to all the people who sacrificed their time and energy to be participants in this study (some more than once!) and those who helped with recruitment. I truly appreciate your willingness to be involved – this thesis is possible because of you!

I can't express how grateful I am for my friends and family for all their encouragement and support despite my frequent complaints and lack of free time! Thank you Daniel for going the extra mile to find people to participate when I needed them at such short notice and Lynnea for spending the better part of a day helping me proof-read! To my dear mum, I don't have the words to say just how much I truly value the consistent encouragement and practical support you have provided me with over the course of the last two years. Without you, I honestly would have given up shortly after I'd started!

Last, but certainly not least, I want to acknowledge my husband Georg. Thank you for always believing in me (especially when I don't believe in myself) and enabling me to pursue my dreams. I am so grateful for everything you've sacrificed during my studies and your unconditional love through it all. I am so lucky to have you!

ABSTRACT

Speech testing is an important part of the audiological test battery as it able to provide an index of a listener's hearing ability beyond what can be revealed with conventional puretone audiometry. Matrix Sentence Tests (MSTs) assess speech understanding in noise and are thought to better approximate the hearing deficits a person might experience in 'real world' situations. The dialectical differences of New Zealand English necessitated the creation of the University of Canterbury Auditory-Visual Matrix Sentence Test (UCAMST; O'Beirne, Trouson, McClelland, Jamaluddin, & Maclagan, 2015; Trounson, 2012), which was adapted from the British English MST to accommodate the unique phonology of New Zealand English and ensure the validity of the measure. As part of a series of studies aimed at progressing the UCAMST towards clinical use, this project sought to continue the process of examining the equivalency across the lists and conditions available for use each of the presentation modes included in the UCAMST. Evaluation with 61 normal hearing participants revealed the sentence lists designed for use in the babble noise condition to be equivalent, however sentences designed for presentation in quiet were significantly different. An additional assessment of 20 normal hearing participants found comparability between numerous condition pairs in terms of the accuracy by which a listener's speech reception threshold is estimated. The data from these participants was also used to cross-validate the UCAMST with two speech audiometry tests routinely carried out in NZ, demonstrating a significant relationship between the speech recognition thresholds of the UCAMST and the meaningful CVC (revised AB) word recognition test. The findings of this study provide evidence for the interchangeable use of sentence lists and conditions in the UCAMST and the potential capacity of the UCAMST to replace the meaningful CVC (revised AB) word recognition test in clinical practice.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABG	Air Bone Gap
AC	Air Conduction
AH/R	Aural Habilitation/Rehabilitation
ANOVA	Analysis of Variance
BC	Bone Conduction
BG	Background
BKB	Bench Kowal Bamford
CHI	Conductive Hearing Impairment
CVC	Consonant Vowel Consonant
dB	Decibels
dB A	‘A’ weighted Decibels
HA	Hearing Aid
HAPI	Hearing Aid Performance Inventory
HF	High Frequency
HI	Hearing Impairment
HINT	Hearing in Noise Test
HL	Hearing Level
Hz	Hertz
IHCs	Inner Sensor Hair Cells
LF	Low Frequency
MST	Matrix Sentence Test
NH	Normal Hearing
NIHI	Noise-Induced Hearing Impairment
NZ	New Zealand

NZAS	New Zealand Audiological Society
OHCs	Outer Sensory Hair Cells
PI	Performance Intensity
PI_{max}	Maximum Performance Intensity
PT	Puretone
QuickSIN	Quick Speech in Noise
SD	Standard Deviation
$s50_{\mathrm{test}}$	Test Specific Function
SIN	Speech in Noise
SNHI	Sensorineural Hearing Impairment
SNR	Signal-to-Noise Ratio
SPIN	Speech Perception in Noise
SPL	Sound Pressure Level
SPSS	Statistical Package for the Social Sciences
SRT	Speech Reception Threshold
TM	Tympanic Membrane
TURMatrix	Turkish Matrix Sentence Test
UCAMST	University of Canterbury Auditory-Visual Matrix Sentence Test

NOMENCLATURE

In accordance with the World Health Organisation's recommendations for a shift in the conceptualisation of disability towards health and functioning (World Health Organisation, 2002), this study exclusively uses the term "hearing impairment" in place of "hearing loss".

Additionally, in an effort to maintain a person-centred perspective, the current study refers to hearing impairment as an experienced deficit instead of what might be construed as a diagnostic label. For example, the phrase "a person with a hearing impairment" is used instead of "a hearing impaired person". Such measures are paid particular attention in this research in order to portray a more holistic view of health which incorporates biological, individual and social components.

To recognition of partnership as a key construct which forms the basis for professional relationships in audiological contexts, people seeking audiology services are herein referred to as "clients" and not "patients".

These substitutions are made with the hope that this research might more accurately and appropriately describe the experience of having a hearing impairment and acknowledge the client's authority over and active participation in the management of their own hearing health. Additionally, the nomenclature in this study is intentionally used to maintain an anti-stigmatising approach when discussing hearing impairment.

CHAPTER 1 *LITERATURE REVIEW*

1.1 Background

Hearing impairment (HI) is a significant health concern worldwide (Tucci, Merson, & Wilson, 2010). The latest data from the World Health Organisation (2017) estimates 328 million adults have at least a moderate HI, equating to approximately one third of all people over the age of 65 affected. This pattern is also widespread in New Zealand (NZ) where the prevalence of HI was reported to be as high as 28.5% of adults over the age of 65 years, according to the latest census data (Statistics New Zealand, 2014). Coincidently, this percentage is estimated to increase, particularly for adults in the 60 years and older age bracket, due to NZ's aging population (Exeter et al., 2015). This is of utmost concern as a HI can have far reaching implications in many areas of a person's life including their safety, employability, and also their physical, psychological, and social health and wellbeing.

Personal safety can become an issue for a person with a HI, as reduced hearing sensitivity may make it more difficult to hear alarms or warnings in emergency situations, whether that be inside the home, in traffic, or out in the community (Ohene-Djan, Hersh, & Naqvi, 2010; Picard et al., 2008). Having a HI has also been connected with non-participation in education and employment and those with a HI in the workforce are more likely to be over-represented within the low income wage bracket (Garramiola-Bilbao & Rodríguez-Álvarez, 2016; Hogan, O'Loughlin, Davis, & Kendig, 2009; Järvelin, Mäki-torkko, Sorri, & Rantakallio, 1997).

In addition to this, having a HI has been inexplicably linked with physical functioning. Indeed, self-reported ability to participate in activities of daily living, lower-extremity mobility and levels of physical activity were decreased for people with a HI compared with people who had normal hearing (NH; Chen et al., 2015; Chen, Genther, Betz, & Lin, 2014; Gispen, Chen, Genther, & Lin, 2014). HI has also been associated with numerous physical health conditions including dementia, cardiovascular disease, diabetes, and arthritis (Cruickshanks & Wichmann, 2012; Fritze et al., 2016; Stam et al., 2014).

Beyond the practical and physical implications that a HI can have on a person's functioning in their everyday life, it may also influence a person's psychosocial health and wellbeing. Indeed, having a HI has been linked to decreased social participation as the person affected may distance themselves from interpersonal interactions to avoid embarrassment and listening fatigue (Mick, Kawachi, & Lin, 2014; Mikkola et al., 2015). This phenomenon may also contribute to the higher rates of anxiety and depression experienced by people with a HI (Contrera et al., 2017; Kim et al., 2017). In combination, these factors result in a decreased perceived quality of life for people who have a HI compared to those with NH (Lessoway, 2014).

The impact of a HI is not just limited to the individual affected. It can also involve friends, family and significant others through third party disability (Scarinci, Worrall, & Hickson, 2009). Having a significant other with a HI can lead to anxiety and frustration for family members and spouses. These feelings can arise from difficulties communicating in a variety of situations when the listening environment is less than ideal and as a result having to repeat things they or others have said (Scarinci, Worrall, & Hickson, 2008; Wallhagen, Strawbridge, Shema, & Kaplan, 2004). This can put a strain on marital relationships and restrict the couple's ability to engage with others socially (Armero, 2001; Scarinci et al., 2008). For this reason and those described above, it is essential that every action is taken to provide effective rehabilitation for people who have a HI in order to secure the best possible outcomes for themselves and their families.

Audiologists strive to minimise the impact that a HI can have through the provision of aural habilitation/rehabilitation (AH/R). This is most frequently administered through the prescription of amplification which is delivered via a hearing aid (HA; Chisolm et al., 2007; Kelly-Campbell & Lessoway, 2015). In order for audiologists to do this effectively, it is important that they obtain hearing test results that accurately represent a person’s hearing thresholds; most commonly this is tested using puretone audiometry (PT) and plotted on an audiogram. These results can then be used to estimate how much amplification a person with a HI may need depending on the amplification prescription algorithm which is employed.

While the audiogram provides information about a person’s hearing acuity and is useful for the prescription of amplification, literature has found that neither the configuration nor degree of HI is predictive of HA satisfaction and perceived benefit (Bertoli, Bodmer, & Probst, 2010; Vestergaard Knudsen, Öberg, Nielsen, Naylor, & Kramer, 2010). This phenomenon could be due to the fact that the stimulus used in PT audiometry has limited translatability to the listening environments in the ‘real world’. Indeed, these listening environments can be characterised by a multitude of sounds including environmental and speech sounds, for which speech tests are better designed to measure a person’s hearing deficits (Dietz et al., 2014; Ozimek, Kutzner, Sek, & Wicher, 2009).

Therefore, it is common practice in NZ, and internationally, to carry out speech testing to augment the audiogram and reveal a person’s ability to detect and recognise speech (Boothroyd, 1968; Boothroyd & Nittrouer, 1988; Hamid & Brookler, 2006).

Speech tests form a valuable part of the audiological test battery as they require the listener to undertake tasks which are necessary for communication in the 'real world'; that is, the person is asked to identify words which are presented either in quiet or in the presence of background noise. The results from the speech test can then be interpreted by the audiologist to assist in determining the type of HI a person may have and inform their decision making regarding potential candidacy for trialling a HA (Hall, 1983; Niemeyer, 1976). For this reason, speech recognition tests have become an integral part of audiological practice in NZ and throughout the world, and form the basis of this study. Following the development of a new matrix sentence test for NZ, this project seeks to further evaluate the new tool and cross-validate its performance against other speech tests commonly used in audiological practice in NZ in order to progress it towards routine research and clinical use.

In order to discuss the new proposed method of assessment, it is important to first understand the structure and function of the normal auditory system and the anatomy of hearing impairment, which is outlined in the following sections.

1.2 The anatomy of hearing

The auditory system consists of four parts: the outer-ear, the middle-ear, the inner-ear, and the auditory nervous system (Møller, 2013). These parts all have a unique role in the transformation of an audible stimulus into the sensation of hearing. The acoustic stimulus is essentially a travelling pressure wave characterised by two important features; the first being the frequency of the wave, measured in Hertz (Hz), and the second being the amplitude, typically measured in decibels (dB) relative to some standard reference (Moore, 2012).

The function of the outer-ear is to collect the sounds in the environment and transfer them down the external auditory canal towards the middle-ear. This is made possible due to the specific features of the external, partially cartilaginous pinna and concha. These features act like a funnel which, together with the resonant attributes of the external auditory canal, results in an increased sound pressure level (SPL) at the tympanic membrane (TM; Pickles, 2012). The outer-ear also provides important information for sound localisation achieved via analysis of differences in the timing and intensity of the acoustic stimuli obtained from each ear (Grothe, Pecka, & McAlpine, 2010).

The next portion of the auditory system is the middle-ear. This is a cavity that is separated from the outer-ear by the TM and houses the ossicular chain. This ossicular chain is comprised of three small bones: the malleus, the incus, and the stapes–which are supported in space by the middle-ear muscles (Møller, 2013). The TM is attached to the manubrium of the malleus at the umbo and connects to the oval window of the cochlea by the footplate of the stapes. Sound waves from the outer-ear are transmitted into the middle-ear through the TM along the ossicular chain. The main function of the middle-ear is to perform acoustic impedance matching to minimise the amount of acoustic energy that is reflected at the oval window. Impedance matching is achieved through pressure transformation and the lever action of the ossicles (Pickles, 2012; Tonndorf & Khanna, 1966).

The third part of the auditory pathway is the inner-ear, the anterior portion of which is the cochlea (Pickles, 2012). The cochlea is separated into three fluid-filled canals: scala tympani and scala vestibuli, which contain the high-sodium perilymph, and scala media, which contains the high-potassium endolymph. Scala media is separated from scala tympani by Reissner’s membrane and from scala vestibuli by the basilar membrane. The organ of Corti sits on the basilar membrane and contains the sensory hair cells which are distributed tonotopically along the membrane, with high frequencies situated at the basal end and low frequencies at the apical end (Shera, 2015). There are two types of sensory hair cells; the outer hair cells (OHCs) and the inner hair cells (IHCs). The electromotile OHCs are arranged in three to four rows along the outer edge of the basilar membrane and actively increase the amplitude of basilar membrane vibration. The IHCs are arranged in a single row along the inner edge of the basilar membrane and are largely responsible for neurotransmission onto afferent neurones (Yost, 2007). When a vibratory sound stimulus enters the cochlea through the round window it results in a transverse wave that travels along the basilar membrane and vibrates maximally in the area which corresponds to the characteristic frequency of the sound stimulus (Ni, Elliott, Ayat, & Teal, 2014). The OHCs contribute to the basilar membrane vibration amplitude through electro-mechanical transduction. This comes about via the contraction and elongation of the voltage-sensitive motor protein prestin in response to cyclical changes in intracellular potential produced by mechanoelectrical transduction (Nowotny & Gummer, 2006). While this is occurring, the IHCs also carry out mechanoelectrical transduction, wherein the cyclical changes in membrane potential alter the intracellular calcium concentration and lead to the release of the neurotransmitter glutamate onto afferent Type I auditory nerve fibres (Fettiplace & Kim, 2014).

The fourth and final part of the auditory system is the auditory nervous system which is comprised of the auditory nerve, various relay nuclei, and the auditory cortex of the brain. The neural pathway from the cochlea to the brain consists of type I primary afferent neurones which synapse with the IHCs, and type II afferent neurones which synapse with the OHCs. The type I afferents constitute 95% of the total afferent neurones and can be further categorised according to their spontaneous rate. These neurones are arranged tonotopically and carry electrical signals to the cochlear nucleus. They do this via the vestibule-cochlear nerve, also known as the VIIIth cranial nerve or the auditory nerve (Humphries, Liebenthal, & Binder, 2010; Musiek & Baran, 2007). Along the pathway to the auditory cortex, sound is analysed at four major points or nuclei: the cochlear nucleus, superior olivary complex, the central nucleus of the contralateral inferior colliculus, and the lateral portion of the ventral division of the contralateral medial geniculate body (Pickles, 2012). These nuclei are involved in signal analysis and processing to aid in sound localisation, frequency discrimination and binaural hearing (Pickles, 2012). Finally, information is received by the auditory cortex which is theorised to carry out a range of different functions including analysis of complex sounds, sound localisation, ear selection, response inhibition, identification of stimulus, discrimination of temporal patterns, and concept formation (Pickles, 2012).

1.3 Hearing impairment

In order to explore the types of tests and assessments used to diagnose a HI, it is vital to have an understanding about how a HI manifests along the auditory pathway and the effect this has on a person’s audiometric performance. HI is an expansive concept which encompasses a range of different pathologies and is characterised by a partial or complete inability to hear. Pathologies most commonly involve an interruption or impairment of any of the structures along the auditory pathway from the outer-ear through to the auditory cortex. Depending on the location of injury, specific features are detectable when assessed audiometrically, which has led to the current practice of classifying a HI as either being conductive, sensorineural, or mixed. A conductive HI (CHI) can be caused by disorders which result in the non-transmission of sound stimuli to the cochlea, whereas a sensorineural HI (SNHI) may be caused by disorders which affect the functioning of the sensory cells or the auditory nervous system. A HI is considered mixed when it contains CHI and SNHI components, suggesting pathology in multiple areas along the auditory pathway (Møller, 2013).

1.3.1 Conductive hearing impairment

The aetiology of a CHI can be attributed to a variety of different disorders, which may result in the obstruction of the external auditory canal, imbalanced pressurisation of the middle-ear space, inflammatory processes, and/or disruption of the ossicular chain (Møller, 2013). Conditions affecting the external auditory canal which could result in a measurable CHI can include external auditory canal atresia, collapsing canals, and impacted cerumen. These may occur as a result of a congenital abnormality, genetic predisposition, and/or age related changes in the structure of the outer-ear (Randolph & Schow, 1983; Subha & Raman, 2006; Zhang et al., 2016). A perforation in the TM or fluid in the middle-ear space due to otitis media with effusion, can also present as a CHI (Butler & Williams, 2003; Ristovska, Jachova, Filipovski, & Atanasova, 2016). Otitis media with effusion is characterised by inflammation causing a negative pressure in the middle-ear space that may lead to a retracted TM when the absolute pressure within the middle-ear is less than atmospheric pressure (Mansour, Magnan, Haidar, & Nicolas, 2015). In severe cases this can cause the erosion of the ossicles and which will exacerbate a CHI (James et al., 2012). A cholesteatoma is a lesion of the ear made up of a mass of keratinising epithelium, which can develop as a result of, and concurrent with, otitis media with effusion and retraction pockets in the TM (Caponetti, Thompson, & Pantanowitz, 2009). If untreated it can cause a CHI through erosion of the ossicles and progress to a SNHI if the inner-ear becomes involved (Bhutta, Williamson, & Sudhoff, 2011; Kumar, Seshaprasad, & Ahmed, 2016). An interruption or fixation of the ossicular chain is another pathology which decreases the transmission of sound to the cochlea and can result in a CHI (Møller, 2013). This occurs with ossicular discontinuity when the ossicles become detached through a physical trauma to the ear or skull, and when the ossicles become immobilised due to a genetic connective disorder known as otosclerosis (Farahmand et al., 2016; Thys & Camp, 2009). Depending on the aetiology, a CHI can be improved or even completely resolved through medical management and/or surgery which is why a CHI is often considered temporary.

Audiometrically, a CHI is characterised by elevated air-conduction (AC) thresholds with better underlying hearing, measurable via bone-conduction (BC). Therefore, an air-bone gap (ABG) in the audiogram of ≥ 15 dB is commonly seen, and is the requirement for classifying a HI as being conductive or sensorineural. LF and high frequency (HF) hearing thresholds can be involved depending on the aetiology (Farahmand et al., 2016; Kumar et al., 2016; Park et al., 2015; Ravicz, Rosowski, & Merchant, 2004; Ristovska et al., 2016). With respect to otosclerosis, a "Carhart's notch" may be present at 2 kHz on the audiogram, observed as a decrease in the BC thresholds at that frequency by approximately 30 dB (Kashio et al., 2011). Speech discrimination performance may be affected to varying degrees depending on the nature of the CHI. For example a client with a flat CHI may demonstrate no change in their maximum score if the stimulus presentation level is high enough, and speech-in-noise understanding may be unchanged if the speech component is at audible levels (Hsieh, Lin, Ho, & Liu, 2009; Nia & Bance, 2001)

1.3.2 Sensorineural hearing impairment

SNHI occurs when there is injury to the sensory hair cells in the cochlear, the neural pathway from the cochlea to the brain, and/or the central auditory nervous system. SNHI differs from CHI as it is usually permanent and intractable (Møller, 2013). Within the cochlea, mostly OHCs but also IHCs can become biologically damaged through exposure to excessively loud noise, resulting in what is known as a noise-induced hearing impairment (NIHI; Straatman, Lea, Le, & Westerberg, 2017). Ototoxic medications such as aminoglycosides, macrolides, glycopeptide antibiotics, and chemotherapeutic agents can all cause a SNHI via damage to sensory hair cells and tissue degeneration in the cochlea (Taneja, Taneja, Varshney, & Varshney, 2015). A SNHI can also develop as part of a disease process such as Meniere's disease (associated with a fluctuating SNHI; Sajjadi & Paparella, 2008), or an infection, which may be congenital, in the case of perinatal cytomegalovirus (Foulon, 2008), or acquired as part of the pneumococcal meningitis sequelae for example (Perny et al., 2016). Along with the brain and other organs embedded in the skull, the cochlea can become damaged by a blow to the head. This happens most commonly with transverse fractures as opposed to longitudinal fractures of the temporal bone which are more commonly associated with producing a CHI (Singh, Singh, & Singh, 2013). Decreased hearing sensitivity can occur as a result of age-related changes or presbycusis. Presbycusis is multi-factorial but generally involves the degeneration of the stria vascularis—the battery of the cochlea—leading to a loss of HF hearing sensitivity (Gates & Mills, 2005). Further, research has long shown that the auditory nerve and auditory cortex can also be implicated in presbycusis, which may account for the reduced word recognition scores of those affected (Frisina, 1997).

SNHI caused by damage to the OHCs and IHCs influences cochlear tuning (Moore, 2001). Indeed, the psychophysical tuning curves measured for individuals with a SNHI are broader and less defined compared with NH listeners (Carney & Nelson, 1983; Kluk & Moore, 2004; Moore, 2012). The decreased frequency selectivity of the cochlea for clients with a SNHI results in reduced speech discrimination scores and poorer performance for tests assessing speech understanding in the presence of background noise (Moore, 2001; Ostler, & Crandell, 2001).

The neural pathway connecting the cochlea with the central auditory nervous system is another potential area of pathology which can translate into a measurable SNHI. Tumours of the vestibular nerve—known as a vestibular schwannoma—is one example of a type of tumour which can lead to reduced hearing sensitivity through the compression of the neighbouring cochlear nerve (Gimsing, 2010; Sauvaget, Kici, Kania, Herman, & Tran Ba Huy, 2005). These tumours commonly develop sporadically in older adulthood or as part of neurofibromatosis type II, an autosomal dominant disease (Neff, Welling, Akhmametyeva, & Chang, 2006). Pathology of the central auditory nervous system which could result in a SNHI may arise from tumours (Møller, 2013), or a stroke or haemorrhage in the brain via ischaemia and nerve cell damage (Bamiou et al., 2012; Kamat, Kalani, Metreveli, Tyagi, & Tyagi, 2015; Vos, Greebe, Visser-Meily, Rinkel, & Vergouwen, 2017). A HI can also develop in hyperbilirubinemia as high serum concentrations of bilirubin is neurotoxic and thus damaging to the structures involved with auditory processing (Shapiro & Popelka, 2011).

Audiometrically, the presentation of SNHI will differ depending on the pathological cause of the HI and the affected area along the auditory pathway. NIHI may produce an audiogram that is notched at 4 kHz dropping to an average ≤ 75 dB. This is due to the combined resonant frequencies of the outer-ear, middle-ear and the external auditory canal (Straatman et al., 2017). In contrast, a HI which developed due to ototoxicity will show decreases in the HF hearing followed by LFs, with the degree of HI sustained being dose and duration dependent (Taneja et al., 2015). In the case of Meniere’s disease, SNHI may have a LF reverse sloping configuration (Sajjadi & Paparella, 2008). For cytomegalovirus, HI may be variable with thresholds ranging from mild to profound, unilaterally or bilaterally (Foulon, 2008). The degree of HI associated with pneumococcal meningitis is related to the severity of the infection (Perny et al., 2016). Similarly, head injuries may result in a flat or HF HI depending on the degree of damage sustained during impact (Singh et al., 2013). A deterioration of the HF hearing occurs first which then progresses to the lower frequencies, with reduced word recognition being the audiometric hallmark of presbycusis (Gates & Mills, 2005). SNHI due to a vestibular schwannoma can be gradual or sudden, however, the HI is commonly asymmetrical in the case of sporadically occurring vestibular schwannomas and bilateral for people who have neurofibromatosis type II (Cousins & Prasad, 2008). People who have a vestibular schwannoma may also present with an abnormal pattern of acoustic reflexes and reduced word recognition performance. Tumours of the central auditory nervous system will vary in the extent to which they impact a person’s hearing thresholds; however those affected may have reduced scores in low redundancy speech tests (Møller, 2013). When the SNHI is caused by either a stroke or haemorrhage in the brain, audiometrically there may be a decrease in hearing thresholds and their performance in low redundancy speech tests may be negatively affected (Bamiou et al., 2012; Vos et al., 2017). As hyperbilirubinemia has a neurotoxic effect in the central auditory nervous system, a person who has a HI as a result of this pathology may experience difficulties with auditory processing tasks, delayed auditory-evoked brainstem response latencies, and elevated hearing thresholds (Jiang, 2007). HF HI and reduced speech discrimination may also be a symptom (Møller, 2013).

1.3.3 Mixed hearing impairment

A HI can be classified as mixed when it reveals audiometric results consistent with both CHI and SNHI pathologies. Indeed, CHI and SNHI are not mutually exclusive and it is possible for a person to experience both types of HI concurrently.

Audiometrically, a mixed HI will present as thresholds outside of the normal range where the measured BC and the AC thresholds are not significantly different and there is a significant ABG at one or more (but not all) of those frequencies.

The following section discusses the role of audiology for the assessment of the different types of HI described above.

1.4 Audiology

Over the past century, audiology has transitioned from using tuning forks to modern audiometers for HI identification, from the use of horns for the provision of amplification to analogue HAs and then to digital HAs, and from in-the-field training provided in military hospitals to audiologists having access to postgraduate level education in universities (Katz, 2015). The unique, adaptable nature of audiology as a profession, as seen throughout history, has cultivated and refined the definition of an audiologist into what it is today: an autonomous hearing healthcare provider (New Zealand Audiological Society [NZAS], 2009). As described by the New Zealand Audiological Society (NZAS, 2009, p. 1), audiologists “specialise in the prevention, identification, assessment, diagnosis, management and treatment of disorders of the auditory, balance and other related neural systems” working with both adults and children in the private or public health sector. In addition to this, audiologists provide AH/R and seek to prevent HI through education and the provision of hearing protection equipment. In order to fulfil their role and provide the most appropriate services, audiologists utilise a range of different tests and assessments—which are outlined in the following sections—to obtain hearing health information from clients.

1.5 The audiological test battery

The audiological test battery is a series of tests and assessments commonly used by audiologists to ascertain the presence, type, and severity of a HI. In NZ, adult diagnostic assessments usually begin with a thorough case history and otoscopy. This is typically followed by the audiological test battery consisting of PT audiometry, immittance testing, and speech testing. Auditory evoked potentials and otoacoustic emissions may also be measured if appropriate, depending on the details of the case (NZAS, 2008). An audiological test battery approach is important because consistencies or inconsistencies between the results obtained can assist an audiologist's decision making with regards to identifying the presence and nature of a HI (Kreisman, Smart, & John, 2015). The concept of comparing test results for consistency was first introduced by Jerger and Hayes (1976a), and proposes that the results from one test should be cross-checked against the results of another to improve the accuracy of correctly identifying a hearing related disorder (R. Turner, 2003). Although originally developed for paediatric audiometry, it is now implemented in all areas of diagnostic audiology. While is it outside the scope of practice for an audiologist to diagnose the pathology which may be causing the HI, by analysing the results of different tests, an ear, nose and throat specialist or medical doctor will be better equipped to determine the area of the auditory pathway which may be affected, and identify the potential conditions that the measured HI may be a symptom of. Similarly, certain patterns of results, which can indicate a serious medical disorder and necessitate an urgent referral, might only become apparent in the context of the audiological test battery and may be missed if less comprehensive testing were carried out (Kreisman et al., 2015). In the previous section, this study detailed the audiometric findings which would be associated with different pathologies of the auditory system. The following section outlines the tests used to obtain these findings.

1.5.1 Pure tone audiometry

PT audiometry is a key component of the audiological test battery used to identify the quietest sound a person can hear at a particular frequency. The utility of PTs for evaluating a HI was first discovered in the 19th century using tuning forks, and to this day PT testing is still carried out via an audiometer since the advent of the modern audiometer in 1937 (Feldmann, 1997; Sente, 2004). The stimuli used in PT audiometry are acoustic sinusoidal longitudinal waves which are characterised by two factors: the frequency, measured in Hz; and the amplitude, measured in dB or SPL (Moore, 2012). These elicit a very simple response from the auditory system that is frequency specific, and are therefore ideal for use in ascertaining the minimum detectable level or absolute threshold at which a person is able to perceive a sound of a particular frequency in quiet conditions (Moore, 2012). The procedure most commonly used for determining a client's absolute threshold in adult diagnostic audiology in NZ is a modified Hughson and Westlake technique as described by Carhart and Jerger (1959). While the human auditory system is capable of discerning frequencies from 20 to 20,000 Hz, hearing acuity is most sensitive in the range of 500 to 8,000 Hz with the frequencies crucial for speech understanding residing in the 100 to 6,000 Hz bracket (French & Steinberg, 1947; Schlauch & Nelson, 2015). Conventionally, PT testing uses an audiometer to assess the frequency range of 250 to 8,000 Hz for AC and 500 to 4,000 Hz for BC in a sound treated room or booth (NZAS, 2016b).

1.5.1.1 Air conduction

AC PT testing assesses the pathway by which sound is conveyed to the cochlea when conducted via the air. In this circumstance, an acoustic stimulus is able to travel through the external auditory canal, down the ossicles, and directly into the cochlea via the oval window. Earphones and speakers are the two main types of transducers which can be used to transmit the acoustic stimulus to the ear (Schlauch & Nelson, 2015). Supra-aural and insert earphones are the most commonly used earphones in current audiological practice in NZ and present the stimuli directly to the external auditory canal and proximal to the tympanic membrane respectively. Speakers which present the stimulus in a sound field are unable to obtain separate ear information, and introduce the potential for a sound to be modified by the unique properties of the outer-ear (Van Wanrooij & Van Opstal, 2005). Regardless of the transducer used, the absolute thresholds obtained for AC are recorded on the audiogram for each ear—or binaurally if tested in the sound-field—and classified according to the degree of HI. The classification scale used in clinical audiological practice in NZ has been adapted from Goodman’s 1965 version (as cited in Clarke, 1981) and determines absolute thresholds from -10 to 15 dB as being indicative of NH; 16 to 25 dB as a slight HI; 26 to 40 dB as a mild HI; 41 to 55 dB as a moderate HI; 56 to 70 dB as a moderately severe HI; 71 to 90 dB as a severe HI; and ≥ 91 dB as a profound HI (NZAS, 2016b).

1.5.1.2 Bone conduction

BC PT testing assesses the transmission of sound to the cochlea using a vibratory stimulus coupled to the head with a bone vibrator. There are four distinct mechanisms which allow BC absolute thresholds to be observed: osseo-tympanic mode, compression/distortion mode, inertial mode, and more recently the non-osseous mode (Sohmer, Freeman, Geal-Dor, Adelman, & Savion, 2000; Tonndorf, 1966). The effectiveness of these individual modes depends on the appropriate positioning of the bone conductor, which is most commonly placed on the mastoid process for audiological testing as this has been found to give lower absolute thresholds. Due to the fact that there is an assumed 0 dB interaural attenuation for BC PT testing, results obtained are assumed to be indicative of the absolute threshold in the better hearing ear, and contralateral masking will need to be used when there is a significant AGB of ≥ 15 dB to reveal the thresholds of the poorer hearing ear in cases of an asymmetrical HI (Yacullo, 1996).

1.5.2 Immittance tests and auditory evoked potentials

Other assessments such as tympanometry, acoustic reflex tests, auditory evoked brainstem responses, and otoacoustic emissions provide important diagnostic information specific to different areas of the auditory system. While these tests are part of the audiological test battery and are useful for cross checking other audiological findings (R. Turner, 2003), they are less relevant to the current study and are therefore not referred to herein.

1.5.3 Speech Audiometry

Speech is a term used to describe the production of sounds and is the vessel for spoken language and communication (Moore, 2012). It can be sub-divided into three units: words, syllables, and phonemes (Peeva et al., 2010). From as early as the 19th century, speech has been recognised for its utility in assessing auditory function and it was not long before Fletcher and Steinberg (1929) developed the first method for creating and implementing speech audiometry assessment materials (Wilson & McArdle, 2005). Speech as a stimulus for audiological assessment has greater face validity for clients whose chief complaints relate to an inability to understand speech and is better suited to reveal pathologies beyond the cochlea which may involve impairment in the processing of auditory information (McArdle & Hnath-Christolm, 2015). As outlined in previous sections of this study, a HI can be classified in terms of the assumed location and type of the causative pathology, be that conductive or sensorineural in origin. In addition to this, a HI can be categorised depending on the characteristic components exhibited (Carhart, 1951). The audibility component describes a HI which is distinguished by a decrease in the acuity and/or rise in the attenuation of an acoustic signal. In contrast, the distortion component is typified by an experienced reduction in the clarity and increase in the distortion of an acoustic signal (Plomp, 1978). Speech audiometry is able to differentiate between these HI components, act as a cross check for other tests in the audiological test battery, and contribute valuable information about a client’s ability to recognise speech as a function of intensity (Dirks, Kamm, Bower, & Betsworth, 1977; Ventry & Chaiklin, 1965; Wilson & McArdle, 2005).

Speech recognition testing is the main assessment used in speech audiometry and involves presenting a speech stimulus at different intensities to a client who is tasked with identifying what they heard (McArdle & Hnath-Christolm, 2015). One method to ensure an authentic reflection of a client’s speech recognition abilities involves using stimuli with a low level of predictability which is phonetically and phonemically balanced (Egan, 1948). A low level of predictability can be achieved by assessing a client’s speech recognition using monosyllabic words as opposed to nonsense syllables or multiple words in a sentence (McArdle & Hnath-Christolm, 2015). Phonetic and phonemic balance is maintained when there is a proportionally correct representation of the phonetic sound components found in the language of the speech test, and when words are structured ‘consonant-vowel-consonant’ (CVC), with the total number of phonemes occurring in the word list being equally distributed and weighted (Lehiste & Peterson, 1959).

Multiple factors have been found to impact the accuracy of speech audiometry including calibration, familiarity, list length, mode of presentation, use of a carrier phrase, instructions given to the client, and client specific features (McArdle & Hnath-

Christolm, 2015). The language spoken by the client is an important consideration for speech audiometry. In fact, even within the same language, test materials in a different dialect or accent can result in a small, yet significant, measurable decrease in test performance (Gordon-Salant, Yeni-Komshian, & Fitzgibbons, 2010; Quar, Soli, Chan, Ishak, & Abdul Wahat, 2017). Indeed, NZ English has unique characteristics, such as raised vowels, which separate it from British or American English and for this reason speech audiometry in NZ uses speech stimuli recorded from a native NZ English speaker in order to minimise errors in identification related to differences in pronunciation (Gordon, 2004; Maclagan & Hay, 2007; Purdy, Arlington, & Johnstone, 2000).

In NZ, speech audiometry is categorically performed in the absence of background (BG) noise using monosyllabic meaningful CVC words as outlined above (Boothroyd, 1968; Boothroyd & Nittrouer, 1988; NZAS, 2016a). Testing in quiet is diagnostically advantageous as it ensures the results obtained are reflective of a client's performance in optimal listening conditions. The test is comprised of lists of 10 words which are presented following the carrier phrase "Say [the word] _________" and following each word the client is required to repeat what they heard, even if they only heard a portion of that word. The client's responses are scored phonemically and tallied to calculate a percentage of correct speech recognition for the intensity at which the list was presented. This method of scoring is valuable in a clinical setting as it increases the potential test size, allowing for subtle differences in performance to be revealed without increasing the assessment time (Boothroyd, 1968). Presentation intensities should be selected with the intention to define a performance intensity (*PI*) function and enable the identification of the maximum performance intensity (*PI*max) and half-peak level (NZAS, 2016a). The *PI*max describes the highest speech recognition percentage attainable for an individual client, and the half-peak level is the level at which the client achieves half of their *PI*max percentage. The need for accuracy is paramount when obtaining a *PI*max, which can be compounded by the apparent variation in the *PI*max for individuals. Thus research suggests that stimuli be presented over multiple intensities in order to reveal the *PI* function in sufficient detail (Beattie & Raffin, 1985; Beattie & Zipp, 1990).

In addition to providing useful diagnostic information, the findings of speech audiometry can also be used in AH/R (Boothroyd, 2008). It is not unusual for audiologists to counsel clients about their HI by explaining aspects of their speech test results and use the *PI* function to indicate whether a client would be a suitable candidate for amplification. This practice may be based on the reasoning that if a client’s recognition probability at full audibility during a speech assessment is high, then when provided with adequate amplification they should theoretically have better speech understanding and thus derive benefit from wearing HAs (C. Turner & Brus, 2001). However, there is controversy regarding the efficacy of increasing audibility in an effort to improve speech intelligibility as this presents a very simplistic approach to candidacy and assuming HA benefit (Ching, Dillon, & Byrne, 1998; Dillon, 2012; Karimi, Ashrafi, Khosravi, Shahidipour, & Vafaee, 2008).

Despite the widespread use and value of speech recognition testing in quiet, disadvantages of this presentation mode are apparent in literature (Beattie, Barr, & Roup, 1997; Wilson, 2011; Wilson, McArdle, & Smith, 2007). In particular, the inability of speech testing in quiet to predict the communicative performance of a client in their ‘real world’ environment is a fundamental limitation. This can lead to inconsistencies between the client’s speech recognition scores in the clinical environment, which may be adequate, and their subjective experience of listening difficulties in the presence of BG noise throughout their daily life (Beattie et al., 1997). This discrepancy can be exacerbated by tests which use monosyllabic stimuli, as clients are unable to make use of contextual cues to assume what a misheard word may be in relation to the rest of the sentence as they would in normal communication (Bagley, 1900).

The main consequence of the inability of speech in quiet testing to approximate a client's communication performance in the 'real world' is that it largely invalidates the use of this type of test as an indication of potential benefit or an outcome measure for audiological interventions. HAs are the most commonly prescribed intervention recommended by audiologists to assist clients to meet their communication requirements. HAs represent a significant financial investment for some clients and cost has been identified a significant barrier to HA uptake (Jenstad & Moon, 2011). Therefore, the ratio of cost to perceived benefit and measures to validate this is an important consideration for ethical audiological practice (NZAS, 2014). Although, it is tempting for audiologists to assume a relationship between improved speech recognition results due to an increase in audibility in the clinical setting and potential benefit from amplification for clients in the 'real world', research has shown that aided speech performance cannot be reliably predicted from the assessment of speech in quiet alone (Hoppe, Hast, & Hocke, 2014; Thümmler, Liebscher, & Hoppe, 2016). Additionally, speech audiometry in quiet does not fully account for the variety in the subjective perceived benefit of using HAs, therefore making it inappropriate for independent use as an outcome measure (Brännström, Lantz, Nielsen, & Olsen, 2014)

Another disadvantage of speech recognition using monosyllabic stimuli in quiet is its lack of sensitivity (Beattie et al., 1997). This predominantly impacts the ability of the test to distinguish between the performance of clients with NH and those with a mild

HI (Carhart, 1965; Speaks, Karmen, & Benitez, 1967). In such cases, ceiling effects can be encountered where the client may achieve scores similar to those of a person with NH (Beattie et al., 1997). Therefore, the test may inadequate reflect the communication difficulties experienced by a person who has a HI to a lesser degree. For this reason, and those described previously, many researchers recommend that speech audiometry should include an assessment of speech recognition in quiet as well as in noise to supplement the audiological test battery (Beattie et al., 1997; Carhart & Young, 1976; Dirks, Morgan, & Dubno, 1982).

Due to the shortcomings of assessing speech recognition in quiet, which are outlined above, the following section explores the advantages of speech in noise (SIN) testing and potential reasons for the apparent lack of up-take of such tests in the clinical and research environment.

1.6 Speech in noise testing

In 1970 Carhart and Tillman acknowledged the increased communicative difficulty that is experienced by those with a HI which occurs when listening in the presence of BG noise, and concluded that an assessment of speech recognition in noise should be universally implemented in audiological practice. Indeed, evidence reveals that listeners with a SNHI perform 10 to 15 dB worse than those with NH when tasked with recognising speech in BG noise (Carhart & Tillman, 1970; Groen, 1979; Olsen & Carhart, 1967). Despite the apparent need for SIN tests, integration of SIN assessments has not yet been fully adopted into routine audiological practice. Reasons for this may be due to the audiologist’s unfamiliarity with the delivery, scoring and interpretation of results obtained from SIN tests, as well as time constraints during audiological appointments (Wilson et al., 2007).

The omission of SIN testing as a regular part of the audiological test battery is unfortunate as the results obtained are clinically valuable for a variety of purposes. Diagnostically, SIN assessments can be used to help elucidate the location of pathology and differentiate between peripheral and central lesions along the auditory pathway (Morales-Garcia & Poole, 1972). They can also be used to assist an audiologist's decision making regarding the most appropriate ear to amplify in unilateral fittings where the findings from PT audiometry and tests of word recognition in quiet are similar (Beattie & Warren, 1982). In addition to this, SIN testing can be used to compare different HAs or assess the functionality of features included in digital HAs (Jerger & Hayes, 1976b).

The following section highlights the parameters that constitute SIN assessments making them advantageous for clinical use.

1.6.1 Psychophysical parameters

Similar to speech audiometry presented in quiet, SIN assessments are characterised by psychophysical parameters. Instead of a *PI* function, SIN tests utilise a psychometric function to define the relationship between a client's performance on a psychophysical task and a physical aspect of the stimulus. The psychophysical task in SIN tests is the percentage correct score a listener is awarded and the stimulus aspect is the signal-to-noise ratio (SNR; MacPherson & Akeroyd, 2014). The psychometric function is usually sigmoidal in shape and when fit to experimental data collected, allows for the estimation of two parameters: threshold and slope (Treutwein, 1995). The threshold—or speech reception threshold (*SRT*)—is the stimulus level required to obtain a 50% correct score and the slope refers to the gradient at a particular SNR. Put more simply, the slope can be considered the rate at which performance changes relative to changes in the SNR of the stimulus (Gilchrist, Jerwood, & Ismaiel, 2005; MacPherson & Akeroyd, 2014).

The slope of the psychometric function is an important feature in the measurement of speech intelligibility, as it is able to determine the amount of perceptual benefit a listener could expect from an increase in the SNR (MacPherson & Akeroyd, 2014). This is indicated by the incline of the slope. For example, a steeper slope suggests greater speech intelligibility would be achieved with a relatively small increase in the SNR, whereas for a shallower slope, the same increase in SNR would result in a smaller perceptual benefit. This information is valuable in the clinical setting, as the slope of the psychometric function varies greatly due to individual and test factors (MacPherson & Akeroyd, 2014). Therefore, the perceived benefit a client can anticipate from a HA's ability to increase the SNR in a listening environment will differ, and an audiologist can tailor their counselling about the expected improvement the HA will likely afford a client when listening in noise.

In addition to this, the slope of the psychometric function can be used as an indication of the sensitivity of the measure (Ozimek, Warzybok, & Kutzner, 2010). That is, the slope of the function at the point of the *SRT* can be used to check the validity of the *SRT* itself (Klein, 2001). Due to the inverse relationship between the *SRT* and the slope, a higher slope value is associated with a more reliable *SRT* and therefore reduces the number of trials necessary (Ozimek et al., 2010). This is valuable in a clinical context as it can inform a clinician's decision making regarding the most appropriate tool to use for assessment to ensure findings are obtained efficiently in terms of time and accuracy (MacPherson & Akeroyd, 2014)

1.6.2 Masking

As previously mentioned in this study, one of the advantages of carrying out SIN audiometry is that it more closely resembles clients' 'real world' listening environments (Francart, van Wieringen, & Wouters, 2011). This is achieved through the simultaneous presentation of acoustic masking noise with the stimulus. The two predominant types of masking noise used for SIN testing are continuous speech-shaped noise and babble noise (Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004). The selection of masking most appropriate for use in SIN testing is a contentious issue within literature, as outlined below (Francart et al., 2011).

1.6.2.1 Continuous speech-shaped noise

Continuous speech-shaped noise is a masker which has the same spectral content as conversational speech. The concept behind the development of this type of masking is known as energetic masking, which occurs when the acoustic energy of the masker is able to render the target stimulus inaudible when occupying the same critical bands at the same time (Durlach et al., 2003). The advantage of this type of masking is the low level of variability because the masking is constant and therefore the effectiveness of the masker is assumed to be consistent for the duration of the presentation (Killion et al., 2004). The clinical benefit of this is an improvement in the reproducibility of the test and therefore increased confidence in the results obtained (Killion et al., 2004).

Another advantage of continuous speech-shaped noise for SIN testing is the enhanced sensitivity of the masker. Research has found that continuous speech-shaped noise as a masker is able to produce psychometric functions with steeper slopes (Wagener & Brand, 2005). This conclusion was reinforced by Francart et al.’s research which compared static and fluctuating maskers like babble noise (2011). The sensitivity of a measure is an important consideration for audiologists, as test results are a crucial component upon which clinical decision making is based and erroneous scores on a SIN test can lead to inappropriate recommendations for clients.

1.6.2.2 Babble

In contrast with continuous speech-shaped noise, babble noise masking is a stimulus which is usually comprised of multiple competing talkers; typically four talkers are used to adequately simulate the listening context of a small social gathering (Killion et al., 2004). The premise behind babble noise is to more accurately approximate a client's 'real world' listening environment, and reveal the communication difficulties they may experience therein. This results in higher face validity for SIN tests which utilise babble noise masking (Killion et al., 2004)

During babble noise masking, both the noise and the target stimulus can be audible, however, a listener may perform poorly as they struggle to differentiate between the desired stimulus and the masker (Myerson et al., 2016). Despite this, research has found babble noise is better able to discriminate between the different degrees of HI as a function of the *SRT* compared with static maskers (Francart et al., 2011). This may occur as a result of temporal processing due to the fluctuating nature of babble noise and the increased amplitude modulation of this type of noise compared with other stationary types of masking (Bacon, Opie, & Montoya, 1998; Hopkins & Moore, 2009). The amplitude modulation in the envelope of a sound arises predominantly from the gaps between syllables and words which produces a higher SNR in these temporal dips and allows a listener to briefly "glimpse" the signal (Bacon et al., 1998; Hopkins & Moore, 2009; Howard-Jones & Rosen, 1993; Peters, Moore, & Baer, 1998, p. 578). This phenomenon, known as "masking release", has been found to typically affect those with NH (Bacon et al., 1998, p. 549; Hopkins & Moore, 2009). Listeners with a HI however, show little, if any improvement in performance related to masking release, which further supports the use of babble noise as an appropriate masker for clinical assessments as it better approximates the performance a client with a HI would have in the ‘real world’ (Bacon et al., 1998; Francart et al., 2011).

1.6.3 Measures of Speech in Noise Testing

As the need for audiometric speech assessment in the presence of BG noise became increasingly apparent, various SIN measures were developed to supplement the audiological test battery (Wilson & McArdle, 2005). A clinically valuable characteristic of these measures is the ability of the assessment to be undertaken in both the aided and unaided condition. This has potential applications for HA outcome evaluations depending on the sensitivity of the measure (Mendel, 2007; Taylor, 2003). This study has already summarised how the slope of a psychometric function may be indicative of the potential perceptual benefit an individual client can expect with an increase in SNR. However, the ability of SIN tests to predict the total benefit derived from HA use (herein referred to as HA benefit) is separate from perceptual benefit and includes both the objective improvement in speech intelligibility and the subjective appraisal of perceived benefit. This concept of total benefit is a controversial subject in literature and will be explored forthwith. A key factor in opposition to the use of SIN testing for the assessment of HA benefit is the lack of direct measurement—and therefore prediction—of the subjective HA benefit of a client (Cord, Leek, & Walden, 2000; Grunditz & Magnusson, 2013). Despite this, one study compared the performance of different SIN measures including the Speech Perception in Noise (SPIN; Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984), Hearing in Noise Test (HINT; Nilsson, Soli, & Sullivan, 1994), and the Quick Speech in Noise (QuickSIN; Killion et al., 2004) test with the Hearing Aid Performance Inventory (HAPI; Walden, Demorest, & Hepler, 1984) which is a subjective measure of HA performance (Mendel, 2007). This study found significant correlations between the HAPI and both the SPIN and the QuickSIN, suggesting that these measures of SIN testing are capable of predicting HA benefit (Mendel, 2007).

Another important consideration for SIN audiometry is the process in which thresholds are obtained. Currently, methods for threshold seeking in SIN tests use either a fixed or adaptive procedure (Taylor, 2003). Each procedure has respective advantages and disadvantages which influence the appropriateness of the test for clinical use as outlined below.

1.6.3.1 Fixed

SIN tests which assess a listener's word recognition performance at a static SNR are referred to as fixed measures (Taylor, 2003). An advantage of a fixed measure test, such as the SPIN (Bilger et al., 1984), is that it provides a percentage correct score which makes interpreting and explaining the assessment to the client easier and improves face validity (Taylor, 2003). In the SPIN, the listener is required to repeat the final monosyllabic word in the sentence sequence, which was designed to have equal probability of being either highly predictable or contextually neutral (Kalikow, Stevens, & Elliott, 1977). Another key asset of this test is the ability to select the desired SNR, which gives the audiologist greater control over the presentation of SNRs which would be most suitable for a client's individual difficulties in BG noise (Bilger et al., 1984). This characteristic of fixed measures can also be a disadvantage because determining the SNR most appropriate for assessment can be difficult at times and lead to the results being inaccurately over- or under-stated (Taylor, 2003). Audiologists can minimise the likelihood of this, however, by assessing multiple SNRs (Taylor, 2003). Research has revealed that the volume of conversational speech increases in response to increased BG noise (Taylor, 2003). In 1977 Pearsons et al. (as cited in Taylor, 2003) stipulated that in the presence of 55, 65 and 75 dB SPL BG noise, the average level of conversational speech was 61, 68, and 74 dB SPL (ie. +6, +3, and -1 dB SNR) respectively. This indicates the potential for fixed SNR measures to reveal a client's performance in a clinical environment that would accurately reflect their speech recognition capacity in the 'real world' in listening conditions which are "relatively easy", "moderately difficult", and "challenging" (Taylor, 2003, p. 32).

1.6.3.2 Adaptive

Adaptive SNR measures are SIN tests which assess the speech-to-noise ratio whilst the intensity level of either the speech or noise is varied based on the listeners response (Taylor, 2003). The HINT is an example of an adaptive measure which uses modified BKB sentences (Bench, Kowal, & Bamford, 1979) in the presence of speech-shaped noise at a fixed intensity of 65 dB SPL (Nilsson et al., 1994). The presentation of the stimulus, however, varies in steps of 2 dB to reveal the SNR at which 50% of sentences are repeated correctly, known as the reception threshold for sentences (Taylor, 2003). In order for a response to be considered correct, a listener must identify all the key words in the sentence (Taylor, 2003). The QuickSIN is another adaptive measure which also uses sentence stimuli to assess word recognition performance in the presence of four-talker babble noise (Killion et al., 2004). Unlike the HINT, the QuickSIN varies the level of the BG noise automatically (in 5 dB steps beginning at a SNR of +25 dB SNR) whilst presenting the target stimulus at a fixed intensity which can be selected by the audiologist (Taylor, 2003). The presentation level recommended by the test suggests 70 dB hearing level (HL) should be used for listeners with NH and a volume deemed loud, but comfortable for listeners with a HI to ensure the audibility of speech cues in the sentence (Killion et al., 2004). Each sentence contains five key words for which a listener is awarded one point for each one correctly identified. The total points obtained in the test is then calculated and subtracted from a reference value of 25.5 to give the client's SNR loss (Killion et al., 2004).

The ability of an adaptive SIN measure to determine the SNR loss is a clinically worthwhile tool in AH/R (Taylor, 2003). SNR loss is defined as the increase in SNR that is necessary for a client to identify 50% of the words in a sentence correctly (Killion et al., 2004) and can be used to compare how well a client can recognise words in the presence of BG noise relative to someone with NH (Taylor, 2003). This can be a valuable tool for counselling a client and also may account for the variation in the perceived HA benefit for clients with similar audiometric configurations (Taylor, 2003).

In addition to this, adaptive measures like the QuickSIN can also be useful in assessing features of HAs. Many digital HAs have directional microphones which are able to decrease gain for sounds received by the backwards facing microphone in relation to the sound received by the front facing microphone in an effort to increase the SNR and make it easier for a client to hear sounds directly in front of them (Dillon, 2012). The QuickSIN can be used to verify a HAs ability to do this when speakers are orientated 180 degrees apart to the front and back of the listener and the test is undertaken in the aided condition (Killion et al., 2004).

1.6.4 Stimuli: Word versus Sentence

The stimulus used by SNR measures, be it monosyllabic words or sentences, is another important consideration for SIN tests. Decision making regarding the appropriateness of stimuli for use in SNR measures typically involves evaluating the purpose of the test and the cognitive status of the listener (Wilson, 2003). Monosyllabic words are useful for determining a client's word recognition ability in the presence of BG noise without the advantage of contextual cues to help deduce words in an utterance which may be unintelligible (Ozimek et al., 2009). However, monosyllabic word identification may not be reflective of ‘real world’ listening situations where listening and responding to spoken sentences forms the basis of communication. For this reason, many SNR measures utilise sentence material that allow listeners to make use of the contextual cues which are exploited in everyday conversation (Ozimek et al., 2009). Thus, the degree of face validity for sentence stimuli is higher as it more closely approximates the communication deficits experienced by a client (Killion et al., 2004).

Another advantage of sentence stimuli for SNR measures is its ability to test multiple speech sounds in a single presentation (Hochmuth et al., 2012). This stimulus also typically results in speech intelligibility psychometric functions with steeper slopes, which indicates greater accuracy with respect to estimating the *SRT* (Hochmuth et al., 2012). Therefore, the clinical efficacy for using sentence stimuli is high as it produces reliable results in a time efficient manner (Kollmeier & Wesselkamp, 1997; Nilsson et al., 1994).

Despite the clinical value of using sentence material for SNR measures, research has revealed that there are many factors beyond HI which affect a listener’s ability to hear speech in the presence of BG noise (Wilson, 2003). The cognitive ability of the listener has been identified as a key component which influences their performance in a SIN assessment, with greater implications for sentence recognition as opposed to other stimuli (Cervera, Soler, Dasi, & Ruiz, 2009; McArdle, Wilson, & Burks, 2005; Wilson et al., 2007). Researchers have suggested prior assessment of a client’s ability to complete the test should be carried out to decrease the likelihood of a listener’s working memory and the increased cognitive load associated with SNR measures using sentences as stimuli erroneously influencing the validity of the results obtained (Craik, 1994; Kramer, Zekveld, & Houtgast, 2009; McArdle et al., 2005; Wilson et al., 2007).

In combination, the capacity of sentence stimuli to approximate the communication environment of the ‘real world’ and its ability to test multiple speech sounds in a single presentation are desirable features of SNR measures which make it ideal for use clinically. Indeed, sentence SNR measures are almost universally supported in speech audiometry because of time-efficiency and the fact that sentence stimuli is thought to better reflect the hearing deficits experienced by a person in the presence of noise which is considered valuable to the process of AH/R (Dietz et al., 2014). However, as performance in these tests can be influenced by cognitive factors, audiologists need to consider the appropriateness of the measure for use on a case-by-case basis (Wilson et al., 2007).

1.6.4.1 Sentence Tests

As the efficacy of sentence stimuli for use in SIN testing became widely recognised, two distinct categories of sentence SNR measures became apparent. The first, known as ‘Plomp-type tests’, are characterised by their ability to reflect conversational speech in the ‘real world’ (Dietz et al., 2014). These tests comprise lists of sentences containing meaningful and phonemically balanced material which have a low predictability, as no consistent grammatical structure is maintained (Plomp & Mimpen, 1979). The HINT and the QuickSIN described previously in this study are examples of Plomp-type SNR measures which use sentences as stimuli (Killion et al., 2004). The HINT and the QuickSIN have been widely received by the audiological community, with the older HINT having been translated into multiple languages including Cantonese (Wong & Soli, 2005) and NZ English (Hope, 2010).

The other type of SNR measure which uses sentence stimuli is the Matrix Sentence Test (MST). This test, first developed for the Swedish language (Hagerman, 1982), utilises a matrix of words five columns across, with each column containing 10 names, verbs, numbers, adjectives, and objects respectively. The items included for use in the matrix were selected with the intention that, when combined, they would produce a phonemically balanced and grammatically correct five-word sentence (Hagerman, 1982). Generating material for the stimuli used in this type of test often involves recording sentences with minimal co-articulation to enable individual words to be cut, synthesised, and edited to ensure equal difficulty across sentences is maintained (Akeroyd et al., 2015). Altogether these aspects guarantee that the stimuli used in this type of test are semantically different, yet syntactically identical (Akeroyd et al., 2015). The identical structure of the sentences allows for a practically unlimited stock of speech material as words can be randomly selected from each column to create new sentences (Hagerman, 1982).

1.6.4.2 Matrix sentence tests

There are numerous advantages which make MSTs more favourable for use in the clinical and research setting in comparison with other speech audiometry measures. A particularly desirable feature of MSTs is that the vast number of potential sentence stimuli in a MST can reduce redundancy in the assessment (Hagerman, 1982). This is important, considering the current procedure to test speech recognition in NZ uses CVC words comprised of 10 lists of 10 words (Boothroyd & Nittrouer, 1988). In a clinical case where repeated testing is required, this may create bias in the results obtained from a particular client if they become familiar with and memorise the test stimuli. Another advantage of MSTs is that the sentence stimuli has very low predictability, as the matrix is designed so that every item in each of the columns would make grammatical sense regardless of the preceding or following words (Hagerman, 1982). This allows for a better approximation of a client's authentic speech recognition ability in the 'real world' without the use of contextual cues which could potentially cause variability and bias in the test results (Hochmuth et al., 2012).

For the reasons described above, MSTs have received attention internationally and have been developed in various languages including Finish (Dietz et al., 2014), Russian (Warzybok et al., 2015), Spanish (Hochmuth et al., 2012), Danish (Wagener, Josvassen, & Ardenkjær, 2003), German (Wagener, Brand, & Kollmeier, 1999), Polish (Ozimek et al., 2010), French (Jansen et al., 2012), Dutch (Houben et al., 2014), and English (S. Hall, 2006). As the procedures for the development of MSTs are highly consistent globally, it is possible to obtain an index of a client's speech intelligibility performance in noise internationally (Akeroyd et al., 2015).

1.6.4.3 Presentation mode: Open vs Closed

The mode in which stimuli is presented is another key consideration for MSTs. Many published MSTs (such as Dietz et al., 2014; Hochmuth et al., 2012; Ozimek et al., 2010; Wagener et al., 2003) can be delivered bi-modally via an open set mode and a closed set mode. In an open set mode, a listener is required to verbally repeat what they heard for each sentence presentation, which is scored by an audiologist. A closed set mode takes advantage of the uniform nature of the MST to enable multiple choice self-administration. The closed set form of assessment requires the client to independently identify the words they heard in a sentence from the items included in the matrix, which is displayed on a touch screen for example. The benefit of closed set testing is that it negates the need for audiologist supervision, enabling more efficient use of time during the audiology appointment (Ozimek et al., 2010). Eliminating the audiologist's role in scoring the test may be preferable, as it reduces the potential for bias to be introduced into the test results due to administrator oversight or miscalculation. While this is certainly advantageous, the impact of the closed set presentation mode on a listener's performance is a controversial issue in literature. For example, the Polish MST found no difference in the performance of listeners in the open or closed set mode (Ozimek et al., 2010). However, another study found a significant difference in the *SRT* for the open and the closed set condition, with better speech intelligibility measured for the closed set mode (Hochmuth et al., 2012). Hochmuth et al. (2012) concluded that this finding may be the result of neglecting to provide training, and thus listeners were able to perform better in the closed set mode by utilising the visual cues on the touch screen, leading to the finding of non-equivalence between conditions (Tye-Murray, 2015).

1.7 The Development of the University of Canterbury Auditory-Visual Matrix Sentence Test

1.7.1 Rationale

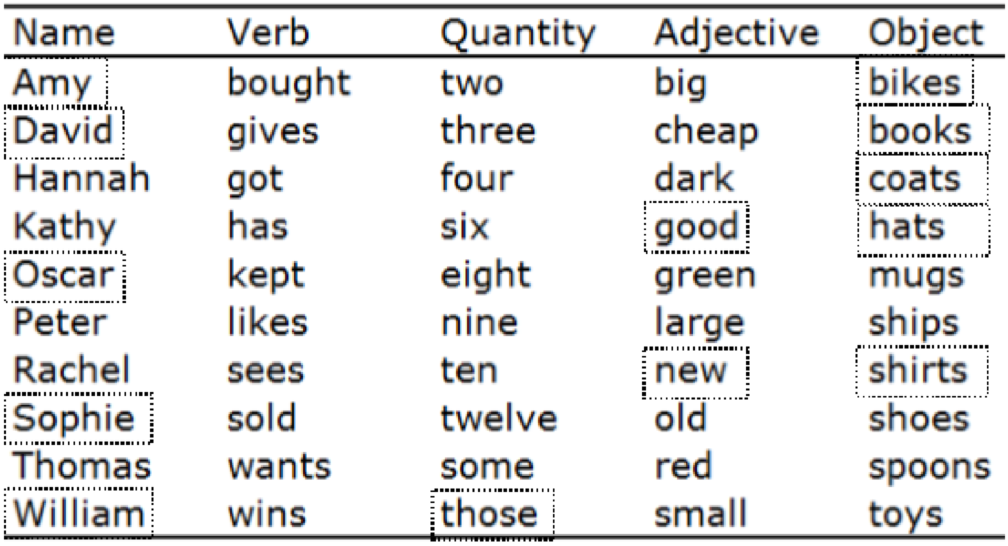
As outlined previously in this study, speech audiometry in NZ currently assesses word recognition in quiet using monosyllabic meaningful CVC words (Boothroyd, 1968; Boothroyd & Nittrouer, 1988). While word recognition tests provide information which is valuable clinically, its ability to predict a listener's ability to understand speech in the presence of BG noise is inadequate (Beattie et al., 1997). For this reason, and the advantages already described, it is recommended that a MST using NZ English be developed to complement the audiological test battery (Trounson, 2012). NZ English has unique phonology which invalidates the use of the British or Australian MST in a NZ context as it may compromise the validity and reliability of the test, especially for listeners who have SNHI when listening in the presence of BG noise (Gordon, 2004; Maclagan & Hay, 2007; Purdy et al., 2000; Zokoll et al., 2013). Thus, the University of

Canterbury Auditory-Visual Matrix Sentence Test (UCAMST) was designed by Trounson and O’Beirne (O’Beirne et al. 2015; Trounson, 2012)

A unique feature of the UCAMST is the (optional) visual component which is currently the only MST (in its New Zealand English and Malay versions) that allows for the presentation of stimuli in the auditory-alone, visual-alone or auditory-visual mode (Jamaluddin, 2016; Trounson, 2012). The rationale for this is the improvement in speech intelligibility when listeners can both see and hear a speaker as opposed to hearing only, especially at low SNRs (Grant, Walden, & Seitz, 1998; Sumby & Pollack, 1954). Also, the assessment of speech recognition in noise using an auditory-visual mode of presentation is thought to improve how natural a speech stimulus sounds (Mattheyses, Latacz, & Verhelst, 2009). Indeed, the audio-visual presentation mode allows listeners to exploit both visual and auditory cues as they would during communication in the ‘real world’, and is therefore able to provide a closer approximation of a client’s overall ability to understand speech. Thus, the ability of the UCAMST to evaluate speech recognition in all three presentation modes may be useful in diagnostic audiology (Tye-Murray, Sommers, & Spehar, 2007).

1.7.2 Matrix development

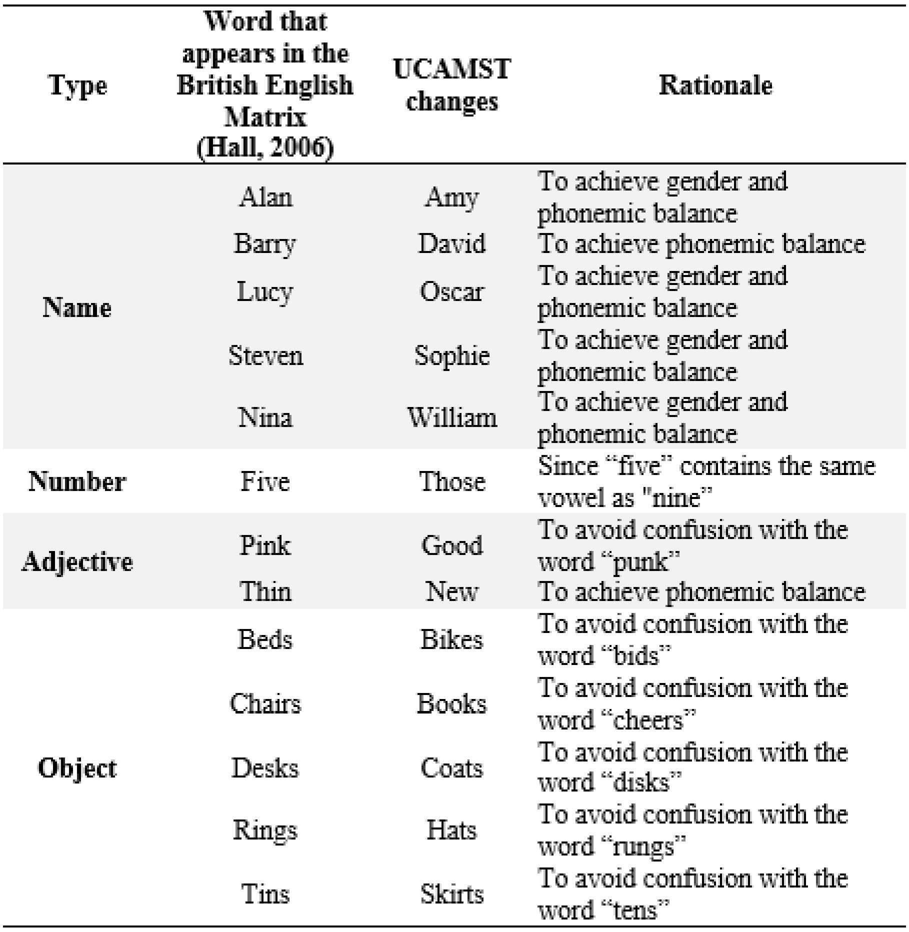
Trounson (2012) adapted the British English MST in order to create a SIN assessment better suited for use by NZ-native English speakers (S. Hall, 2006). Figure 1 illustrates the words replaced from the British English MST and the resultant base matrix for the UCAMST (S. Hall, 2006; Trounson, 2012).



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 1. Word matrix for the UCAMST. *Retrieved from Trounson (2012, p. 24). Note. The hashed boxes depict words replaced from the British MST (S. Hall, 2006).*

Substitutions were made for words which would cause vowel confusion and to achieve phonemic balance using the NZHINT (Hope, 2010) as a basis for comparison (Trounson, 2012). Various names included in the British English MST were also changed in order to attain an equal number of names associated with each gender (Trounson, 2012). Figure 2 displays the rationale for each of the substitutions made (Stone, 2016).

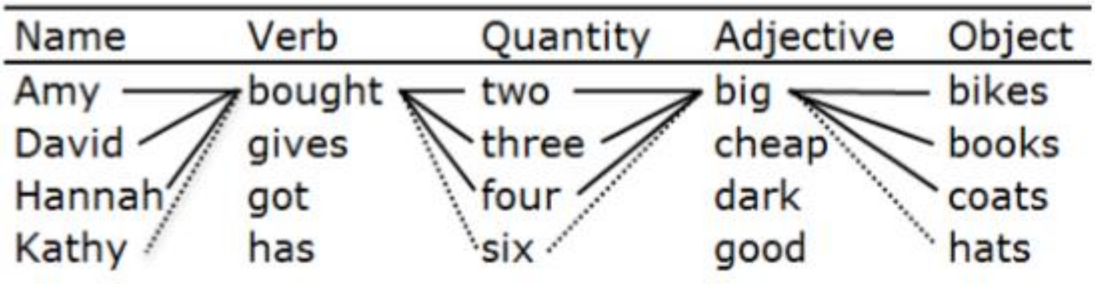


Type	Word that appears in the British English Matrix (Hall, 2006)	UCAMST changes	Rationale
Name	Alan	Amy	To achieve gender and phonemic balance
	Barry	David	To achieve phonemic balance
	Lucy	Oscar	To achieve gender and phonemic balance
	Steven	Sophie	To achieve gender and phonemic balance
	Nina	William	To achieve gender and phonemic balance
Number	Five	Those	Since "five" contains the same vowel as "nine"
Adjective	Pink	Good	To avoid confusion with the word "punk"
	Thin	New	To achieve phonemic balance
Object	Beds	Bikes	To avoid confusion with the word "bids"
	Chairs	Books	To avoid confusion with the word "cheers"
	Desks	Coats	To avoid confusion with the word "disks"
	Rings	Hats	To avoid confusion with the word "rungs"
	Tins	Skirts	To avoid confusion with the word "tens"

Figure 2. Rationale for the changes made to the British English MST in the development of the UCAMST. *Retrieved from (Stone, 2016, p. 23). Note. British English MST from S. Hall (2006) with information obtained from Trounson (2012).*

1.7.3 Generating, Recording and Editing Sentences

The UCAMST was originally produced in 2011 and has been developed in a manner consistent with the recommendations described by Akeroyd et al. (2015) and in keeping with the methodology used by other MSTs internationally. In order to create the stimuli for the UCAMST, an adult, native NZ English speaking actress with a verified accent read 100 five-word sentences. The sentences' sequence followed the pattern: name, verb, number, adjective, noun and were produced by combining the items in the matrix in a process identical to that carried out by the Danish MST (Wagener et al., 2003).



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

Figure 3. UCAMST sentence generation pattern. *Retrieved from Trounson (2012, p. 27).*

The pattern of pairing words seen above was used in order to achieve 10 different realisations of a word. This procedure has become widely accepted internationally for generating sentence material in MSTs, and is advantageous as it allows for co-articulation and thus preserves the natural prosody of the sentence (Dietz et al., 2014; Hochmuth et al., 2012; Houben et al., 2014; Ozimek et al., 2010; Wagener et al., 2003). Following recording, 400 word fragments were obtained which could be recombined and synthesised to produce 100,000 distinct sentences (Trounson, 2012). Sentences were analysed and those which sounded unnatural were removed from the final lists. Although numerous measures were taken to ensure both the visual and audio material recorded was as natural as possible, as described by Trounson (2012), there were occasionally noticeable jerk artefacts (referred to as “judder”) in some of the visual recordings. This was due to the position of the actress’ head appearing mismatched across the recombined visual fragments.

1.7.4 Selecting Sentence Stimuli

As a result of the marked judder in the visual recordings of the UCAMST, McClelland (2014) conducted a study to quantify the noticeability of the judders in the synthesised sentences for observers. To subjectively assess judder noticeability, 18 participants with NH were requested to rate how obvious judders appeared between fragment transitions on a scale from 0 (indicating no apparent judder) to 10 (highly noticeable judder) for synthesised and non-synthesised (i.e. control conditions) sentences (McClelland, 2014).

Multiple comparisons were made across sentences to produce a final repertoire of audio-visual stimulus material composed of control sentences and sentences that were evaluated to have the least amount of judder (McClelland, 2014).

1.7.5 Generation of Masking Noise

Masking noise for use in the UCAMST includes continuous speech-shaped noise and six-talker babble noise (herein referred to as “constant noise” and “babble noise” respectively). The constant noise was developed using the originally recorded sentences spoken by the actress and randomly superimposing this 10,000 times to produce a spectrally matched masker. Spectral matching occurs when the spectra of the masker and the stimulus are almost identical, and is thought to be advantageous in ensuring the validity of the test through maintaining a stable SNR (King, 2011). The babble noise used in the UCAMST was originally produced as part of a previous thesis using three male and three female native NZ English speakers (Spencer, 2011). Each speaker was recorded reading 20, 6- to 10-word, semantically anomalous sentences which were edited and combined together into a single sound file (Spencer, 2011).

1.7.6 Normalisation

Normalisation was the necessary next step in the development of the UCAMST, which has been carried out by multiple MSTs internationally (Hagerman, 1982; Hochmuth et al., 2012; Houben et al., 2014) and recommended by Akeroyd et al. (2015). The purpose of normalisation is to achieve a high level of homogeneity and ensure equally difficulty across the stimulus material included in the test (Akeroyd et al., 2015; Wagener et al., 2003). The process suggested by Akeroyd et al. (2015) to for normalisation requires generating a speech intelligibility function for each word intended for use in the test and adjusting the level at which it is presented (within a specified limit). This is done in order that the intelligibility functions obtained are as similar as possible and are therefore more likely to be equally difficult.

For the normalisation of the UCAMST, McClelland (2014) recruited 18 participants with NH to listen to 400 sentences in constant and babble noise at various SNRs to delineate adequate psychometric functions. This was then fit to a logistical model to obtain fragment or word specific intelligibility functions to give valuable information for comparison, such as the pre-normalisation midpoint (L_{mid} or 50% correct point). This process revealed the 15 fragments designed for use in constant noise needed to be removed from the stimulus material, however the remaining 385 were able to be adequately fit to the model and a mean pre-normalised L_{mid} of -10.3 dB SNR[1] was obtained (± 2.1 dB standard deviation [SD]; McClelland, 2014). This same process was carried out for words designed for use in constant noise and adjustments were made to the word-specific SI functions to ensure better overlap between the mean post-normalised L_{mid} for fragments and words. The predicted difference in the post-normalised functions as a result of word normalisation for the constant noise condition is demonstrated in Figure 4 (McClelland, 2014).

---

[1] SNR values obtained from McClelland (2014) have been corrected by Stone (2016) due to inaccuracies in the measurement of signal intensity recording as described in the following section.

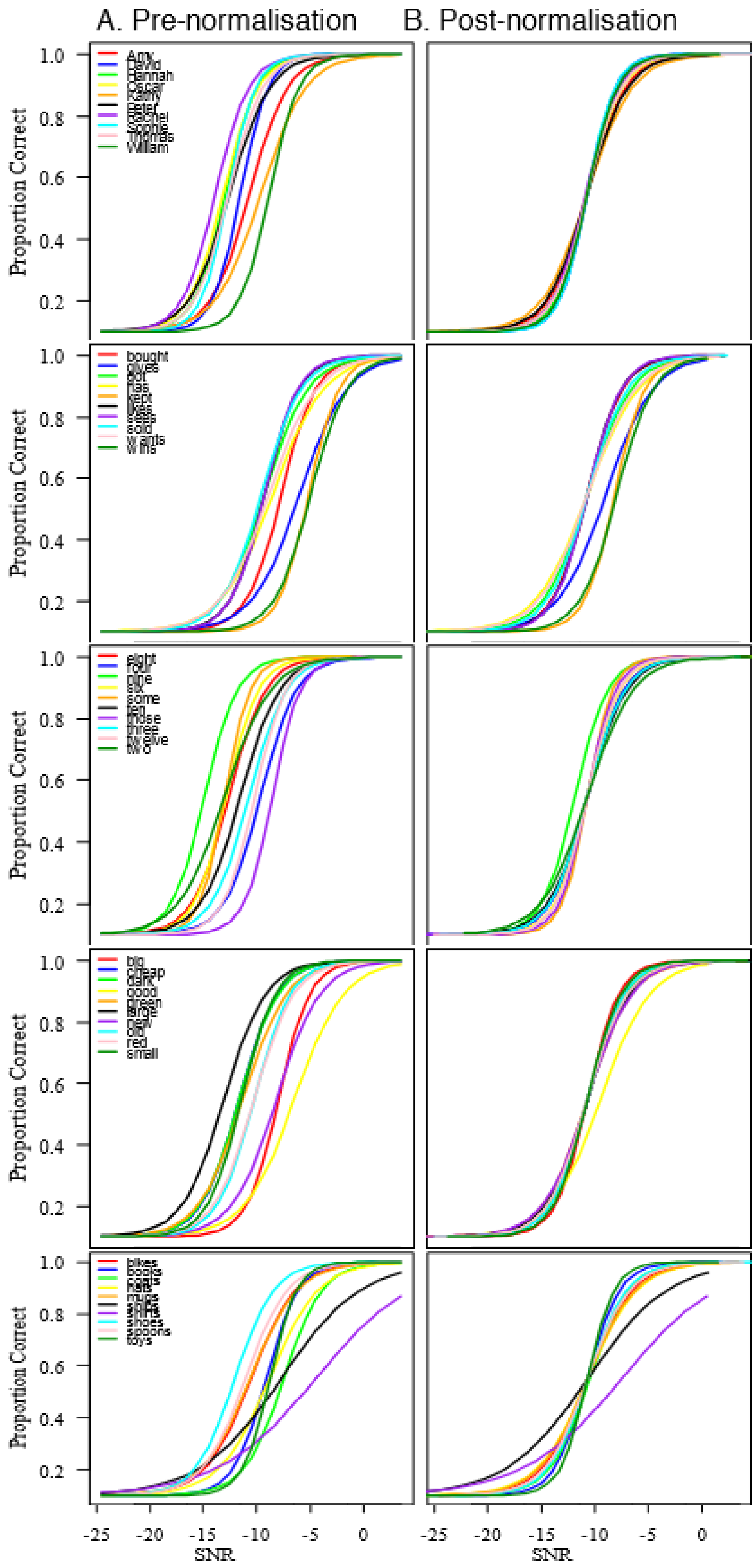


Figure 4. Pre-normalisation and predicted post-normalisation intelligibility functions for words presented in constant noise. *Retrieved from (Stone, 2016, p. 31). Note. Figure originally adapted from (McClelland, 2014, p. 82).*

The procedure as described above was then carried out for the stimuli designed for use in babble noise. Fragment-specific normalisation revealed 47 fragments were unable to fit the model and were subsequently discarded (McClelland, 2014). The prevailing fragments were found to have a pre-normalisation L_{mid} of -11.0 dB SNR (±2.9 dB [SD]). This same process was conducted a final time for the words designed for use in babble noise and adjustments were made to the word-specific SI functions to ensure better overlap between the mean post-normalised L_{mid} for fragments and words. The predicted difference in the post-normalised functions as a result of word normalisation for the babble noise condition is demonstrated in Figure 5 (McClelland, 2014). The reduced overlap for the predicted post-normalisation function for the babble noise condition was deemed to be a result of the extensive adjustments carried out in an attempt to achieve equivalence (McClelland, 2014). One issue that became known following normalisation is that there were inaccuracies in the method through which the SPL was measured. This caused the signal to be erroneously recorded as quieter than in reality by 3.85 dB SPL for both noise conditions. Subsequently, 3.85 dB has been retrospectively added by Stone (2016) to values obtained from McClelland (2014) to account for this.

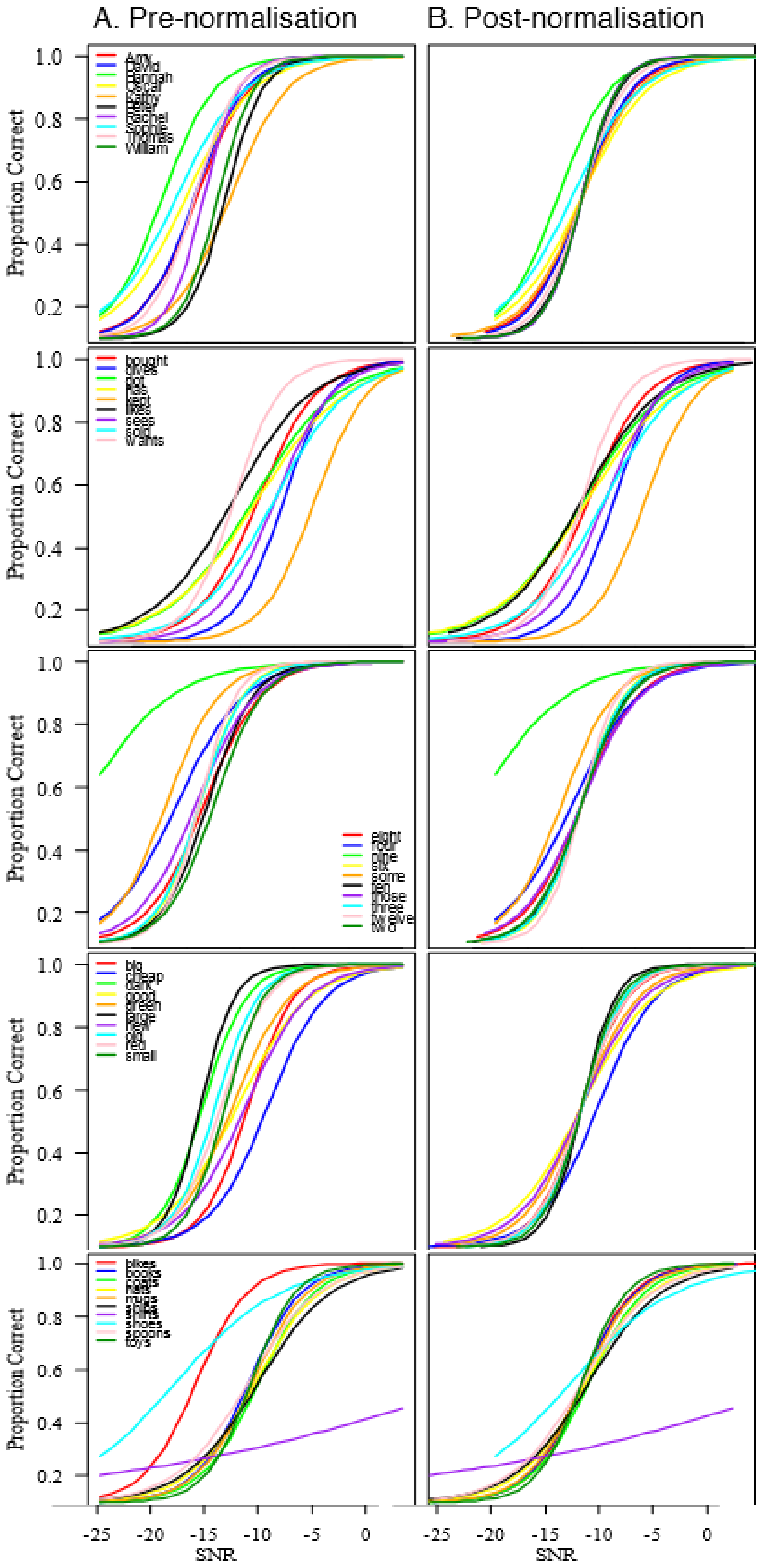


Figure 5. Pre-normalisation and predicted post-normalisation intelligibility functions for words presented in constant noise. *Retrieved from (Stone, 2016, p. 33). Note. Figure originally adapted from (McClelland, 2014, p. 87).*

After normalisation has occurred it is possible to construct sentence lists which are equally difficult (Akeroyd et al., 2015). Homologous lists will improve the validity of the test by ensuring that a client's measured performance is a true reflection of their ability to understand speech in background noise and not due to discrepancies in the difficulty of the test lists (Wagener et al., 2003). Normalisation of the test material is also important, as it improves the sensitivity of the test in determining the *SRT* by allowing smaller differences in performance to be measured and thus more reliable results obtained (Brand & Kollmeier, 2002).

1.7.7 Generation of Sentence Lists

As part of a previous study, McClelland (2014) developed 30 base lists of 20 sentences for the constant noise and babble noise condition respectively. Preliminary evaluation of normalisation was carried out for these lists and found no significant difference between lists designed for the constant noise or babble noise condition. However, these sentences proposed for use were generated using only the psychometric data for the auditory-alone presentation mode. This is inadequate for the UCAMST as the corresponding visual component for these sentences may have poor quality transitions that would introduce bias in the test results when the visual stimuli is presented (Stone, 2016). Therefore, Stone sought to produce lists with optimal visual quality and minimal SDs for the sentence-specific slopes between lists (2016). This required utilising software to generate sentences and compare the "pixel difference value" or "judder tiers" of the sentences produced (Trounson, 2012). Sentences were rejected by the software depending on the magnitude of judder present at the transition point between video frames, resulting in 20 unique lists for each noise condition with considerably lower SDs. Sentences for the constant noise condition had one occurrence of each word in the matrix for each list. The words "wins" and "shirts" were removed from the babble noise condition as McClelland (2014) found these words produced abnormal psychometric functions; thus for the lists in the babble noise condition, one word from each of the verb and noun columns were randomly repeated twice for each list (Stone, 2016). Upon further analysis of the resultant lists for subjective judder by two observers, eight additional lists were rejected, leaving a total of 16 lists of 10 sentences for each noise condition (Stone, 2016).

1.7.8 Evaluation of Normalisation

The next stage in the progression of a speech measure towards clinical use is what is known as the evaluation of normalisation (Akeroyd et al., 2015). This process involves comparing the slope and *SRT* derived from the test specific function ($s50_{test}$) of the sentences to ensure that lists included in the test are equally difficult, and therefore reveal the sensitivity and reliability of the measure (Akeroyd et al., 2015). This was carried out for the UCAMST by presenting the sentence stimuli fixed at 65 dB SPL and the noise at two different SNRs in order to achieve a ‘pair of compromise’, which is thought to most accurately estimate the *SRT* and the $s50_{test}$ (Ozimek et al., 2010; Stone, 2016). The equation used by Stone (2016) to delineate the $s50_{test}$ for sentence materials in the UCAMST was adapted from Hochmuth et al. (2012) and involves the modification of the mean word-specific function and the SD of the *SRT*s as seen in Figure 6 below.

$$s50_{test} = \frac{s_{word}}{\sqrt{1 + \frac{16 s_{word}^{2} \times \sigma_{Lmid}^{2}}{\left(\ln\left(2e^{\frac{1}{2}} - 1 + 2e^{\frac{1}{4}}\right)\right)^{2}}}}$$

Figure 6. s50test equation used in the UCAMST. *Retrieved from Stone (2016, p. 34). Note: "s50test = test-specific speech recognition curve; sword = slope of the word-specific intelligibility function; σ = SD of t word-specific Lmid measures" (Stone, 2016, p. 34).*

Stone (2016) collected data from a total of 42 participants in the open and closed set condition for constant and babble noise. The participants' performance was recorded for 180 sentence presentations; the first 20 of which were practice sentences, and the following 160 constituted the 16 lists designed for the UCAMST. Half of the sentences in each list were randomly assigned the presentation level of -14.3 dB SNR or -7.6 dB SNR to ensure an equal distribution of sentences at each SNR (Stone, 2016). Results from Stone's study (2016) revealed that there was no significant difference in the slope and *SRT* of the lists designed for use in constant noise for the open set and closed set condition. For the babble noise however, while no significant difference existed for the *SRT*, a significant difference was apparent for the slope of both the open set and closed set conditions. The findings also revealed a significant difference in the anticipated estimates of the slope and *SRT* for each noise condition available in the UCAMST (Stone, 2016). In addition to this, the UCAMST was found to differ significantly from other MSTs included for analysis except for the Danish MST (Wagener et al., 2003) with respect to the *SRT* (Stone, 2016). A key limitation of the study by Stone (2016) is that a malfunction in the software allowed the lists designed for use in constant noise to be presented in the babble noise condition, which may account for the lack of equivalence for some of the findings obtained (Stone, 2016). Thus, it has been recommended that the evaluation of normalisation be repeated for babble noise paired with the correct stimuli to accurately assess the equivalence of the lists for that type of noise. This will be investigated as part of the current study.

1.7.9 Validation

Cross-validating a MST involves comparing its performance with other speech tests, in the same language and other languages, to ensure comparability across tests (Akeroyd et al., 2015). Validation of the UCAMST has been partially completed by Stone (2016) for international MSTs. However, no investigation has been undertaken to compare the performance of the UCAMST against speech tests currently used in NZ. This study will seek to accomplish this for the meaningful CVC (revised AB) word recognition test (Boothroyd, 1968; Boothroyd & Nittrouer, 1988) and the QuickSIN (Killion et al., 2004).

1.8 Study Rationale

Interestingly, since MSTs were first utilised in audiology by Hagerman (1982) its application has predominantly focused on SIN testing. The benefits of testing speech intelligibility in quiet and using sentence materials as the target speech stimuli have already been discussed earlier in this study. In light of this, it is prudent to investigate the capacity of the UCAMST for the assessment of speech understanding in the absence of competing BG noise. However, before this can be addressed, it is necessary to ensure the equivalency of the sentence lists intended for presentation in quiet, as list homogeneity is not guaranteed when stimuli designed for presentation in noise are adopted for use in quiet without modification (Loven & Hawkins, 1983; Nilsson et al., 1994). Therefore, the current study seeks to continue the process of evaluating normalisation for the babble noise—as recommended by Stone (2016)—and in quiet presentation mode to ensure list equivalence and thus the reliability and sensitivity of the UCAMST. This research will also compare the performance of the UCAMST against the meaningful CVC (revised AB) word recognition test (Boothroyd, 1968; Boothroyd & Nittrouer, 1988) and the QuickSIN (Killion et al., 2004). Validation is the necessary next step to further develop the UCAMST towards clinical and research use in NZ as part of the University of Canterbury Adaptive Speech Test platform (Akeroyd et al., 2015; O'Beirne, McGaffin, & Rickard, 2012).

In keeping with previous studies, this research will only involve presentation of the UCAMST in the auditory-alone condition, as findings from the Malay version of the UCAMST (Jamaluddin, 2016) revealed that participants became reliant on the visual cues provided in the auditory-visual mode for sentences with poor SNRs. This increased the difficulty with which a psychometric function could be obtained for the auditory-visual condition and necessitated the removal of the visual component for normalisation (McClelland, 2014).

1.9 Aims and research questions

This thesis aims to continue the process of generating test lists which are appropriate for use in each of the presentation modes included in the UCAMST and evaluate the difficulty of such lists. In order to evaluate list equivalence, this study sought to answer two primary questions:

(1) Are the stimulus lists designed for use in the open set, babble noise; closed set, babble noise; open set, in quiet; and closed set, in quiet condition equivalent with regards to:

    a. Slope of the psychometric function (herein referred to as slope)
    b. The SNR at which *SRT* is estimated (herein referred to as *SRT*)

(2) Is there a difference between the slope and *SRT* of the six test conditions (i.e. closed set, constant noise; open set, constant noise; closed set, babble noise; open set, babble noise; closed set in quiet; open set in quiet)

This research also aims to compare the performance of the UCAMST against other speech tests routinely used in audiology clinics in New Zealand. In order to verify this, this study will answer the subsequent question:

(3) Is there any correlation between results from the UCAMST and those from the following:

    a. Meaningful CVC (revised AB) word recognition test, and
    b. The QuickSIN

1.10 Hypotheses

Based on the findings of previous research the following hypotheses were proposed for the current study:

For research question (1):

That no significant differences would be found between the stimulus lists in the closed set, babble noise; open set, babble noise; the closed set, quiet condition; and the open set quiet condition respectively for:

a. Slope, and
b. SRT

For research question (2):

That no significant difference would be found between the six test conditions (i.e. open set, constant noise; closed set, constant noise; open set, babble noise; closed set, babble noise; open set, in quiet; closed set in quiet) with regards to the:

a. Slope, and
b. SRT

For research question (3):

That no significant correlation exists between the results from the UCAMST in the open/closed set quiet conditions and those from the meaningful CVC (revised AB) word recognition test; and no significant correlation exists between the results from the UCAMST open/closed set constant or babble conditions and those from the QuickSIN.

CHAPTER 2 *METHODS*

2.1 Introduction

As described, the purpose of the current study is to determine the list equivalence for stimuli designed for use with babble noise and in quiet in order to reveal the reliability and sensitivity of the UCAMST. Additionally, this research seeks to cross-validate the UCAMST with the meaningful CVC (revised AB) word recognition test (Boothroyd, 1968; Boothroyd & Nittrouer, 1988) and the QuickSIN (Killion et al., 2004) to establish whether there is comparability across speech tests used in NZ. The following chapter will be separated accordingly into two parts: the first outlines the methodology employed to collect data used for the evaluation of the lists designed for babble noise and in quiet, and the second details the methodology by which data was obtained for the investigation of equivalence across the different UCAMST conditions and for the validation of the UCAMST.

As the current research is part of a series of studies aimed at developing and progressing the UCAMST towards research and clinical use, ethical approval was already given by the University of Canterbury Human Ethics Committee for the testing required to evaluate normalisation and covered the testing required for the validation of the UCAMST also (Approval number HEC 2014/49, see Appendix A). All procedures implemented for data collection in this study were conducted in accordance with those proposed in the ethics application.

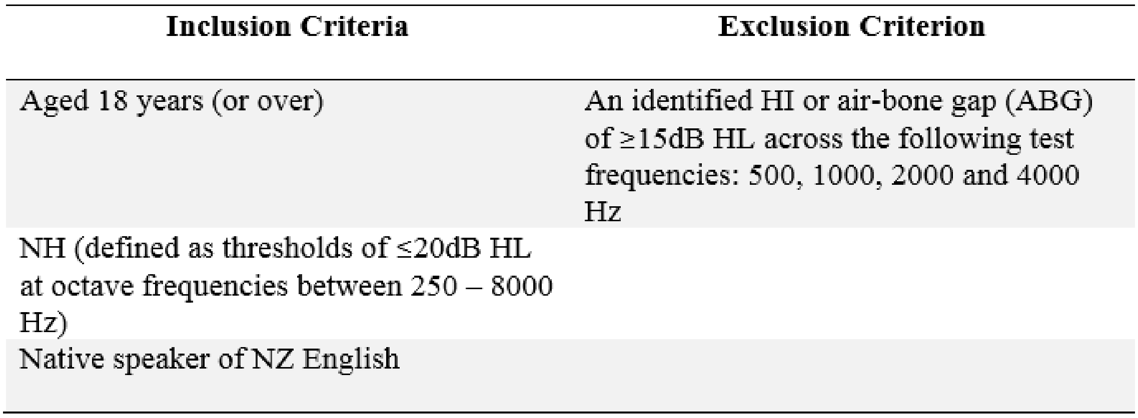
The inducement offered to all participants in part one and two of this study was a $10 Motor Trade Association voucher.

2.2 Part 1: Evaluation of normalisation for babble noise

2.2.1 Participants

2.2.1.1 Recruitment

Prior to commencing data collection it was determined that 30 participants would be necessary to complete the evaluation of normalisation for sentences designed for presentation in the babble noise condition. This number of participants was selected as it would allow for a minimum of seven approximations of the *SRT* for each SNR within each list in both the open and closed presentation mode. Recruitment was carried out via the circulation of advertisements and an email invitation. As demonstrated in Appendix B, the invitations included a short description of the purpose of the study, the eligibility requirements necessary for participation as well as what would be involved in the testing. Preceding any testing, potential candidates were provided with an information sheet, see Appendix C, which explained the procedure and possible risks involved in the study. Following this, informed consent was obtained (please refer to Appendix D for the consent form). As this part of the current research is a repetition of the evaluation of normalisation partially carried out by Stone (2016), it was necessary to replicate the inclusion criteria and testing conditions as closely as possible to avoid introducing bias in the study and to ensure comparability across results. Thus, once informed consent was received, screening was carried out via an interview and a hearing test identical to the procedures followed by Stone (2016) to ensure participants met the inclusion criteria as detailed below in Figure 7 .



Inclusion Criteria	Exclusion Criterion
Aged 18 years (or over)	An identified HI or air-bone gap (ABG) of ≥15dB HL across the following test frequencies: 500, 1000, 2000 and 4000 Hz
NH (defined as thresholds of ≤20dB HL at octave frequencies between 250 – 8000 Hz)	
Native speaker of NZ English	

Figure 7. Inclusion and Exclusion criteria. *Retrieved from Stone (2016).*

The age requirement of 18 years or older was selected on the basis that the task to be completed in this study was considered to be of a high cognitive load and the testing procedure necessitated sustained concentration over an extended period of time which may have been difficult for children and adolescents (Betts, McKay, Maruff, & Anderson, 2006). As the UCAMST was developed for use in NZ, it was also essential that candidates recruited for the evaluation of normalisation were native speakers of NZ English to preserve the validity of the results presented in this study (Akeroyd et al., 2015). Having NH with no identified CHI was necessary to be eligible to participate, as any form of HI, including an ABG—which may indicate the presence of underlying middle ear pathology—would be a potential source of bias in the data otained (Akeroyd et al., 2015).

2.2.1.2 Demographics

A total of 31 NH listeners were recruited for the evaluation of normalisation for babble noise and were tested in two blocks, in which they were assigned semi-randomly to the open set or closed set condition (assignment to the open set and closed conditions will be outlined in a following section). The participant demographics for each presentation mode are outlined in Table 1.

Table 1. Participant demographics for normalisation in babble noise.

	n	***M* age (years)**	***M* PTA**		**Gender**
			R	**L**	
Closed-set Babble	15	31.47	4.11	3.17	*n* M = 15 *n* F = 16
Open-set Babble	16	29.13	3.76	3.85	
Total	31	30.30	3.94	3.51	

Note. n = number of participants; M = mean; PTA = puretone average; R = right ear; L = left ear; M = males; F = females.

2.2.2 Stimuli

The stimulus for presentation was fixed at 65 dB SPL in the presence of babble noise. Lists were presented at four SNRs in total. The first block of testing used -14.3 and -7.6 dB SNR as Stone’s (2016) research indicated these levels would best approximate the pair of compromise. However, preliminary analysis of the data revealed that the task was too difficult and the resultant participant performance was too low to adequately fit a *PI* function and accurately estimate the *SRT*. Thus, a second block of testing using -7.6 and -3.7 dB SNR was carried out three weeks later to supplement the data previously obtained. For both blocks of testing, the SNRs were randomly assigned to half of the lists for each condition to ensure that the SNRs were equally distributed across sentences.

2.2.3 Experimental instrumentation

Participants were first screened in a sound treated room at Bay Audiology (Palmerston North). Audiometric testing took place in an adjoining sound attenuated test booth by means of a calibrated Otometrics Aurical audiometer (Natus; Taasrup, Denmark). PTs were delivered via Telephonics TDH-39P supra-aural headphones (Griffon Corp; New York, USA) for AC octave frequencies ranging from 250 - 8,000 Hz and via a RadioEar B-71 BC headband (Minnesota, USA) for the octave frequencies of 500 - 4,000 Hz.

These devices were worn by the participants who would press a push button linked to the audiometer to indicate they heard a tone.

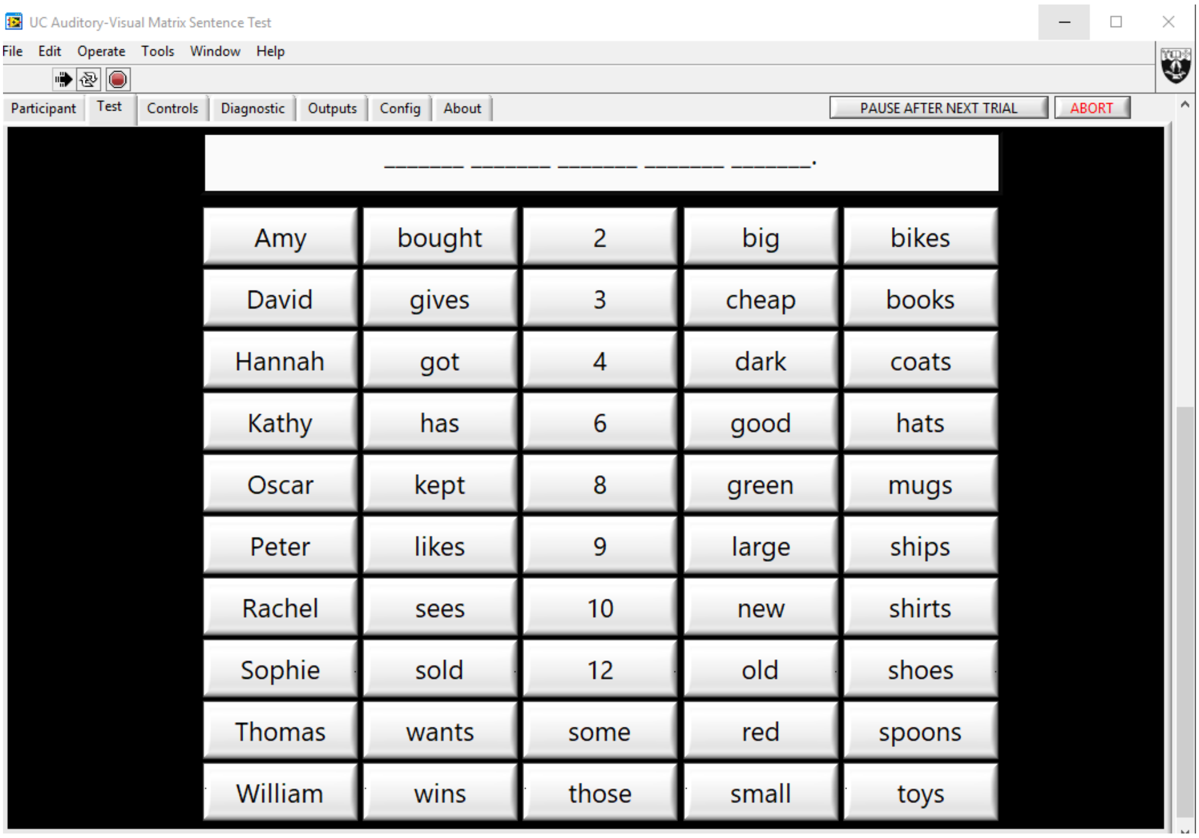
Experimental testing was carried out in the same location as the initial screening and participants were seated alone or with the researcher, depending on whether they were assigned the closed set or the open set condition respectively. The UCAMST software was developed using LabVIEW (National instruments; Texas, USA) and installed on a Hewlett Packard (California, USA) laptop. The sentence stimuli and babble noise was presented simultaneously through Sennheiser (Sennheiser electronic GmbH & Co.KG, Germany) HD280 Pro circumaural headphones (64 Ω impedance) connected to the SBX Prostudio SoundBlaster sound card (Creative Labs, Singapore). Microsoft Excel 2013 was used to investigate the data and generate intelligibility functions for comparison. All statistical analyses were undertaken using the IBM Statistical Package for the Social Sciences (SPSS; version 23; IBM Corp.; New York., USA).

2.2.4 Experimental procedures

As mentioned above, prior to any procedures taking place, informed consent was received from each listener. Following this, screening was carried out to determine whether each candidate was eligible to participate in the study. Screening involved an interview comprised of several questions to assess whether the participant had any known HI or difficulty understanding speech in BG noise, any history with the ear, nose and throat specialist or tinnitus, and any recurrent or recent ear infections. Otoscopic examination was then undertaken to exclude the presence of any cerumen occluding the external ear canal, as this may impact the audiometric information obtained. Participants were then seated in the sound treated test booth, as described previously, to have their AC and BC hearing tested. Listeners were instructed that they would hear a series of tones at various pitches and must respond by pressing the button each time they heard the tone, even if was barely audible. The audiometric results were explained to each participant at this time and any participants who were identified to have a HI which excluded them from further involvement in the study were provided information with respect to follow-up measures (see Appendix C for further details).

The experimental procedure for participants differed depending on which condition they were assigned to undertake. For the listeners in the first block, the presentation mode was assigned randomly, alternating the open and closed set condition for each successive listener. However, in the second block, 10 participants were re-recruited from the first block, as the stringent time frame in which access to the test equipment was possible made recruitment of new participants impracticable. Therefore, in order to minimise the potential bias retesting may introduce, it was necessary to present the test for these participants in the same open or closed set condition which they had been assigned originally.

Participants assigned to undertake the test in the closed set condition were seated alone in the quiet, sound treated room in front of the laptop. These listeners were verbally instructed that they would hear a sentence of varying volumes in the presence of babble noise through the headphones and that their task was to identify all the words in that sentence from the matrix displayed on the screen using the mouse. The layout of the matrix visible for participants following each sentence presentation is demonstrated in Figure 8.



Amy	bought	2	big	bikes
David	gives	3	cheap	books
Hannah	got	4	dark	coats
Kathy	has	6	good	hats
Oscar	kept	8	green	mugs
Peter	likes	9	large	ships
Rachel	sees	10	new	shirts
Sophie	sold	12	old	shoes
Thomas	wants	some	red	spoons
William	wins	those	small	toys

Figure 8. Closed set UCAMST matrix depicted after each trial.

Participants were encouraged to make an educated guess should they be uncertain of the words in the sentence in order to progress to the next trial.

The experimental procedure for testing in the open set condition was similar to that of the closed set condition, with the exception being the presence of the researcher to score participant responses. Listeners assigned to the open set condition were seated in the same quiet, sound treated room with the researcher, however in this case they were seated facing away from the laptop screen. The verbal instructions that were given to the participant regarding what they would hear through the headphones were the same as those given to participants in the closed set condition, although open set listeners were instructed to verbally repeat the sentence which was displayed on the screen for the researcher to score as depicted in Figure 9.

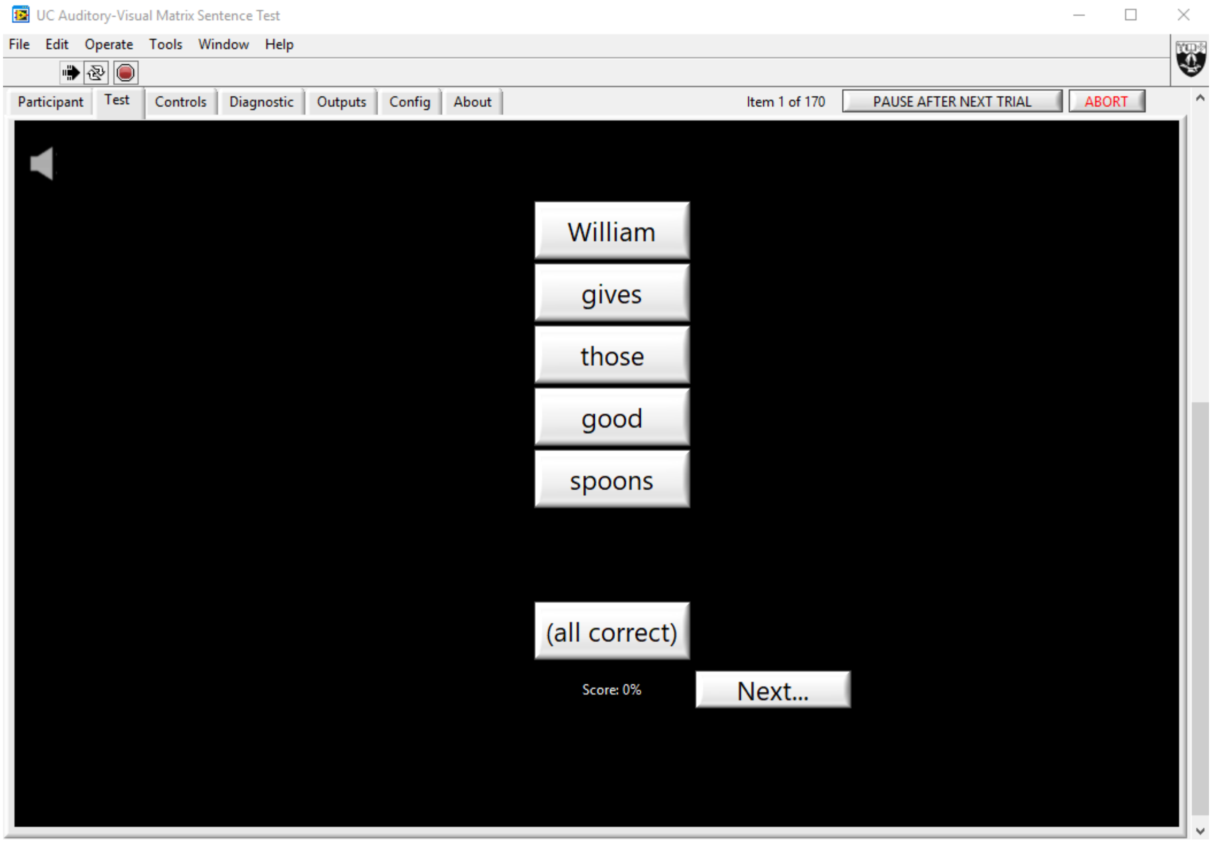


Figure 9. Open set UCAMST matrix used by researcher to identify words repeated by participants.

Participants in this condition were also encouraged to make an educated guess in the instance that they were uncertain of words presented in the sentence.

For both the closed and open set conditions, sentences were scored by the number of words correctly identified, as opposed to fragment or sentence scoring, to give a total score out of five for each sentence. Regardless of condition, each participant was presented 20 practice sentences as a training set to enable the listener to become familiar with the test format and task. 160 test sentences (i.e. 16 lists) were then presented to a total of 21 listeners in the first block of testing and 20 listeners in the second block of testing respectively and their performance recorded. The resultant data from these trials was then used for analysis in the current study. From beginning to end, the complete experimental procedure for each participant took approximately 60 minutes, irrespective of condition.

2.3 Part 1: Evaluation of normalisation for in quiet

2.3.1 Participants

2.3.1.1 Recruitment

Prior to commencing data collection it was determined that 30 participants would be necessary to complete the evaluation of normalisation for sentences designed to be presented in the quiet condition. This number of participants was selected as it would allow five approximations of the *SRT* at each presentation level for each condition within each list. Recruitment was carried out in the same way as it was for the evaluation of normalisation for babble noise; with the circulation of advertisements and email invitations throughout the University of Canterbury community (Christchurch, NZ). Informed consent was obtained using the same information sheet and consent form as for the evaluation of normalisation of the babble noise condition because the potential risks and the experimental procedure were inherently similar. Eligibility to participate in the research was identical to the inclusion and exclusion criteria outlined in Figure 7 for the reasons described previously in this study.

2.3.1.2 Demographics

A total of 30 NH listeners were recruited for the evaluation of normalisation of the UCAMST sentences for the quiet presentation mode and were assigned randomly to the open or closed set condition. The participant demographics for each condition are outlined in Table 2.

Table 2. Participant demographics for the evaluation of normalisation in quiet.

	n	***M* age (years)**	***M* PTA**		**Gender**
			R	**L**	
Closed-set Quiet	15	26.45	2.00	1.72	*n* M = 8 *n* F = 22
Open-set Quiet	15	33.22	3.50	2.67	
Total	30	29.84	2.75	2.20	

Note. CQ = Closed set, Quiet; OQ = Open set, Quiet; n = number of participants; M = mean; PTA = puretone average; R = right ear; L = left ear; M = males; F = females.

2.3.2 Stimuli

The stimuli originally designed for presentation in constant noise was used to evaluate normalisation in quiet. The rationale behind this stems from the finding that constant noise is able to act as an energetic masker which occupies the same auditory filters as the target stimuli, thus rendering the target stimuli completely inaudible depending on the SNR presented (Durlach et al., 2003; Myerson et al., 2016). Therefore, the efficacy of constant noise as a masker would have a similar effect to presenting the target stimuli at decreasing levels until it is barely audible, if not inaudible. As a result of this, it was deemed appropriate for the researcher to present sentences optimised for use in constant noise in the quiet condition. There is limited literature available regarding procedures for the evaluation of normalisation for the quiet presentation mode. However, one MST which has reported evaluating this condition, found no significant difference between the list equivalence results using stimuli optimised for presentation in constant speech-shaped noise and stimuli optimised for presentation in quiet (Theelen-van den Hoek, Houben, & Dreschler, 2014).

To investigate which intensity levels to present, a pilot study involving 30 NH adults was carried out. Participants for the pilot study were recruited in the same way as described in the previous section and had their hearing tested at the University of Canterbury Speech and Hearing Clinic in a sound proof booth via a calibrated

Grason-Sadler (Minnesota, USA) GSI audiometer for the frequency range of 250 – 8,000 Hz. Pilot demographics are outlined in Table 3.

Table 3. Pilot participant demographics.

	N	***M* age (years)**	***M* Threshold (2 kHz)**		**Gender**
			R	**L**	
	30	25.75	-0.83	-1.17	Male = 4 Female = 26
Total	30	25.75	-1.00		

Note. N = number of participants; M = mean; R = right ear; L = left ear.

According to current NZAS best practice guidelines, initial presentation of speech stimuli for speech testing to define a *PI* function should be set above the listeners 2,000 Hz threshold and that a minimum of three intensity levels should be presented (NZAS, 2016b). Therefore, based on the average threshold of -1.00 dB at 2,000 Hz across both ears for NH listeners in the pilot study, the three intensity levels of 25, 15, and 5 dB A were selected for use to approximate a *PI* function as accurately and reliably as possible.

2.3.3 Experimental instrumentation

Experimental instrumentation for the evaluation of normalisation for the quiet presentation mode was similar to that described for babble noise, with the exception that testing took place at the University of Canterbury Speech and Hearing Clinic (Christchurch, NZ). Participants were first screened in a research laboratory at the University of Canterbury Department of Communication Disorders (Christchurch, NZ). Audiometric testing took place in an adjoining sound attenuated test booth by means of a calibrated Grason-Sadler GSI audiometer (Minnesota, USA). PTs were delivered via Telephonics TDH-50P supra-aural headphones (Griffon Corp; New York, USA) for AC octave frequencies ranging from 250 – 8,000 Hz and via a RadioEar B-71 BC headband (Minnesota, USA) for the octave frequencies of 500 – 4,000 Hz. These devices were worn by the participants who would respond by pressing a push button linked to the audiometer to indicate they heard a tone.

Experimental testing was carried out in the sound attenuated test booth where the participant was seated alone to minimise masking from external sources of noise, such as the fluorescent lights and fan. The UCAMST software was installed on a Hewlett-Packard (California, USA) desktop computer connected to an ēlo touch sensitive monitor (ēlo ET1715L, Tyco Electronics; CA, USA). The sentence stimuli were presented through Sennheiser HD280 Pro 64 Ω circumaural headphones (Sennheiser electronic GmbH & Co.KG, Germany) connected to the SBX Prostudio SoundBlaster external sound card (Creative Labs; Singapore). Microsoft Excel 2013 was used to investigate the data and generate intelligibility functions for comparison. All statistical analyses were undertaken using the IBM SPSS (version 23, IBM Corp.; New York., USA).

2.3.4 Experimental procedures

The screening procedure for the evaluation of normalisation for the quiet presentation mode was identical to that followed for babble noise and therefore won’t be repeated here.

The response mode of the UCAMST was allocated by alternating the open and closed set condition with each successive listener. Participants assigned to undertake the test in the closed set condition were seated alone in the quiet test booth facing the touch receptive monitor. These listeners were verbally instructed that they would hear a series of sentences at varying levels through the headphones and their task was to identify all the words in each sentence from the matrix displayed on the screen by selecting the corresponding words heard from the appropriate columns. This could be done using the mouse which was placed on the table in front of them or their finger on the touch screen. The layout of the matrix became visible for participants following each sentence presentation and is illustrated in Figure 8.

The experimental procedure for testing in the open set condition was similar to that of the closed set condition, except that in this case the screen in front of the listener was blank. The verbal instructions given to the participant regarding what they would hear through the headphones was the same as those given to participants in the closed set condition, although open set listeners were instructed to repeat the sentence aloud. The researcher seated outside the sound attenuated test booth listened for the participants responses from the booth relayed via the audiometer and scored these on the computer screen which was not visible to the participant. The matrix used by the researcher to score participant responses is depicted in Figure 9.

All participants, irrespective of condition, were encouraged to make an educated guess in the instance that they were uncertain of words presented in the sentence in order to progress to the next trial. Word scoring was again used to give a total score out of five for each sentence. Before scores were recorded, each participant was presented 20 practice sentences as a training set to enable the listener to become familiar with the test format and task. 150 test sentences (i.e. 15 lists) were then presented to 30 participants and their performance documented. The resulting data from these trials was then used for analysis in the current study. The complete experimental procedure for each participant took approximately 60 minutes, irrespective of condition.

2.4 Part 2: Evaluation of condition equivalence and validation

2.4.1 Participants

2.4.1.1 Recruitment

Prior to commencing data collection it was determined that 21 participants would be necessary to evaluate condition equivalence and validate the UCAMST. This number of participants was selected following a calculation using data analysis software (G*Power, version 3.1.9.2; Düsseldorf, Germany), and was based on the findings of a previous thesis which revealed a significant relationship between the UCAMST and the QuickSIN (Andre, 2016). Recruitment was carried out in a manner identical to the procedure used for the recruitment of participants for the evaluation of normalisation for the quiet presentation mode. Informed consent was obtained using the same information sheet and consent form as for the evaluation of normalisation, as the potential risks and the experimental procedures across tests were inherently similar. The inclusion and exclusion criteria which is outlined in Figure 7 was again used for the validation of the UCAMST.

2.4.1.2 Demographics

A total of 21 NH listeners were recruited for the validation of the UCAMST. The participant demographics for each condition are outlined in Table 4.

Table 4. Validation participant demographics.

	n	**Age (years)**	**QuickSIN SNR loss (dB SNR)**	***SRT* AB words (dB)**	**PTA**		**Gender**
					R	**L**	
	1	18.54	1.5	11.9	4.17	2.50	*n* M = 5 *n* F = 16
	2	24.17	0.5	6.7	0.00	-0.83	
	3	29.11	0.5	7.7	7.50	4.17	
	4	25.18	0.5	9.1	5.00	6.67	
	5	22.92	0.5	1.5	-0.83	1.67	
	6	24.09	1.5	0.0	-2.50	-1.67	
	7	20.87	0.5	1.8	2.50	3.33	
	8	29.11	0.5	8.3	2.50	1.67	
	9	22.65	1.5	-11.7	-6.67	-4.17	
	10	29.02	-0.5	8.1	4.17	1.67	
	11	23.03	-1.5	2.8	0.83	2.50	
	12	21.87	0.5	9.9	5.83	5.83	
	13	20.91	0.5	8.9	2.50	5.00	
	14	28.45	5.5	-2.2	-1.67	-5.83	
	15	26.77	0.5	-0.2	0.00	-0.83	
	16	26.14	0.5	0.7	-3.33	-2.50	
	17	21.61	0.5	9.8	9.17	5.83	
	18	21.46	0.5	1.1	-2.50	-0.83	
	19	24.41	0.5	1.0	0.83	0.83	
	20	37.20	0.5	7.6	2.50	3.33	
	21	23.75	2.5	6.9	1.67	-0.83	
Total	21	24.82	0.83	4.27	1.51	1.31	

Note. N = number of participants; SNR = signal-to-noise ratio loss; dB = decibels; SRT = speech reception threshold; PTA = puretone average; R = right ear; L = left ear; M = males; F = females.

2.4.2 Stimuli

To cross-validate the UCAMST against the QuickSIN and meaningful CVC (revised AB) word recognition test, an adaptive method of presenting sentence stimuli was implemented. The software was programmed to execute dual adaptive tracks over a total of 28 sentence trials to calculate the SNRs at which a participant would score 20% and 80% correctly, resulting in 14 sentences being presented for each adaptive track respectively. In this procedure, both the constant or babble BG noise remained fixed at 65 dB SPL and the target sentence was varied accordingly.

2.4.3 Experimental instrumentation

Experimental instrumentation for the participants' screening and HT was identical to that described for the evaluation of normalisation for the quiet presentation mode. Listeners recruited for the validation of the UCAMST were also required to undertake speech testing. The meaningful CVC (revised AB) word recognition test and the QuickSIN were assessed in a sound attenuated test booth by means of a calibrated Grason-Sadler GSI audiometer (Minnesota, USA) with an attached compact disc player. Speech test material was delivered via Telephonics TDH-50P supra-aural headphones (Griffon Corp; New York, USA) worn by the participants who would respond verbally to indicate what they heard. The experimental instrumentation for the UCAMST component of this study was identical to that carried out for the evaluation of normalisation for the quiet presentation mode and therefore won't be repeated here.

2.4.4 Experimental procedures

The screening procedure for Part 2 was identical to that carried out in Part 1. In summary, informed consent was gained from each candidate followed by a brief interview and explanation of what would be involved in the testing process. An otoscopic examination and a hearing test was then performed and the results explained to the participant.

Validation of the UCAMST required all participants to undertake the meaningful CVC (revised AB) word recognition test, the QuickSIN, and the UCAMST.

These tests were presented in succession in a randomised sequence. As a result of this the experimental procedure for each participant was unique and thus the process carried out for each test will be described below, in no particular order.

2.4.4.1 Meaningful CVC (revised AB) word recognition test

Prior to testing word recognition, listeners were verbally instructed that they would hear two words at varying volumes; the first, the carrier phrase: “say”, followed by another word which they must repeat out loud. Participants were encouraged to make an educated guess for each trial if they were uncertain and that they could respond with a sound should they only hear part of a word. The supra-aural headphones were then fitted to the listener and the audiometer calibrated for the meaningful CVC (revised AB) word recognition test using the appropriate track on the compact disc. Lists one to four were presented binaurally at 50, 20, 10 and 5 dB respectively. Participant responses were scored phonemically and awarded points according to the algorithm depicted in the Table 5. Points were tallied to give a percentage correct score for each intensity level that was interpreted for the participant with regards to their hearing thresholds and recorded for analysis in the current study.

Table 5. Scoring algorithm for phonemes in the meaningful CVC (revised AB) word recognition test

***n* Phonemes correct**	***n* Points awarded**
0	0
1	3
2	7
3	10

Note. N = number.

2.4.4.2 QuickSIN

For the QuickSIN, participants remained seated in the sound attenuated booth and were instructed by the researcher that they would hear a sentence in the presence of BG noise through the headphones. Their task was to repeat the sentence out loud. Listeners were encouraged to make an educated guess if they were uncertain of words in the sentence, especially as the test progressed and became more difficult. Participants were fitted with the supra-aural headphones and the audiometer was calibrated for the QuickSIN test using the appropriate track on the compact disc. The presentation level was set to 70 dB HL, as recommended for NH listeners, and routed to be delivered binaurally (Killion et al., 2004). Two practice lists were presented to familiarise the participant with the test and then their responses were recorded for list one. One point was awarded for each correctly identified key word repeated. The total points earned across all five sentences of list one was then subtracted from 25 to produce the participant's SNR loss. The SNR loss was interpreted for the listener using the information found in Table 6. (Etymotic Research Incorporation, 2006) and recorded for analysis in this research.

Table 6. Categories of SNR loss.

SNR loss	Degree
0-3	Normal
3-7	Mild
7-15	Moderate
>15	Severe

Note. SNR = signal-to-noise ratio.

2.4.4.3 UCAMST

The experimental procedure for the UCAMST involved delivering the test adaptively in all available modes. Hence, each participant was required to undertake the test in the open and closed set condition in the quiet presentation mode as well as in the presence of babble noise, and constant noise respectively. Tests were presented sequentially with brief instructions given before each task.

Participants were seated in the sound attenuated test booth facing the touch-screen monitor and were verbally instructed that they would hear a sentence through the headphones which would either be in constant noise, babble noise, or in quiet. Directions given to listeners varied depending on the condition of the listener's subsequent task. For closed set tests, participants were asked to identify what they heard for each sentence trial from the matrix displayed on the monitor by selecting the corresponding words they heard from the appropriate columns using the mouse placed on the table in front of them or their finger on the touch screen. The layout of the matrix became visible for participants following each sentence presentation and is illustrated in Figure 8. Alternatively, if the participant's task was to complete the open set condition for a presentation mode, the monitor in front of them was left blank and they were requested to repeat the sentence out loud for the researcher on the other side of the test booth to score on the computer screen outside the participant's field of vision. The matrix used by the researcher to score participant responses is depicted in Figure 9.

All participants, regardless of condition, were encouraged to make an educated guess if they were unsure of words presented in the sentence in order to progress to the next trial. Word scoring was again used to give a total score out of five for each sentence. The first closed set task, irrespective of condition or where it occurred in the sequence of tasks, was presented as a training set to allow participants to become familiar with responding independently on the touch screen. Results from the training sets were discarded and the data from the following closed set tests (including a repetition of the condition which was presented for training purposes) were recorded for analysis alongside the results from the open set tests. Each test was comprised of 28 sentences to adaptively ascertain the SNR at which the listener correctly identified either 80% or 20% of the words in a particular trial.

Rest breaks were encouraged throughout the appointment due to the high level of concentration required and the increased cognitive load associated with performing the tasks necessary for the cross-validation of the UCAMST. Excluding such breaks, the complete experimental procedure took approximately 75 minutes.

2.5 Statistical analyses

Prior to completing any analyses, the data was first examined for normal distribution using the Shapiro Wilks test and also examined for bias in the form of significant outliers, skewness, and kurtosis. The outcomes of these investigations revealed multiple violations to the assumption of normality for several conditions and variables throughout the data. Therefore, it was determined that non-parametric analyses would need to be used for hypothesis testing in this study. A Kruskal-Wallis analysis and a univariate analysis of variance (ANOVA) were used for the evaluation of hypothesis one. Equivalence across conditions was calculated for hypothesis two using a Friedman's related-measures ANOVA followed by a Wilcoxon signed rank test for pair-wise comparisons. For the investigation of hypothesis three, two-tailed bivariate correlations were performed using the Spearman's (non-parametric) correlation coefficient.

A statistical $p \geq 0.05$ value was selected for use for the investigation of hypotheses one and two. This value was also used for the comparison of the UCAMST and the meaningful CVC (revised AB) word recognition test. For the repeated correlations carried out between variables of the UCAMST and the QuickSIN, a less conservative Bonferroni correction factor of ≥ 0.01 was selected for use interpreting the *p*-values for the correlations to minimise the likelihood of making a type I error.

CHAPTER 3 *RESULTS*

3.1 Introduction

The findings of this study are reported in three sections. In the first section, the equivalence of the lists designed for the open set and closed set condition are examined for the babble noise and quiet presentation mode of the UCAMST. The second section evaluates the UCAMST for equivalence across each of the conditions. The third section compares the UCAMST in the quiet presentation mode with the meaningful CVC (revised AB) word recognition test, as well as the UCAMST babble noise and constant noise condition with the QuickSIN to determine if there are any correlations between the variables of interest.

As previously described, prior to completing any analyses for hypothesis testing, the data was first examined for normal distribution using the Shapiro Wilks test and also examined for bias in the form of significant outliers, skewness, and kurtosis. As significant bias was revealed in several conditions and variables throughout the data, non-parametric testing was exclusively used in this study.

3.2 List equivalence results

Prior to performing any hypothesis testing analyses, descriptive statistics were examined for the data collected for the evaluation of normalisation. These figures are provided in Table 7. Two data points from list one of the closed set quiet condition were omitted from the data set as they were biased by administrator miscalculation and contributed to erroneously poor psychometric functions and were considerably different from the remainder of the data. The missing values were excluded test-wise in each of the subsequent analyses detailed in this section.

Table 7. Means and Standard deviations of the slope and *SRT* of the lists designed for the UCAMST’s open set and closed set, babble noise and quiet condition.

	Condition															
	Closed Set, Babble Noise				Open Set, Babble Noise				Closed Set, Quiet				Open Set, Quiet			
	Slope (%/dB)		***SRT* (dB SNR)**		**Slope (%/dB)**		***SRT* (dB SNR)**		**Slope (%/dB)**		***SRT* (dB A)**		**Slope (%/dB)**		***SRT* (dB A)**	
List	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD
1	8.84	3.14	-8.34	1.50	12.99	2.74	-6.60	1.00	14.77	18.30	-5.28	2.13	21.66	22.17	-2.08	2.42
2	8.53	4.44	-8.38	1.00	13.17	3.31	-7.34	1.18	6.07	1.88	-4.74	1.89	14.26	5.90	-1.45	1.54
3	9.99	3.49	-8.92	1.46	10.15	2.49	-7.46	1.08	13.23	17.85	-4.84	1.80	12.68	4.71	-0.73	0.92
4	8.90	2.42	-8.64	1.42	12.40	2.93	-7.22	1.37	7.76	1.43	-4.42	1.98	11.51	5.90	-2.66	2.13
5	9.51	2.13	-8.87	0.63	12.62	4.59	-7.34	1.28	5.32	1.32	-3.76	2.52	15.65	5.59	-1.70	1.46
6	8.28	2.44	-8.36	1.00	10.64	4.34	-7.63	1.48	13.20	9.89	-4.91	1.78	11.90	5.99	-0.57	1.95
7	9.65	7.22	-8.39	0.88	13.13	5.13	-6.88	0.47	7.16	2.38	-2.86	1.16	9.24	2.69	-2.94	1.53
8	11.28	4.57	-9.21	0.85	14.15	3.23	-7.18	0.77	7.00	2.48	-4.05	2.16	15.08	5.44	-2.17	1.30
9	7.77	2.39	-8.80	1.62	10.69	2.43	-6.92	0.94	8.75	2.35	-4.57	1.95	12.68	4.74	-0.61	1.90
10	8.49	2.31	-8.43	1.07	12.35	3.97	-6.71	0.76	8.22	2.74	-4.46	1.55	12.35	4.88	-2.58	2.08
11	10.63	3.41	-9.00	1.31	13.20	2.84	-7.17	0.84	8.24	1.81	-4.08	1.84	17.96	3.98	-1.68	1.10
12	8.75	2.96	-8.35	1.14	13.95	6.06	-6.90	1.04	7.28	2.69	-4.41	2.29	13.35	6.46	-0.92	2.35
13	11.12	5.80	-7.97	1.38	10.72	3.45	-6.20	1.22	6.68	1.63	-2.88	2.35	10.78	6.70	-1.48	1.48
14	8.09	1.64	-8.88	1.50	14.47	5.25	-6.66	1.01	8.73	1.81	-4.55	2.11	12.13	3.86	-2.33	1.67
15	10.12	4.64	-8.87	1.27	10.63	3.23	-7.48	1.25	12.51	1.49	-4.42	2.21	16.21	6.63	-1.44	1.48
16	9.70	4.21	-8.50	1.55	11.19	3.64	-6.88	1.01	5.59	1.32	-5.09	2.36	8.12	3.95	-0.66	3.25
M	9.35	3.58	-8.62	1.22	12.28	3.73	-7.04	1.04	8.78	4.46	-4.33	2.01	13.47	6.22	-1.63	1.79

Note. M = mean, SD = standard deviation.

Hypothesis one states that no significant differences would be found between the stimulus lists in the closed set, babble noise; open set, babble noise; closed set, quiet condition; and open set quiet condition respectively for slope and *SRT*. The results of the Kruskal-Wallis analysis and the Univariate analysis of variance (ANOVA) aimed at testing this hypothesis can be seen in Table 8. The significance values obtained from the Kruskal-Wallis test were asymptotic as exact scores were unable to be executed in SPSS.

Table 8. KWH, P-values, and η^2 for the Kruskal-Wallis one-way ANOVA in each of the four conditions.

Condition	**Variable**	**KWH**	***p***	**η^2**
Closed set,	Slope	9.72	0.84	0.07
Babble noise	*SRT*	13.78	0.54	0.06
Open set,	Slope	18.16	0.23	0.12
Babble noise	*SRT*	17.28	0.30	0.12
Closed set,	Slope	46.11	<0.01	0.12
Quiet	*SRT*	19.12	0.21	0.11
Open set,	Slope	35.35	<0.01	0.17
Quiet	*SRT*	24.47	0.06	0.15

Note. Degrees of freedom = 15; KWH = Kruskal-Wallis H.

The Kruskal-Wallis ANOVA for the closed set babble condition revealed that there were no significant differences between the lists designed for this condition with regards to slope, $KWH = 9.72$, $p = 0.84$, $\eta^2 = 0.07$, and *SRT*, $KWH = 13.78$, $p = 0.54$, $\eta^2 = 0.06$. The same is true for the lists designed for the open set babble condition for slope, $KWH = 18.16$, $p = 0.23$, $\eta^2 = 0.12$, and *SRT*, $KWH = 17.28$, $p = 0.30$, $\eta^2 = 0.12$. These findings are illustrated in Figure 10 and Figure 11 where there is minimal variation between the lists. The ANOVA also revealed that there were no significant differences between the quiet open set condition lists and quiet closed set condition lists with regards to the *SRT*, $KWH = 24.47$, $p = 0.06$, $\eta^2 = 0.15$, and $KWH = 19.12$, $p = 0.21$, $\eta^2 = 0.11$ respectively. Lists designed for presentation in the open set and closed set quiet condition, however, were found to differ significantly with regards to slope, p = < 0.01. These findings are demonstrated in Figure 12 and Figure 13 where the variation between lists is apparent.

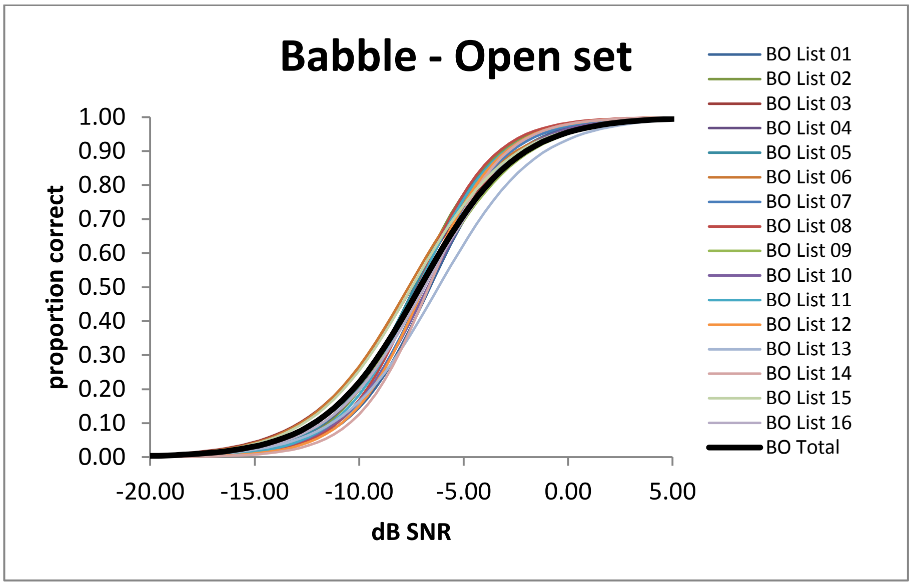


Figure 10. Intelligibility functions generated for the lists designed for use in the open set babble noise condition.

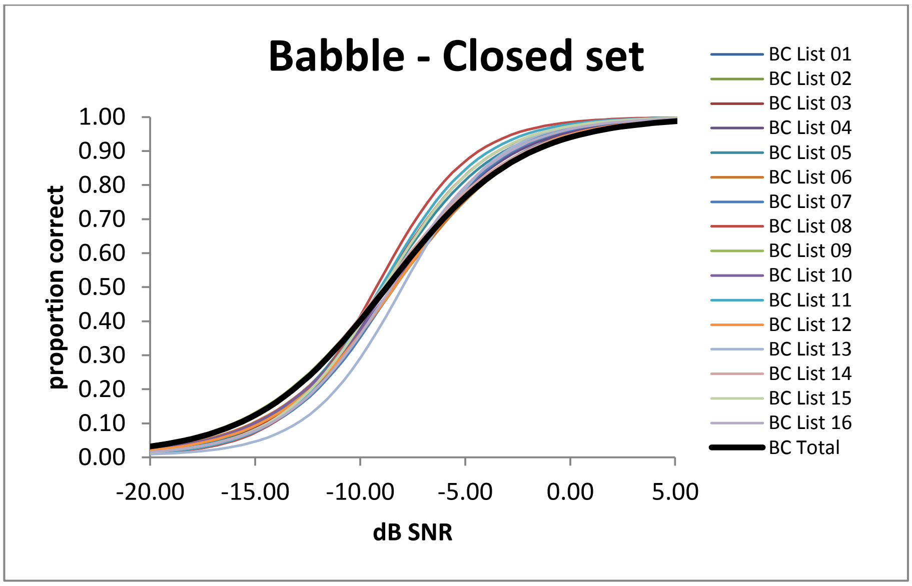


Figure 11. Intelligibility functions generated for the lists designed for use in the closed set babble noise condition.

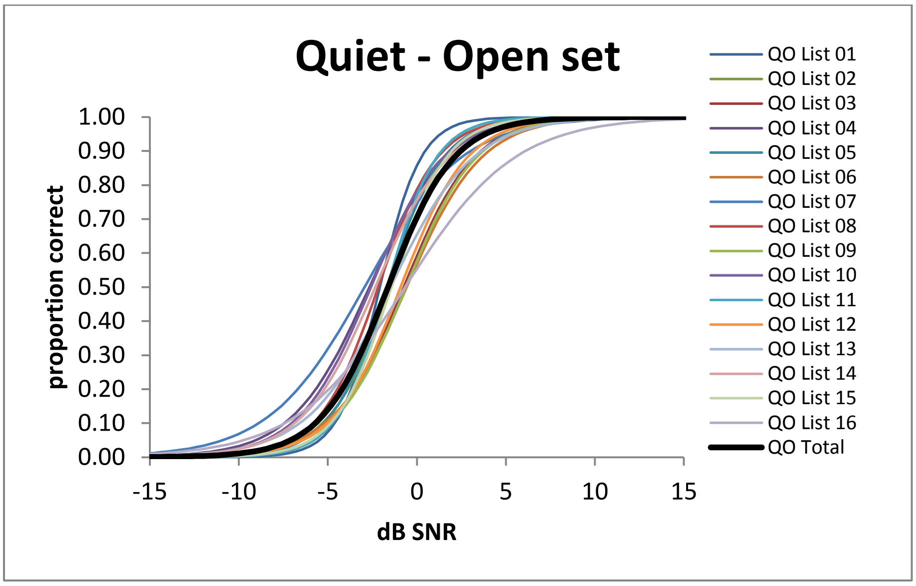


Figure 12. Intelligibility functions generated for the lists designed for use in the open set quiet condition.

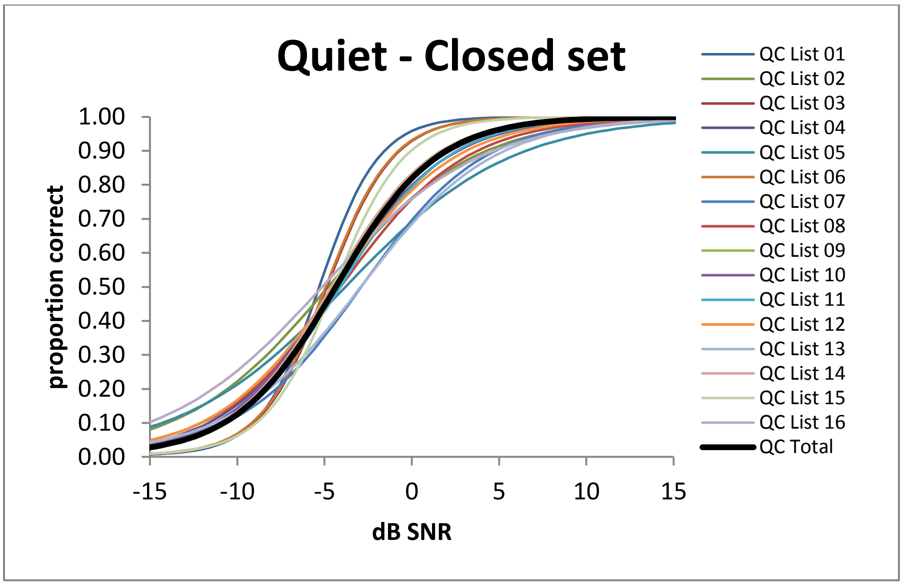


Figure 13. Intelligibility functions generated for the lists designed for use in the closed set quiet condition.

3.3 Condition equivalence results

Prior to performing any analyses for testing the second hypothesis of this study, the descriptive statistics of the data collected for determining condition equivalence was examined, the outcome of which is outlined in Table 9. Eight data points were omitted from the data set as they were biased by administrator miscalculation and produced erroneously poor psychometric functions which were substantially different from the remainder of the data. These missing values were excluded list-wise in each of the following analyses detailed in this section.

Table 9. Means and Standard deviations of the slope and *SRT* of lists designed for each condition in the UCAMST.

Condition	**Variables**			
	Slope (%/dB)		***SRT* (dB SNR or dB A)**	
	M	SD	*M*	SD
Closed set, Constant noise	14.51	6.24	-8.394	1.19
Open set, Constant noise	21.09	14.41	-7.48	0.94
Closed set, Babble noise	12.17	9.81	-8.91	1.87
Open set, Babble noise	10.18	4.39	-6.87	1.60
Closed set, Quiet	8.85	2.97	2.31	1.33
Open set, Quiet	9.56	5.51	4.97	1.47

Note. M = mean, SD = standard deviation.

Hypothesis two states that no significant difference would be found between the six test conditions (i.e. open set, constant noise; closed set, constant noise; open set, babble noise; closed set, babble noise; open set, in quiet; closed set in quiet) with regards to the slope, and *SRT*. To investigate this, a Friedman's related-measures ANOVA was performed. As outlined in Table 10, the findings of this test revealed at least one condition to be significantly different from the others with regards to the slope, $\chi^2 = 20.38$, $p = < 0.01$, and the *SRT*, $\chi^2 = 45.33$, $p = < 0.01$. The significance values obtained from the Friedman's test were asymptotic as exact scores were unable to be executed in SPSS.

Table 10. χ^2, and P-values for the Friedman's related-measures ANOVA for the slope and *SRT* of the open and closed set babble, constant, and quiet condition.

Variable	χ^2	***p***
Slope	20.38	<0.01
SRT	45.33	<0.01

Note. Degrees of freedom = 5.

These results were followed-up with a Wilcoxon signed rank test to conduct pairwise comparisons; the outcome of which is displayed in Table 11 and Table 12. Exact values of significance were able to be obtained for both analyses (detailed below).

Table 11. Z-values of the Wilcoxon signed rank test for the slope across test conditions.

	Closed set, Constant noise	**Closed set, Quiet**	**Open set Babble noise**	**Open set, Constant noise**	**Open set, Quiet**
Closed set, Babble noise	**-1.81 (0.07)**	**-1.60 (0.11)**	**-0.78 (0.44)**	**-1.91 (0.06)**	**-1.82 (0.07)**
Closed set, Constant noise		-2.59	-2.16	**-1.36 (0.17)**	-2.17
Closed set, Quiet			**-1.16 (0.24)**	-3.30	**-0.52 (0.60)**
Open set, Babble noise				-2.59	**-0.45 (0.65)**
Open set, Constant noise					-2.95

Note. All tests were significant with p = < 0.05 except where noted in parentheses; non-significant p-values are indicated in ***Bold****.*

The findings from the Wilcoxon signed rank analysis displayed in Table 11 reveal that there are significant differences ($p = < 0.05$) between six of the conditions with regards to slope. However, nine conditions were found to be equivalent as indicated in bold in the table above.

To summarise these results, the closed set babble condition was found to be equivalent with the four following conditions respectively: closed set constant noise, $\chi^2 = -1.81$, $p = 0.07$; open set constant noise, $\chi^2 = -1.91$, $p = 0.06$; closed set quiet, $\chi^2 = -1.60$, $p = 0.11$; and open set quiet $\chi^2 = -1.82$, $p = 0.07$. The open set babble noise condition was separately equivalent with the three following conditions: closed set babble noise, $\chi^2 = -0.78$, $p = 0.44$; open quiet, $\chi^2 = -0.45$, $p = 0.65$; and the closed set constant noise, $\chi^2 = -1.16$, $p = 0.24$. In addition to the pairings already described above, no significant differences were found between the closed set constant noise condition and the open set constant noise condition, $\chi^2 = -1.36$, $p = 0.17$, which is also true of the open set quiet condition and the closed set quiet condition, $\chi^2 = -0.52$, $p = 0.60$. The equivalence pairs described are with respect to the slope of the individual conditions. The equivalence of conditions regarding *SRT*, however, is less prevalent and is outlined in Table 12.

Table 12. Z-values of the Wilcoxon signed rank test for the <u>*SRT*</u> across test conditions.

	Closed set, Constant noise	**Closed set, Quiet**	**Open set Babble noise**	**Open set, Constant noise**	**Open set, Quiet**
Closed set, Babble noise	**-0.31 (0.77)**	-3.30	-2.48	-2.12	-3.41
Closed set, Constant noise		-3.18	-2.50	-2.78	-3.30
Closed set, Quiet			-3.30	-3.30	-2.76
Open set, Babble noise				-0.78	-3.41
Open set, Constant noise					-3.41

Note. All tests were significant with p = < 0.05 except where noted in parentheses; non-significant p-values are indicated in ***Bold****.*

The findings of the Wilcoxon signed rank test for comparing the conditions of the UCAMST with respect to the *SRT*, revealed significant differences between all the conditions analysed with the exception of the closed set constant and babble noise condition, $\chi^2 = -0.31$, $p = 0.77$.

The results from the analyses in this section is visualised in Figure 14 which graphically displays the difference between the psychometric functions for each of the conditions. The difference between the quiet condition versus the constant and babble noise condition is particularly apparent in this figure.

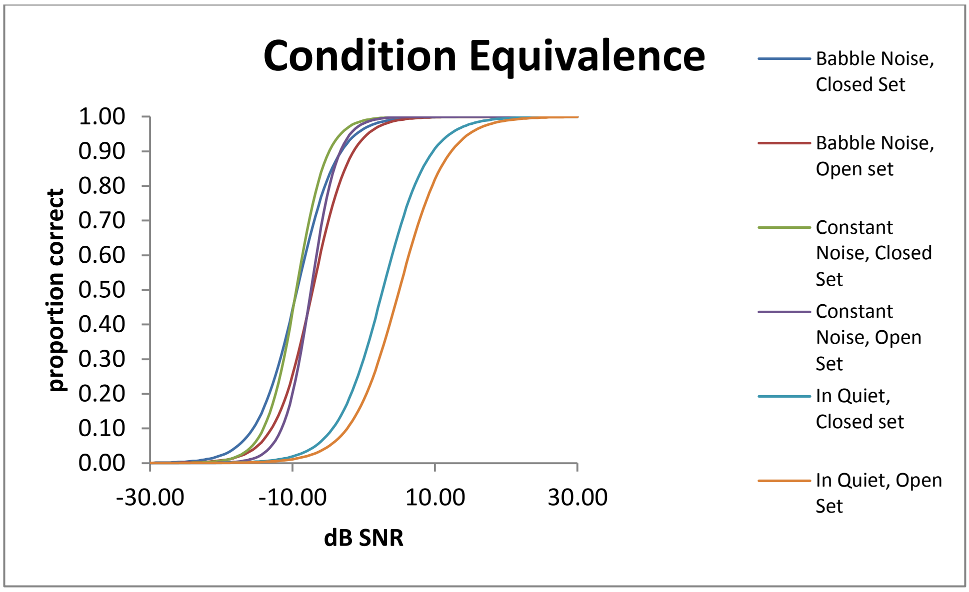


Figure 14. Intelligibility functions generated for each condition in the UCAMST.

3.4 Comparison across tests results

Prior to any hypothesis testing, descriptive statistics were inspected for the variables under investigation in this section. These values are provided in Table 13. The data obtained from participant 9 was removed from the data set as their performance in several of the tests administered resulted in psychometric functions that were substantially different from those derived from the other participants which considerably altered the mean and SDs of the data set. The removal of this participant's data is justified on the basis that the listener had extensive previous experience with the assessments administered and therefore, familiarity with the stimuli may have led to the false enhancement of their performance for the tests delivered.

Table 13. Means and Standard deviations of the variables for conditions in the UCAMST and for the meaningful CVC (revised AB) word recognition test and the QuickSIN.

Condition	Variable	Units	*M*	SD
Closed set, Babble noise	**Slope**	%/dB	9.58	1.74
	SRT	dB SNR	-9.47	1.80
	SNR 80%		-5.73	1.64
	SNR 20%		-13.22	2.23
Open set, Babble noise	**Slope**	%/dB	10.60	4.20
	SRT	dB SNR	-6.80	2.12
	SNR 80%		-3.21	2.32
	SNR 20%		-10.38	2.34
Closed set, Constant noise	**Slope**	%/dB	12.78	2.97
	SRT	dB SNR	-9.59	1.02
	SNR 80%		-6.73	1.06
	SNR 20%		-12.44	1.36
Open set, Constant noise	**Slope**	%/dB	15.80	8.79
	SRT	dB SNR	-7.26	1.50
	SNR 80%		-4.74	1.35
	SNR 20%		-9.76	1.92
Closed set, Quiet	**Slope**	%/dB	8.45	1.52
	SRT	dB A	2.27	2.13
	Level 80%		6.49	2.65
	Level 20%		-1.98	1.80
Open set, Quiet	**Slope**	%/dB	8.40	3.37
	SRT	dB A	5.02	1.80
	Level 80%		9.62	2.82
	Level 20%		0.41	1.64
Meaningful CVC (revised AB) word recognition test	**Slope**	%/dB	6.03	2.68
	SRT	dB A	5.07	4.28
QuickSIN	**SNR loss**	dB SNR	0.80	1.34

Note. M = mean, SD = standard deviation.

Hypothesis three states that no significant correlation exists between the UCAMST in the open and/or closed set quiet condition and the meaningful CVC (revised AB) word recognition test, and no significant correlation exists between the UCAMST open and/or closed set constant or babble condition and the QuickSIN. To investigate this two-tailed bivariate correlations were performed using the Spearman's non-parametric) correlation coefficient. The outcomes of these analyses are detailed in Table 14, Table 15, and Table 16.

Table 14. Correlations between the open set and closed set, quiet UCAMST conditions and the meaningful CVC (revised AB) word recognition test with regards to slope.

			Meaningful CVC (revised AB) word recognition test
Variable	**UCAMST Condition**		**Slope**
Slope	**Open set, Quiet**	r_s	> 0.01
		p	1.00
	Closed set, Quiet	r_s	> 0.01
		p	1.00

Note. *r_s = Spearman's Rho correlation coefficient; p = probability value.*

The findings from Table 14 reveal that there are no significant correlations ($p \geq 0.05$) between the meaningful CVC (revised AB) word recognition test and the UCAMST in the open set and closed set quiet condition with regards to slope.

Table 15. Correlations between the open set and closed set, quiet UCAMST conditions and the meaningful CVC (revised AB) word recognition test with regards to *SRT*.

			Meaningful CVC (revised AB) word recognition test
Variable	**UCAMST Condition**		***SRT***
SRT	**Open set, Quiet**	r_s	0.35
		p	0.14
	Closed set, Quiet	r_s	0.56
		p	0.01

Note. *r_s = Spearman's Rho correlation coefficient; p = probability value.*

The results, shown in Table 15, indicate that, with regards to *SRT*, there is no significant correlation ($p \geq 0.05$) between the meaningful CVC (revised AB) word recognition test and the UCAMST in the open set quiet condition. However, there is a significant correlation between the *SRT* of the closed set quiet condition and the *SRT* of the meaningful CVC (revised AB) word recognition test $r_s(18) = 0.56$, $p = 0.01$. This relationship can be visualised in Figure 15.

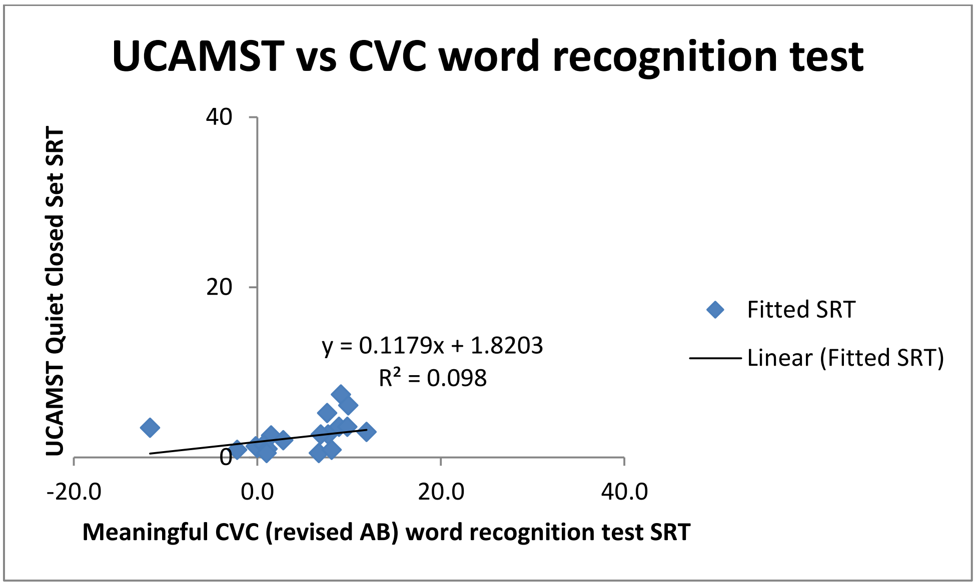


Figure 15. Correlation between the *SRTs* of the UCAMST presented in quiet in the closed set condition and the meaningful CVC (revised AB) word recognition test.

Table 16. Correlations between the open and closed set, babble and constant UCAMST conditions and the QuickSIN.

UCAMST Condition	Variable		QuickSIN SNR loss
Closed set, Babble noise	***SRT***	r_s	0.37
		p	0.11
	SNR 80%	r_s	0.42
		p	0.07
	SNR 20%	r_s	0.26
		p	0.27
Open set, Babble noise	***SRT***	r_s	0.20
		p	0.40
	SNR 80%	r_s	0.19
		p	0.41
	SNR 20%	r_s	0.16
		p	0.50
Closed set, Constant noise	***SRT***	r_s	0.29
		p	0.22
	SNR 80%	r_s	0.18
		p	0.44
	SNR 20%	r_s	0.33
		p	0.15
Open set, Constant noise	***SRT***	r_s	0.27
		p	0.25
	SNR 80%	r_s	0.36
		p	0.12
	SNR 20%	r_s	0.10
		p	0.67

Note. r_s *= Spearman's Rho correlation coefficient; p = probability value, SNR 80% = the SNR at which an 80% correct score is achieved; SNR 20% = the SNR at which a 20 % correct score is achieved.*

As repeated correlations are known to artificially inflate the $1 - \alpha$ level, a slightly less conservative Bonferroni correction factor of ≥ 0.01 was selected for use in interpreting the *p*-values for the correlations detailed above in Table 16. Using this correction factor, no significant correlations were found between the *SRT*, SNR 80%, or SNR 20% of the different conditions of the UCAMST and the QuickSIN SNR loss.

3.5 Summary

This section provides a brief summary of the main results of the current study:

(1) The lists in the babble noise condition were revealed to be equivalent with respect to the slope and *SRT*. The lists in the quiet condition were found to be equivalent with regards to the *SRT* but not slope.

(2) Analyses of the six UCAMST conditions show significant differences between the derived slope and *SRT* from the lists which comprise each condition. Pairwise comparisons indicate greater equivalence between conditions with reference to the slope as opposed to the *SRT*.

(3) Comparisons between the UCAMST and meaningful CVC (revised AB) word recognition test demonstrate a significant correlation between the estimated *SRT* of the closed set quiet condition and the meaningful CVC (revised AB) word recognition test. No significant correlations were present between the UCAMST constant noise and babble noise condition and the QuickSIN.

CHAPTER 4 *DISCUSSION*

4.1 Introduction

The aim of this research was to evaluate the equivalence of the lists and conditions available for use in the UCAMST in terms of the expected estimations for slope and *SRT*. Additionally, this study sought to cross-validate the UCAMST with the meaningful CVC (revised AB) word recognition test and the QuickSIN. The results from the analyses recorded in Chapter 3 are explored in the following sections, and discussed with respect to other relevant findings documented in literature. The limitations encountered in this project are also reviewed, along with recommendations for potential areas for future research.

4.2 Equivalence measures

4.2.1 List equivalence

The first hypothesis this research sought to address concerned list equivalence and predicted that there would be no significant difference between the lists designed for presentation in babble noise, and those presented in quiet. Although this hypothesis had already been explored in a previous study on the UCAMST by Stone (2016) for babble noise and constant noise, a malfunction in the software during experimental procedures led to the recommendation that the evaluation of normalisation process for babble noise be repeated before the development of the UCAMST could progress further. The results from the current research revealed that there is no significant difference between the lists designed for presentation in babble noise with respect to the expected estimations of the slope and *SRT* for each list. Moreover, this finding was maintained irrespective of whether the test was delivered in the open set or closed set condition.

The list equivalence for the quiet presentation mode was also investigated. The findings from the analyses documented in the previous section demonstrate that there were no significant differences between the lists presented in quiet with regards to the *SRT* in both the open set and closed set condition. However, equivalence was not wholly achieved for the quiet presentation mode, as significant differences were revealed for the slope which was derived from the psychometric functions generated for each list in the open set and closed set condition.

As a result of these findings, hypothesis one may be considered to be partially true. It was correct in the assumption that there would be no significant difference between the lists comprised in the babble noise presentation mode, but was proven incorrect with the finding of significant differences for the slope of the lists selected for presentation in quiet. As mentioned previously, establishing equivalency across test lists strengthens the reliability and sensitivity of the measure (Akeroyd et al., 2015). The failure to do so for the slope of the quiet condition is critical as the implications of non-equivalent lists could be fluctuations of the derived slope of a listeners performance, depending on the list presented.

Consequently, these findings have implications for the use of the UCAMST in practice clinically or in the research environment. Indeed, equivalency across list slopes can, for all intents and purposes, be perceived as an indicator of equal difficulty across sentence material. Thus, a clinician or researcher can confidently administer the test and obtain a *SRT* that is an accurate and reliable representation of a listeners performance regardless of the list presented. Such interpretation is possible in the babble noise condition but not in the quiet presentation mode. However, with further in-depth analysis, the results obtained in this research may act as the foundation for the generation of lists for the quiet condition which are more likely to be equivalent.

Due to the similarities in the test procedure and analyses in the current study, it is possible to evaluate the findings of this research with the results reported by Stone (2016). For each condition under investigation, both studies reported the partial eta squared effect size, η^2, which, in this case, is a measure of the amount of variance in the slope or *SRT* which is accounted for by the condition the test was presented in, either the open set or the closed set. The η^2 for babble noise reported by Stone (2016) ranged from 0.13 – 0.15 for the closed set condition and 0.12 – 0.17 for the open set condition. In contrast, the η^2 results obtained in this study were smaller and less spread, with values ranging from 0.06 – 0.07 for the closed set and 0.12 for the open set condition. A low effect size is desirable when seeking to attain equivalence as it indicates that the variation, or lack thereof, is predominantly due to the homogeneity of lists and not attributable to the condition in which the test was presented.

In addition to the effect size, Stone (2016) also reported that the lists for babble noise were significantly different with regards to slope in the closed set, $\chi^2 = 31.74$, $p = < 0.01$, and open set condition, $\chi^2 = 34.27$, $p = < 0.01$. These findings differ to the results obtained for babble noise in the current study which found no significant difference for the slope of the lists in babble noise regardless of condition. The dissimilarity in the equivalence values recorded in Stone’s study and the current research provides evidence towards the conclusion that the significant differences between lists revealed in the previous study may be accredited to the malfunction in the software which allowed non-optimised stimuli to be presented.

The results of this research can also be compared with those described in other MSTs. While the UCAMST is one of the pioneers of the quiet presentation mode, and MSTs using babble noise as a masker are less prevalent than constant noise, there are several studies which have detailed the outcomes of the process of the evaluation of normalisation, including the work by Jamaluddin (2016) and Ozimek et al. (2010). Jamaluddin (2016) developed the Malay version of the UCAMST and evaluated normalisation for babble noise in the closed set condition via two presentation methods: adaptive measurements and fixed SNR levels. The fixed SNR levels method is similar to the experimental procedure utilised in Part 1 of the current study and found no significant differences between the lists with respect to the expected estimations of slope and *SRT*. The evaluation of normalisation for the Polish MST, as described by Ozimek et al. (2010), also found the lists designed for presentation in the babble noise open set condition to be equivalent with respect to slope and *SRT*. The mean values obtained for the *SRT* and slope are as follows for the open set babble condition of the Polish MST, -9.6 dB SNR ± 0.2 and 17.7 ± 1.6 %/dB (Ozimek et al., 2010), and the UCAMST, -7.04 ± 10.9 dB SNR and 12.28 ± 3.93 %/dB, respectively. For the closed babble condition of the Malay and New Zealand English versions of the UCAMST, the independent mean values recorded for *SRT* and slope are -10.1 ± 0.2 dB SNR and 14.9 ± 1.2 %/dB (Jamaluddin, 2016), and -8.62 ± 1.24 dB SNR and 9.35 ± 3.82 %/dB. Although analysing these figures for equivalence is outside the scope of this study, it is obvious that the mean scores for the New Zealand English version of the UCAMST are lower than those reported in the international MSTs considered here. This phenomenon could be due to aspects of the experimental procedure which involved repeat presentation of the -7.6 dB SNR and repeat testing of 10 participants.

As documented in Chapter 2 of this study, the experimental procedure for the evaluation of normalisation of lists in the babble noise condition was separated into two blocks. The first block administered the test using the fixed presentation levels of -14.3 and -7.6 dB SNR, however the selected SNRs made the task too difficult and the resultant participant performance was inadequate for the purpose of fitting a *PI* function. Therefore, a second block of testing was undertaken using the easier fixed presentation levels of -7.6 and -3.7 dB SNR. Interestingly, participant performance was higher for the -7.6 dB SNR in the second block as opposed to the first block of testing, with the mean score for the open set condition increasing from 36.1% ± 14.5% in the first block to 53.6% ± 14.8% in the second block. This occurrence was also apparent in the closed set babble noise condition where the average scores increased from 44.3% ± 16.9% in the first block to 62.8% ± 12.3% in the second block for the same -7.6 dB SNR.

A potential reason for the difference in participant performance may be related to the perceived difficulty of the task. There is evidence in literature which suggests that in examination environments there is a relationship between knowledge of the material and the number of questions neglected to answer or ‘skipped’, and that participants are more likely to skip a question than hazard a guess when they are uncertain of the answer (Baldiga, 2014). Thus, considering that block one contained a SNR which may be considered particularly challenging (i.e. -14.3 dB SNR), it is possible that participants were less certain of what they heard and therefore less willing to guess even for the comparably easier SNR of -7.6 dB SNR in the first block. In contrast, participants may have been more willing to guess in the second block of testing as the task was not as challenging and they were therefore able to achieve higher scores for the SNR of -7.6 dB SNR as word scoring would have allowed them to accrue points for each correctly repeated word in the sentence.

Alternatively, the elevated scores seen in block two may be due to the inclusion of 10 participants who had previously taken part in the testing for block one. As mentioned previously in Chapter 2 of this study, the time restraints for the availability of audiological equipment made recruiting 20 new participants impracticable. Therefore

10 volunteers were re-recruited and retested for the evaluation of normalisation for babble noise. To minimise the bias that repeat testing may introduce into the data set, re-recruited participants undertook the test in the same condition that they were presented initially. There is extensive evidence in literature supporting what is known as the practice effect which describes the improvement in performance for cognitive tests that occurs as a result of repeated exposure to the test materials (Duff, Callister, Dennett, & Tometich, 2012). Such effects are usually considered a source of error as familiarity with the assessment stimuli can cause falsely inflated test scores. Many factors influence the strength of the practice effect including the retest interval (Tombaugh, 2005), the inherent susceptibility of the test used (Basso, Carona, Lowery, & Axelrod, 2002), and individual considerations such as the age and health status of the participant (Calamia, Markon, & Tranel, 2012).

Like all other MSTs, a key advantage of the UCAMST is that the sentence stimulus design (i.e. name, verb, number, adjective, object) ensures the predictability of the test is relatively low. In addition to this, participants recruited for the evaluation of normalisation were presented a total of 160 sentence trials over the course of approximately 45 minutes, therefore reducing the likelihood of a re-recruited listener memorising test materials which they heard during block one of testing. Unfortunately, further investigation into the extent of the practice effect associated with UCAMST was beyond the scope of this study due to the limited number of re-recruited participants and stringent time constraints. However, it is an important issue which will be given further consideration in later sections with respect to the limitations of this study and possible future research directions for the UCAMST.

4.2.2 Condition equivalence

The second hypothesis proposed in this research concerned condition equivalence and predicted that there would be no significant difference between the six test conditions available in the UCAMST with regards to the expected estimations of slope and *SRT*. Like list equivalence, this aspect of the UCAMST has already been investigated by Stone (2016). However, due to the issues encountered during the data collection phase of Stone's study, as described previously, re-analysis of condition equivalence was warranted. The results of the current study found significant differences between at least one of the conditions for both slope and *SRT* when compared using a Friedman's analysis. However, when followed up with the Wilcoxon analysis, equivalence was revealed between multiple pairs of conditions.

The majority of equivalent pairs were found with respect to the derived slope, as opposed to the *SRT* where only the closed set babble and constant noise conditions revealed no significant differences. This finding is to be expected considering that the *SRT*s obtained from the presentation of the UCAMST in noise and in quiet are in different units. Regarding the estimation of slope, however, equivalent pairs naturally grouped into one of two categories; that is, pairs within a condition or pairs across conditions. For the former category, this research found there to be no significant difference between the open set and the closed set for each type of noise (i.e. babble, constant, and quiet) respectively. The latter category for slope can be further divided into three subcategories: closed set pairs, open set pairs, and opposite pairs. For the closed set pairs, the babble noise condition was found to be separately equivalent with the constant noise and the quiet condition. For the open set presentation mode, no significant differences were found for one pair only: the babble noise and the quiet condition. Lastly, the opposite pairs were comprised of the closed set babble noise condition which was equivalent with the open set constant noise and the open set quiet condition separately. Also in this subcategory, no significant differences were revealed between the open set babble noise and the closed set constant noise condition.

In light of these findings, hypothesis two may be considered to be partially true. While the premise that equivalence would be maintained across all conditions was unmet, a total of 10 equivalence pairs were revealed, which constitutes one third of the total number of comparisons possible. Obtaining equivalence within and across conditions is valuable in the clinical and research environment as it allows for the tests to be used interchangeably with comparable accuracy. This can enable practicing audiologists to provide a more targeted approach to assessment by allowing them to present the test in the type of noise most troublesome to a particular listener, thus enhancing the face validity of the evaluation. Another clinical advantage of condition equivalence is the ability to compare differences in performance across tests to help pinpoint any specific listening disabilities an individual may have.

An interesting finding revealed in this study is that of the equivalent pairs within a condition. Indeed, as no significant differences were detectable between the open set and closed set for each condition, it can be surmised that, regardless of presentation mode, lists for each respective condition are equally difficult in terms of the derived slope and therefore can be used interchangeably. This discovery is novel for the UCAMST, as previous investigation by Stone (2016) documented participant's performance in the open set condition to be significantly worse than that in the closed set condition. The discrepancies between the current research's results and that of Stone (2006) may be due to the presentation of non-optimised lists for babble noise and the findings for constant noise concerning list equivalence. Stone's (2016) evaluation of condition equivalence was carried out using the data collected for the analysis of list equivalence for both the babble and constant noise condition. While no significant differences were documented for both the slope and *SRT* of lists in the open set and closed set constant noise condition, a post-hoc analysis revealed there to be insufficient power to identify such a difference, should it exist. Therefore, there may have been enough variation between the lists in the constant noise open set and closed set conditions to cause them to be significantly different.

Literature regarding the equivalence across presentation modes for MSTs internationally is divided. The Spanish MST found significantly higher estimations of *SRT* for the closed set condition as opposed to the open set condition (Hochmuth et al., 2012), whereas the Polish MST found that performance remained equivalent across presentation modes (Ozimek et al., 2010). Equivalence across presentation modes has been theorised to be the result of extensive training, which, in the case of the Polish MST, was carried out over hour long sessions until participant responses became stable (Hochmuth et al., 2012; Ozimek et al., 2010; Stone, 2016). In contrast with the Polish MST, the Spanish MST presented a total of 120 practice sentences to allow for familiarisation of the task using an adaptive procedure and double lists (i.e. 20 sentences; Hochmuth et al., 2012).

As described previously in this study, training for the evaluation of equivalence across conditions in the UCAMST was carried out using an adaptive procedure for a total of 28 sentence trials prior to the presentation of the initial closed set condition. The particular condition presented for the training set was identical to the first closed set condition trial, which was randomly allocated for each participant. Therefore, given that the number of sentences presented for training in the UCAMST was less than a quarter of the number administered in the Spanish MST, it appears that the length of training has less of an impact on the equivalence measured across conditions than previously estimated. However, the effectiveness of the training provided was not measured in this study or any previous research on the UCAMST, and therefore it is possible that familiarisation during the practice phase of the assessment may not have been adequate for stabilising performance on the UCAMST. This issue will be considered further as a limitation of the study.

In addition to the within-condition findings for the open set and closed set presentation mode, this study also contributes valuable insights regarding the equivalent pairs revealed across conditions with respect to the expected estimations of slope. The versatility of the closed set babble noise condition is particularly notable, as it is separately equivalent with the open set and closed set constant noise and quiet conditions respectively. Equivalence amongst the condition combinations stated is largely unexpected, as the types of noise used for masking, or lack thereof, are vastly different in terms of their interaction with the target stimulus (Lidestam, Holgersson, & Moradi, 2014). Indeed, constant noise acts as an energetic masker by occupying the same auditory filters as the desired speech signal, therefore rendering it inaudible (Myerson et al., 2016). In contrast, babble noise functions as an informational masker in which both the sentence stimuli and the masking noise are both audible, creating uncertainty when differentiating between the two. As a result of this, NH listeners are able to make use of the temporal gaps in the masking noise to glimpse the target signal and therefore categorically perform better in this type of noise (Peters et al., 1998; Van Engen & Chandrasekaran, 2012). However, this concept is multifaceted and affected by various factors such as the number of talkers present in the babble noise (Hornsby, Ricketts, & Johnson, 2006; Simpson & Cooke, 2005).

The number of equivalent pairs found across conditions in the current research differs from the results documented by Stone (2016) which reported no significant differences between the derived slope of the closed set constant noise and the open set babble noise condition. Despite this, the trend observed by Stone for the slopes of the constant noise conditions versus the babble noise conditions is also apparent in this study. There was general tendency for steeper slopes to be measured for the open set compared with the closed set condition; however, an exception to this is seen in babble noise, where the open set condition was marginally shallower, 20.18 %/dB, than the closed set condition, 12.17 %/dB. Consistent with other findings described in literature, the constant noise had the steepest derived slopes overall, followed by babble noise, and then the quiet presentation mode (Francart et al., 2011).

The relative slopes of the various conditions of the UCAMST is an important feature for evaluation, as its gradient is a means by which the sensitivity of the test can be measured (Theunissen, Swanepoel, & Hanekom, 2009). A steeper slope is thought to more accurately and efficiently estimate the *SRT*, as a small change in the SNR will result in a large change in the expected performance of a listener (Ozimek et al., 2010). This is advantageous in the clinical environment where time constraints and repeat testing necessitate timely and precise assessment of a listeners SIN understanding. Although the constant noise condition has a higher degree of sensitivity, babble noise as a masker has greater face validity for listeners who struggle to hear speech in an everyday conversational context where multiple speakers are talking at once (Wilson et al., 2007). Therefore, it is sensible that the selection of the most appropriate UCAMST condition to administer for a particular listener be considered on a case-by-case basis and guided by the intention behind delivering the test.

4.3 Comparison across test types

The final hypothesis proposed in this study predicted that there would be no significant correlations between the UCAMST and the meaningful CVC (revised AB) word recognition test, and the UCAMST and the QuickSIN respectively. Comparisons with other speech tests is an important aspect of the validation phase in the development of a new MST and is recommended to help ensure comparability across assessments (Akeroyd et al., 2015). Having analogous tests presents the opportunity for the substitution of one test with another with minimal to no loss of information. This may be advantageous in the clinical environment where time constraints make performing repeated assessments using multiple tests undesirable, if not impracticable.

No significant correlations were found in the current study between the meaningful CVC (revised AB) word recognition test and the UCAMST in the quiet presentation mode with regards to slope. However, there was a significant correlation between the *SRT* of the UCAMST closed set quiet condition and the *SRT* of the meaningful CVC (revised AB) word recognition test. Comparison of the UCAMST and the QuickSIN found no significant correlations between any of the variables included for analysis.

When these results are viewed altogether, the overwhelming lack of significant correlations provides reasonable justification to accept hypothesis three as being almost completely correct in its assumptions bar one significant finding. This particular exception confirms the existence of a moderate positive relationship between the *SRT*s of the UCAMST in the quiet closed set condition and the meaningful CVC (revised AB) word recognition test. This correlation reveals the potential ability of the UCAMST's closed set quiet condition to be able to replace the meaningful CVC (revised AB) word recognition test in routine clinical assessment.

It is interesting that the association between the meaningful CVC (revised AB) word recognition test was evident with the closed set, as opposed to the open set, quiet condition of the UCAMST. This finding may be attributed to the differences in the scoring procedure and target stimuli for the respective tests. The phonemic scoring used by the meaningful CVC (revised AB) word recognition test makes this assessment comparatively more difficult than the UCAMST as each trial had only four potential levels at which points could be awarded (i.e. 0, 3, 7, or 10 points) as opposed to six levels for the UCAMST (i.e. 0, 1, 2, 3, 4 or 5 points). In addition to this, the meaningful CVC (revised AB) word recognition test presents monosyllabic words as the target stimuli in the open set condition, which was anecdotally more challenging than the forced multiple-choice delivery of the closed set condition of the UCAMST.

Although this research determined that for the expected estimations of slope, the lists in the open set and closed set conditions of the UCAMST were equally difficult, the derived *SRT* of the closed set quiet condition was relatively lower than of the open set: 2.27 ± 2.13 dB A and 5.02 ± 1.80 dB A respectively. This suggests that the closed set presentation mode made it easier for participants to identify the words in the sentence correctly and consequently, they performed better on average in this condition, which is consistent with the results reported in the study by Hochmuth et al. (2012). In combination, these findings and the factors described above, suggest that the relative difficulty of the tests described may be a determinant of significant correlations between the meaningful CVC (revised AB) word recognition test and the UCAMST quiet condition.

To the best knowledge of the author, there is presently no research in literature comparing the meaningful CVC (revised AB) word recognition test used in New Zealand with other speech tests, let alone the UCAMST. However, a previous study by

Andre (2016) compared the open set constant noise condition with the QuickSIN and reported a statistically significant correlation between the SNR in which a listener was expected to score 20% of the sentence correctly in the UCAMST and the SNR loss measured by the QuickSIN. As described earlier, the results obtained in current research did not reproduce this finding. Even if this study were to use the standard limit for statistical power of 0.05 like Andre (2016), instead of the less conservative correction factor of 0.01, inspection of the results would still reveal non-significant correlations. The apparent discrepancies between the results obtained for the UCAMST with respect to its relationship with the QuickSIN may be attributed to the variation in the demographics of participants involved and the differences in the experimental procedures implemented.

Indeed, the objective of Andre's research was to investigate the impact of auditory-visual integration on hearing aid benefit using the data obtained from 11 participants with a downward-sloping HI (2016). SIN understanding was assessed adaptively over a total of 15 sentences using the UCAMST with the initial intensity noise level determined using a loudness scaling procedure, and six to eight trials presented prior to the actual test for task training. This is in contrast to the current study, which analysed the data collected from 20 NH participants who undertook the UCAMST adaptively for all six modes, having 28 sentence trials presented in each condition and completed a training set of 28 sentences prior to any closed set testing. Therefore it is conceivable that such differences could have contributed to the lack of congruent findings across the studies. This may be considered a reasonable finding given the inclusion criteria in this research of having NH may have limited the range of results available for comparison due to factors, such as the noise floor, disproportionately influencing the performance of NH listeners. A broader spread of data which can be obtained from participants with varying degrees of HI may enable higher correlations to be revealed similar to those found previously (Andre, 2016).

Andre also proposed that a correlation between the UCAMST and the QuickSIN would likely be replicated in a NH cohort (2016). This assumption was based on the relative proximity of Stone's (2016) reported average SNR at which a 20% correct score was obtained and a calculated estimate of the same; the actual values were recorded as -11.6 dB SNR and -14.5 dB SNR respectively. The results in the current research were notably higher than either of the figures listed, with a mean of -9.76 ± 1.92 dB SNR which may further explain why this study was unable to replicate the significant correlation previously demonstrated between the UCAMST and the QuickSIN.

While it is more prevalent in literature for new MSTs to be compared with already established international MSTs during the validation phase of development, some researchers also acknowledge the importance of investigating the relationship a new MST may have with accepted and routine speech assessments used in clinical practice. The Turkish MST, otherwise known as the TURMatrix (Zokoll et al., 2013), sought to examine this relationship by evaluating the association between the TURMatrix and the Turkish HINT (Cekic & Sennaroglu, 2008) for the constant noise condition and the quiet presentation mode. Findings from this analysis revealed that the MST was able to achieve lower *SRT*s with greater efficiency compared to the HINT test for both conditions considered (Zokoll et al., 2013).

Although a NZHINT test has been developed, it is not available for use clinically, as the uptake of any kind of SIN test by audiologists in NZ has been slow, with most clinicians in favour of assessing speech understanding using monosyllabic words in quiet (Hope, 2010). If SIN testing is carried out, the QuickSIN is most commonly selected for use, due to its ability to capture a listener's individual SNR loss in a time efficient manner, despite the fact that the American English stimuli used may lead to erroneous results for native NZ English speakers (Killion et al., 2004). Due to the mainstream implementation of the meaningful CVC (revised AB) word recognition test and, to a lesser degree, the QuickSIN, it was determined that these measures would be most appropriate for comparison with the UCAMST. However, other SIN tests such as the NZHINT and the NZ dichotic digits test (King, 2011) are valuable tools which may demonstrate greater similarities with the UCAMST, as they were each specifically designed for use in the NZ audiological context. This presents the opportunity for additional validation of the UCAMST to investigate the comparability across these tests, which will be considered in a later section of this study regarding possible areas for future research.

4.4 Summary

One of the purposes of this study was to redress the issues previously encountered during the evaluation of the UCAMST. The results obtained for the analysis of equivalence within and across conditions of the UCAMST were encouraging and largely consistent with other findings in literature. The relationship between the UCAMST and other speech tests routinely used in audiology clinics in NZ was also of interest for the validation of the measure. Investigation of the data to this end revealed predominantly non-significant results contrary to the correlations documented in another study involving the UCAMST (Andre, 2016), with one exception being a correlation between the *SRT*s of the meaningful CVC (revised AB) word recognition test and the UCAMST in the closed set quiet condition.

4.5 Study limitations

Although every effort was made to maintain scientific rigor, shortcomings of the experimental design and stringent time constraints contributed to the emergence of limitations within this study which may have influenced the findings obtained. These will be examined in the following sections to allow for impartial interpretation of the results presented in this research. Subsequent projects should be designed in light of these considerations to avoid the limitations of the current methodology.

4.5.1 The sample

4.5.1.1 Sample size

Although this study was able to obtain the intended number of participants to test each hypothesis respectively, a larger sample size would have benefited this research for three distinct reasons. The first is related to the administrator miscalculation which led to the exclusion of several data points for the investigations concerned with evaluating the UCAMST, and an instance where familiarity with the test necessitated the removal of the total results obtained from one participant for validation analyses. Examination of this data revealed it to be unrepresentative of the rest of the sample and inclusion would erroneously introduce bias into the results; thus justifying its rejection.

A larger sample size would have also been advantageous for the evaluation of normalisation of the lists designed for presentation in babble noise. In this condition, 10 participants were re-tested, as the stringent time period restricting access to the experimental equipment impeded the recruitment of new volunteers. This factor alone is a potential source of bias by contributing to the practice effect and will accordingly be examined further in a later section (Duff et al., 2012).

The validation stage is the third aspect of this study that would have profited from a larger sample size. This is because a broader pool of participant data would allow for in-depth follow-up investigations of significant findings to be carried out, such as a partial correlation or regression analysis (Algina & Olejnik, 2003). This would have been particularly useful for determining the relative contribution of the variables in the unanticipated discovery of the relationship between the meaningful CVC (revised AB) word recognition test and the UCAMST in the closed set quiet condition. The strict time constraints of this study made undertaking additional recruitment and further evaluation of this finding impracticable, however this does present the opportunity for exploration in successive research projects.

4.5.1.2 Generalisability

The generalisability of the sample used for analysis is another potential limitation of this study. Due to the inclusion criteria of having NH and the fact that advertisement for recruitment was predominantly conducted throughout the University of Canterbury community (Christchurch, NZ) the resultant participant demographic of age was relatively narrow. Indeed, the average age of the participants tested in any capacity in the current research was 27.7 years old. Although this did not appear to have a profound impact on the results obtained, it is impossible to predict whether the findings would be generalisable for older adults where working memory may have a more substantial influence (van Rooij & Plomp, 1990). The effect of working memory on the performance of different age groups becomes particularly important in light of the equivalence revealed between the slope of the respective open set and closed set conditions for each of the presentation modes available in the UCAMST. This unaddressed variable and the resultant possible implications on the appropriateness the presentation mode selected for use is an area for future research and will be considered further in a subsequent section.

4.5.2 Methodology

4.5.2.1 Training effect

The undetermined effect of training is another limitation of this study. Internationally, many MSTs have explored the phenomena of the differences in a listener's performance recorded for the first list presented, in contrast with the last list presented (Dietz et al., 2014; Hochmuth et al., 2012; Wagener et al., 2003). The improvement in the scores recorded is thought to be the result of increased familiarity with the required task which develops as the test progresses (Hochmuth et al., 2012). Literature has reported that a training set of 20 to 40 practice sentences may be adequate to stabilise participant performance (Dietz et al., 2014; Hochmuth et al., 2012). The current research presented a total of 20 sentence trials in accordance with this finding and other recommendations (Akeroyd et al., 2015) however the effectiveness of a training set of this size for the UCAMST is yet to be established. Although the impact of this may be minimal when assessing SIN understanding adaptively, insufficient familiarisation of the test when stimuli is presented at fixed SNRs may introduce bias in the findings obtained. Therefore, it is advisable that the training effect specific to the UCAMST be determined to ensure the validity of the measure and progress it towards clinical and research use.

4.5.2.2 Repeat testing

As mentioned previously in this study, a total of 10 participants were re-recruited for the evaluation of normalisation of the babble noise condition. Repeat testing of these listeners increases the likelihood of the practice effect erroneously influencing the findings obtained. Despite taking steps to reduce the potential for this, such as ensuring both presentations of the UCAMST occurred in the same conditions and the randomisation of the lists presented, it is still possible that the results obtained for the second block of testing may be biased by inaccurately better representations of performance for SIN understanding. In addition to this, the time constraints which dictated the availability of the experimental equipment required for testing meant that the interval between exposures to the UCAMST was only three weeks long for repeat listeners. This is unfortunate, as literature has demonstrated a relationship between the length of time between test administrations and the extent of the practice effect measured (Bartels, Wegrzyn, Wiedl, Ackermann, & Ehrenreich, 2010; Benedict & Zgaljardic, 1998).

The Australian MST examined the practice effect for listeners who had been re-recruited from the optimisation phase for participation in the evaluation of normalisation stage of the measure's development and found significant differences between the estimated *SRT*s of experienced and naïve subjects (Kelly et al., 2017). While it would have been interesting to carry out a similar investigation to determine the extent of the practice effect for the UCAMST, the repeated testing in this project was the consequence of an unexpected complication in the experimental procedure and therefore the methodology of this research was not appropriately designed to allow for adequate analysis of these findings with respect to the sample size and scope of the current study. However, this may have important implications for the use of the UCAMST clinically and for research purposes, and is a factor which should be explored further in future research.

4.6 Beyond the current study: Future research directions

The limitations of the current study highlight several unresolved issues which would benefit from additional investigation in order to progress the development of the

UCAMST to a stage where it can be successfully mainstreamed for research and clinical use. Although these persisting areas of interest are beyond the scope of this project, they will be outlined in the following sections to help direct future research on the UCAMST towards the ultimate goal of integration within the audiological test battery in NZ.

4.6.1 Comparing the UCAMST with the NZHINT

The NZHINT (Hope, 2010) is a test which uses NZ English sentences as the target stimuli presented concurrently with constant masking noise. There are multiple similarities in the methodology used for the development of the NZHINT and the UCAMST with regards to the phonemic balancing of lists and the equalisation of sentence intelligibility (Hope, 2010; McClelland, 2014). Like the UCAMST, the NZHINT also employs adaptive procedures to determine a listener's *SRT*. As both tests appear to resemble each other in various aspects, it would be useful to explore the relationship between them to ascertain the possibility of interchangeable use. Cross-validation of the UCAMST with the NZHINT is an important area for future research, as it could produce evidence for decreasing redundancy across SIN measures available in the NZ audiological context.

4.6.2 Examining the application of the auditory-visual mode

The current research explored aspects of the UCAMST in the auditory-alone mode with the rationale that exclusive auditory evaluation would minimise reliance on visual cues for poor SNRs and increase the validity of the findings (Jamaluddin, 2016). The study by Andre (2016) attempted to explore the relationship between the estimated auditory-visual enhancement, which was calculated using the auditory-visual and auditory-alone modes of the UCAMST and various HA outcome measures. However, findings were inconclusive due to low statistical power. As recognised by Andre (2016), examination of the utility of UCAMST for the assessment of auditory-visual integration has important implications for AH/R. Indeed, as auditory-visual integration is a mechanism used by most people during communication (Tye-Murray et al., 2007), measures employing such stimuli are uniquely positioned to evaluate audiological interventions and potentially predict HA benefit. Further investigation in this area would add to the body of literature about the appropriateness of UCAMST for application in AH/R.

4.6.3 Exploring the influence of the practice effect

The practice effect is a feature of cognitive tests which can also influence MSTs (Duff et al., 2012). Indeed, the improvement in performance for experienced, as opposed to naïve participants during the development of the Australian English MST was attributed to practice effects (Kelly et al., 2017). Unfortunately, similar investigations specific to the UCAMST were unable to be fulfilled as such analyses were beyond the scope of this study as described previously. The degree to which the practice effect influences the findings of a particular measure is of concern in an audiological context, as the sometimes fluctuating and progressive nature of HI often requires repeated assessments which may occur over a short period of time in some cases. Therefore, evaluating the extent of the practice effect for the UCAMST over different intervals of time will affect the confidence with which the test results obtained can be interpreted and is thus an important area for future research.

4.6.4 Piloting with diverse demographics

To date, participants involved in the experimental procedures for the development of the UCAMST have predominantly been young NH adults, with the exception of Andre's work involving the UCAMST (2016). While this relatively homologous sample was adequate for the construction of the UCAMST, listeners who do not fall within this demographic—with respect to age and hearing status—may perform differently. This is an area which needs further exploration, as outlined below.

4.6.4.1 Individuals in different age groups

Research has shown that the ability to understand SIN is age dependant to a certain extent (Humes, 2015). The observed changes in working memory which occur throughout adulthood has been found to be an important determinant of performance in some SIN assessments (Arehart, Souza, Baca, & Kates, 2013; Arlinger, Lunner, Lyxell, & Pichora-Fuller, 2009; Foo, Rudner, Ronnberg, & Lunner, 2007; Rudner, Lunner, Behrens, Thorén, & Rönnberg, 2012). One aspect of Andre's research on the UCAMST (2016) sought to investigate, and if necessary control for, the impact of working memory. However, low statistical power meant no significant findings were able to be obtained in that study. Going forward, the impact of working memory within the UCAMST is a particular area of interest, given the equivalent findings revealed in the current project. Although interchangeable use of the open set and closed set conditions may be indicated for listeners similar to the NH young adult participants involved in this study, the same may not be true for older adults. For these clients, cognitive factors may result in their performance being measured inaccurately as poorer in the closed set condition—as opposed to the open set condition—due to the greater intellectual burden associated with self-scoring. As the majority of people with a HI requiring audiological services are over the age of 65 years, the need for accurate and efficient measures of speech understanding for this age group is paramount (Newman & Sandridge, 2004). Therefore, further investigation on the UCAMST should seek to quantify the impact of working memory on performance to ensure the reliability of the measure for use with clients of different ages.

Future research on the UCAMST for the purpose of optimising its utility for different age groups should not be limited to older adults alone. Indeed, the features of the UCAMST which make it clinically advantageous for adults may also make it suitable for adaptation for use in paediatric audiology (Ozimek, Kutzner, & Libiszewski, 2012). A concurrent study by Foreman (in progress) sought to develop a MST suitable for the assessment of NZ paediatric population. This was evaluated alongside the parent UCAMST to investigate the equivalence across sentence lists and conditions for both the visual and auditory-visual mode. The work by Foreman (in progress) is of particular interest, as it will establish the knowledge base concerning the utility of the UCAMST for the paediatric population and potentially extend the clinical applicability of the measure.

4.6.4.2 Individuals with a HI

As a SIN assessment tool, the aim of the UCAMST is to be able to evaluate the speech understanding of clients with varying degrees of HI. Literature has demonstrated the increased difficulty that listeners with a HI experience when tasked with recognising speech in BG noise (Peters et al., 1998; Wilson et al., 2007). The difference in performance is thought to be attributable to the type of masking noise used, with greater distinctions between the performance of listeners with NH and HI revealed for babble noise compared with other types of noise (Peters et al., 1998; Wilson et al., 2007). To date, the majority of the research carried out for the purpose of developing the UCAMST involved participants with NH. The final phase of the procedure for the generation of new MSTs recommended by Akeroyd et al. (2015) was completed in this study with the validation of the UCAMST. Therefore, the logical next step to prepare the UCAMST for clinical and research use is to collect normative data for the measure in each mode for listeners with differing degrees of HI. This would allow for the development of an index against which a given client’s performance can be compared to obtain a relative measure of the listener’s ability to understand speech in the presence of BG noise.

4.7 Concluding remarks

This research is part of a larger series with the shared objective to design, create and develop a MST suitable for audiological use in NZ. The current project sought to re-evaluate the equivalence of the lists and conditions available in the UCAMST in order to redress the limitations encountered in a previous study. The results from this research found equal difficulty for the lists designed for use in the babble condition, regardless of presentation mode. However, the same was not found to be true for the quiet condition. Subsequent analysis of the six different conditions available in the UCAMST revealed equal difficulty, and thus the potential for interchangeable use for multiple pairs of conditions with regards to the expected estimation of the slope variable. Cross-validation of the UCAMST with the meaningful CVC (revised AB) word recognition test and the QuickSIN was another aim of this study, and analyses to this end revealed the potential for the UCAMST to replace the meaningful CVC (revised AB) word recognition test in practice.

The findings of this study present valuable insights about speech understanding in BG noise and provide evidence for the appropriateness of the UCAMST for use within the NZ audiological context. This is a welcome revelation, as the validity and sensitivity of MSTs make them a useful tool for the assessment of speech understanding in noise. Furthermore, the configurability of the UCAMST is advantageous in both the clinical or research context, as the method in which the test is delivered can be customised as appropriate for a given listener. It is hoped that subsequent research on this topic will involve the collection of normative data for the UCAMST to further progress the measure towards the ultimate goal of nationwide adoption and integration within the NZ audiological test battery.

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*(2), 17-22. doi: 10.3109/14992027.2015.1030513

Algina, J., & Olejnik, S. (2003). Sample size tables for correlation analysis with applications in partial correlation and multiple regression analysis. *Multivariate Behavioral Research, 38*(3), 309-323. doi: 10.1207/S15327906MBR3803_02

Andre, E. R. (2016). Is auditory-visual integration a factor in hearing aid outcomes? a thesis submitted in partial fulfilment of the requirements for the degree of Master of Audiology at the University of Canterbury: University of Canterbury.

Arehart, K. H., Souza, P., Baca, R., & Kates, J. M. (2013). Working memory, age, and hearing loss: Susceptibility to hearing aid distortion. *Ear and Hearing, 34*(3), 251-260. doi: 10.1097/AUD.0b013e318271aa5e

Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, M. K. (2009). The emergence of cognitive hearing science. *Scandinavian Journal of Psychology, 50*(5), 371-384. doi: 10.1111/j.1467-9450.2009.00753.x

Armero, O. E. (2001). Effects of denied hearing loss on the significant other. *The Hearing Journal, 54*(5), 44-47. doi: 10.1097/01.HJ.0000294841.86637.5d

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), 549-563. doi: 10.1044/jslhr.4103.549

Bagley, W. C. (1900). The apperception of the spoken sentence: A study in the psychology of language. *The American Journal of Psychology, 12*(1), 80-130.

Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science, 60*(2), 434-448. doi: 10.1287/mnsc.2013.1776

Bamiou, D., Werring, D., Cox, K., Stevens, J., Musiek, F. E., Brown, M. M., & Luxon, L. M. (2012). Patient-reported auditory functions after stroke of the central auditory pathway. *Stroke; A Journal of Cerebral Circulation, 43*(5), 1285-1289. doi: 10.1161/STROKEAHA.111.644039

Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: A longitudinal study on frequent repetitive cognitive testing. *BMC neuroscience, 11*(1), 118-118. doi: 10.1186/1471-2202-11-118

Basso, M. R., Carona, F. D., Lowery, N., & Axelrod, B. N. (2002). Practice effects on the WAIS-III across 3- and 6-month intervals. *The Clinical Neuropsychologist, 16*(1), 57-63. doi: 10.1076/clin.16.1.57.8329

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*(3), 153-164. doi: 10.3109/03005364000000018

Beattie, R. C., & Raffin, M. J. M. (1985). Reliability of threshold, slope, and PB max for monosyllabic words. *Journal of Speech and Hearing Disorders, 50*(2), 166-178.

Beattie, R. C., & Warren, V. G. (1982). Relationships among speech threshold, loudness discomfort, comfortable loudness, and PB max in the elderly hearing impaired. *The American journal of otology, 3*(4), 353-358.

Beattie, R. C., & Zipp, J. A. (1990). Range of intensities yielding PB max and the threshold for monosyllabic words for hearing-impaired subjects. *Journal of Speech and Hearing Disorders, 55*(3), 417-426.

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

Benedict, R. H. B., & Zgaljardic, D. J. (1998). Practice effects during repeated administrations of memory tests with and without alternate forms. *Journal of Clinical and Experimental Neuropsychology, 20*(3), 339-352. doi: 10.1076/jcen.20.3.339.822

Bertoli, S., Bodmer, D., & Probst, R. (2010). Survey on hearing aid outcome in Switzerland: Associations with type of fitting (bilateral/unilateral), level of hearing aid signal processing, and hearing loss. *International Journal of Audiology, 49*(5), 333-346. doi: 10.3109/14992020903473431

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. doi: 10.1080/09297040500488522

Bhutta, M. F., Williamson, I. G., & Sudhoff, H. H. (2011). Cholesteatoma. *British Medical Journal, 342*(1), 1088-1088. doi: 10.1136/bmj.d1088

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*(1), 32-48.

Boothroyd, A. (1968). Developments in speech audiometry. *International Audiology, 7*(3), 368-368. doi: 10.3109/05384916809074343

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

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

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

Brännström, K. J., Lantz, J., Nielsen, L. H., & Olsen, S. Ø. (2014). Prediction of IOI-HA scores using speech reception thresholds and speech discrimination scores in quiet. *Journal of the American Academy of Audiology, 25*(2), 154-163. doi: 10.3766/jaaa.25.2.4

Butler, C. C., & Williams, R. G. (2003). The etiology, pathophysiology, and management of otitis media with effusion. *Current Infectious Disease Reports, 5*(3), 205-212. doi: 10.1007/s11908-003-0075-x

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: Meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist, 26*(4), 543-570. doi: 10.1080/13854046.2012.680913

Caponetti, G., Thompson, L. D., & Pantanowitz, L. (2009). Cholesteatoma. *Ear, Nose, & Troat Journal, 88*(11), 1196.

Carhart, R. (1951). Basic principles of speech audiometry. *Acta Oto-Laryngologica, 40*(1-2), 62.

Carhart, R. (1965). Monaural and binaural discrimination against competing sentences. *The Journal of the Acoustical Society of America, 37*(6), 1205. doi: 10.1121/1.1939552

Carhart, R., & Jerger, J. (1959). Preferred method for clinical determination of pure-tone thresholds. *Journal of Speech and Hearing Disorders, 24*(4), 330. doi: 10.1044/jshd.2404.330

Carhart, R., & Tillman, T. W. (1970). Interaction of competing speech signals with hearing losses. *Archives of Otolaryngology, 91*(3), 273.

Carhart, R., & Young, L. (1976). Development of test procedures for evaluation of binaural hearing aids. A final report. *Bulletin of Prosthetics Research*, 9.

Carney, A. E., & Nelson, D. A. (1983). An analysis of psychophysical tuning curves in normal and pathological ears. *The Journal of the Acoustical Society of America, 73*(1), 268-278. doi: 10.1121/1.388860

Cekic, S., & Sennaroglu, G. (2008). The Turkish Hearing in Noise Test. *International Journal of Audiology, 47*(6), 366-368. doi: 10.1080/14992020802056027

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. doi: 10.1037/a0014321

Chen, D. S., Betz, J., Yaffe, K., Ayonayon, H. N., Kritchevsky, S., Martin, K. R., . . . Lin, F. R. (2015). Association of hearing impairment with declines in physical functioning and the risk of disability in older adults. *Journals of Gerontology: Medical Sciences, 70*(5), 654-661. doi: 10.1093/gerona/glu207

Chen, D. S., Genther, D. J., Betz, J., & Lin, F. R. (2014). Association between hearing impairment and self- reported difficulty in physical functioning. *Journal of the American Geriatrics Society, 62*(5), 850-856. doi: 10.1111/jgs.12800

Ching, T. Y. C., Dillon, H., & Byrne, D. (1998). Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification. *The Journal of the Acoustical Society of America, 103*(2), 1128-1140. doi: 10.1121/1.421224

Chisolm, T., Johnson, C., Danhauer, J., Portz, L., Abrams, H., Lesner, S., . . . Newman, C. (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

Clarke, J. G. (1981). Uses and abuses of hearing loss classification. *American Speech and Hearing Association, 23*, 493-500.

Contrera, K. J., Betz, J., Deal, J., Choi, J. S., Ayonayon, H. N., Harris, T., . . . Lin, F. R. (2017). Association of hearing impairment and anxiety in older adults. *Journal of Aging and Health, 29*(1), 172-184. doi: 10.1177/0898264316634571

Cord, M. T., Leek, M. R., & Walden, B. E. (2000). Speech recognition ability in noise and its relationship to perceived hearing aid benefit. *Journal of the American Academy of Audiology, 11*(9), 475-483.

Cousins, V. C., & Prasad, J. (2008). Asymmetrical hearing loss. *Australian Family Physician, 37*(5), 312-320.

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

Cruickshanks, K., & Wichmann, M. (2012). Hearing impairment and other health conditions in older adults: Chance associations or opportunities for prevention? *Seminars in Hearing, 33*(3), 217-224. doi: 10.1055/s-0032-1315720

Dietz, A., Buschermöhle, M., Aarnisalo, A., Vanhanen, 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*(7), 728-737. doi: 10.3109/00016489.2014.898185

Dillon, H. (2012). *Hearing aids* (2nd ed.). Sydney, Australia: Boomerang Press.

Dirks, D. D., Kamm, C., Bower, D., & Betsworth, A. (1977). Use of performance-intensity functions for diagnosis. *Journal of Speech and Hearing Disorders, 42*(3), 408-415.

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*(2), 114.

Duff, K., Callister, C., Dennett, K., & Tometich, D. (2012). Practice effects: A unique cognitive variable. *The Clinical Neuropsychologist, 26*(7), 1117-1127. doi: 10.1080/13854046.2012.722685

Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003). Note on informational masking. *The Journal of the Acoustical Society of America, 113*(6), 2984-2987. doi: 10.1121/1.1570435

Egan, J. P. (1948). Articulation testing methods. *The Laryngoscope, 58*(9), 955.

Etymotic Research Incorporation. (2006). *QuickSIN speech-in-noise test*. Retrieved from https://www.etymotic.com/downloads/dl/file/id/259/product/159/quicksin_user_manual.pdf

Farahmand, R. B., Merchant, G. R., Lookabaugh, S. A., Röösli, C., Ulku, C. H., McKenna, M. J., . . . Nakajima, H. H. (2016). The audiometric and mechanical effects of partial ossicular discontinuity. *Ear and Hearing, 37*(2), 206-215. doi: 10.1097/AUD.0000000000000239

Feldmann, H. (1997). History of the tuning fork II: Evolution of the classical experiments by Weber, Rinne and Schwabach. *Laryngo-Rhino-Otologie, 76*(5), 318.

Fettiplace, R., & Kim, K. X. (2014). The physiology of mechanoelectrical transduction channels in hearing. *Physiological reviews, 94*(3), 951.

Fletcher, H., & Steinberg, J. C. (1929). Articulation testing methods. *The Bell System Technical Journal, 8*(4), 806-854. doi: 10.1002/j.1538-7305.1929.tb01246.x

Foo, C., Rudner, M., Ronnberg, J., & Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *Journal of the American Academy of Audiology, 18*(7), 618-631. doi: 10.3766/jaaa.18.7.8

Foreman, P. (in progress). *Development and preliminary evaluation of the University of Canterbury Paediatric Auditory-Visual Matrix Sentence Test* (Master’s Thesis), University of Canterbury, Christchurch, New Zealand.

Foulon, I. (2008). A 10-year prospective study of sensorineural hearing loss in children with congenital cytomegalovirus infection. *Journal of Pediatrics, 153*(1), 84-88. doi: 10.1016/j.jpeds.2007.12.049

Francart, T., van Wieringen, A., & Wouters, J. (2011). Comparison of fluctuating maskers for speech recognition tests. *International Journal of Audiology, 50*(1), 2-13. doi: 10.3109/14992027.2010.505582

French, N. R., & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America, 19*(1), 90-119. doi: 10.1121/1.1916407

Frisina, R. D. (1997). Speech recognition in noise and presbycusis: Relations to possible neural mechanisms. *Hearing Research, 106*(1-2), 95-104. doi: 10.1016/S0378-5955(97)00006-3

Fritze, T., Teipel, S., Óvári, A., Kilimann, I., Witt, G., & Doblhammer, G. (2016). Hearing impairment affects dementia incidence. An analysis based on longitudinal health claims data in Germany. *Public Library of Science, 11*(7), e0156876. doi: 10.1371/journal.pone.0156876

Garramiola-Bilbao, I., & Rodríguez-Álvarez, A. (2016). Linking hearing impairment, employment and education. *Public Health, 141*, 130-135. doi: 10.1016/j.puhe.2016.09.013

Gates, G. A., & Mills, J. H. (2005). Presbycusis. *The Lancet, 366*(9491), 1111-1120. doi: 10.1016/S0140-6736(05)67423-5

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

Gimsing, S. (2010). Vestibular schwannoma: When to look for it? *The Journal of Laryngology and Otology, 124*(3), 258-264. doi: 10.1017/S0022215109991423

Gispen, F. E., Chen, D. S., Genther, D. J., & Lin, F. R. (2014). Association between hearing impairment and lower levels of physical activity in older adults. *Journal of the American Geriatrics Society, 62*(8), 1427-1433. doi: 10.1111/jgs.12938

Gordon-Salant, S., Yeni-Komshian, G. H., & Fitzgibbons, P. J. (2010). Recognition of accented English in quiet and noise by younger and older listeners. *Journal of the Acoustical Society of America, 128*(5), 3152-3160. doi: 10.1121/1.3495940

Gordon, E. (2004). *New Zealand English: Its origins and evolution*. Cambridge, UK: Cambridge University Press.

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. *Journal of the Acoustical Society of America, 103*(5), 2677-2690. doi: 10.1121/1.422788

Groen, J. J. (1979). Social hearing handicap: Its measured by speech audiometry in noise. *Internaltional Journal of Audiology, 8*, 182-183.

Grothe, B., Pecka, M., & McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiological reviews, 90*(3), 983-1012. doi: 10.1152/physrev.00026.2009

Grunditz, M., & Magnusson, L. (2013). Validation of a speech-in-noise test used for verification of hearing aid fitting. *Hearing, Balance and Communication, 11*(2), 64-71. doi: 10.3109/21695717.2013.782135

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

Hall, J. (1983). Diagnostic applications of speech audiometry. *Seminars in Hearing, 4*(03), 179-203. doi: 10.1055/s-0028-1091425

Hall, S. (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.

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

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*(7), 536-544. doi: 10.3109/14992027.2012.670731

Hogan, A., O'Loughlin, K., Davis, A., & Kendig, H. (2009). Hearing loss and paid employment: Australian population survey findings. *International Journal of Audiology, 48*(3), 117-122. doi: 10.1080/14992020802449008

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.

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. doi: 10.1121/1.3037233

Hoppe, U., Hast, A., & Hocke, T. (2014). Speech perception with hearing aids in comparison to pure-tone hearing loss. *HNO, 62*(6), 443.

Hornsby, B. W. Y., Ricketts, T. A., & Johnson, E. E. (2006). The effects of speech and speechlike maskers on unaided and aided speech recognition in persons with hearing loss. *Journal of the American Academy of Audiology, 17*(6), 432-447. doi: http://dx.doi.org/10.3766/jaaa.17.6.5

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

Howard-Jones, P. A., & Rosen, S. (1993). The perception of speech in fluctuating noise. *Acustica, 78*(5), 258-272.

Humes, L. E. (2015). Age-related hearing loss, noise-induced hearing loss, and speech-understanding performance. *Journal of the Acoustical Society of America, 137*(4), 2248-2248. doi: 10.1121/1.4920200

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *Neuroimage, 50*(3), 1202-1211. doi: 10.1016/j.neuroimage.2010.01.046

Hsieh, D. L., Lin, K. L., Ho, J. H., & Liu, T. C. (2009). Hearing in noise test in subjects with conductive hearing loss. *Journal of the Formosan Medical Association, 108*(12), 937-942. doi: 10.1016/S0929-6646(10)60006-X

Jamaluddin, S. A. (2016). *Development and evaluation of the digit triplet and auditory-visual matrix sentence tests in Malay* (Doctoral Thesis). University of Canterbury, Christchurch, New Zealand. Retrieved from https://ir.canterbury.ac.nz/handle/10092/12288

James, A. L., Papsin, B. C., Trimble, K., Ramsden, J., Sanjeevan, N., Bailie, N., & Chadha, N. K. (2012). Tympanic membrane retraction: An endoscopic evaluation of staging systems. *The Laryngoscope, 122*(5), 1115-1120. doi: 10.1002/lary.23203

Jansen, S., Luts, H., Wagener, K., Kollmeier, B., Rio, M., Dauman, R., . . . 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. doi: 10.3109/14992027.2011.633568

Järvelin, M., Mäki-torkko, E., Sorri, M. J., & Rantakallio, P. T. (1997). Effect of hearing impairment on educational outcomes and employment up to the age of

25 years in Northern Finland. *British Journal of Audiology, 31*(3), 165-175. doi: 10.3109/03005364000000019

Jenstad, L., & Moon, J. (2011). Systematic review of barriers and facilitators to hearing aid uptake in older adults. *Audiology Research, 1*(1S), e25-e25. doi: 10.4081/audiores.2011.e25

Jerger, J., & Hayes, D. (1976a). The cross-check principle in pediatric audiometry. *Archives of Otolaryngology, 102*(10), 614.

Jerger, J., & Hayes, D. (1976b). Hearing aid evaluation - clinical experience with a new philosophy. *Archives of Otolaryngology, 102*, 214-225.

Jiang, Z. D. (2007). Changes in brainstem auditory evoked response latencies in term neonates with hyperbilirubinemia. *Pediatric Neurology, 37*(1), 35-41. doi: 10.1016/j.pediatrneurol.2007.03.006

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*(5), 1337-1351. doi: 10.1121/1.381436

Kamat, P. K., Kalani, A., Metreveli, N., Tyagi, S. C., & Tyagi, N. (2015). A possible molecular mechanism of hearing loss during cerebral ischemia in mice. *Canadian Journal of Physiology and Pharmacology, 93*(7), 505. doi: 10.1139/cjpp-2014-0489

Karimi, L. J., Ashrafi, M., Khosravi, E., Shahidipour, Z., & Vafaee, F. (2008). Hearing screening in the elderly and evaluating the need for hearing aid. *Audiology, 16*(2), 38-46.

Kashio, A., Ito, K., Kakigi, A., Karino, S., Iwasaki, S.-I., Sakamoto, T., . . . Yamasoba, T. (2011). Carhart notch 2-kHz bone conduction threshold dip: a nondefinitive predictor of stapes fixation in conductive hearing loss with normal tympanic membrane. *Archives of Otolaryngology - Head & Neck Surgery, 137*(3), 236.

Katz, J. (2015). A brief introduction to clinical audiology and this handbook. In J. Katz (Eds.), *Handbook of clinical audiology* (7th ed., pp. 3-7). Philadelphia, PA: Lippincott Williams and Wilkins.

Kelly-Campbell, R., & Lessoway, K. (2015). Hearing aid and hearing assistance technology use in Aotearoa/New Zealand. *International Journal of Audiology, 54*(5), 308-315. doi: 10.3109/14992027.2014.979952

Kelly, H., Lin, G., Sankaran, N., Xia, J., Kalluri, S., & Carlile, S. (2017). Development and evaluation of a mixed gender, multi-talker matrix sentence test in Australian English. *International Journal of Audiology, 56*(2), 85-91. doi: 10.1080/14992027.2016.1236415

Killion, M., Niquette, P., Gudmundsen, G., Revit, L., & 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 I), 2395-2405. doi: 10.1121/1.1784440

Kim, S. Y., Kim, H., Park, E., Joe, J., Sim, S., & Choi, H. G. (2017). Severe hearing impairment and risk of depression: A national cohort study. *PLoS One, 12*(6). doi: 10.1371/journal.pone.0179973

King, S. M. (2011). *Development and evaluation of a New Zealand Digit Triplet Test for auditory screening* (Master's Thesis). University of Canterbury, Christchurch, New Zealand. Retrieved from https://ir.canterbury.ac.nz/handle/10092/5679

Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics, 63*(8), 1421-1455. doi: 10.3758/BF03194552

Kluk, K., & Moore, B. C. J. (2004). Factors affecting psychophysical tuning curves for normally hearing subjects. *Hearing Research, 194*(1), 118-134. doi: 10.1016/j.heares.2004.04.012

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

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

Kreisman, B. M., Smart, J. L., & John, A. B. (2015). Diagnostic audiology. In J. Katz (Eds.), *Handbook of clinical audiology* (7th ed., pp. 119-133). Philadelphia, PA: Lippincott Williams and Wilkins.

Kumar, R., Seshaprasad, & Ahmed, M. (2016). Audiological evaluation in patients of CSOM with cholesteatoma. *Journal of Evidence Based Medicine and Healthcare, 3*(30), 1344-1347. doi: 10.18410/jebmh/2016/307

Lehiste, I., & Peterson, G. E. (1959). Linguistic considerations in the study of speech intelligibility. *Journal of the Acoustical Society of America, 31*(3), 280. doi: 10.1121/1.1907713

Lessoway, K. (2014). *Perception of quality of life for adults with hearing impairment in Aotearoa/New Zealand* (Master's Thesis). University of Canterbury, Christchurch, New Zealand. Retrieved from https://ir.canterbury.ac.nz/handle/10092/9599

Lidestam, B., Holgersson, J., & Moradi, S. (2014). Comparison of informational vs. energetic masking effects on speechreading performance. *Frontiers in Psychology, 5*(639), 639. doi: 10.3389/fpsyg.2014.00639

Loven, F. C., & Hawkins, D. B. (1983). Interlist equivalency of the CID W-22 word lists presented in quiet and in noise. *Ear and Hearing, 4*(2), 91-97. doi: 10.1097/00003446-198303000-00005

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. doi: 10.1017/S0954394507070020

MacPherson, A., & Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends in hearing, 18,* 233121651453772. doi: 10.1177/2331216514537722

Mansour, S., Magnan, J., Haidar, H., & Nicolas, K. (2015). *Tympanic membrane retraction pocket: Overview and advances in diagnosis and management*. New York, NY: Springer.

Mason, M. J. (2016). Structure and function of the mammalian middle ear: Inferring function from structure. *Journal of Anatomy, 228*(2), 300-312. doi: 10.1111/joa.12316

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, 2009*(1), 1-12. doi: 10.1155/2009/169819

McArdle, R. A., & Hnath-Christolm, T. (2015). Speech audiometry. In J. Katz (Eds.), *Handbook of clinical audiology* (7th ed., pp. 61-76). Philadelphia, PA: Lippincott Williams and 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. doi: 10.3766/jaaa.16.9.9

McClelland, A. (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 https://ir.canterbury.ac.nz/handle/10092/10456

Mendel, L. L. (2007). Objective and subjective hearing aid assessment outcomes. *American Journal of Audiology, 16*(2), 118-129. doi: 10.1044/1059-0889(2007/016)

Mick, P., Kawachi, I., & Lin, F. R. (2014). The association between hearing loss and social isolation in older adults. *Otolaryngology–Head and Neck Surgery, 150*(3), 378-384. doi: 10.1177/0194599813518021

Mikkola, T. M., Portegijs, E., Rantakokko, M., Gagné, J., Rantanen, T., & Viljanen, A. (2015). Association of self-reported hearing difficulty to objective and perceived participation outside the home in older community-dwelling adults. *Journal of Aging and Health, 27*(1), 103-122. doi: 10.1177/0898264314538662

Møller, A. R. (2013). *Hearing: Anatomy, physiology, and disorders of the auditory system* (3rd ed.). San Diego, CA: Plural Pub.

Moore, B. C. J. (2001). Dead regions in the cochlea: Diagnosis, perceptual consequences, and implications for the fitting of hearing aids. *Trends in Amplification, 5*(1), 1-34. doi: 10.1177/108471380100500102

Moore, B. C. J. (2012). *An introduction to the psychology of hearing* (6th ed.). Bingley, England: Emerald Group Publishing.

Morales-Garcia, C., & Poole, J. P. (1972). Masked speech audiometry in central deafness. *Acta Oto-Laryngologica, 74*(5), 307.

Musiek, F. E., & Baran, J. A. (2007). *The auditory system: Anatomy, physiology and clinical correlates*. Boston: Pearson.

Myerson, J., Spehar, B., Tye-Murray, N., Van Engen, K., Hale, S., & Sommers, M. S. (2016). Cross-modal informational masking of lipreading by babble. *Attention, Perception, & Psychophysics, 78*(1), 346-354. doi: 10.3758/s13414-015-0990-6

Neff, B. A., Welling, D. B., Akhmametyeva, E., & Chang, L. S. (2006). The molecular biology of vestibular schwannomas: Dissecting the pathogenic process at the molecular level. *Otology & Neurotology, 27*(2), 197-208. doi: 10.1097/01.mao.0000180484.24242.54

Newman, C. W., & Sandridge, S. A. (2004). Hearing loss is often undiscovered, but screeing is easy. *Cleveland Clinic Journal of Medicine, 71*(3), 225-232. doi: http://dx.doi.org/10.3949/ccjm.71.3.22

Ni, G., Elliott, S. J., Ayat, M., & Teal, P. D. (2014). Modelling cochlear mechanics. *BioMed Research International, 2014,* 150637. doi: 10.1155/2014/150637

Nia, J., & Bance, M. (2001). Effects of varying unilateral conductive hearing losses on speech-in-noise discrimination: An experimental study with implications for surgical correction. *Otology and Neurotology, 22*(6), 737-744. doi: 10.1097/00129492-200111000-00005

Niemeyer, W. (1976). Speech audiometry and fitting of hearing aids in noise. *Audiology, 15*(5), 421-427.

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. doi: 10.1121/1.408469

Nowotny, M., & Gummer, A. W. (2006). Electromechanical transduction: Influence of the outer hair cells on the motion of the organ of Corti. *HNO, 54*(7), 536.

New Zealand Audiological Society. (2008). *Standard audiological assessment*. Retrieved from https://www.audiology.org.nz/userfiles/file/pdf/Std%20Audio%20Assment.pdf

New Zealand Audiological Society. (2009). *Scope of practice for audiologists*. Retrieved from https://www.audiology.org.nz/userfiles/file/pdf/Scope%20Practice%20Audios.pdf

New Zealand Audiological Society. (2014). *Code of ethics*. Retrieved from https://www.audiology.org.nz/Userfiles/file/Code%20of%20Ethics%2011%20December%202014%20(1).pdf

New Zealand Audiological Society. (2016a). *NZAS best practice guidelines speech audiometry in diagnostic hearing assessment for young people and adults*. Retrieved from https://www.audiology.org.nz/mnzas-standards-of-practice.aspx

New Zealand Audiological Society. (2016b). *NZAS best practice guilelines adult pure tone audiometry*. Retrieved from https://www.audiology.org.nz/mnzas-standards-of-practice.aspx

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). doi: 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.

Ohene-Djan, J., Hersh, M., & Naqvi, S. (2010). Road safety and deaf people: The role of the police. *Journal of Prevention & Intervention in the Community, 38*(4), 316-331. doi: 10.1080/10852352.2010.509023

Olsen, W. O., & Carhart, R. (1967). Development of test procedures for evaluation of binaural hearing aids. *Bulletin of Prosthetics Research, 10*(7), 22-49.

Ostler, D. A., & Crandell, C. C. (2001). The effects of symmetrical and asymmetrical sensorineural hearing loss on speech perception in noise. *The Journal of the Acoustical Society of America, 109*(5), 2503-2503. 10.1121/1.4744918

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

Ozimek, E., Kutzner, D., Sek, A., & Wicher, A. (2009). Polish sentence tests for measuring the intelligibility of speech in interfering noise. *International Journal of Audiology, 48*(7), 433-443. doi: 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*(6), 444-454. doi: 10.3109/14992021003681030

Park, H., Hong, S. N., Kim, H. S., Han, J. J., Chung, J., Suh, M., . . . Lee, J. H. (2015). Determinants of conductive hearing loss in tympanic membrane perforation.

Clinical and Experimental Otorhinolaryngology, 8(2), 92-96. doi: 10.3342/ceo.2015.8.2.92

Peeva, M. G., Guenther, F. H., Tourville, J. A., Nieto-Castanon, A., Anton, J., Nazarian, B., & Alario, F. X. (2010). Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage, 50*(2), 626-638. doi: 10.1016/j.neuroimage.2009.12.065

Perny, M., Roccio, M., Grandgirard, D., Solyga, M., Senn, P., & Leib, S. L. (2016). The severity of infection determines the localization of damage and extent of sensorineural hearing loss in experimental pneumococcal meningitis. *The Journal of neuroscience, 36*(29), 7740-7749. doi: 10.1523/JNEUROSCI.0554-16.2016

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. *Journal of the Acoustical Society of America, 103*(1), 577-587. doi: 10.1121/1.421128

Picard, M., Girard, S. A., Courteau, M., Leroux, T., Larocque, R., Turcotte, F., . . . Simard, M. (2008). Could driving safety be compromised by noise exposure at work and noise-induced hearing loss? *Traffic Injury Prevention, 9*(5), 489-499. doi: 10.1080/15389580802271478

Pickles, J. O. (2012). *An introduction to the physiology of hearing* (4th ed.). Bingley, England: Emerald Group Publishing.

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

Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology, 18*(1), 43.

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.

Quar, T. K., Soli, S. D., Chan, Y. F., Ishak, W. S., & Abdul Wahat, N. H. (2017). The effect of non-native and non-regional speech testing on a multi-lingual population. *International Journal of Audiology, 56*(2), 92-98. doi: 10.1080/14992027.2016.1210828

Randolph, L. J., & Schow, R. L. (1983). Threshold inaccuracies in an elderly clinical population: Ear canal collapse as a possible cause. *Journal of Speech and Hearing Research, 26*(1), 54-58.

Ravicz, M. E., Rosowski, J. J., & Merchant, S. N. (2004). Mechanisms of hearing loss resulting from middle-ear fluid. *Hearing Research, 195*(1), 103-130. doi: 10.1016/j.heares.2004.05.010

Ristovska, L., Jachova, Z., Filipovski, R., & Atanasova, N. (2016). Correlation between tympanic membrane perforation and hearing loss. *The Journal of Special Education and Rehabilitation, 17*(1/2), 36-49. doi: 10.19057/jser.2016.2

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

Sajjadi, H., & Paparella, M. M. (2008). Meniere's disease. *The Lancet, 372*(9636), 406-414. doi: 10.1016/S0140-6736(08)61161-7

Sauvaget, E., Kici, S., Kania, R., Herman, P., & Tran Ba Huy, P. (2005). Sudden sensorineural hearing loss as a revealing symptom of vestibular schwannoma. *Acta Oto-Laryngologica, 125*(6), 592-595. doi: 10.1080/00016480510030246

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

Scarinci, N., Worrall, L., & Hickson, L. (2009). The ICF and third-party disability: Its application to spouses of older people with hearing impairment. *Disability & Rehabilitation, 31*(25), 2088-2100. doi: 10.3109/09638280902927028

Schlauch, R. S., & Nelson, P. (2015).Puretone evaluation. In J. Katz (Eds.), *Handbook of clinical audiology* (7th ed., pp. 29-47). Philadelphia, PA: Lippincott Williams and Wilkins.

Sente, M. (2004). The history of audiology. *Medicinski Pregled, 57*(11-12), 611.

Shapiro, S. M., & Popelka, G. R. (2011). Auditory impairment in infants at risk for bilirubin-induced neurologic dysfunction. *Seminars in Perinatology, 35*(3), 162-170. doi: 10.1053/j.semperi.2011.02.011

Shera, C. A. (2015). The spiral staircase: Tonotopic microstructure and cochlear tuning. *The Journal of Neuroscience, 35*(11), 4683-4690. doi: 10.1523/JNEUROSCI.4788-14.2015

Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *Journal of the Acoustical Society of America, 118*(5), 2775-2778. doi: http://dx.doi.org/10.1121/1.2062650

Singh, G., Singh, B., & Singh, D. (2013). Prospective study of ‘otological injury secondary to head trauma’. *Indian Journal of Otolaryngology and Head & Neck Surgery, 65*(S3), 498-504. doi: 10.1007/s12070-011-0404-6

Sohmer, H., Freeman, S., Geal-Dor, M., Adelman, C., & Savion, I. (2000). Bone conduction experiments in humans – a fluid pathway from bone to ear. *Hearing Research, 146*(1), 81-88. doi: 10.1016/S0378-5955(00)00099-X

Speaks, C., Karmen, J. L., & Benitez, L. (1967). Effect of a Competing Message on Synthetic Sentence Identification. *Journal of Speech and Hearing Research, 10*(2), 390.

Spencer, G. A. (2011). *Effects of speaker age on speech understanding and listening effort in older adults* (Master's Thesis). University of Canterbury, Christchurch, New Zealand.

Stam, M., Kostense, P. J., Lemke, U., Merkus, P., Smit, J. H., Festen, J. M., & Kramer, S. E. (2014). Comorbidity in adults with hearing difficulties: Which chronic medical conditions are related to hearing impairment? *International Journal of Audiology, 53*(6), 392-401. doi: 10.3109/14992027.2013.879340

Statistics New Zealand (2014). *Disability survey: 2013*. Retrieved from http://www.stats.govt.nz/browse_for_stats/health/disabilities/DisabilitySurvey_HOTP2013/Commentary.aspx

Stone, J. (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

Straatman, L. V., Lea, J., Le, T. N., & Westerberg, B. (2017). Current insights in noise-induced hearing loss: A literature review of the underlying mechanism, pathophysiology, asymmetry, and management options. *Journal of Otolaryngology - Head & Neck Surgery, 46*(1). doi: 10.1186/s40463-017-0219-x

Subha, S. T., & Raman, R. (2006). Role of impacted cerumen in hearing loss. *Ear, Nose, & Throat Journal, 85*(10), 650-650.

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

Taneja, M. K., Taneja, V., Varshney, H., & Varshney, J. (2015). Ototoxicity, drugs, chemicals, mobile phones and deafness. *Indian Journal of Otology, 21*(3), 161. doi: 10.4103/0971-7749.161014

Taylor, B. (2003). Speech-in-noise tests: How and why to include them in your basic test battery. *Hearing Journal*, *56*(1), 40-46.

Theelen-van den Hoek, F. L., Houben, R., & Dreschler, W. A. (2014). Investigation into the applicability and optimization of the Dutch matrix sentence test for use with cochlear implant users. *International Journal of Audiology, 53*(11), 817-828. doi: 10.3109/14992027.2014.922223

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. doi: http://dx.doi.org/10.3109/14992020903082088

Thümmler, R., Liebscher, T., & Hoppe, U. (2016). Influence of hearing aids on monosyllabic test score and subjective everyday hearing. *HNO*, *64*(8), 595.

Thys, M., & Camp, G. V. (2009). Genetics of otosclerosis. *Otology & Neurotology, 30*(8), 1021.

Tombaugh, T. N. (2005). Test-retest reliable coefficients and 5-year change scores for the MMSE and 3MS. *Archives of Clinical Neuropsychology, 20*(4), 485-503. doi: 10.1016/j.acn.2004.11.004

Tonndorf, J. (1966). Bone conduction. Studies in experimental animals. *Acta Oto-Laryngologica*, Suppl 213:211.

Tonndorf, J., & Khanna, S. M. (1966). Quality of impedance matching by the middle ear of cats. *The Journal of the Acoustical Society of America, 40*(5), 1263. doi: 10.1121/1.1943112

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research, 35*(17), 2503-2522. doi: 10.1016/0042-6989(95)00016-S

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

Tucci, D., Merson, M. H., & Wilson, B. S. (2010). A summary of the literature on global hearing impairment: Current status and priorities for action. *Otology & Neurotology, 31*(1), 31-41. doi: 10.1097/MAO.0b013e3181c0eaec

Turner, C., & Brus, S. L. (2001). Providing low- and mid-frequency speech information to listeners with sensorineural hearing loss. *The Journal of the Acoustical Society of America, 109*(6), 2999-3006. doi: 10.1121/1.1371757

Turner, R. (2003). Double checking the cross-check principle. *Journal of the American Academy of Audiology, 14*, 269-277.

Tye-Murray, N. (2015). *Foundations of aural rehabilitation: children, adults, and their family members* (4th ed.). Stamford, Australia: Delmar.

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

Van Engen, K., & Chandrasekaran, B. (2012). Sentence recognition as a function of the number of talkers in competing multi-talker babble. *The Journal of the Acoustical Society of America, 132*(3), 1969-1969. doi: 10.1121/1.4755263 van Rooij, J. C. G. M., & Plomp, R. (1990). Auditive and cognitive factors in speech perception by elderly listeners, ii: Multivariate analyses. *The Journal of the Acoustical Society of America, 88*(6), 2611-2624. doi: 10.1121/1.399981

Van Wanrooij, M. M., & Van Opstal, A. J. (2005). Relearning sound localization with a new ear. *Journal of Neuroscience, 25*(22), 5413-5424. doi: 10.1523/JNEUROSCI.0850-05.2005

Ventry, I. M., & Chaiklin, J. B. (1965). Evaluation of pure tone audiogram configurations used in identifying adults with functional hearing loss. *Journal of Auditory Research, 5*, 212-218.

Vestergaard Knudsen, L., Öberg, M., Nielsen, C., Naylor, G., & Kramer, S. E. (2010). Factors influencing help seeking, hearing aid uptake, hearing aid use and satisfaction with hearing aids: A review of the literature. *Trends in Amplification*, *14*(3), 127-154. doi: 10.1177/1084713810385712

Villchur, E. (2000). *Acoustics for audiologists*. San Diego, CA: Singular Pub. Group.

Vos, E. M., Greebe, P., Visser-Meily, J. M. A., Rinkel, G. J. E., & Vergouwen, M. D. I. (2017). Subjective hearing impairment after subarachnoid haemorrhage : Prevalence and risk factors. *Journal of the Neurological Sciences, 372*, 184-186. doi: 10.1016/j.jns.2016.11.062

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., Brand, T., & Kollmeier, B. (1999). Development and evaluation of a German speech intelligibility test. Part III: Evaluation of the Oldenburg sentence test. *Z Audiology, 38*, 86-95.

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.

Walden, B. E., Demorest, M. E., & Hepler, E. L. (1984). Self-report approach to assessing benefit derived from amplification. *Journal of Speech and Hearing Research, 27*(1), 49-56.

Wallhagen, M. I., Strawbridge, W. J., Shema, S. J., & Kaplan, G. A. (2004). Impact of self-assessed hearing loss on a spouse: A longitudinal analysis of couples. *The Journals of Gerontology: Series B, 59*(3), S190-S196. doi: 10.1093/geronb/59.3.S190

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*(sup2), 35-43. doi: 10.3109/14992027.2015.1020969

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. (2011). Clinical experience with the words-in-noise test on 3430 veterans: comparisons with pure-tone thresholds and word recognition in quiet. *Journal of the American Academy of Audiology, 22*(7), 405-423. doi: 10.3766/jaaa.22.7.3

Wilson, R. H., & McArdle, R. (2005). Speech signals used to evaluate functional status of the auditory system. *Journal of Rehabilitation Research and Development, 42*(4 Suppl 2), 79. doi: 10.1682/JRRD.2005.06.0096

Wilson, R. H., McArdle, R. A., & Smith, S. L. (2007). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research*, *50*(4), 844. doi: 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. doi: 10.1097/00003446-200506000-00004

World Health Organisation. (2002). *Towards a common language for functioning, disability and health ICF*. Retrieved from http://www.who.int/classifications/icf/training/icfbeginnersguide.pdf

World Health Organisation (2017). *Deafness and hearing loss*. Retrieved from http://www.who.int/mediacentre/factsheets/fs300/en/

Yacullo, W. (1996). *Clinical masking procedures*. Boston, MA: Allyn & Bacon.

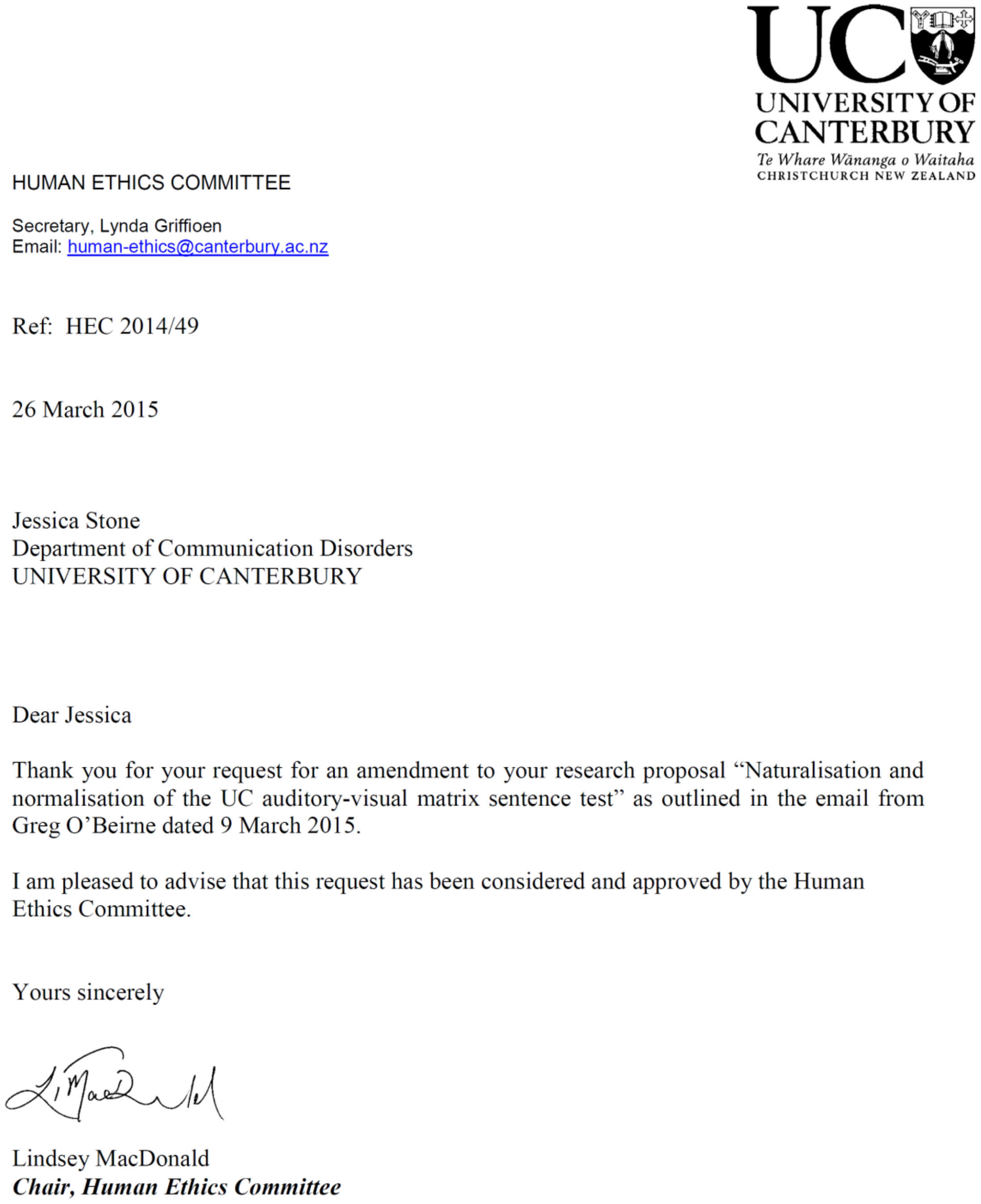
Yost, W. A. (2007). *Fundamentals of hearing: An introduction* (5th ed.). San Diego, CA: Academic Press.

Zhang, L., Gao, N., Yin, Y., Yang, L., Xie, Y., Chen, Y., . . . Zhang, T. (2016). Bone conduction hearing in congenital aural atresia. *European Archives of Oto-Rhino-Laryngology, 273*(7), 1697-1703. doi: 10.1007/s00405-015-3727-1

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

APPENDIX A: *ETHICAL APPROVAL*

Correspondence between the Human Ethics Committee and Stone (2016) confirming the ethical approval pertaining to the current thesis.



UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Lynda Griffioen
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2014/49

26 March 2015

Jessica Stone
Department of Communication Disorders
UNIVERSITY OF CANTERBURY

Dear Jessica

Thank you for your request for an amendment to your research proposal "Naturalisation and normalisation of the UC auditory-visual matrix sentence test" as outlined in the email from Greg O'Beirne dated 9 March 2015.

I am pleased to advise that this request has been considered and approved by the Human Ethics Committee.

Yours sincerely

Lindsey MacDonald
Chair, Human Ethics Committee

APPENDIX B: *RECRUITMENT*

B.1 Advertisement utilised for the recruitment phase of this project.

VOLUNTEERS NEEDED

TO HELP DEVELOP A NZ AUDITORY-VISUAL SPEECH TEST

We are looking for participants that:
ARE **18 YEARS OF AGE OR OLDER**
HAVE **NORMAL HEARING**
ARE **NATIVE SPEAKERS OF NEW ZEALAND ENGLISH**
HAVE **NO CHRONIC DEXTERITY ISSUES**

We are developing an exciting new speech test called the UC Auditory-Visual Matrix Sentence Test (UCAMST). The UCAMST will use both auditory and visual stimuli to diagnose hearing loss in NZ.

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

You will be needed for a **ONE HOUR SESSION**, and during this time you will:
GET A FREE HEARING CHECK
RECEIVE A $10 FUEL VOUCHER
GET A FIRST-HAND LOOK AT THE NEW SPEECH TEST
HELP DEVELOP AN EXCITING NEW SPEECH TEST FOR USE IN NEW ZEALAND

For more information or to be involved in this project,
please contact **Amber Ripberger** at
amber.ripberger@pg.canterbury.ac.nz or text/call **02102513130**

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

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber 02102513130
amber.ripberger@pg.canterbury.ac.nz

Speech test
Amber: 02102513130
amber.ripberger@pg.canterbury.ac.nz

B.2 Email invitation circulated during the recruitment phase of this project.

Hi Everyone,

Volunteers are needed!

The UC Auditory-Visual Matrix Sentence test is an exciting new speech test that uses both auditory and visual cues in the diagnosis of hearing loss. Our goal is to further develop the test for use in NZ audiology clinics in the future.

If you:

- Are **18 years of age or older**
- Have **normal hearing**
- Are a **native speaker of NZ English**
- Have **no chronic dexterity issues**

Then I 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 one hour and during this time you will:

- **Get a free hearing check**
- **Receive a $10 petrol voucher**
- **See this exciting new speech test first-hand**
- **Help to develop an exciting new speech test for use in NZ audiology clinics in future**

For more information please email Amber Ripberger:

amber.ripberger@pg.canterbury.ac.nz or phone/txt 02102513130

APPENDIX C: *INFORMATION SHEET*

Information sheet given to participants prior to obtaining informed consent.

Information Sheet

Full Project Title:	**Further Evaluation and Validation of the University of Canterbury Auditory-Visual Matrix Sentence Test**
Principle Researcher:	Amber Ripberger, 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 an auditory-visual speech test in NZ English to supplement the information gathered from other tests typically used in audiology. The study aims to assess the difficulty of the sentence lists to ensure each of the lists are of equal difficulty.

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
- Have normal hearing
- Be a native NZ English speaker
- Have no current middle ear pathology (i.e. ear infections)

Prior to any testing, you will be asked for a history of your hearing health, which ethnic group you belong to, and you ears will be examined. You will then undergo a hearing check to determine your hearing ability (i.e. whether you have normal hearing or whether you have a hearing impairment and, if so, to what degree), alternatively if you have an audiologist-completed audiogram dated within six months you will not be required to undergo this check. I will inform you of the results of your hearing test and, 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. In the event of an unexpected diagnosis of a hearing loss, a full audiological assessment will be offered at the University of Canterbury Speech and Hearing Clinic free of charge. If you choose to follow up with your GP, this will be at your own expense. If a conductive hearing loss were to be identified during the hearing check, you will receive a $10 fuel voucher for your time.

Following these checks, the study will begin. Shorts sentences being read in quiet or in noise will be presented to you. The words will change in loudness and may at times be difficult for you to hear. After each sentence had been read, you will be asked to identify what you thought you heard. The study will require a maximum of 2 hours for your time.

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 telephone number 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 $10, as well as a free hearing check.

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

Signing this indicated 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 at any stage 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,

Amber Ripberger
2nd year MAud student
Department of Communication Disorders
Email: amber.ripberger@pg.canterbury.ac.nz
Phone: 0275836343

Greg O'Beirne
Primary research supervisor and 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 3 364 2987 ext. 94313

Rebecca Kelly-Campbell
Secondary research supervisor and 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: +64 3 364 2987 ext. 7077

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)

APPENDIX D: *CONSENT FORM*

Consent form signed by all participants in this study.

Consent Form for Persons Participating in Research Studies

Full Project Title: *Further Evaluation and Validation 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 the 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:

____

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.