The Influence of Individual and Linguistic Factors and The Segmentation Strategies of Older Listeners in Adverse Conditions

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

by Phoebe Anna Robinson

2021

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to thank my primary supervisor Professor Megan McAuliffe for her endless support and enthusiasm during the writing of this thesis. I would also like to thank my secondary supervisor Dr. Annalise Fletcher for her support, especially with problem solving in R! Finally, I would like to thank my classmates, friends and family for always being there for me.

Abstract

Purpose: Firstly, this study aimed to examine the influence of listener-based behavioural characteristics and the statistical properties of language on older listeners' speech perception in adverse conditions. Next, it aimed to determine the extent to which older listeners relied on syllabic stress cues to segment the speech stream in these conditions.

Method: One hundred and three older healthy listeners repeated six syllable phrases presented at four signal-to-noise ratios (SNRs) – -5, -2, 1, and 4 dB. The phrases were orthographically transcribed and scored for recognition accuracy using a binary correct/incorrect coding at word level. Behavioural tests examining participants' hearing acuity, vocabulary knowledge, working memory, non-verbal intelligence and processing speed were also conducted. Each phrase was further coded based on the lexical boundary error (LBE) patterns evidenced within the orthographically transcribed phrases—as an indicator of the participants' lexical segmentation strategies.

Results: As listening conditions deteriorated, older listeners' speech recognition accuracy decreased. Those who exhibited higher hearing thresholds and poorer working memory scores also exhibited reduced speech perception accuracy (when analyses controlled for vocabulary knowledge, non-verbal intelligence and processing speed). The statistical properties of language had the largest effect on word recognition scores. Words with higher lexical frequencies and larger phonological neighbourhoods were associated with greater word recognition accuracy. Older listeners exhibited more predicted errors than unpredicted errors in their speech segmentation patterns—demonstrating a reliance on syllabic cues to segment the speech stream in adverse conditions. Degree of hearing loss (i.e., normal hearing, slight, and mild to moderate) did not appear to effect segmentation strategies.

Conclusions: While older listeners' speech perception accuracy is influenced by working memory and audibility, implicit linguistic knowledge also plays a considerable role

in the ability to comprehend speech in adverse conditions. The degree of hearing loss, at least that measured by the current study, did not appear to influence segmentation strategies to a notable degree.

1.1 Introduction

Typically, the ability to perceive speech in quiet conditions does not markedly decline with age (Gordon-Salant, 2005; Rooij et al., 1989). However, in adverse listening conditions the difference between older and younger adults' ability to perceive speech becomes evident (Humes, 1996; Plomp, 1986). Research has shown that audibility is the primary contributor to the perception difficulties experienced by older adults in degraded listening environments (Frisina & Frisina, 1997; Humes, 2007). However, it is widely recognised that the degree of difficulty experienced by older adults considerably outweighs that which might be expected based on their audiogram (Bergman, 1971). Previous studies have found that cognitive factors also play a significant role on an older listener's speech recognition accuracy in adverse conditions (Moore et al., 2014; Mukari et al., 2014). Cognition measures that have been investigated in the literature include working memory, long-term memory, short-term memory, non-verbal intelligence, vocabulary knowledge and processing speed (Heinrich et al., 2015; Ingvalson et al., 2017; Schoof & Rosen, 2014).

Recently, studies have gone beyond examination of individual factors affecting speech recognition to investigate the contribution of implicit linguistic knowledge (e.g., phonological similarity and word frequency) to speech recognition accuracy in adverse conditions (Fletcher et al., 2019; McCreery et al., 2019). Words that occur more commonly in the language are recognised with greater accuracy, as are those with sparse phonological neighbourhoods (the number of words that sound similar to the target word) (Luce & Pisoni, 1998; Storkel, 2013). These studies have commonly focused on young adults' perception of speech—there has been limited research examining the effect of implicit linguistic knowledge on older adults' speech recognition in adverse conditions. Therefore, this thesis investigates the listener-based and linguistic factors that influence older listeners' speech recognition and, subsequently examines the process that underpins this.

The first section of this thesis reviews the difficulties faced by older adults as they perceive speech in adverse conditions—specifically examining the individual factors of audibility, vocabulary knowledge, working memory, processing speed and non-verbal intelligence. Next, this section will address the limited research on the influence that linguistic factors, such as word frequency and phonological neighborhood density, have on speech perception in adverse conditions.

Following this, the strategies that older listeners employ to segment the speech stream into its component words in adverse conditions will be examined. In particular, the process of lexical segmentation is investigated. Lexical segmentation is defined as the process of segmenting the continuous speech stream into word units (Jusczyk & Luce, 2002). Prior research has shown that, in English, listeners pay attention to the presence of strong syllables as a cue to segment the speech stream in adverse conditions. This is known as the metrical segmentation strategy, or MSS (Cutler & Norris, 1988). If listeners employ the MSS in their word segmentation, they will relate stressed syllables to the beginning of a lexical word, and unstressed syllables to the beginning of a grammatical word (Woodfield & Akeroyd, 2010). Research has shown that, in adverse conditions, listeners exhibit increased reliance on stressed syllables to aid their segmentation decision—particularly as the signal becomes increasingly degraded (Mattys et al., 2005). Furthermore, it appears that both audibility and age may affect a listener's preference towards segmentation strategies (McAuliffe et al., 2013).

1.2 Speech Perception in Aging

Older and younger listeners commonly exhibit similar ability to perceive speech in quiet conditions (Fostick et al., 2013). However, as conditions deteriorate the effect of ageing becomes evident (Humes, 1996; Plomp, 1986). Research suggests that this difference in perception in adverse conditions is due to differences in both individual behavioural-based factors, and implicit linguistic knowledge, in older adults.

1.3 Factors Underpinning Older Listeners Speech Recognition Difficulties

When listening in adverse conditions, listeners use different strategies to identify the difference between the acoustic source and the perceptual experience of that sound. Bregman (1990) was the first to propose that in environments with competing background noise, listeners use auditory scene analysis to segregate the signal from the noise—enabling comprehension. To do so, the listener begins by separating the auditory signal into elements, then sorting these elements into groups from each acoustic source. This grouping process is influenced by either inherent knowledge, or learnt skills though listening and language experience. Therefore, there is a higher demand on a listener's cognitive abilities when listening to speech in adverse conditions, compared to when listening in optimal conditions (Pichora-Fuller, 2009).

Since Bregman (1990), the literature investigating how a listener perceives speech in adverse conditions has broadened. More recently, Cooke (2006) theorised a glimpsing model, in which it was proposed that listeners make use of glimpses in the time-frequency plane, when the target speech is not being masked, to aid their recognition accuracy. This is supported by the fact that much of the speech signal is redundant, meaning there is no effect on intelligibility when the signal is degraded to a certain degree (Lippmann, 1996).

One of the key factors affecting speech recognition accuracy is the degree of degradation of the speech signal. Indeed, there is a direct relationship between increasing SNR and a listener's speech perception accuracy decreasing (McAuliffe et al., 2013; Shojaei et al., 2016). This is consistent with the glimpsing model by Cooke (2006), which states that the number of glimpses available in the time-frequency plane has a positive relationship to speech intelligibility. In the past, the difference between older and younger adults' perception ability was attributed to a decline in hearing acuity that occurs with age (Humes et al., 1994). However, in more recent years, it has been noted that the difficulty experienced by older adults is evident even in those with normal hearing (Fostick et al., 2013; Parbery-Clark et al., 2011). This suggests that factors other than audibility play a role in this decline. Such factors may include vocabulary knowledge, working memory, non-verbal intelligence, and processing speed. Additionally, implicit linguistic knowledge, such as word frequency and phonological similarity, have been shown to influence an older listener's speech perception in adverse conditions. The following sections will review previous research that has examined the effect of both individual and linguistic factors on older listeners' speech perception accuracy.

1.3.1 Individual Factors

The importance of individual factors such as cognition as an influencing factor for older adults' speech perception in adverse conditions is widely recognised, secondary to the effect of hearing loss (Akeroyd, 2008; Frisina & Frisina, 1997; Humes, 2007). It seems that younger adults rely on their hearing acuity, however older adults utilise more cognitive support, including working memory, cognitive flexibility, receptive vocabulary and inhibitory control (Ingvalson et al., 2017). Furthermore, the influence of cognition on speech perception accuracy varies at different levels of degradation (Mattys et al., 2005). At mid ranges of intelligibility, factors such as cognition and hearing play a considerably greater role compared to when the signal is easier to understand (Fletcher et al., 2019).

This literature review will further examine these factors that contribute towards older and younger listeners' ability to perceive speech at different levels of adverse conditions including: audibility, vocabulary knowledge, working memory, non-verbal intelligence and processing speed.

1.3.1.1 Audibility. One of the main contributing factors to older adults' difficulties comprehending speech in adverse conditions is audibility (Frisina & Frisina, 1997; Humes, 2007). Population trends show that as we age our hearing thresholds worsen, resulting in age related hearing loss, known as presbycusis (Working Group on Speech Understanding and Aging, 1988). This can be due to the involvement of one or several factors to do with the aging of the hearing system, involving damage to hair cells in the cochlea over time, a reduction in blood supply to the cochlea (metabolic presbycusis) and a reduction in neural firing due to a loss of nerve fibres (Sprinzl & Riechelmann, 2010). Presbycusis typically affects high frequency hearing initially, often resulting in decreased clarity when listening to speech. Due to the low frequencies generally remaining unaffected this often results in no issues with the volume of the speech. Such people will report that they can hear the speech but have difficulty with understanding.

Clarity becomes particularly important when attempting to perceive speech in adverse conditions (Janse & Adank, 2012). This is a reason behind the correlation between increasing age and a decline in speech perception accuracy. Frisina and Frisina (1997) investigated the contribution of presbycusis towards a listener's speech recognition accuracy in adverse conditions. The study included younger participants, and older participants with a high frequency hearing loss. It was found that the older adults with a hearing loss performed more poorly than older adults with normal hearing when listening in adverse conditions. Increasing age was also a contributing factor towards a listener's decline in speech perception ability. Similarly, Humes (2007) reported that audibility was the main contributing factor to an older adult's speech recognition ability in adverse conditions.

1.3.1.2 Vocabulary Knowledge. Vocabulary knowledge involves the listener's knowledge of words, including their phonetic properties, meaning and spelling, and the knowledge of the semantic and grammatical relationship that each word has with other words

(Moghadam et al., 2012). It is directly related to a person's language experience and age is commonly positively correlated with greater vocabulary knowledge (McAuliffe et al., 2013; Neger et al., 2014; Sheldon et al., 2008). Also, native speakers also show better vocabulary knowledge compared to non-native speakers due to their experience with the language, aiding in more accurate speech perception (Mattys et al., 2005).

There are several reasons why the extent of a listener's vocabulary knowledge is an important factor in the success of the listener's ability to accurately perceive speech in adverse conditions (Banks et al., 2015). Firstly, when listening to speech in adverse conditions a listener integrates prior vocabulary knowledge through top-down mechanisms with the sensory input, which influences the resulting perception (Sohoglu et al., 2012). In adverse conditions this is particularly useful to the listener—enabling them to restore the degraded input, resulting in more accurate perception (Benard et al., 2014; Fletcher et al., 2019). It is hypothesized that listeners with a stronger vocabulary knowledge are better able to use available glimpses in the signal to form a lexical hypothesis, therefore improving the accuracy of their speech perception (Fletcher et al., 2019).

Several studies have shown that listeners with a stronger vocabulary knowledge are better able to perceive degraded speech than those with weaker vocabulary knowledge (Ingvalson et al., 2017; McAuliffe et al., 2013; Munson, 1999). The effect of better vocabulary knowledge leading to speech perception accuracy appears to be evident regardless of how the speech signal is degraded. For example, when speech is presented in noise, this effect is seen at moderate signal to noise ratios in which listeners are still able to utilise the available cues (Fletcher et al., 2019). Additionally, vocabulary knowledge is an indicator of a listener's ability to understand accented speech—a naturally occurring adverse listening condition (Banks et al., 2015; Janse & Adank, 2012). There is also a relationship between listeners' vocabulary knowledge and their ability to adapt to degraded speech over time and to learn from repeated exposure (Neger et al., 2014).

The effect of vocabulary knowledge appears to remain important regardless of a listener's age. This is because both older and younger listeners who have better vocabulary knowledge have shown increased ability in recognising accented speech. For example, older adults with higher vocabulary knowledge have shown greater adaptation to unfamiliar accents over the course of time as compared to those with poorer vocabulary knowledge (Janse & Adank, 2012). This provides evidence of a greater ability to perceptually learn.

1.3.1.3 Working Memory. Working memory is defined as the ability to simultaneously store and process task relevant information (Daneman & Carpenter, 1980). It is an active process that involves the ability to switch attention, inhibit irrelevant information, resource share and update long term memory (Singh et al., 2018). Therefore, tests of working memory will include all of these components so that they can be analysed as a whole. When measuring the effect of working memory on speech perception accuracy in adverse conditions, depending on the type of speech test being used, different aspects of cognition will contribute in differing amounts (Heinrich et al., 2015).

Working memory is highly correlated with language perception particularly in adverse conditions (Rönnberg et al., 2008). The Ease of Language Understanding model (ELU-model) proposed by Rönnberg et al. (2008) shows that in optimal conditions, minimal processing effort is required—the process occurs efficiently, and without the need of explicit processing and storage. This is because the signal matches the listeners' existing phonological representations, which is the listeners' knowledge of the sounds that make up the language. However, in adverse listening conditions when speech is presented to a listener that does not match their existing phonological representations, working memory is required to attempt to make sense of this mismatch of information.

Commonly, listeners with an increased working memory capacity are more accurate at perceiving speech in noise—however, these same effects are not always evident in the perception of accented speech (Akeroyd, 2008; Banks et al., 2015; Desjardins & Doherty, 2013). It appears that strong vocabulary knowledge, supported by an increased working memory capacity, is more important in efficient perceptual adaptation to accented speech. Working memory has been found to influence accurate speech perception in both older and younger listeners. For example, a recent study by Kim et al. (2020) showed that working memory capacity is an accurate predictor of speech recognition in in noise and fast speech conditions for older listeners. This relationship was evident when age and hearing sensitivity were controlled for. As adverse conditions became harder (increasing noise and time compression) this relationship remained consistent. This indicated that there was no change in the demand on a listener's working memory capacity as listening conditions became more difficult. However, these findings do not follow the ELU-model by (Rönnberg et al., 2008) which indicates that the demand on working memory increases as listening conditions become more difficult.

The effects of working memory in older listeners are also evident in the speech disorder literature. Ingvalson et al. (2017) found that older listeners with an increased working memory capacity were able to recognise dysarthric speech more accurately than those with a lower working memory capacity. Furthermore, it was observed that there was an interaction between working memory and hearing. Older listeners with increased working memory capacity and better hearing thresholds were most accurate when perceiving the dysarthric speech, compared to older adults with a lower working memory capacity and worse hearing thresholds. It is also recognised that this finding does not match the ELU model by (Rönnberg et al., 2008) due to the correlation found between an older listener's working memory and hearing. Ingvalson et al. (2017) reported that the relationship between working memory and speech recognition in adverse conditions was not found in a younger adult group. This finding may suggest that the listening task required less cognitive capacity for the younger listeners compared to the older listener group. Several other recent studies have challenged the degree to which working memory is important in the perception of degraded speech for younger listeners with normal hearing acuity (Füllgrabe & Rosen, 2016; Koeritzer et al., 2018). Hence, there is some debate about the degree to which working memory contributes to a listener's speech recognition accuracy in adverse conditions regardless of age (Zekveld et al., 2013).

Working memory capacity is also known to have more of an effect on speech perception accuracy in low-context environments. An experiment by Nagaraj (2017) examined the relationship that working memory has on speech perception for sentences with and without context in older adults with a hearing impairment. When sentences with context were presented in multitalker babble and both age and audibility were accounted for, working memory showed no significant contribution toward speech recognition accuracy. However, when listeners were presented with low-context sentences in multitalker babble, better working memory resulted in more accurate speech recognition. This suggests that working memory has a role in accurate speech recognition for low context sentences but may not affect the perception of semantically rich content.

1.3.1.4 Non-verbal Intelligence. Non-verbal intelligence is a measure of a person's intelligence without the influence of language. In some early studies, intelligence was quantified as a cumulative measurement encompassing both verbal and non-verbal intelligence. However, it became clear that the difference between an individual's verbal and non-verbal intelligence scores can be significant due to differing strengths in language abilities (Gundersen & Feldt, 1960). Therefore, to rule out the possibility of a listener achieving better speech perception scores due to a strength in language, measurements of non-verbal intelligence are often used to isolate the effect of general intelligence (while reducing the influence of vocabulary and language experience).

There has been conflicting evidence in the literature regarding the relationship between non-verbal intelligence and a listener's speech perception. Studies have shown that non-verbal intelligence is a significant contributing factor towards a listener's speech perception in noise (Heinrich et al., 2015; Humes, 2002; Kilman et al., 2015). For example, Humes (2002) found that, after audibility, the second main contributing factors to the speech perception accuracy of older hearing aid wearing listeners were non-verbal and verbal intelligence. Furthermore, it was observed that when effects of non-verbal intelligence were controlled for, age did not significantly affect performance. This suggested that it is not increasing age that negatively impacts our speech recognition abilities, but declining non-verbal intelligence that typically occurs as we age. Humes (2002) also included a measure of vocabulary and working memory by using the WAIS-R created by Wechsler (1981) but found that the intelligence measurements accounted for more variance in speech perception accuracy. However, a measure of processing speed was not included in this study. Similarly, Kilman et al. (2015) found a relationship between adults' non-verbal intelligence and working memory when perceiving speech in noise. When a participant scored well in both non-verbal intelligence and working memory testing, they had better speech recognition scores. However, individual processing speed was not considered when examining the relationship between individual factors and the ability to perceive speech in adverse conditions.

Additionally, a study by Heinrich et al. (2015) found that for older adults' with a mild hearing loss, non-verbal intelligence scores were a predictive factor to their success in perceiving speech in noise. Older adults with high non-verbal intelligence scores had better speech perception accuracy than older adults with lower non-verbal intelligence scores.

However, Heinrich et al. (2015) did not include a measure of vocabulary, although measures of processing speed and working memory were included. The inclusion of different cognitive measurements across studies makes the relationship between non-verbal intelligence and speech perception in adverse conditions more difficult to interpret, as intelligence measurements may be correlated with vocabulary, working memory and processing speed. This means that any effects of intelligence test scores could be due to correlations with more specific cognitive skills, rather than an effect of general intelligence alone.

It appears that a relationship between non-verbal intelligence and accuracy in speech perception may emerge with age. This is evident through studies that show no effect of non-verbal intelligence on a younger listener's speech perception accuracy (Fletcher et al., 2019; Tamati et al., 2013). However, there are inconsistencies within the literature due to some studies not accounting for the correlation between non-verbal intelligence, processing speed, vocabulary and working memory. When these factors are accounted for, studies may be less likely to establish a relationship between speech perception skills and a listener's non-verbal intelligence level (Fletcher et al., 2019). Therefore, this relationship will be investigated further in the current study.

1.3.1.5 Processing Speed. Processing speed refers to the ability to process simple information rapidly (Ebaid et al., 2017; Kaufman et al., 2010). Typically, processing speed declines with age which may result in increased difficulty when listening in adverse conditions (Schoof & Rosen, 2014).

In an optimal listening environment, in both low and high contextual situations, there is little demand on processing speed as listening effort is minimal (Pichora-Fuller, 2003). However, as listening conditions deteriorate, this places considerable strain on processing. A similar strain on processing is seen when the listener has a reduced auditory processing ability, resulting in a reduction in clarity of the perceived signal. The contribution of processing speed in speech perception has produced differing results in literature depending on a listener's age and the listening conditions.

When a listener is presented with noise-vocoded speech, the speech perception threshold is found to be determined by both processing speed and age (Neger et al., 2014). Increasing age and a slow processing speed is found to decrease a listener's ability to accurately perceive noise-vocoded speech. Similar results are found in the literature where there is a significant positive relationship between older adults' processing speed and executive function with speech recognition performance in noise (Dryden et al., 2017; Woods et al., 2013). In contrast, Sommers and Danielson (1999) found that processing speed did not contribute to older adults' ability to perceive speech in noise in both high and low predictability situations.

1.3.2 Summary of Individual Difference Findings.

There is a wide range of literature attempting to explore the relationships between the individual factors described previously, and their influence on a listener's speech perception in adverse conditions. Many studies have refined the population that they are examining by only including listeners with normal hearing (e.g., Banks et al., 2015; Fletcher et al., 2019), or listeners with some degree of hearing loss (e.g., Humes, 2002). Other studies included listeners with normal hearing and listeners with a hearing loss (e.g., Torkildsen et al., 2019).

In the literature, there is also variation involving the type of speech perception tests used. Firstly, studies varied in the adverse conditions that were used. Some studies tested participants with accented speech (e.g., Banks et al., 2015; Janse & Adank, 2012), while other studies used dysarthric speech (e.g., Ingvalson et al., 2017), or speech in noise (e.g., O'Neill et al., 2019; Torkildsen et al., 2019). Using different speech tests could introduce a significant amount of variation to the individual influencing factors on perception. This is important to consider as some listeners may have more experience at listening to accented speech or speech in noise in comparison to others (producing unexplained variance). Some studies such as Banks et al. (2015) attempted to account for this by creating a new accent, meaning all listeners would be equally unfamiliar with it, while also making the speech have low intelligibility.

Another issue is that each of the speech perception tests used very different stimuli which creates variation amongst results. Many of the studies in the literature assessed speech perception accuracy by using sentences without context (e.g., Fletcher et al., 2019; Ingvalson et al., 2017; Torkildsen et al., 2019). High and low predictability sentences were also used (e.g., Frisina & Frisina, 1997). Another variation of testing sentence perception was to use nonsense syllables with carrier phrases (e.g., Woods et al., 2013). Heinrich et al. (2015) included a test of phoneme discrimination, whereas Torkildsen et al. (2019) included a monosyllabic test. It is important to be aware of the variation that this introduces when making comparisons between studies.

When examining the individual factors that affect speech perception in adverse conditions it is important that the correlations between individual factors are recognised and are therefore measured. For example, a person with a high non-verbal intelligence is also more likely to have a high working memory, vocabulary and processing speed. Therefore, when studies do not include all of these measures (e.g., Heinrich et al., 2015; Humes, 2002; Kilman et al., 2015; Torkildsen et al., 2019), we must consider how other individual factors might be influencing their results. For example, if it is concluded that non-verbal intelligence is a contributing factor to speech perception in adverse conditions, but processing speed is not measured and accounted for, then the reported relationship could be partly attributable to processing speed, rather than a more general effect of non-verbal intelligence.

The current study addresses these limitations by investigating the effect that individual difference measures have on older listeners' ability to perceive speech in adverse conditions. These measures include hearing acuity, vocabulary knowledge, non-verbal intelligence, working memory, and processing speed. Each of these will be examined under different levels of signal degradation to explore the relationship that noise has with accuracy in speech perception.

1.3.3 Linguistic Factors

While individual factors such as hearing loss and working memory may influence older listeners' speech perception accuracy, the linguistic composition of the message also plays an important role. Two key lexical cues are phonological similarity and word frequency. The next part of this review will discuss their effects.

Phonological similarity is defined as the number of words that sound similar to the target word (Storkel, 2013). Phonological similarity affects the way in which a listener recognises a word while also affecting the production and their acquisition of nonsense and real words. Phonological neighbours are words that are able to be formed through an addition, deletion, or substitution of a phoneme (Zhang et al., 2019). If a word has phonological neighbours that occur more frequently in the language than the target word, this makes the word even harder to identify.

Word frequency is how often a word occurs in a language (Brysbaert et al., 2018). Words that occur more frequently in language, otherwise known as high frequency words, can be processed more easily than low frequency words. This is known as the word frequency effect. The effect of word frequency is seen in word naming tasks and lexical and semantic decision making (Yonelinas, 2002). It is also seen in in memory tasks such as when a listener is required to remember a list of words. Low frequency words are more difficult to recall compared to high frequency words (Almond et al., 2013). The word frequency effect is thought to differ between listeners due to each listener being exposed to different words more frequently throughout their lives compared to others (Brysbaert et al., 2018).

1.3.3.1 Context. Some studies investigating differences in word frequency effect will use sentences or words with context (Cop et al., 2015). It is argued that this is appropriate as it mimics a more natural environment where listeners are able to integrate semantic contextual cues in recognition. However, without contextual cues, studies can examine the ability of participants to make use of glimpses in the speech signal and how this aids their speech recognition without the aid of contextual cues (Krull et al., 2013). This enables the researcher to isolate the participant's ability to listen and recognise the word rather than to problem solve by using contextual cues. There is also an argument that contextual cues become less important as speech is degraded. For example, Mattys et al. (2005) proposes that depending on how degraded the signal is, listeners will rely on different segmentation cues. In good listening conditions listeners will rely on context, however when conditions become more adverse listeners will rely on lexical cues, then segmental and prosodic cues.

In everyday conversation, listeners are easily able to fill in misrepresentations of the signal when context is provided (Mattys et al., 2012). However, investigations of word frequency effects have employed a range of testing conditions which create variation in the cues which a listener can rely on. Some studies have used isolated words (e.g., Krull et al., 2013) while others have used sentences both with (e.g., Cop et al., 2015) or without context (e.g., Fletcher et al., 2019). Each situation provides the listener with different amounts of cues, whether it is semantic cues in a contextual sentence, or lexical, segmental and supra-segmental cues that remain in sentences without context. In some cases, to better understand the effect of neighbourhood density and phonological frequency, nonwords are used (e.g., Storkel, 2013). The use of non-word stimuli eliminates both lexical and contextual cues leaving only segmental and supra-segmental cues. This further reduces the ecological validity of the experiment. In the present study non-sense sentences will be made up of real English words. This allows for generalisation of the results to every day listening environments, while also reducing the influence of context. This will also enable comparison with our existing data in young adults.

1.3.3.2 The Neighbourhood Activation Model. Luce and Pisoni (1998) aimed to establish a relationship between a listener's knowledge of the statistical properties of language and their ability to perceive speech in adverse conditions. The variables examined were the number of words in a neighbourhood, the frequency of word occurrence in the language, and how phonetically similar the words were. Overall, the experiments concluded that the number and nature of words in a phonological neighbourhood affects the speed and accuracy of word recognition. This led to the creation of the Neighbourhood Activation Model (NAM).

The NAM states that speech perception accuracy is affected by the number of words in an individual's lexicon that are similar to the target word, which is consistent with the effects of word frequency and phonological similarity (Fletcher et al., 2019; Luce & Pisoni, 1998). It also recognises that words with high frequency phonological neighbours are more difficult to recognise than words with phonological neighbours that occur less frequently in the language. The NAM begins with the activation of acoustic-phonetic pattern activation following a stimulus. These patterns activate despite the possibility of the stimulus being a nonword. The acoustic-phonetic patterns then activate word decision units if they correspond to words in a listener's memory. This activation of the word units is known as neighbourhood activation. Following this, the word units use higher level lexical information from the long-term memory which is relevant to the words in which they relate to. It is at this stage in which word frequency becomes an influence. At the same time, any relevant information from the short-term memory is also incorporated to identify each word.

1.3.3.3 The Effect of Implicit Linguistic Knowledge on Speech Recognition. A listener's knowledge of the statistical properties of language, specifically word frequency and phonological similarity, aids recognition accuracy in adverse conditions (Fletcher et al., 2019). This is evident through studies examining the differences in native and non-native listeners' speech perception accuracy in adverse conditions (Cutler et al., 2008; Cutler et al., 2004). Results suggest that native listeners show an advantage over non-native listeners in these conditions due to their implicit understanding of the statistical properties of the language. An experiment by Cooke et al. (2008) also showed that native listeners were able to perceive speech in adverse conditions more accurately than non-native listeners. This difference in perception ability was attributed to native listeners having an implicit ability to make use of specific cues which aid in accurate perception. Phonological similarity and word frequency have proven to be such factors that contribute to this difference.

However, more recently in the literature there is debate about the relationship between neighbourhood density and neighbourhood frequency effects due to differing results being found. A study by Vitevitch and Sommers (2003) found that words with high neighbourhood frequencies result in faster and more accurate identification due to the significant processing advantage that is evident compared to words with low neighbourhood frequencies. It is thought that when a listener is undergoing word identification that multiple words are activated in a person's memory, and in turn influence the speed and accuracy of speech production. This is thought to be due to high frequency words receiving greater amounts of activation via phonological nodes, therefore increasing the likelihood of the target word becoming activated. Additionally, the word frequency of the neighbours of the target word has been proven to influence the accuracy and speed of speech production.

On the contrary, it has also been demonstrated that words that have high density neighbourhood frequencies are recognised less accurately and more slowly than words that have sparser neighbourhood frequencies (Luce & Pisoni, 1998). It is suggested that this is due to the increased cognitive load that processing a word with more phonological neighbours requires. It is likely that the effect of neighbourhood density may be dependent on the level of speech signal degradation, with processing advantages more apparent only when the word can be easily distinguished from neighbouring words with similar phonetic-acoustic properties.

1.3.3.4 Varying SNR and Age Effects. It has been shown that word frequency and phonological similarity are both accurate predictors of speech recognition for younger listeners when speech is presented in different levels of noise (Fletcher et al., 2019). At lower signal to noise ratios the effect of word frequency diminishes, which indicates there is a point where there is a decreased likelihood of identifying high frequency words. Phonological similarity has been found to be a negative predictor of word recognition accuracy at many signal to noise ratios, with particularly strong negative effects at lower ratios such as -2 and -5 (Fletcher et al., 2019). This suggests that listeners are still able to make use of phonetic cues even when the signal is highly degraded.

There has been minimal research regarding how these linguistic factors affect speech perception as people age, or in those with hearing loss. Almond et al. (2013) found that, in a word recall task, older adults acquired high frequency words better than low frequency words, an effect not evident in younger adults. This finding was attributed to low frequency words being less interconnected in neuronal networks and therefore more prone to atrophy with age (Almond et al., 2013). The older adults also showed a deficit in consolidation and encoding compared to younger adults (Almond et al., 2013). This suggests that word frequency has an effect on encoding abilities, therefore suggesting that for older adults the encoding of high frequency words is much greater than low frequency words.

1.3.3.5 Word Frequency and Vocabulary Knowledge. The word frequency effect is also likely to be prominent at different frequency ranges for people with different levels of language exposure (Brysbaert et al., 2018). Early research suggested a relationship between word frequency effect and vocabulary size in which there is a larger word frequency effect for listeners with a smaller vocabulary size (Preston, 1935). This means that listeners with a smaller vocabulary size will recognise words that occur less frequently in the language with less accuracy compared to a person with a larger vocabulary size. This effect has also been observed in second language speakers by Cop et al. (2015). It is evident that the frequency effect is considerably more distinct for a listener's second language suggesting that it is vocabulary size and exposure, not intelligence that determines the differences seen in word frequency effect. Cop et al. (2015) created a model which predicted the relation between word frequency effects for people with different word exposure levels using accuracy and response times. This was represented by a logarithmic curve, rather than a direct linear relationship. Furthermore, there was larger variance found in low frequency words with some words being recognised as quickly as the high frequency words. This is due to the word frequency effect accounting for an estimated 30-40% of variance in word recognition testing. It is thought that this is because low frequency words have a relation to high frequency words through compounding, inflection and derivation, therefore these words are recognised quickly by being split into these components. Alternatively, some low frequency words are widely known but are not often spoken. One limitation of the theories behind the word frequency effect is that it is assumed that a person is equally likely to remember different words each time they are exposed to them. However, this is not always the case, as for some words a listener may only need to be exposed once to remember the word for the rest of their life. Hence, the number of encounters with words may not be the optimal measure of word knowledge.

Overall, there has been differing results in the literature when determining the interactions between neighbourhood frequency and the speed of recognition in adverse listening conditions.

1.4 Aims and Hypotheses

Multiple studies have shown that as a signal becomes more degraded this negatively effects an older listener's speech perception accuracy (Fletcher et al., 2019; Mattys et al., 2012; Shojaei et al., 2016). Previous literature also shows a clear effect of cognition on an older listener's ability to perceive speech in adverse conditions (Fletcher et al., 2019; Heinrich et al., 2015; Wingfield et al., 2005). Additionally, studies have demonstrated the effect of implicit linguistic knowledge, specifically word frequency and phonological similarity, on speech perception in adverse conditions (Almond et al., 2013; Fletcher et al., 2019; Luce & Pisoni, 1998). The first component of the current study focuses on the contributions of individual behavioural factors (e.g., audibility, vocabulary knowledge, working memory, non-verbal intelligence and processing speed) and implicit linguistic knowledge (i.e., word frequency and phonological similarity) on speech perception in ageing. It is hypothesised:

1. That as listening conditions worsen and SNR decreases, older listeners speech perception accuracy will also decrease (Fletcher et al., 2019; Mattys et al., 2012).

a. Older listeners with larger vocabularies will have higher word recognition scores compared to listeners with less extensive vocabularies (Banks et al., 2015; Fletcher et al., 2019; Ingvalson et al., 2017; McAuliffe et al., 2013).

b. Older listeners who score highly in working memory testing will show increased accuracy when perceiving speech in adverse conditions compared to those with lower working memory scores (Akeroyd, 2008; Heinrich et al., 2015; Nagaraj, 2017).

c. It is hypothesised that older listeners with a better processing speed will be able to recognise words in adverse conditions better than older listeners who scored less well in processing speed measures (Neger et al., 2014; Woods et al., 2013).

d. There will be no relationship between older listeners' scores in non-verbal intelligence and their ability to perceive speech in adverse conditions when working memory, vocabulary and processing speed are controlled for (Fletcher et al., 2019; Torkildsen et al., 2019).
e. High frequency words will be recognised by older adults more accurately compared to low frequency words (Almond et al., 2013; Fletcher et al., 2019).
f. Words with dense neighbourhoods will be recognised less accurately by older adults than words with sparse neighbourhoods (Luce & Pisoni, 1998).

1.5 Perception in Continuous Speech

When a listener hears speech in an everyday context the sound is a continuous acoustic signal, unlike the gaps that are consistently present in a written format. A listener must employ strategies to distinguish one word from the next to perceive the sentence. This is known as lexical segmentation and is defined as the process of segmenting the continuous speech stream into word units (Jusczyk & Luce, 2002).

1.5.1 Segmentation Strategies

Cutler and Butterfield (1992) analysed error patterns and found that when listening in adverse conditions listeners will insert boundaries before strong syllables to produce lexical words and delete boundaries before weak syllables. This is due to the high frequency of words in the English language that begin with a strong syllable. When a boundary is inserted before a weak syllable this produces a grammatical word.

Since Cutler and Butterfield (1992), the strategies that listeners use to segment speech at different signal to noise ratios has been further investigated. Mattys et al. (2005) discovered that, depending on how degraded the signal was, listeners will rely on different segmentation cues. This led to a hierarchical model of segmentation, as described previously in this review, which suggests that in good listening conditions listeners will rely on context for segmentation. However, as listening conditions become more adverse, segmentation becomes hierarchically driven, where preference is given to lexical cues, then segmental and prosodic cues.

The first strategy used is lexically (knowledge) driven segmentation in which segments that make lexical sense are favoured over segmental and prosodic segmentation cues. This strategy is favoured at low signal to noise ratios when the speech cues are audible above the noise, and in dysarthric speech (Liss et al., 1998; Smith et al., 1989). It is theorised that this involves the listener identifying words sequentially then placing boundaries between words that make lexical sense (Gow & Gordon, 1995). This recognition occurs as soon as a listener is able to distinguish the word from other words with the same onset. However, this theory hypothesises that listeners will predict a word offset allowing them to identify the following words onset. This was proved unreliable by Luce (1986) as it was found that listeners can determine words before their offsets only 40% of the time. Several other models of lexical segmentation have also been proposed including Klatt (1979) LAFS strategy and the McClelland and Elman (1986) TRACE model. The LAFS system accounts for expected phonological and acoustic properties of English which play a role when processing the speech signal. The TRACE model is described as representing speech perception by interactive activation. These interactions occur across a number of simple processing units which result in a perception of speech.

Another strategy is for listeners to use sublexical (signal) cues derived from the signal. Such cues include metrical stress, phonotactic regularities, and acoustic-phonetic variants. One such sublexical strategy that is commonly documented is the MSS (Cutler & Norris, 1988). The MSS involves listeners relating stressed syllables to the beginning of a lexical word and unstressed syllables to the beginning of a grammatical word (Woodfield & Akeroyd, 2010). This results in a commonly observed pattern where listeners insert word boundaries before stressed syllables and delete word boundaries before weak syllables (Cutler & Butterfield, 1992). When a listener encounters an implausible phoneme sequence, otherwise known as a phonotactic trough, this indicates that there should be a word boundary separating this sequence.

1.5.3 Segmentation in Adverse Conditions

The MSS is relied upon particularly when speech becomes degraded, due to the robust nature of the metric cues which allows listeners to still perceive speech despite the adverse conditions (Mattys et al., 2005). It is used in such a situation where a listener is no longer able to rely on lexical-semantic segmentation. From a limited amount of research, it is evident that age and audibility have an effect on how listeners preference one form of segmentation strategy over another when listening in adverse conditions (McAuliffe et al., 2013; Woodfield & Akeroyd, 2010). The following sections will review this.

1.5.3.1 The Effect of Age on Segmentation. It has been hypothesised that with age a listener's segmentation ability diminishes due to the increased cognitive load that occurs when listening to speech in noise (Wingfield et al., 2005; Woodfield & Akeroyd, 2010). When a listener is experiencing heightened cognitive load there is an increased reliance on lexical-segmental cues (Mattys et al., 2009). Older adults with presbycusis have difficulties in noise with and without amplification (Plomp, 1986). In addition to a loss of hearing sensitivity, aging results in a decline in the auditory system in general, and higher cognitive processing such as memory, attention, and cognitive slowing (Schneider et al., 2002). A reduction in temporal processing and intensity discrimination, seen in older adults, also impacts the detection of prosodic stress and therefore reduces segmentation ability. Overall, both older and younger listeners show consistency in the type of errors that are made, with insertion of a boundary before a strong syllable (IS) being the most common, deletion of a boundary before a weak syllable (DW) being the next most common, then insertion of a boundary before a weak syllable (IW) and deletion of a boundary before a strong syllable (DS) being the least common (McAuliffe et al., 2013; Woodfield & Akeroyd, 2010). These errors are typical of the MSS which was consistently used by both older and younger adults across all SNRs presented (McAuliffe et al., 2013; Woodfield & Akeroyd, 2010). The use of the MSS becomes more evident at higher SNR's compared to the low SNR's where words are difficult to distinguish. Additionally, in the study conducted by McAuliffe et al. (2013), it was found that the younger listeners showed a significantly stronger adherence to the MSS, compared to the older listeners. This was interpreted to mean that individuals with more language experience (i.e., older listeners) were better able to use cue redundancies in the speech signal to form a lexical hypothesis and therefore more accurately perceive dysarthric speech without relying on the MSS .

Differences in specific error patterns were also observed in McAuliffe et al. (2013), as younger listeners tended to delete word boundaries before weak syllables more regularly compared to the older listeners. Older listeners, however, were more likely to insert word boundaries before weak syllables compared to the younger listeners. There were no significant differences between the older and younger listeners inserting or deleting boundaries before strong syllables.

1.5.3.2 The Effect of Audibility on Segmentation. There are limited studies in the literature which examine segmentation abilities in older adults with normal hearing and older hearing-impaired adults. One such study by Woodfield and Akeroyd (2010) examined segmentation ability in young normal hearing, older normal hearing, and older hearing-impaired listeners. It was found that there was no difference in the use of the MSS across each of the groups. This was contrary to what was hypothesised by the authors and suggests that listeners rely on metrical prosody for segmentation in adverse conditions regardless of age or hearing impairment. This suggests that metrical cues are robust, as there as use of these cues at signal to noise ratios well below what was needed for full recognition of a sentence.

A similar study by Choe et al. (2012) found that there was a relationship between the use of stress cues in segmentation and a listener's word recognition ability. The study divided the participants into groups according to their word recognition score which resulted in two groups labelled as 'Better Performing' and 'Poorer Performing'. Results showed that the participants with high word recognition scores labelled as 'Better Performing' did not employ syllabic strength cues when determining both the high and low intelligibility phrases. In contrast the 'Poor Performing' group used a higher proportion of predicted LBEs when listening to the low intelligibility phrases. This shows that the 'Poor Performing' group were more reliant on prosodic cues to segment speech even when listening conditions were poor. Choe et al. (2012) suggests that in relation to the hierarchical model of segmentation by Mattys et al. (2005) that 'Better Performing' participants are still able to segment words when they are phonemically ambiguous in poor listening conditions and therefore do not rely on syllabic cues unlike 'Poor Performing' participants. The current study will build upon the findings from Choe et al. (2012) by investigating if factors such as age and degree of hearing loss influence this difference in ability to segment speech between individuals.

1.5.4 Considerations for the Present Study

It is observed that younger listeners are more likely to give an answer when tested, which needs to be accounted for when determining comparisons between older and younger listeners' segmentation abilities (McAuliffe et al., 2013; Woodfield & Akeroyd, 2010).

Furthermore, it is recognised in the work by Woodfield and Akeroyd (2010) that older listeners (aged 56-65 years) may not have shown any difference in their use of the MSS compared to younger adult listeners because they were a 'young' older adult group. The same age range for the older listeners was used in the study by McAuliffe et al. (2013) where the older adults mean age was 64.8 years with good hearing for their age. To the researcher's knowledge, there is a lack of literature that has investigated an even older age group to examine the effect of age on the use of the MSS.

In previous literature, participant numbers are small, with listener groups containing around 15-20 listeners (McAuliffe et al., 2013; Woodfield & Akeroyd, 2010). Studies with a larger number of participants are better representative of the variation that is seen in the population and therefore have more ecological variability. The present study has 103 participants in the younger listeners group and 103 participants in the older adult group.

1.6 Aims and Hypotheses

The second part of the current study aims to investigate the interactions between age, degree of hearing loss, LBE and the patterns in which errors are made. The MSS ratio will be determined for each group to establish whether there is a relationship between age and the use of the MSS ratio.

From these aims it is hypothesised:

1. The differences in younger and older adult's segmentation ability are largely unexplored in the literature.
    a. That both the older and younger groups will employ the MSS, exhibiting a high number of IS and DW errors compared to IW and DS errors (McAuliffe et al., 2013; Woodfield & Akeroyd, 2010).
    b. That both the older and younger adults will rely on the MSS equally, showing no difference in reliance between the two groups (McAuliffe et al., 2013; Woodfield & Akeroyd, 2010).
    c. Additionally, that older adults are more likely to insert a word boundary compared to younger adults. Also, that younger adults are more likely to delete a word boundary compared to older adults (LaCross et al., 2016; McAuliffe et al., 2013).

d. Furthermore, it is hypothesised that older listeners with poorer hearing will be more reliant on the MSS compared to older listeners with better hearing (McAuliffe et al., 2013).

Method

2.2 Participants

One hundred and three older listeners (43 male and 60 female) participated in the current study. All listeners were native speakers of English and were aged between 60-83 years (average age 69.42). Prior to participation, all listeners passed a screening test for cognitive impairment, the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005). Listeners were recruited through a list of volunteers held by the NZ Institute of Language, Brain and Behaviour, social media, researchers' friends and family, and prior participant databases.

The study was approved by the University of Canterbury Human Ethics Committee. Participants provided both written and verbal consent to take part in the study. These data were collected as part of a research grant held by Professor Megan McAuliffe.

2.3 Overview

The current study has two primary areas of focus. The first is to examine the influence of individual factors and the statistical properties of language on older listeners' speech perception in adverse conditions. The second focus is to investigate how older individuals rely on syllabic stress cues to segment the speech stream in adverse listening conditions.

The first section begins by examining the contributions that specific individual factors have on speech perception in adverse conditions. This includes the effect of audibility, vocabulary knowledge, working memory, non-verbal intelligence, and processing speed. Each listener participated in a series of assessments to provide measures of listener-based factors. Listeners were then tested under different levels of signal degradation to determine the relationship between noise level and accuracy in speech perception. This section of the study also examines the effect that implicit linguistic knowledge, such as word frequency and phonological similarity, has on a listener's ability to perceive speech in adverse conditions.

This part of the study uses the same method as Fletcher et al. (2019) to assess the influence of individual and implicit linguistic knowledge, which is described in further detail in the following section.

The second section of this study explores the relationship between LBEs for older and younger listeners. To do this LBE patterns are analysed. A group of younger listeners are included in this section for the LBE initial analysis only—for comparative purposes. These are the same younger participants reported in Fletcher et al. (2019). This section also examines how hearing influences older listeners' segmentation strategies. To do this, participants are split into groups according to their level of hearing including normal hearing, slight hearing loss and mild to moderate hearing loss.

2.3.1 Individual Measures

Participants were assessed using a number of individual tests including audibility, vocabulary knowledge, working memory, non-verbal intelligence, and processing speed. The method for each of these assessments are outlined below.

2.3.1.1 Audibility. Participants hearing thresholds from 0.25-8kHz were measured using the GSI 61 two channel audiometer with Telephonics supraaural headphones within a soundproof booth. A pure tone average of 500, 1000, 2000, and 4000Hz was calculated for each participant with the air conduction thresholds of the better hearing ear. Following this, participants were placed into three groups depending on their pure tone average including: normal hearing (-10 to 15 dB HL), slight hearing loss (16 to 25 dB HL), and mild to moderate hearing loss (26 to 50 dB HL). This was based on the Goodman (1965) scale for classification of degree of hearing loss.

2.3.1.2 Working Memory. Working memory was assessed using the Digit Span test, a sub-test of the Wechsler Adult Intelligence Scale Fourth Edition (Wechsler et al., 2008). A reading span test (RSPAN) was also used to assess working memory which was adapted from

Tompkins et al. (1994). The digit span was administered to participants following the WAIS-IV manual instructions (Wechsler et al., 2008). The test is comprised of three components. The first of which is the forward digit span which requires the participants to repeat the items in the order in which they were presented. The second is the backward digit span in which the participants must repeat the items in reverse order. The third is the sequence digit span which requires the participants to repeat the items back in order from lowest to highest. Each span is scored out of 16 resulting in the highest possible total score being 48.

The RSPAN test was used in conjunction with the WAIS-IV as a measure of verbal working memory. This use of the RSPAN compliments the WAIS-IV as a measure of working memory as it provides a more accurate representation of the influence of working memory on speech recognition compared the independent interpretation of the WAIS-IV score (Akeroyd, 2008). Participants were given written instructions for the RSPAN test presented on a screen with practice stimuli. The test consisted of 12 sets in total in which there were three sets of two sentences, three sets of three sentences, three sets of four sentences, and three sets of five sentences (which made a total of 42 sentences). Participants were instructed that the number of sentences in each set would increase as the test progressed. They were required to repeat back each sentence presented on the screen then state whether each one was ‘true’ or ‘false’. They then had to remember the final word of each sentence, and when the set was complete, repeat back the final words in the order in which they were presented. The tester controlled the presentation of each sentence. The tester scored the total amount of words recalled with a maximum possible score of 42. The RSPAN was programmed in DirectRT (Jarvis, 2010).

2.3.1.3 Vocabulary Knowledge. The Peabody Picture Vocabulary Test (PPVT- IV, (Dunn & Dunn, 2007) was used to assess a participant’s vocabulary by assessing their ability to listen and understand single words (McAuliffe et al., 2013). The test was administered according to standard testing procedures. It consisted of nineteen sets of twelve items in which the participants were required to select which picture out of four which best represented each item. The raw scores were converted to the standardized scores after the test was administered.

2.3.1.4 Non-verbal Intelligence. Non-verbal intelligence was tested using two subtests of the WAIS-IV, block design and matrix reasoning (Wechsler et al., 2008). Procedures were consistent with the WAIS-IV manual protocols (Wechsler et al., 2008). For the block design test, participants were required to reconstruct a picture using red and black blocks within a time limit. The matrix reasoning task required participants to complete two types of matrix, the first being a two-by-two matrix and a series matrix. Participants were given two samples as an example. They were then required to complete either the matrix or the series using one of five options. A point was given for each correct item with a raw total of 26.

2.3.1.5 Processing Speed. A measure of each participant's processing speed was obtained to account for individual variability in the participants' ability to process information rapidly (Ebaid et al., 2017). Two subtests from WAIS-IV, symbol search and coding, were used to assess processing speed (Wechsler et al., 2008). Procedures were consistent with the WAIS-IV manual protocols (Wechsler et al., 2008). Both tests were administered using pen and paper. The symbol search task required participants to find a specified symbol within a time limit from a group of other symbols. A maximum raw score of 60 was possible. The coding task required participants to copy down a symbol which was paired with a number within a time limit. A total score of 135 was possible.

2.4 Speech Perception Task

The speech perception task was one of the primary dependent variables of this study.

2.4.1 Speech Stimuli

Participants were presented with 128 phrases consisting of three to five words which were six syllables long, without context and with alternating stress contrasts (64 strong weak and 64 weak strong) consistent with previous research from this laboratory (e.g., Fletcher et al., 2019; McAuliffe et al., 2013). The phrases were from a pool of 160 phrases in which none of the words were repeated. For the 358 words in the phrases, both lexical frequency and phonological similarity was obtained using the English Lexicon Project (ELP) (Balota et al., 2007). The English Lexicon Project gathered data for 40,481 words and 40,481 non-words. This was implemented by six universities from 816 participants. This resulted in a standardized dataset which contains both behavioral and descriptive information for each word used and is available at elexicon.wustl.edu.

The current study examined both word frequency and phonological similarity for each sentence in the speech perception task. There are multiple ways in which phonological similarity and word frequency can be measured. For the present study, two methods from the ELP were used: the Log Hyperspace Analogue to Language (HAL) frequency norms was used as a measure of word frequency, and the Phonological Levenstein Distance (PLD) was used as a measure of phonological similarity (Lund & Burgess, 1996). The HALincludes approximately 160 million words from Usenet news groups which encompasses a wide range of topics, and the text is conversational to provide an accurate representation of everyday language. The PLD was used as a measure of phonological similarity. More commonly in research, phonological similarity is measured by neighbourhood density, which represents the number of words that differ from the target word by a single sound substitution (Luce & Pisoni, 1998; Vitevitch & Sommers, 2003). However, since so many words in our lexicon have no direct phonological neighbours, counting the number of words that differ by a single sound substation can fail to characterize differences in the phonology of a large numbers of words (Vitevitch & Sommers, 2003). For this reason, the PLD measurement was chosen. The

PLD calculates the phonological distance between two words by finding the minimum number of additions, substitutions, or deletions that result in the initial word becoming the final word (Fontan et al., 2016). This is repeated for every word in the ELP, then the average of the target words closest 20 phonological neighbours is calculated (Yap & Balota, 2009). Hence, PLD can be considered a continuous measurement. Unfortunately, minimal research has been carried out using the PLD to date. Suárez et al. (2011) investigated the influence that phonological similarity and word frequency have on speech perception by only using lexical hermits which are words with no phonological neighbours (therefore minimising the generalisability of their findings). The PLD in the ELP is based on the Unisyn Lexicon which was developed by the Centre for Speech and Technology Research at the University of Edinburgh (Centre for Speech Technology Research, n.d.).

2.4.2 Recording the Stimuli

The 128 phrases were spoken by eight healthy New Zealand English speakers (four female and four male). Each phrase was recorded twice by each speaker in a soundproof booth. The phrases were presented for ten seconds on a PowerPoint presentation on a screen in front of the speaker. Speakers were instructed to speak in their normal voice and the speed at which they spoke was directed by listening to four example recordings. Short monaural audio digital recordings (44.1 kHz sampling rate, 16-bit quantization) were directly captured to a compact flash memory card via Earthworks M30 desk microphone situated 30cm to the side of the talker coupled to a TASCAM HD-P2 portable stereo recorder. If a recorded phrase contained an error, speakers were asked to record that phrase again.

2.4.3 Stimuli Selection and Counterbalancing

Firstly the audio files were divided into individual files manually using MATLAB (Mathworks, n.d.). For each phrase the best, most natural recording was chosen. If differences were unclear between multiple recordings the first recording was chosen.

Following this, the recorded phrases were mixed by noise matching the long-term average speech spectrum of that talker. The different masker levels (signal to noise ratios, SNRs) were established as per Sinex (2013). The level of the recorded phrases remained constant at 65dB A while the noise varied as per Gilbert et al. (2013).

2.4.4 Pilot Study to Select Degradation Levels

A small pilot study was conducted to establish which SNR levels would yield a range of speech perception accuracies from participants. The pilot study consisted of 5 adults, all of whom had hearing within normal limits. Each listener was presented with all 128 phrases at four different SNR levels chosen from 14 levels ranging consecutively from -7 to +6 dB. Results indicated that the four SNRs to use for the main study should be -5,-2, +1, and +4. This is because they produced results averaging from 28%-83% of words correct per phrase. This was indicated to be appropriate to avoid floor and ceiling effects and to effectively allow for the investigation of the effect of different SNR's on word recognition.

2.4.5 Counterbalancing and Randomising the Speech Stimuli

A list containing the 128 phrases was created for each listener which was divided into four counterbalanced blocks for each of the four SNR levels. The eight talkers (four male and four female) recorded four phrases which were played to the listeners at the four SNR levels. For each SNR level, each talker's phrases were balanced so that one higher average word frequency and one lower average word frequency for each stress pattern was included (strong-weak and weak-strong). The average frequency of the phrase and phrase length were balanced for each by stress pattern, across stress patterns and across speakers. The lists which were generated and selected randomly were rejected if they did not meet the criteria for being balanced.

2.5 Procedure

Testing was carried out at the University of Canterbury Speech Perception Lab in which participants attended two 60-90 minute sessions. The second session took place within one to 21 days of the first session. Participants were rewarded at the commencement of the second session with a $50 voucher. The first session consisted of gathering consent from each participant, retrieving demographic information and half of the speech perception testing and half of the cognition and vocabulary testing. The second session completed the testing by including the remining half of the speech perception testing and the cognition and vocabulary testing. The order of the speech perception testing, and the cognition and vocabulary testing were counterbalanced. Each participant was assigned to one of eight different test orders. Each task was grouped into similar tasks and the WAIS-IV testing was administered as per protocols (Wechsler et al., 2008).

2.5.1 Speech Perception Experiment

Speech perception testing took place in a soundproof booth. Participants were situated at 0° azimuth and 0.5m from the speaker. The speaker was calibrated using a 1kHz tone with a sound level meter (Reed ST-805 Compact Digital). The phrases were set at 65dB at the participants' head height on the system gain. Phrases were on an external soundcard (THX Truststudio PRO) presented to the speaker and amplifier (Crown D-75A). Phrases were presented via free field to ensure a comfort due to the lack of experience taking part in research by some of the older participants.

Participants were instructed that they would hear phrases containing English words which did not make sense. They were told that some sentences would be harder to hear than other, but that they needed to guess an answer if they did not hear. If participants were unable to guess a word, they were instructed to use the filler 'something' as per McAuliffe et al. (2013). Participants were required to repeat the whole phrase with each phrase being played once. Participants had five seconds after each phrase to repeat the sentence back, which was recorded by a Sony IC digital recorder to later be transcribed.

Participants were presented with 10 orthographic examples of each type of phrase, then they were played five audio examples each at +6 dB SNR so that each listener was familiar with the audio that they would be hearing.

For the first session participants heard the first half of the list which they were assigned, this was two SNR sections of 32 phrases. In the second session, participants listened to the remaining two sections of 32 phrases. Between blocks participants were given a short break before beginning the second block.

2.6 Data Analysis

2.6.1 Transcription

The tester orthographically transcribed each phrase spoken by the participants. Transcriptions were in New Zealand English and in lowercase. For words or non-words in which the spelling was unclear a DISC phonetic transcription was added so that other researchers were able to understand the transcription. When a participant used the filler ‘something’ this was transcribed as ‘X’.

To ensure accuracy in transcription, a second transcriber, who was blind to the first transcription, transcribed each phrase. Both transcriptions were automatically checked for matches. The second transcriber then checked the mismatches for any spelling errors or fixed errors and the remaining mismatches were passed on for consensus checking. The consensus check was carried out by a third researcher who chose the correct transcription for each mismatch. Once this was complete a final file was made for scoring and statistical analysis.

2.6.2 Dependent Variables

2.6.2.1 Word Accuracy. A MATLAB script automatically scored the accuracy of the individual words in the phrases (Mathworks, n.d.). This checked whether the words in the transcription matched any of the words in the target phrase. If a word matched this was marked as correct and given a mark of 1, and if a word did not match it was given a mark of 0. There were 37,183 words that were analysed in total (103 listeners that read 361 content words).

2.6.2.2 LBEs. A research assistant classified the occurrence of LBEs for each sentence for each participant. The author was blinded to the group allocations. The LBEs were classed as either predictable or unpredictable errors. A predictable error is when a listener either inserts a lexical boundary before a strong syllable, or when a lexical boundary is deleted before a weak syllable. Unpredictable errors are when a listener inserts a lexical boundary before a weak syllable or deletes a lexical boundary before a strong syllable. Reliance on predictable patterns of errors is language specific and occurs due to the predictable patterns in the English language. The predictable patterns are that a strong syllable has a higher likelihood of being a word onset, therefore errors patterns show some reliance upon this. This is commonly referred to as the MSS in which a ratio can be calculated to assess a listener's reliance on predictable error patterns (Cutler & Norris, 1988). The MSS ratio is calculated by adding together the predictable errors (insertion of a lexical boundary before a strong syllable, and deletion of a lexical boundary before a weak syllable), then dividing by the total number of LBEs. If the ratio is above 0.5 this indicates a stress-based approach to lexical segmentation (Spitzer et al., 2007). There were 300 opportunities for an LBE across all the phrases. If a listener responded with the filler of 'something' then this was not recorded as an LBE, but the remainder of the phrase was still classified if LBEs occurred. All phrases were scored on their accuracy regardless of the number of LBEs made. Intra and inter-coder reliability was checked across 25% of the listeners transcripts.

All errors were recorded which also included addition of a syllable, deletion of a syllable and substitution of a syllable.

2.7 Statistical Analysis

There were two main areas of statistical analysis. First, the effect of individual and lexical characteristics on perception was examined using a binomial mixed effects model, with the glmer function in R version 3.4.1. The dependent variable of word accuracy was accounted for using a binomial logit function (0= incorrect, 1= correct). The model was fit with the maximum likelihood criterion and the random intercepts for each participant, talker, phrase and target word were included in the model. Phrase and target word were a nested random effect. This is due to each word only occurring within one phrase.

When analysing LBEs, t-tests were used to compare both older and younger adults. To analyse error patterns, predicted and unpredicted errors, and reliance on the MSS in older and younger adults, X^2 goodness of fit and t-tests were used. Proportion of expected errors compared to observed errors was compared between older and younger adults and analysed further using X^2 goodness of fit tests. Following this, the influence of audibility on segmentation was assessed by using a one-way between subjects ANOVA with post hoc comparisons using t-tests with a Bonferroni correction to compare reliance on the MSS and the average number of LBEs within each hearing group. Next, the proportion of expected errors compared to observed errors were compared between hearing groups and analysed further using a two-way ANOVA.

Results

3.2 Effect of Individual Listener Factors and Implicit Linguistic Knowledge

The first aim of the study was to examine the effect of individual listener factors and implicit linguistic knowledge on older adults' ability to perceive speech in adverse conditions. To do this, measures of audibility, vocabulary knowledge, working memory, non-verbal intelligence, and processing speed were obtained from each participant. These measures were subsequently included as independent variables in a mixed effect model, with word accuracy as the dependent variable.

3.2.1 Group Level Results on Behavioural Measures

Table 1 details raw scores for the group of older listeners on a range of behavioural tests. This includes measures of hearing, vocabulary knowledge, working memory capacity, non-verbal intelligence and processing speed. As can be seen from the range of scores, participants exhibited some degree of variation in their performance on these tests. Overall, the hearing threshold average was equivalent to the upper end of normal hearing (-10 to 15 dB HL) as per the Goodman (1965) classification. Each of these measures are displayed in Table 1 including test type, raw mean score, raw total and range.

Table 1.

Raw Scores for Older Listeners Individual Measure Tests

Measure	Test	Raw Mean	Raw Total	Range
Best Ear Pure-Tone Average (dB HL)	Pure tone audiometry (average from 500 to 4000Hz)	15.07	NA	-1.25-48.75
Vocabulary Score	Peabody Picture Vocabulary Test-Fourth Edition	215.9	228	178-228
Working Memory Score	Digit Span	28.74	48	10-43
	Reading Span	28	42	12-40
Non-Verbal Intelligence	Block design	45.48	66	20-65
	Matrix reasoning	19.5	26	6-26
Processing Speed	Symbol search	33.67	60	17-54
	Coding	71.12	135	7-120

Note. Raw total is the maximum possible score a listener could obtain on each of the tests.

The raw scores for each of the individual factors were divided by the raw total prior to give a score between 0 and 1. Working memory score was operationally defined as the combined score of the digit span (Wechsler Adult Intelligence Scale-Fourth Edition [WAIS-IV]) and Reading Span tasks, giving a maximum score of 2 (Tompkins et al., 1994; Wechsler et al., 2008). Nonverbal intelligence was defined as the combined score from the block design and matrix reasoning tasks (WAIS-IV), with a maximum score of 2 (Wechsler et al., 2008). Processing Speed was a combined score of the raw scores of the symbol search and coding tasks with the highest possible score of 2 (Wechsler et al., 2008). Vocabulary score is the participants' raw score from the Peabody Picture Vocabulary Test-Fourth Edition, divided by the maximum score possible (a participant's score cannot exceed 1) (Dunn & Dunn, 2007). All measurements were then scaled and centred on their means to be compared within statistical models. This is displayed in Table 2 which shows the mean and range for older adults' vocabulary score, working memory score, non-verbal intelligence, and processing speed. Ceiling effects were observed in the scores for older adults' vocabulary knowledge. This is because the majority of scores were in the upper range above the mean, meaning there was little variance in the third and fourth quartiles.

Additionally, phonological similarity and word frequency were calculated for each word by using the Log Hyperspace Analogue to Language (HAL) frequency norms for word frequency, and PLD for phonological similarity, as stated in the methods section (Fletcher et al., 2019; Lund & Burgess, 1996). These measurements were also scaled and centred on their means.

Table 2.

Scaled and Centered Individual Measures for Older Listeners

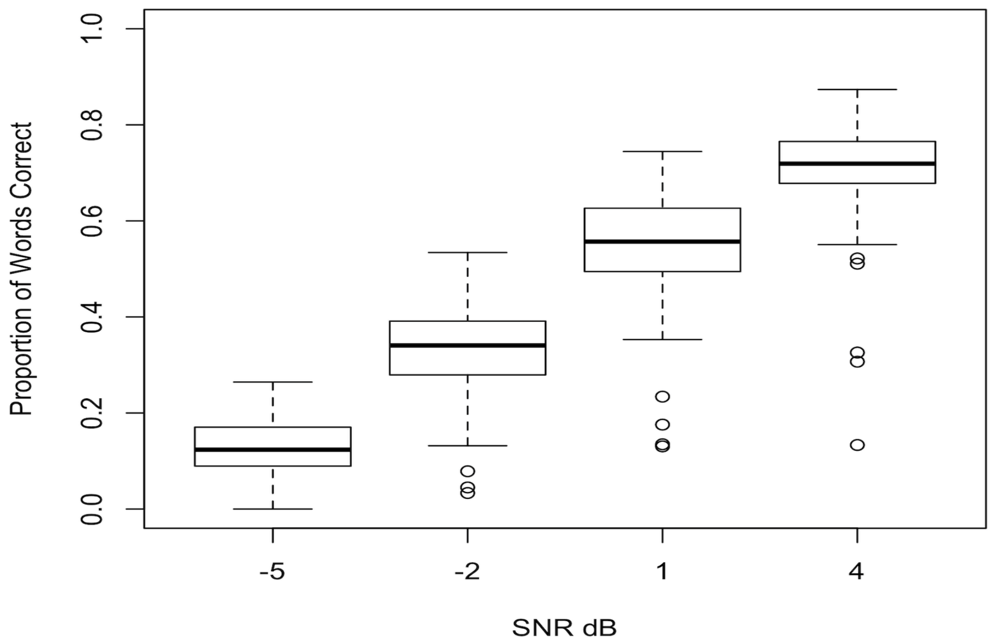
Measure	Scaled and Centered Mean (sd)	Range
Vocabulary Score	0.97 (0.02)	0.93-1.00
Working Memory Score	1.17 (0.20)	0.68-1.57
Non-verbal Intelligence	1.27 (0.25)	0.66-1.80
Processing Speed	0.94 (0.16)	0.61-1.45

3.2.2 Speech Recognition Accuracy Across SNRs

Figure 1 contains the average word recognition score for older listeners across the four SNR conditions. Examination of the figure indicates a similar distribution of performance within each SNR that was tested. As expected, speech perception accuracy decreased as SNR decreases and listening conditions became less favorable.

Figure 1.

Word Recognition Accuracy for Older Listeners at Each SNR Level



Following this, a primary model was constructed to examine the effects of each of the individual and lexical characteristics on the older adults' word recognition ability. This model also accounted for SNR, which was assessed at four different levels. This was treated as a numerical variable due to the equal distance between each SNR level.

To investigate the effect of individual and linguistic differences on word recognition accuracy for older adults, binomial mixed effects modelling was used. Fixed effects included the individual scores for vocabulary, working memory, processing speed and non-verbal IQ. The lexical characteristics of PLD and lexical frequency were also entered as fixed effects, along with the SNR of each sentence. The random effect structure is described in the methods section.

The final statistical model is presented in Table 3. The estimate, standard error, *z*-values and *p*-values are reported for each variable. Table 3 shows that as listening conditions improve, as indicated by a one unit increase in SNR, the accuracy of listeners' word recognition improves significantly ($p < .001$). In addition, both working memory and hearing had significant effects on word recognition. Older adults' who had higher working memory scores exhibited better word recognition accuracy ($b = 0.095$, $SE = 0.047$, $p = .043$). Participants with better hearing also exhibited higher levels of speech recognition ($b = -0.038$, $SE = 0.006$, $p < .001$). Vocabulary, processing speed and non-verbal IQ did not affect word recognition accuracy in older adults ($p > .05$).

Interestingly, word frequency and PLD accounted for more variance in word recognition accuracy than any of the individual factors measured. The higher the PLD value ($b = 0.36$, $SE = 0.050$, $p < .001$) and word frequency score ($b = 0.30$, $SE = 0.053$, $p < .001$), the better listeners were able to recognise words—suggesting that words with higher lexical frequencies and PLD are easier to recognise in adverse listening conditions. SNR had the largest effect on word recognition accuracy with a 1-SD increase compared to the other individual and lexical factors (b = 1.11, SE = 0.014, $p < .001$).

Table 3.

Binomial Mixed Effects Model for Older Adults' Word Recognition

Fixed Effects	Estimate	SE	z	P
(Intercept)	-2.68	0.12	-21.58	< .001
SNR	1.11	0.014	81.60	< .001
Best Ear 500 to 4	-0.038	0.0057	-6.69	< .001
Vocabulary Score	0.050	0.045	1.11	0.27
Working Memory	0.095	0.047	2.026	0.043
Non-Verbal Intelligence	0.0059	0.051	0.12	0.91
Processing Speed	0.0087	0.050	0.18	0.86
PLD	0.36	0.050	7.20	< .001
Word Frequency (logfreqHAL)	0.30	0.053	5.61	< .001

Note: The same fixed effects were significant when a backwards stepwise linear regression was applied to drop the non-significant effects from the model.

3.3 Lexical Segmentation Strategies of Older Adults

The second aim of the study was to investigate how older individuals rely on syllabic stress cues to segment the speech stream in adverse listening conditions. To do this LBEs (LBEs) were analysed. While the study focuses on the performance of older adults, data from a younger participant cohort is included for comparative purposes. Table 4 shows LBE summary data for younger and older participants. Older listeners exhibited a similar average number of LBEs to younger listeners [$t(204) = 1.17$, p = .24]. Next, the patterns of errors were analysed where a high number of predicted errors (insertion of a lexical boundary before a strong syllable, and deletion of a lexical boundary before a weak syllable) would represent a listener's preference towards the MSS. Unpredicted errors consisted of insertion of a lexical boundary before a weak syllable and deletion of a lexical boundary before a strong syllable. The pattern of predicted and unpredicted errors was not uniform across the older and younger listener groups [$X^2\ (1) = 5.07, p = .024$]. This suggested that the older and younger listeners showed a difference in their preference towards segmentation strategies.

These findings were analysed further and it was revealed that both older and younger listeners followed the predicted MSS error pattern, making more predicted errors compared to unpredicted errors [younger participants: $X^2\ (1) = 161.85, p < .001$; older participants $X^2\ (1) = 239.01, p < .001$]. This was also noted in the MSS ratio scores, as both the younger and older participants had an MSS ratio of greater than 0.5, implying the use of the stress-based segmentation when perceiving speech in adverse listening conditions (Spitzer et al., 2007). However, the groups adhered to the MSS in differing strengths, with the older adults showing a significantly stronger reliance on stress-based segmentation than the younger adults [$t(204) = -2.28, p = .024$].

The differences in deletions and insertions between younger and older listeners are also depicted in Table 4. This showed that older adults are significantly more likely to delete a word boundary compared to younger adults [$t(204) = -5.63$, $p < .001$]. In contrast, the younger adults were significantly more likely to insert a word boundary compared to the older adults [$t(204) = -5.63$, $p < .001$].

Table 4.

LBE Summary Data

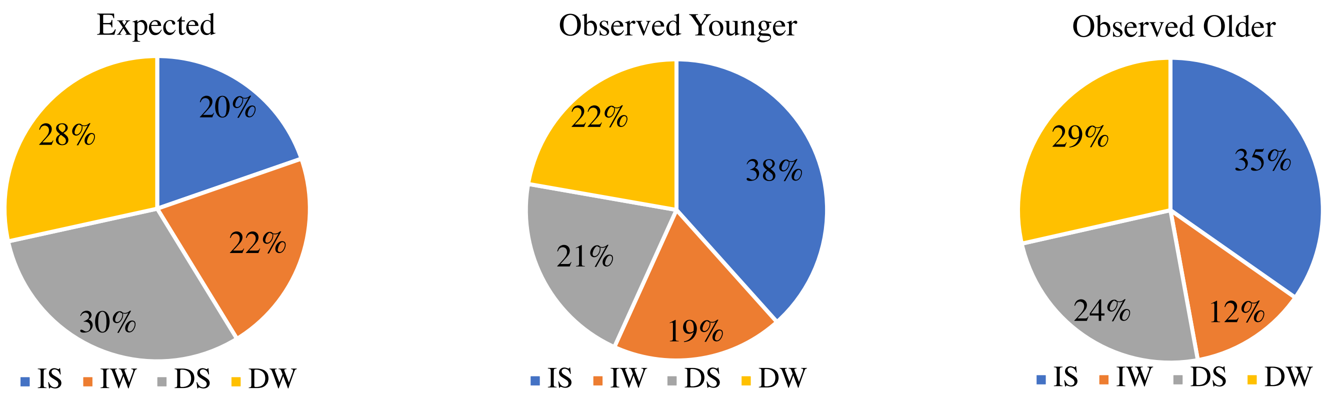
LBEs	Younger	Older
Total Number of Errors (LBE)	3597	3404
Total Number of Predicted Errors	2180	2153
Total Number of Unpredicted Errors	1417	1251
Average Number of Errors Per Listener	34.92 (9.75)	33.05 (13.03)
MSS Ratio	0.61 (0.09)	0.64 (0.11)
Mean Deletions (sd)	0.44 (0.11)	0.54 (0.13)
Mean Insertions (sd)	0.56 (0.11)	0.46 (0.13)

The graphs in figure 2 show how the younger and older adults error distributions coincide with the number of opportunities to make each of the error types. Firstly, the graph on the left 'Expected' represents the expected error distribution, based on the linguistic composition of the experimental phrases. The middle graph is the error distribution of the younger adults, and the graph on the right is the error distribution of the older adults.

Overall examination of the graphs showed that both older and younger listeners tended towards inserting a word boundary when percieving speech in adverse conditions.

Figure 2.

Expected and Observed LBE Outcomes for Older and Younger Participants



Further examination of the type of errors made by the groups indicated that both younger and older participants made significantly more IS errors than expected [younger participants: $X^2\ (1)\ =\ 38.48, p\ <\ .001$; older participants $X^2\ (1)\ =\ 26.58, p\ <\ .001$]. This demonstrated that both listener groups exhibited a tendency toward inserting a lexical boundary in difficult listening conditions—consistent with the MSS. Younger participants showed no significant difference in IW errors made compared to the expected proportion, however older adults made significantly less IW errors than expected [younger participants: $X^2\ (1)\ =\ 1.60, p\ =\ .21$; older participants $X^2\ (1)\ =\ 15.82, p\ <\ .001$]. Younger listeners made significantly less DS errors than expected and the older adults showed no significant difference in DS errors made compared to the expected ratio [younger participants: $X^2\ (1)\ =\ 10.84, p\ <\ .001$; older participants $X^2\ (1)\ =\ 4.16, p\ =\ .04$]. Both the older and younger participants showed no significant difference in the amount of DW errors made compared to what was expected [younger participants: $X^2\ (1)\ =\ 4.85, p\ =\ .028$; older participants $X^2\ (1)\ =\ 0.00077, p\ =\ .98$].

Next, the study investigated whether differences in hearing ability might lead listeners to preference one form of segmentation strategy over another. To do so, the older listeners group was divided into three separate groups based on their average pure tone thresholds between 500Hz and 4kHz in their better hearing ear. The three groups included: normal hearing (-10 – 15 dB HL), slight hearing loss (16 – 25 dB HL), and mild-to-moderate hearing loss (26 – 50 dB HL). The summary data for LBEs for these three groups are shown in Table 5. A one-way between subjects ANOVA was conducted to compare the average number of LBEs made between hearing groups. There was no significant difference in the number of LBEs at the $p < .001$ level for the three groups [$F(2) = 2.31, p = .10$]. Post hoc comparisons using a t-test with Bonferroni correction indicated that the average LBEs for the normal hearing group (2098) was not significantly different from the slight hearing loss group (799) (p = .49) and the mild to moderate hearing loss group (507) (p = .42). The average number of LBEs for the slight hearing loss group were also not significantly different from the mild to moderate hearing loss group (p = .10).

A one-way between subjects ANOVA was also conducted to compare the difference in MSS ratio between hearing groups. There was no significant difference in the MSS ratio at the $p < .001$ level for the three groups [$F(2)= 2.37, p = .096$]. Post hoc comparisons using a t-test with Bonferroni correction indicated that the MSS ratio for the normal hearing group (2098) was not significantly different from the slight hearing loss group (799) (p = .21) and the mild to moderate hearing loss group (507) (p = .54). The MSS ratio for the slight hearing loss group were also not significantly different from the mild to moderate hearing loss group (p = 1.00).

A two-way ANOVA was conducted to determine whether there was a difference in mean deletions between hearing groups which found no significant difference [$F(2) = 0.44, p = 0.64$]. Similarly, no significant difference was found for mean insertions between hearing groups [$F(2) = 0.44, p = 0.64$].

Table 5.

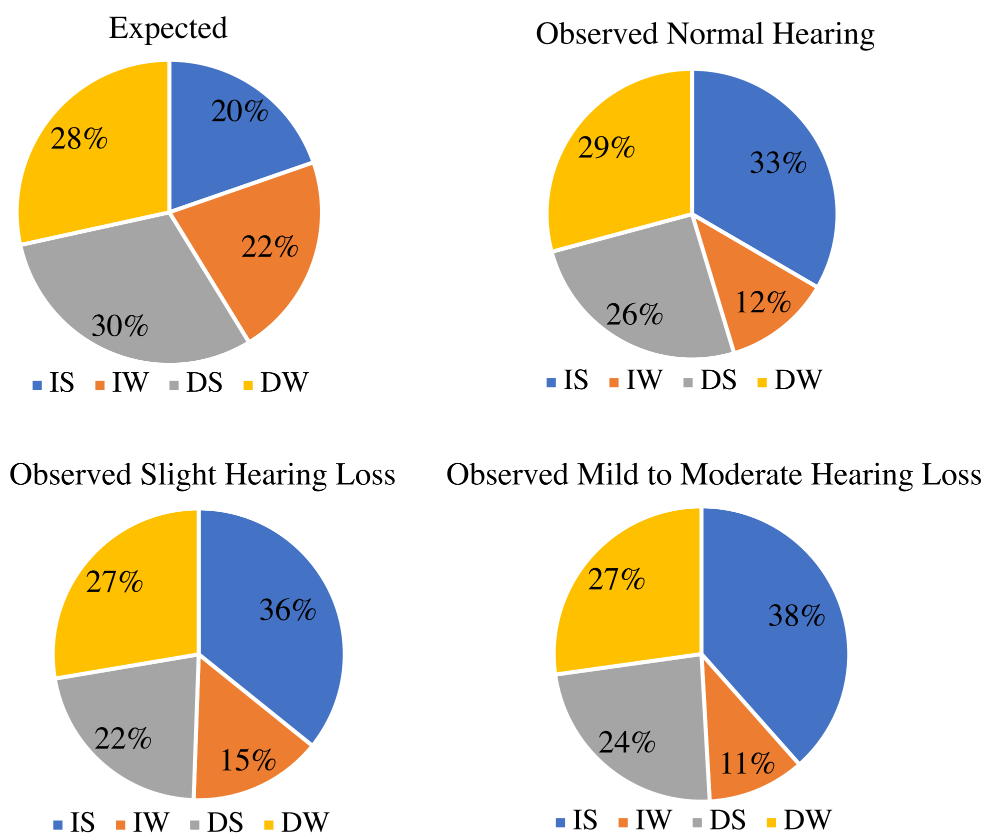
LBEs by Hearing Group

LBEs	Normal Hearing (-10 to 15) (64 participants)	Slight HL (16 to 25) (26 participants)	Mild to Moderate (26 to 50) (13 participants)
Total Number of Errors (LBE)	2098	799	507
Average Number of Errors	32.78 (12.08)	30.73 (13.45)	39 (15.85)
Percentage of Predicted Errors	62.58	63.45	65.68
Percentage of Unpredicted Errors	37.41	36.55	34.32
MSS Ratio (sd)	0.63 (0.47)	0.63 (0.48)	0.67 (0.58)
Mean Deletions (sd)	0.55 (0.47)	0.38 (0.48)	0.67 (0.58)
Mean Insertions (sd)	0.45 (0.47)	0.625 (0.48)	0.33 (0.58)

The graphs in figure three show how the hearing groups error distributions coincide with the number of opportunities to make each of the error types. Firstly, the graph in the top left represents the expected error distribution as explained earlier. The graph in the top right is the error distribution for the normal hearing participants. The bottom left and right graphs represent the error distribution for the slight hearing loss group, and the mild to moderate hearing loss groups respectively.

Figure 3.

Expected and Observed LBE Outcomes for Older Participants Grouped by Hearing



Examination of error types indicated that the normal hearing, slight hearing loss and mild to moderate hearing loss groups made significantly more IS errors than expected [normal hearing participants: $X^2\ (1)\ =\ 22.70, p\ <\ .001$; slight hearing loss participants $X^2\ (1)\ =\ 29.91, p\ <\ .001$; mild to moderate hearing loss participants $X^2\ (1)\ =\ 38.79, p\ <\ .001$]. This demonstrated that all hearing groups exhibited a tendency toward inserting a lexical boundary in difficult listening conditions—consistent with the MSS ratio findings. The participants with normal hearing and mild to moderate hearing loss made significantly less IW errors compared to the expected proportion, and the participants with slight hearing loss showed no significant difference in IW errors [normal hearing participants: $X^2\ (1)\ =\ 17.77, p\ <\ .001$; participants with slight hearing loss $X^2\ (1)\ =\ 8.13, p =\ .0044$; participants with mild to moderate hearing loss $X^2\ (1)\ =\ 23.66, p\ <\ .001$]. None of the hearing groups showed significant differences in DS errors compared to what was expected [normal hearing participants: $X^2\ (1)\ =\ 2.66, p =\ .10$; participants with slight hearing loss $X^2\ (1)\ =\ 8.94, p =\ .0028$; participants with mild to moderate hearing loss $X^2\ (1)\ =\ 5.23, p =\ .0022$]. Similarly, none of the hearing groups showed significant differences in DW errors compared to what was expected [normal hearing participants: $X^2\ (1)\ =\ 0.060, p =\ .81$; slight hearing loss participants $X^2\ (1)\ =\ 0.068, p =\ .79$; mild to moderate hearing loss participants $X^2\ (1)\ =\ 0.17, p =\ .68$].

Discussion

This study aimed to investigate the contribution of individual listener-based factors and implicit linguistic knowledge to older listeners' speech perception in adverse conditions. Additionally, this study aimed to examine whether, and how, older adults relied on syllabic stress cues when segmenting a speech stream in adverse conditions. As expected, as listening conditions worsened, older listeners' speech recognition accuracy decreased. A decrease in speech recognition accuracy was observed in participants who had poorer hearing and working memory scores while controlling for vocabulary knowledge, non-verbal intelligence and processing speed. The statistical properties of language exhibited the largest effect on a listener's speech perception accuracy. Specifically, words that had higher lexical frequencies and larger phonological neighbourhoods resulted in more accurate word recognition scores for participants. Older adults relied on the MSS when segmenting the target speech, also making a higher proportion of insertions than expected. When analysed further, it was revealed that IS errors showed the largest increase in adverse conditions. Furthermore, the degree of hearing loss did not appear to affect listeners' patterns of segmentation of the speech stream.

4.2 Individual Factors Influencing Speech Perception in Adverse Conditions

The first section of this study investigated the contribution of linguistic and individual listener-based factors on older listeners' speech perception in adverse conditions. These individual listener-based factors consisted of: audibility, vocabulary knowledge, working memory, non-verbal intelligence and processing speed. Additionally, the linguistic factors that were investigated were PLD and word frequency.

The results, which examined these effects under four levels of SNR, showed that a strong working memory and better hearing thresholds were associated with higher word recognition scores in older adults. The study also accounted for the influence of differences in other cognitive factors by including measures of non-verbal intelligence, vocabulary knowledge, and processing speed. However, these factors had no influence on older listeners' speech perception accuracy. As expected, the study found that as SNR increased across the four levels, the average word recognition score decreased significantly. These findings were similar to those of Mattys et al. (2012) who conducted a review investigating the effects that adverse conditions have on both older and younger listener's speech recognition abilities. The review showed that, regardless of the type of degradation of the signal, there is a correlation between the degree of degradation and a listener's speech recognition accuracy. Additionally, Shojaei et al. (2016) also showed a direct relationship between older adults' speech perception accuracy and level of SNR. Consistent with the present study it was observed that as SNR increases this significantly decreases speech perception accuracy in older adults.

The findings in the current study can be explained by the glimpsing theory proposed by Cooke (2006). This states that in adverse listening conditions a listener will make use of the glimpses in the time-frequency plane, which occur due to the speech signal containing sparsely distributed high-energy signals. However, at a certain level of degradation, the listener cannot connect the acoustic-phonetic features of speech to their own segmental representations, so a failure of recognition occurs.

The inverse relationship between audibility and older listeners' speech perception accuracy in adverse conditions was also expected. The primary model showed that audibility was the greatest contributor to older adults' speech perception accuracy. This is consistent with the literature which shows that audibility is the main contributing factor towards older listeners' difficulties when listening in adverse conditions (Frisina & Frisina, 1997; Humes, 2007). Frisina and Frisina (1997) found that older listeners with presbycusis had poorer speech recognition accuracy in adverse conditions than older adults with normal hearing.

Also consistent with these findings was a review by Humes (2007) which recognized audibility as the primary contributor to an older listener's speech recognition ability in adverse conditions.

However, the current study also identified a positive relationship between an older listener's working memory and their ability to perceive speech in adverse conditions. This appears to occur due to the mismatch of the signal to the listeners' existing phonological representations when listening in adverse conditions (Rönnberg et al., 2008). The listener is therefore required to rely on explicit processing and storage, which is measured through working memory, to make sense of this mismatch of information. The findings of the current study are consistent with other studies in the literature which also found a relationship between working memory and older listeners' ability to perceive speech in adverse conditions (Ingvalson et al., 2017; Kim et al., 2020). Findings by Ingvalson et al. (2017) showed that older listeners with better hearing and a stronger working memory were able to recognize dysarthric speech more accurately than older listeners with a weaker working memory and poorer hearing. These findings mirror the present study. In addition, a study by Kim et al. (2020) found that perception of speech in noise and time altered conditions is more accurate in those with a stronger working memory than those with a weaker working memory. Consistent with the present study, this relationship remained even when hearing and age were controlled for.

The current study found no effect of vocabulary knowledge, non-verbal intelligence and processing speed. The finding of no effect of vocabulary knowledge was contrary to our hypothesis, and the findings of other similar studies (Banks et al., 2015; Ingvalson et al., 2017; McAuliffe et al., 2013). It was predicted that vocabulary knowledge would have an effect on an older listener's word recognition accuracy, and that a higher vocabulary score would result in better speech perception. It was thought that this was because listeners with a better vocabulary knowledge are better able to use cue redundancies to form a lexical hypothesis, which allows improved speech recognition accuracy (Fletcher et al., 2019). Ingvalson et al. (2017) also observed that older participants with a stronger vocabulary knowledge were able to recognize speech in adverse conditions more accurately than older adults with a weaker vocabulary knowledge. The lack of relationship between vocabulary knowledge and speech recognition accuracy in the current study may be explained by the range of scores that were obtained for the vocabulary test. In this study, vocabulary knowledge was measured using the Peabody Picture Vocabulary test (PPVT-IV) as explained in the methods section (Dunn & Dunn, 2007). Ceiling effects were observed for vocabulary scores as shown in table 1 of the results section. This resulted in a narrow range of scores from 0.93-1.0, with a score of 1 being the maximum score. It is possible that this narrow range did not allow sufficient variation amongst participants resulting in a non-significant relationship.

The present study also showed no relationship between non-verbal intelligence and an older listener's speech recognition accuracy in adverse conditions. The current study is one of the few studies in the literature that uses older participants and includes measurements of non-verbal intelligence and processing speed, in addition to vocabulary, working memory and hearing acuity. The finding of no effect of non-verbal intelligence is consistent with our hypothesis, and the literature investigating this relationship in younger participants (Fletcher et al., 2019; Tamati et al., 2013; Torkildsen et al., 2019). Fletcher et al. (2019), Tamati et al. (2013), and Torkildsen et al. (2019) all used younger participants to investigate this relationship. Each study found that non-verbal intelligence did not contribute to a younger listener's ability to perceive speech in adverse conditions. It is of importance to note that due to the correlation between non-verbal intelligence and vocabulary knowledge, working memory, and processing speed, it is helpful to include multiple cognitive and language measurements when modeling listener perception accuracy. The inclusion of different individual factors in previous models of older listeners' speech perception may account for some of the conflicting findings about the importance of non-verbal intelligence in prior literature (Heinrich et al., 2015; Humes, 2002; Kilman et al., 2015).

Lastly, the finding of no relationship between processing speed and older listeners' speech recognition in adverse conditions is also contrary to the hypothesis and to findings in literature (Neger et al., 2014; Woods et al., 2013). Neger et al. (2014) found that a listeners' speech perception threshold was determined by both processing speed and age. Similarly, Woods et al. (2013) found a positive relationship between both younger and older adults' speech in noise perception accuracy and their processing speed. This relationship was hypothesized in the present study due to the demand on processing speed that occurs when the signal becomes degraded, therefore increasing demand on an older listener's processing abilities (Pichora-Fuller, 2003). Other studies in the literature did not find a relationship between processing speed and a listener's ability to perceive speech in adverse conditions, consistent with the results of the present study (Adank et al., 2009; Sommers & Danielson, 1999). Sommers and Danielson (1999) found no relationship between processing speed and older and younger listeners' ability to perceive speech in high and low predictability situations. Additionally, Adank et al. (2009) found no relationship between processing speed and a younger listener's ability to perceive familiar and unfamiliar accents. Again, some of the different findings across studies may be due to the inclusion of different measurements of listeners' cognitive and language skills.

Although individual factors contributed towards older listeners' word recognition accuracy, their effect was relatively subtle compared to the effect of implicit linguistic knowledge investigated in this study. In the primary model, the significant effect of both PLD and word frequency on older listeners' speech reception accuracy is consistent with the NAM by Luce and Pisoni (1998). The NAM was created by Luce and Pisoni (1998) after a number of experiments that examined the relationship between the statistical properties of language and a listener's ability to perceive speech in adverse conditions. These linguistic properties included the number of words in a neighborhood, word frequency and phonetic similarity. Results showed that the number and nature of words in a phonological neighborhood affect the speed and accuracy of a listener's word recognition. This led to the NAM which states that a listener's speech perception is influenced by the number of words in an individual's lexicon that are similar to the target word. Furthermore, the NAM states that it is harder to identify uncommon words with high frequency phonological neighbours. Therefore, word frequency and PLD affect the speed and accuracy of word recognition which supports the hypothesis of the present study. The NAM model is supported by studies that find a strong relationship between word frequency, phonological similarity, and the likelihood of word recognition in adverse conditions. One such study by Fletcher et al. (2019) found a similar relationship between PLD, word frequency and accuracy of word identification which supports the results of the present study, but with a group of younger adults. Fletcher et al. (2019) investigated this relationship over the same SNR levels as the current study and found that phonological similarity is an accurate predictor of recognition at all SNR's in younger listeners. Additionally, it was found that the effect of word frequency diminishes at lower SNR's. Results of a study by Almond et al. (2013) were also consistent with the present study, finding that older adults were able to recall high frequency words better than low frequency words. However, this effect was not seen amongst younger adults, which is thought to be due to the cognitive decline that occurs with age, which affects neuronal networks that are less interconnected, and therefore the recall of low frequency words. It is interesting to note that the effects of PLD and word frequency in the current study were a very similar size to those reported by Fletcher et al., 2019, providing some evidence that word frequency may have similar effects on older and younger listeners in adverse conditions—though it is acknowledged that the older adults in this study were all required to pass cognitive screenings to meet inclusion criteria.

4.3 Segmentation Strategies in Older Listeners

The second section of this study aimed to examine whether, and how, older adults relied on syllabic stress cues when segmenting a speech stream in adverse conditions. As hypothesized, older and listeners relied on the MSS to segment the target speech—exhibiting a high number of predicted errors (IS and DW) compared to unpredicted errors (IW and DS). A particularly strong preference to IS errors was observed. Furthermore, contrary to what was hypothesized, the older adults showed a significantly stronger reliance on the use of the MSS compared to the younger adults. It was also observed that hearing did not affect older adults' preference towards segmentation strategies.

These findings have similarities to a study by McAuliffe et al. (2013) which examined both older and younger listeners' reliance on the MSS when listening to dysarthric speech. This study also found that both older and younger listeners showed reliance towards the MSS when perceiving the dysarthric speech. As noted by McAuliffe et al. (2013), these findings make sense due to the absence of contextual cues in the speech stream and the lexical 'glimpses' present. This means that the listener is likely to depend on the lexical cues for segmentation, consistent with the hierarchical model of segmentation by Mattys et al. (2005). However, in contrast to the present study, McAuliffe et al. (2013) observed that the older listeners show significantly less reliance on the MSS compared to the younger listeners. Therefore, they concluded that age affects the use of segmentation strategy, and that as we age, we become less reliant on the MSS. The current study showed the opposite effect of age in which increasing age resulted in a stronger reliance on the MSS. However, McAuliffe et al. (2013) recognized that, since audibility and vocabulary were both highly correlated with age, they were unable to conclude whether any individual differences contributed to this difference. Additionally, the older adults in the study by McAuliffe et al. (2013) had good hearing for their age (pure-tone thresholds no greater than 25 dB HL from 500-4000Hz), unlike the present study which encompassed a wider range of hearing (normal to moderate PTA 500-4000Hz average thresholds), which may have influenced this finding.

Further examination of the difference in error types between older and younger listeners showed that older adults are significantly more likely to delete a word boundary compared to the younger adults. This is inconsistent the hypothesis that older adults would be more likely to insert a word boundary compared to the younger adults. This suggests that older adults have a perceptual preference towards multisyllabic words. These findings are inconsistent with results found by McAuliffe et al. (2013) who examined differences in specific error patterns between older and younger listeners. It was found that younger listeners were more likely to delete word boundaries before weak syllables compared to older adults. Furthermore, that older listeners were more likely to insert word boundaries before weak syllables compared to younger adults.

Following this, the effect of hearing thresholds on the older listeners' LBEs was investigated—a relationship that has received little attention in the literature. It was hypothesized that older listeners with poorer hearing would be more reliant on the MSS compared to older listeners with better hearing. Overall, the three groups exhibited a similar reliance on the MSS and a similar average number of LBEs, which is contrary to what was hypothesised. Each of the hearing groups exhibited a lower proportion of deletions and a higher proportion of insertions than expected. All three groups made larger proportion of IS errors compared to what was expected which is consistent with a reliance towards MSS as discussed above. Additionally, the three groups showed no significant difference in insertions and deletions. These findings were consistent with those of Woodfield and Akeroyd (2010) which investigated the segmentation strategies of older listeners with normal hearing and those with hearing loss, and reported no difference in segmentation strategies between the three groups. It appears, from the current findings and those of Woodfield and Akeroyd (2010), that hearing acuity plays only a limited role in our ability to use metrical prosody cues for segmentation. Woodfield and Akeroyd (2010) suggest that this is due to the metrical cues being robust, therefore listeners will rely on them even at signal to noise ratios well below what is needed for full sentence recognition. Additionally, this may be due to the fact that these listeners have many years of experience with English, and therefore can still apply similar processes to parse the speech despite their hearing loss.

4.4 Clinical Implications

This work has clinical importance as its findings can have positive implications for clients with difficulty listening in adverse conditions. This difficulty is commonly seen in audiology clinics and is often met with the solution of a hearing aid. This study has highlighted the importance of considering other factors such as working memory or implicit linguistic knowledge when consulting with such a client. It is not uncommon for a client to complain of difficulties when listening in noise, but then when tested to show a near normal pure tone audiogram. Whether this is due to individual factors, implicit linguistic knowledge or possibly an inability to segment speech, it is important that the client is not discharged from care purely due to their hearing test results. Technology, such as a remote microphone system, may improve the client’s speech perception accuracy in noise by improving the signal to noise ratio. Furthermore, counselling a client about communication strategies that they can employ in adverse conditions may considerably improve their speech perception accuracy.

4.5 Limitations and Future Directions

While this study had a large number of older adults and explored areas of the literature with little research, a number of limitations have been identified. Firstly, the participant group exhibited normal through to moderate hearing ability. Participants with a hearing loss above moderate in their better hearing ear were not included. The present study found significant effects of audibility on an older listener's ability to perceive speech in adverse conditions, and no effect of hearing on segmentation ability. Therefore, it would be interesting if future studies included participants with a wider range of hearing ability.

Furthermore, it would be interesting to explore the relationship between vocabulary knowledge and speech perception in adverse conditions for older adults using a more sensitive measure of vocabulary knowledge. It is thought that the ceiling effects encountered in the Peabody Picture Vocabulary Test (PPVT-IV) from (Dunn & Dunn, 2007) limited the ability to explore this relationship, as no effect of vocabulary knowledge was indicated. It is recognized that a group with a wider range of vocabulary knowledge may also need to be investigated to examine this. In the current study, our participant group may have been skewed toward more active, curious and community-orientated older participants (e.g., people who responded to ads in the newspaper and were interested in attending a university study), and these individuals may not have reflected the general population of people over 60. A wider outreach effort may help to bring in more a varied and representative older participant group. Finally, the findings of the present study need to be replicated, particularly the second section, given the limited amount of literature in this field.

The current study included younger adults, and older adults with a range of hearing impairment from normal to moderate in their better hearing ear (500Hz to 4kHz). However, this study did not include participants with hearing aids which represent a large proportion of the population with hearing loss. Particularly the investigation of individual factors that affect hearing aid user's speech recognition in adverse conditions would be of interest. This is because it is thought that the altered signal that hearing devices provide may cause a mismatch within the cognitive processing pathway (Dryden et al., 2017). Furthermore, it would be interesting to investigate how participants with hearing aids segment speech in adverse conditions, similar to the second section of the current study. Participants with hearing aids were not able to be included in the current study due to scope limitations, however this would be an interesting direction for further research in this field.

4.6 Conclusion

In summary, the current study found that as listening conditions worsened, older listeners' speech recognition accuracy decreased. For older listeners, speech recognition in adverse conditions was influenced by hearing thresholds and working memory. No effect of vocabulary knowledge, non-verbal intelligence or processing speed were found. More significantly, older listeners' speech recognition accuracy was influenced by knowledge of the statistical properties of language, specifically phonological similarity and word frequency. Both older and younger listeners relied on MSS when segmenting the target speech, also making a higher proportion of insertions (specifically IS errors) than expected. Furthermore, there was no significant difference between older listeners divided by hearing groups on their preference towards segmentation strategies.

References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of Familiar and Unfamiliar Native Accents Under Adverse Listening Conditions. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 520-529. https://doi.org/10.1037/a0013552

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology, 47*(S2), S53-S71. https://doi.org/10.1080/14992020802301142

Almond, N. M., Morrison, C. M., & Moulin, C. J. A. (2013). Episodic intertrial learning of younger and older adults: Effects of word frequency. *Aging, Neuropsychology, and Cognition, 20*(2), 174-194. https://doi.org/10.1080/13825585.2012.679914

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445-459. https://doi.org/10.3758/BF03193014

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *The Journal of the Acoustical Society of America, 137*(4), 2015-2024. https://doi.org/10.1121/1.4916265

Benard, M. R., Susanne Mensink, J., & Başkent, D. (2014). Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities. *The Journal of the Acoustical Society of America, 135*(2), EL88-EL94. https://doi.org/10.1121/1.4862879

Bergman, M. (1971). Hearing and Aging: Implications of Recent Research Findings. *Audiology, 10*(3), 164-171. https://doi.org/10.3109/00206097109072554

Bregman, A. S. (1990). *Auditory scene analysis: the perceptual organization of sound*. MIT Press.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science, 27*(1), 45-50. https://doi.org/10.1177/0963721417727521

Centre for Speech Technology Research. (n.d.). *Unisyn lexicon* Retrieved 17 June from http://www.cstr.ed.ac.uk/projects/unisyn/

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

Cooke, M. (2006). A Glimpsing Model of Speech Perception in Noise. *Journal of the Acoustical Society of America, 119*(3), 1562-1573. https://doi.org/10.1121/1.2166600

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America, 123*(1), 414-427. https://doi.org/10.1121/1.2804952

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review, 22*(5), 1216-1234. https://doi.org/10.3758/s13423-015-0819-2

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31*(2), 218-236. https://doi.org/10.1016/0749-596X(92)90012-M

Cutler, A., & Norris, D. (1988). The Role of Strong Syllables in Segmentation for Lexical Access. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 113-121. https://doi.org/10.1037/0096-1523.14.1.113

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466. https://doi.org/10.1016/S0022-5371(80)90312-6

Desjardins, J. L., & Doherty, K. A. (2013). Age-Related Changes in Listening Effort for Various Types of Masker Noises. *Ear and Hearing, 34*(3), 261-272. https://doi.org/10.1097/AUD.0b013e31826d0ba4

Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis. *Trends in Hearing, 21*, 233121651774467-2331216517744675. https://doi.org/10.1177/2331216517744675

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* In TX: Pearson

Ebaid, D., Crewther, S. G., MacCalman, K., Brown, A., & Crewther, D. P. (2017). Cognitive Processing Speed across the Lifespan: Beyond the Influence of Motor Speed. *Frontiers in aging neuroscience, 9*, 62. https://doi.org/10.3389/fnagi.2017.00062

Fletcher, A., McAuliffe, M., Kerr, S., & Sinex, D. (2019). Effects of Vocabulary and Implicit Linguistic Knowledge on Speech Recognition in Adverse Listening Conditions. *American Journal of Audiology, 28*(3S), 742-755. https://doi.org/10.1044/2019_AJA-HEAL18-18-0169

Fontan, L., Ferrané, I., Farinas, J., Pinquier, J., & Aumont, X. (2016). Using phonologically weighted Levenshtein distances for the prediction of microscopic intelligibility.

Fostick, L., Ben-Artzi, E., & Babkoff, H. (2013). Aging and speech perception: beyond hearing threshold and cognitive ability. *Journal of Basic and Clinical Physiology and Pharmacology, 24*(3), 175-183. https://doi.org/10.1515/jbcpp-2013-0048

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

Füllgrabe, C., & Rosen, S. (2016). On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01268

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of PRESTO: A new high-variability sentence recognition test. *Journal of the American Academy of Audiology, 24*(1), 26-36. https://doi.org/10.3766/jaaa.24.1.4

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

Gordon-Salant, S. (2005). Hearing loss and aging: New research findings and clinical implications. *Journal of rehabilitation research and development, 42*(4s), 9. https://doi.org/10.1682/jrrd.2005.01.0006

Gow, D. W., & Gordon, P. C. (1995). Lexical and Prelexical Influences on Word Segmentation: Evidence From Priming. *Journal of Experimental Psychology: Human Perception and Performance, 21*(2), 344-359. https://doi.org/10.1037/0096-1523.21.2.344

Gundersen, R. O., & Feldt, L. S. (1960). The relationship of differences between verbal and nonverbal intelligence scores to achievement. *Journal of Educational Psychology, 51*(3), 115-121. https://doi.org/10.1037/h0046787

Heinrich, A., Henshaw, H., & Ferguson, M. A. (2015). The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00782

Humes, L. E. (1996). Speech understanding in the elderly. *Journal of the American Academy of Audiology, 7*(3), 161.

Humes, L. E. (2002). Factors Underlying the Speech-Recognition Performance of Elderly Hearing-Aid Wearers. *Journal of the Acoustical Society of America, 112*(3), 1112-1132. https://doi.org/10.1121/1.1499132

Humes, L. E. (2007). The Contributions of Audibility and Cognitive Factors to the Benefit Provided by Amplified Speech to Older Adults. *Journal of the American Academy of Audiology, 18*(7), 590-603. https://doi.org/10.3766/jaaa.18.7.6

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., & Lee, L. (1994). Factors Associated With Individual Differences in Clinical Measures of Speech Recognition Among the Elderly. *Journal of Speech and Hearing Research, 37*(2), 465.

Ingvalson, E. M., Lansford, K. L., Fedorova, V., & Fernandez, G. (2017). Receptive Vocabulary, Cognitive Flexibility, and Inhibitory Control Differentially Predict Older and Younger Adults' Success Perceiving Speech by Talkers with Dysarthria. *Journal of Speech, Language, and Hearing Research, 60*(12), 3632-3641. https://doi.org/10.1044/2017_JSLHR-H-17-0119

Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *The Quarterly Journal of Experimental Psychology, 65*(8), 1563-1585. https://doi.org/10.1080/17470218.2012.658822

Jarvis, B. G. (2010). *DirectRT*. In (Version 2010.3.109) Empirisoft Corporation

Jusczyk, P. W., & Luce, P. A. (2002). Speech Perception and Spoken Word Recognition: Past and Present. *Ear and Hearing, 23*(1), 2-40. https://doi.org/10.1097/00003446-200202000-00002

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition, 116*(3), 321-340. https://doi.org/10.1016/j.cognition.2010.05.011

Kilman, L., Zekveld, A. A., Hallgren, M., & Ronnberg, J. (2015). Native and Non-native Speech Perception by Hearing-Impaired Listeners in Noise- and Speech Maskers. *Trends in Hearing, 19*, 233121651557912. https://doi.org/10.1177/2331216515579127

Kim, S., Choi, I., Schwalje, A. T., Kim, K., & Lee, J. H. (2020). Auditory Working Memory Explains Variance in Speech Recognition in Older Listeners Under Adverse Listening Conditions. *Clinical interventions in aging, 15*, 395-406. https://doi.org/10.2147/CIA.S241976

Klatt, D. H. (1979). Speech perception: a model of acoustic–phonetic analysis and lexical access. *Journal of Phonetics, 7*(3), 279-312. https://doi.org/10.1016/S0095-4470(19)31059-9

Koeritzer, M. A., Rogers, C. S., Van Engen, K. J., & Peelle, J. E. (2018). The Impact of Age, Background Noise, Semantic Ambiguity, and Hearing Loss on Recognition Memory for Spoken Sentences. *Journal of Speech, Language, and Hearing Research, 61*(3), 740-751. https://doi.org/10.1044/2017_jslhr-h-17-0077

Krull, V., Humes, L. E., & Kidd, G. R. (2013). Reconstructing Wholes From Parts: Effects of Modality, Age, and Hearing Loss on Word Recognition. *Ear and Hearing, 34*(2), e14-e23. https://doi.org/10.1097/AUD.0b013e31826d0c27

LaCross, A., Liss, J., Barragan, B., Adams, A., Berisha, V., McAuliffe, M., & Fromont, R. (2016). The role of stress and word size in Spanish speech segmentation. *The Journal of the Acoustical Society of America, 140*(6), EL484-EL490. https://doi.org/10.1121/1.4971227

Lippmann, R. P. (1996). Accurate consonant perception without mid-frequency speech energy. *IEEE Transactions on Speech and Audio Processing, 4*(1), 66. https://doi.org/10.1109/TSA.1996.481454

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic Strength and Lexical Boundary Decisions in the Perception of Hypokinetic Dysarthric Speech. *Journal of the Acoustical Society of America, 104*(4), 2457-2466. https://doi.org/10.1121/1.423753

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics, 39*(3), 155-158. https://doi.org/10.3758/BF03212485

Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear & Hearing, 19*(1), 1-36. https://doi.org/10.1097/00003446-199802000-00001

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28*(2), 203-208. https://doi.org/10.3758/BF03204766

Mathworks. (n.d.). *MATLAB*. In

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203-243. https://doi.org/10.1016/j.cogpsych.2009.04.001

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes: Speech Recognition in Adverse Conditions, 27*(7-8), 953-978. https://doi.org/10.1080/01690965.2012.705006

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *Journal of Experimental Psychology: General, 134*(4), 477-500. https://doi.org/10.1037/0096-3445.134.4.477

McAuliffe, M. J., Gibson, E. M. R., Kerr, S. E., Anderson, T., & LaShell, P. J. (2013). Vocabulary influences older and younger listeners' processing of dysarthric speech. *The Journal of the Acoustical Society of America, 134*(2), 1358-1368. https://doi.org/10.1121/1.4812764

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1-86. https://doi.org/10.1016/0010-0285(86)90015-0

McCreery, R. W., Walker, E. A., Spratford, M., Lewis, D., & Brennan, M. (2019). Auditory, Cognitive, and Linguistic Factors Predict Speech Recognition in Adverse Listening Conditions for Children With Hearing Loss. *Frontiers in neuroscience, 13*, 1093. https://doi.org/10.3389/fnins.2019.01093

Moghadam, S. H., Zainal, Z., & Ghaderpour, M. (2012). A Review on the Important Role of Vocabulary Knowledge in Reading Comprehension Performance. *Procedia - Social and Behavioral Sciences, 66*, 555-563. https://doi.org/10.1016/j.sbspro.2012.11.300

Moore, D. R., Edmondson-Jones, M., Dawes, P., Fortnum, H., McCormack, A., Pierzycki, R. H., & Munro, K. J. (2014). Relation between Speech-in-Noise Threshold, Hearing Loss and Cognition from 40–69 Years of Age. *PLoS ONE, 9*(9), e107720. https://doi.org/10.1371/journal.pone.0107720

Mukari, S. Z.-M. S., Wahat, N. H. A., & Mazlan, R. (2014). Effects of ageing and hearing thresholds on speech perception in quiet and in noise perceived in different locations. *Korean journal of audiology, 18*(3), 112-118. https://doi.org/10.7874/kja.2014.18.3.112

Munson, B. (1999). Relationships between expressive vocabulary size and spoken word recognition in children. *The Journal of the Acoustical Society of America, 106*(4), 2245-2246. https://doi.org/10.1121/1.427659

Nagaraj, N. K. (2017). Working Memory and Speech Comprehension in Older Adults With Hearing Impairment. *Journal of speech, language, and hearing research : JSLHR, 60*(10), 2949-2964. https://doi.org/10.1044/2017_JSLHR-H-17-0022

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society, 53*(4), 695-699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Neger, T. M., Rietveld, A. C. M., & Janse, E. (2014). Relationship between perceptual learning in speech and statistical learning in younger and older adults. *Frontiers in Human Neuroscience, 8*, 628-617. https://doi.org/10.3389/fnhum.2014.00628

O'Neill, E. R., Kreft, H. A., & Oxenham, A. J. (2019). Cognitive factors contribute to speech perception in cochlear-implant users and age-matched normal-hearing listeners under vocoded conditions. *The Journal of the Acoustical Society of America, 146*(1), 195-210. https://doi.org/10.1121/1.5116009

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., & Kraus, N. (2011). Musical Experience and the Aging Auditory System: Implications for Cognitive Abilities and Hearing Speech in Noise. *PLoS ONE, 6*(5), e18082. https://doi.org/10.1371/journal.pone.0018082

Pichora-Fuller, K. M. (2003). Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *International Journal of Audiology, 42*(S1), 59-67. https://doi.org/10.3109/14992020309074625

Pichora-Fuller, M. K. (2009). How cognition might influence hearing aid-design, fitting, and outcomes. *The Hearing Journal, 62*(11), 32,34,36,38. https://doi.org/10.1097/01.HJ.0000364274.44847.dc

Plomp, R. (1986). A Signal-to-Noise Ratio Model for the Speech-Reception Threshold of the Hearing Impaired. *Journal of Speech and Hearing Research, 29*(2), 146.

Preston, K. A. (1935, 1935/07/01). The Speed of Word Perception and Its Relation to Reading Ability. *The Journal of General Psychology, 13*(1), 199-203. https://doi.org/10.1080/00221309.1935.9917878

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology, 47*(S2), S99-S105. https://doi.org/10.1080/14992020802301167

Rooij, J. C. G. M. v., Plomp, R., & Orlebeke, J. F. (1989). Auditive and Cognitive Factors in Speech Perception by Elderly Listeners, I: Development of Test Battery. *The Journal of the Acoustical Society of America, 86*(4), 1294-1309. https://doi.org/10.1121/1.398744

Schneider, B. A., Daneman, M., & Pichora-Fuller, M. K. (2002). Listening in Aging Adults: From Discourse Comprehension to Psychoacoustics. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 56*(3), 139-152. https://doi.org/10.1037/h0087392

Schoof, T., & Rosen, S. (2014). The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Frontiers in aging neuroscience, 6*, 307. https://doi.org/10.3389/fnagi.2014.00307

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America, 123*(1), 489-499. https://doi.org/10.1121/1.2783762

Shojaei, E., Ashayeri, H., Jafari, Z., Zarrin Dast, M. R., & Kamali, K. (2016). Effect of signal to noise ratio on the speech perception ability of older adults. *Medical journal of the Islamic Republic of Iran, 30*, 342.

Sinex, D. G. (2013). Recognition of speech in noise after application of time-frequency masks: Dependence on frequency and threshold parameters. *Journal of the Acoustical Society of America, 133*(4), 2390-2396. https://doi.org/10.1121/1.4792143

Singh, K. A., Gignac, G. E., Brydges, C. R., & Ecker, U. K. H. (2018). Working memory capacity mediates the relationship between removal and fluid intelligence. *Journal of Memory and Language, 101*, 18-36. https://doi.org/10.1016/j.jml.2018.03.002

Smith, M. R., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The Perception of Rhythm and Word Boundaries in Noise-Masked Speech. *Journal of Speech and Hearing Research, 32*(4), 912.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 32*(25), 8443-8453. https://doi.org/10.1523/JNEUROSCI.5069-11.2012

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory Processes and Spoken Word Recognition in Young and Older Adults: The Interaction of Lexical Competition and Semantic Context. *Psychology and Aging, 14*(3), 458-472. https://doi.org/10.1037/0882-7974.14.3.458

Spitzer, S. M., Liss, J. M., & Mattys, S. L. (2007). Acoustic Cues to Lexical Segmentation: A Study of Resynthesized Speech. *The Journal of the Acoustical Society of America, 122*(6), 3678-3687. https://doi.org/10.1121/1.2801545

Sprinzl, G. M., & Riechelmann, H. (2010). Current Trends in Treating Hearing Loss in Elderly People: A Review of the Technology and Treatment Options – A Mini-Review. *Gerontology, 56*(3), 351-358. https://doi.org/10.1159/000275062

Storkel, H. L. (2013). A corpus of consonant–vowel–consonant real words and nonwords: Comparison of phonotactic probability, neighborhood density, and consonant age of acquisition. *Behavior Research Methods, 45*(4), 1159-1167. https://doi.org/10.3758/s13428-012-0309-7

Suárez, L., Tan, S. H., Yap, M. J., & Goh, W. D. (2011). Observing neighborhood effects without neighbors. *Psychonomic Bulletin & Review, 18*(3), 605-611. https://doi.org/10.3758/s13423-011-0078-9

Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2013). Some factors underlying individual differences in speech recognition on PRESTO: a first report. *Journal of the American Academy of Audiology 24*(7), 616.

Team, R. (2020). *RStudio: Integrated Development for R*. In RStudio. http://www.rstudio.com/

Tompkins, C. A., Bloise, C. G. R., Timko, M. L., & Baumgaertner, A. (1994). Working Memory and Inference Revision in Brain-Damaged and Normally Aging Adults. *Journal of Speech and Hearing Research, 37*(4), 896.

Torkildsen, J. v. K., Hitchins, A., Myhrum, M., & Wie, O. B. (2019). Speech-in-Noise Perception in Children With Cochlear Implants, Hearing Aids, Developmental Language Disorder and Typical Development: The Effects of Linguistic and Cognitive Abilities. *Frontiers in Psychology, 10*, 2530. https://doi.org/10.3389/fpsyg.2019.02530

Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition, 31*(4), 491-504. https://doi.org/10.3758/BF03196091

Wechsler, D. (1981). Manual for the Wechsler adult intelligence scale—Revised (WAIS-R). In. TX: Psychological Corporation.

Wechsler, D., Coalson, D., & Raiford, S. (2008). Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV). . In *Technical and Interpretative Manual*. TX: Pearson.

Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing Loss in Older Adulthood: What It Is and How It Interacts with Cognitive Performance. *Current Directions in Psychological Science, 14*(3), 144-148. https://doi.org/10.1111/j.0963-7214.2005.00356.x

Woodfield, A., & Akeroyd, M. A. (2010). The Role of Segmentation Difficulties in Speech-in-Speech Understanding in Older and Hearing-Impaired Adults. *Journal of the Acoustical Society of America, 128*(1), EL26-EL31. https://doi.org/10.1121/1.3443570

Woods, W. S., Kalluri, S., Pentony, S., & Nooraei, N. (2013). Predicting the effect of hearing loss and audibility on amplified speech reception in a multi-talker listening scenario. *The Journal of the Acoustical Society of America, 133*(6), 4268-4278. https://doi.org/10.1121/1.4803859

Working Group on Speech Understanding and Aging. (1988). Speech understanding and aging. *The Journal of the Acoustical Society of America, 83*(3), 859-895. https://doi.org/10.1121/1.395965

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60*(4), 502-529. https://doi.org/10.1016/j.jml.2009.02.001

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language, 46*(3), 441-517. https://doi.org/10.1006/jmla.2002.2864

Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Ronnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America, 134*(3), 2225-2234. https://doi.org/10.1121/1.4817926

Zhang, H., Carlson, M. T., & Diaz, M. T. (2019). Investigating the effects of phonological neighbours on word retrieval and phonetic variation in word naming and picture naming paradigms. *Language, cognition and neuroscience, 35*(8), 1-12. https://doi.org/10.1080/23273798.2019.1686529