UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING AND THE ENVIRONMENT

Institute of Sound and Vibration Research

The Effects of Background Sound on Communication: Speech Intelligibility and Reading

by

Hannah Holmes

Thesis for the degree of Doctor of Philosophy

March 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND THE ENVIRONEMENT

INSTITUTE OF SOUND AND VIBRATION RESEARCH

Doctor of Philosophy

THE EFFECTS OF BACKGROUND SOUND ON COMMUNICATION: SPEECH INTELLIGIBILITY AND READING

by Hannah Holmes

Speech intelligibility can vary depending on the characteristics of background sound in which it is presented. Along the auditory pathway interference may occur due to a physically degraded representation of speech at the peripheral level and/or a perceptually degraded representation at higher central cognitive levels. By manipulating background sounds the level at which interference occurs can be considered. In the presence of an interfering single-talker speech background, intelligibility is sometimes improved compared to performance in a stationary noise background thought to be due to an improved physical representation of the target speech.

In children, this improvement is often found to be smaller than that compared to adults, although such findings have not always been reported. What this could suggest however, is that children may be more detrimentally affected by speech backgrounds than adults. This generally has been understood to reflect the maturation of central cognitive processes in children, where speech backgrounds may interfere with speech intelligibility at higher cognitive levels. Recent research however, proposes that this measured improvement may be subject to a signal-to-noise-ratio (SNR) confound which relates to differences in performance with the baseline stationary noise condition from which such improvement is compared. Thus previous findings could have been misinterpreted.

The first of three experiments within this thesis aimed therefore to quantify such speech intelligibility improvement amongst children and adults and to investigate the developmental trajectory over a one year period. Contributing data from large samples, it was found that the intelligibility improvement with speech backgrounds was significantly smaller in children aged 5-6 years compared to adults, and although performance got better amongst children one year later the improvement remained significantly smaller.

The second experiment aimed to further understand the effect speech backgrounds have on speech intelligibility in children by exploring the validity of found child/adult differences. This concerned the SNR confound and used analyses inspired by previous research to take into account baseline stationary noise differences. It was concluded that the difference between children and adults' intelligibility improvement with speech backgrounds lessened, yet a difference remained suggesting fundamental differences in the effects speech backgrounds have on children compared to adults. Such results may have implications for children listening in noisy classrooms.

Since speech backgrounds may interfere with speech intelligibility at higher cognitive levels, the final experiment attempted to further understand the mechanisms involved by aiming to tap into cognitive demands arising from communication in realistic listening situations. In order to do this, a reading paradigm was utilised to track eye movements during reading in the presence of differing background sounds. Whilst input from visual and auditory modalities may not cause either any physical degradation at peripheral levels, interference might occur at higher cognitive levels when processing language. Only one other study to date has used eye tracking technology to provide online insights into cognitive processing difficulty when reading amongst background speech. Therefore the final aim was to determine how different background sounds disrupt the reading process. It was found in stark contrast to speech intelligibility findings, a single-talker speech background caused more interference to the reading task compared to a stationary noise background in adult participants, suggesting that speech intelligibility measurements may not provide any information about cognitive load in everyday communication conditions.

It is concluded that children are more detrimentally affected by speech backgrounds in speech intelligibility tasks compared to adults and that speech backgrounds interfere at higher cognitive levels invoking complex cognitive processes. Further research is needed to establish how speech backgrounds affect children during reading which could have important implications for noise levels in classroom settings.

Table of Contents

List of tables

List of figures

DECLARATION OF AUTHORSHIP

I, Hannah Holmes declare that this thesis entitled “The Effects of Background Sound on Communication: Speech Intelligibility and Reading” and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission:

Signed:..

Date:...

Acknowledgements

First and foremost I would like to express my deepest gratitude and thanks to my PhD supervisor Daniel Rowan, without whom none of this would have been possible. Thank you for the writing of MATLAB codes used in all of the experiments within this PhD. For all of your encouragement support and continuous belief and for all you have done for me, I thank you.

I would also like to acknowledge Helen Cullington for her valuable guidance, feedback and advice throughout this whole process.

My appreciation is extended to Simon Liversedge for providing his expertise in the field of Psychology and providing me with the resources to enable data collection.

To Hazel Blythe I thank you for your invaluable academic support and knowledge and for answering my many questions in our very valuable meetings.

I am indebted to Hayward Godwin, Gemma Fitzsimmons and Charlotte Riggs for their programming of the reading experiment and for facilitating my research, not to mention their expertise, advice and support provided, together with countless hours they spent helping me with recruitment and data collection. I also extend my thanks to the whole of the Psychology department at the University of Southampton for all of their support.

I would also like to thank Ruth Pickering and Brian Glasberg for all their help with statistics.

I am extremely grateful to Kelly Thrift and Kirsty Walker for their help collecting data from children at the schools and to all the staff, parents and pupils of Shirley Infant and Shirley Junior Schools for their participation, hospitality and for allowing me to spend so much time collecting data.

Much appreciation goes to my examiners Steve Bell and Stuart Rosen for their valuable suggestions improving the quality of my work.

My heartfelt thanks goes to Mark for his love, patience and emotional support.

I dedicate this thesis to my Mum and Dad for their love and support and for all they have done for me.

List of Abbreviations

SNR	Signal to noise ratio
SRT	Speech reception threshold
CRM	Coordinate response measure
LTASS	Long term average speech spectrum
AFC	Alternative forced choice
CV	Consonant-vowel
ELU	Ease of language understanding
RAMBPHO	Rapid automatic multimodal bound phonologic
ISO	International organisation for standardisation
BPVS	British picture vocabulary scale
PPVT	Peabody picture vocabulary test
MLE	Maximum likely estimate
ANOVA	Analysis of variance
FMB	Fluctuating masker benefit
BKB	Bamford, Kowel and Bench
IHR	Institute of hearing research
RMS	Root mean square
CI	Confidence interval
FM	Frequency modulation
TOWRE	Test of word reading efficiency
SWE	Sight word efficiency
PDE	Phonemic decoding efficiency
M	Mean
SE	Standard error
IEEE	Institute of electrical and electronic engineers
SS	Speech-shaped

Chapter 1: Introduction

Chapter 1: Introduction

Communication in everyday situations commonly presents us with the task of listening to a target speaker amongst an acoustic mix of various background sounds. Our auditory system must separate out simultaneous signal components to decode meaning from a target of interest and suppress interfering sounds (Garadat & Litovsky 2007). Background sounds can come in various forms and are simply sounds which are not the object of desired attention. Such sounds thus could include a background which is unchanging (i.e. steady-state) or a background which changes spectrally and/or temporally (i.e. fluctuating). It may also comprise of an interfering talker or a group of several interfering talkers. Background sounds therefore may have various acoustic as well as linguistic properties which could impinge upon the intelligibility of speech in different ways.

The interfering effect of a steady-state noise on speech intelligibility is generally found to be stronger than that of a fluctuating single-talker background. This is thought to be due to its greater concentration of spectral and temporal acoustical energy physically overlapping with the acoustic properties of the target speech at the peripheral level (Festen & Plomp 1990). The interfering effect of a two-talker background however is often found to be greater than steady state noise which cannot be explained by its energy content alone, since it is more sparse than steady state noise (Rosen et al. 2013; Freyman et al. 2004). Such findings have thus been interpreted to indicate a greater involvement of higher level cognitive processes in backgrounds containing linguistic properties compared to those without. This indicates that interference from background sounds may occur anywhere along the auditory pathway from peripheral to central cognitive levels and by manipulating background sounds we could better understand the mechanisms involved.

One important example where background sounds may cause interference to speech intelligibility is amongst school aged children in the classroom setting. With several overlapping acoustic signals and reverberations of similar spectral and temporal qualities entering the ear from various locations, classroom listening conditions may be especially complex (Garadat & Litovsky 2007). Within the one classroom children may be required to listen to a teacher whilst other groups of children are practising reading aloud or simply involved in discussion. In a survey of noise levels inside the classrooms of 140 primary schools in London (Shield & Dockrell 2004), it was found that most noise comes from the voices of other children described as “classroom babble”. Depending on the nature of the lesson, and the activities being carried out “classroom babble” was found to vary over a 20 dB (A) range during the day with the average levels of occupied classrooms being at 72 dB (A). Such listening environments may be disruptive for children particularly, since several studies have shown children to require more favourable signal-to-noise ratios (SNR) than adults in order to achieve the same speech intelligibility levels with both steady-state noise backgrounds (Elliott 1979; Papso & Blood 1989; Nittrouer & Boothroyd 1990; Fallon et al. 2000; Hall et al. 2002; Litovsky 2005; Wightman & Kistler 2005; Johnstone & Litovsky 2006; Hall et al. 2012; Bonino et al. 2013; Leibold & Buss 2013) and fluctuating speech backgrounds (Hall et al. 2002; Litovsky 2005; Johnstone & Litovsky 2006; Bonino et al. 2013; Leibold & Buss 2013). Moreover, this difference in performance between children and adults observed with speech backgrounds has been found by some studies to be more pronounced than those differences seen with steady speech-shaped noise backgrounds (Hall et al. 2002; Bonino et al. 2013; Leibold & Buss 2013). It is of relevance to note however, this interaction is not always reported (Litovsky 2005) and small sample sizes have often been used. Furthermore, the findings of previous studies are difficult to compare since complex and sometimes cognitively demanding speech tasks with various complex backgrounds have been used. The cognitive skills of children in comparison to adults have also often been overlooked, so child/adult differences may be affected by such variables. The first aim of this thesis therefore is to establish the existence of any child/adult differences contributing large samples using a cognitively undemanding task, with simple background sounds and taking into account knowledge of children's vocabulary.

The finding that the discrepancy between children's and adults' speech intelligibility is larger in speech backgrounds than in steady noise backgrounds has generally been interpreted to be due to developmental differences concerning cognitive factors, attention and speech and language skills (McCreery et al. 2010). That interpretation combined with knowledge of the acoustic environment within the classroom has been taken to indicate that learning at school may be more difficult than realised previously (Leibold et al. 2007). However, that interpretation is questioned when generalising recent developments in the understanding of differences in susceptibility to fluctuating vs. steady backgrounds between hearing and hearing-impaired adult populations (Bernstein 2012). Bernstein (2012) showed that the difference between populations with the steady noise condition confounds the statistical interaction between populations and backgrounds with the most common way it is investigated. There is reason to believe that the previously reported interaction between children and adults with different backgrounds could also have arisen from this confound owing to the methods and analyses that have been used to investigate it to date. The second aim of this thesis is therefore to address this outstanding issue to determine if children and adults really do differ in terms of their ability to cope with speech background interference in speech intelligibility tasks.

Whilst speech backgrounds encapsulate realistic everyday listening conditions, and since this particular type of background interference may invoke complex cognitive processes, it is of interest to understand further the cognitive involvement. In listening studies it is difficult to separate out sensory, cognitive and linguistic interactions between the target and background (Kidd et al. 2007). It is thought beyond peripheral levels, that auditory language comprehension follows the same process as written language comprehension (Cutler & Clifton 2000). Such processing relies on working memory whereby the processing of speech is automatic and obligatory so may compete for resources (Baddeley 2012). By adopting a reading paradigm and investigating the effects background sounds have on the normal reading process therefore, we may be better able to understand how background speech interferes with speech intelligibility at higher cognitive levels when processing language. Although interference at the peripheral, sensory level will not occur when reading in background sounds, interference at higher cognitive levels seems to occur, at least with some background sounds (Martin et al. 1988; Sörqvist et al. 2010). Findings from reading tasks could thus be used to advance our understanding of why child/adult differences exist, if they do, to look further into the effects of how speech backgrounds interfere at higher cognitive levels, and understand the effects these background sounds have on communication more generally. The third aim of this thesis is therefore to investigate the effect of background sounds, especially speech, on both language processing through listening and reading.

Chapter 2 consists of a literature review to summarise how background sounds with various properties may affect speech intelligibility differently particularly with both children and adults. It aims to explain how interference from background sounds may occur with reference to theoretical frameworks and how such differences between children and adults with different background sounds could have been misinterpreted. Finally it aims to summarise theories on reading and compile the evidence on background sound interference to further advance our understanding of different background sounds on communication more generally.

Chapter 3 consists of an initial experiment to establish child/adult differences using a standard clinical speech intelligibility task with different background sounds to contribute data from

 Chapter 1: Introduction

large samples to quantify any child/adult differences and to investigate developmental issues. Chapter 4 consist of a revised experiment to further explore if such child/adult differences are authentic or due to confounding effects of the methods and analyses. Chapter 5 consists of an eye tracking experiment to determine the interfering effect different background sounds have on the reading process and written language processing and chapter 6 summaries the findings from the experimental work combined providing some conclusions with reference to implications and future research and identifies novel contributions to the literature.

The primary contribution to knowledge made by this thesis is a clearer understanding of child/adult differences seen in speech intelligibility tasks with noise and speech backgrounds. The findings of this thesis show that children (5-8 years) require improved SNRs to achieve the same intelligibility levels as adults with noise backgrounds even with simple tasks and simple background sounds and when vocabulary is controlled for. This child/adult difference is even larger with speech backgrounds and is found to be partly but not entirely due to confounding effects of the methods and analyses. Reading in background sound may provide information about how background speech interferes with the processing of language at higher cognitive levels and further our understanding of the cognitive involvement when listening to speech amongst interfering talkers. Backgrounds with distinguishable linguistic content were found to interfere the most with the normal process of reading where reading was slowed down suggesting higher level processing difficulties. Such a task may be used as a tool to investigate cognitive linguistic interference and findings have implications especially for children learning in noisy environments. See Chapter 6 for a more comprehensive overview of the original contributions made by this thesis.

Aspects of this study have been reported at various audiology research meetings:

- *Holmes H, Rowan D and Cullington HE. Speech perception in different masker types: how can differences between children and adults be interpreted? Poster presented at The Third Joint Annual Conference, Experimental and Clinical Short Papers Meetings of the British Society of Audiology. September 2012; Nottingham, UK.*

- *Holmes H, Rowan D and Cullington HE. Effects of different masker types on speech perception in children and adults. Poster presented at The Fourth Joint Annual Conference, Experimental and Clinical Short Papers Meetings of the British Society of Audiology. 2013 Sept 4 – 6; Keele, UK.*

- *Holmes H, Rowan D, Blyth HI, Fitzsimmons G, Godwin HJ, Liversedge SP, Riggs CA and Cullington HE. Reading in background noise. Poster presented at The Fifth Joint Annual Conference, Experimental and Clinical Short Papers Meetings of the British Society of Audiology. 2014 Sept 1 – 3, Keele, UK.*

Chapter 2: **Background**

2.1 Introduction

Our ability to distinguish between sounds enables us to listen to one voice amongst a background of noise or other talkers (Miller 1947). This phenomenon has long been referred to as the cocktail party problem (Cherry 1953). In a visual context it is clear how one object may obstruct the view of another. In an auditory context however this is slightly more complicated since acoustic signals are additive and combine together in a mixture (Cooke 2006). Plomp (1977) calculated that the signal to noise ratio at the average cocktail party is around 0 dB when the talker is situated 0.7 m from the listener. Whilst this listening situation may appear particularly challenging, it is a level deemed sufficient to follow the conversation (Miller 1947).

Masking is the term used to describe how one sound (i.e. the target) may be degraded by the presence of another sound (i.e. the masker) (Durlach 2006). When measuring the effects of masking, behavioural tests usually document any increase in threshold calculated for the detection or recognition of the target sound when the masker is present (Moore 2008). By measuring how much speech we are able to understand in background sound, we can investigate how people communicate in daily listening situations to further understand the mechanisms involved in coping with background sound (Mattys et al. 2012). Measures of speech intelligibility often involve recording the percentage of words correctly recalled after aural presentation. The signal to noise ratio may be changed and performance recorded whereby listeners achieve a range of scores from 0-100% with the 50% correct point considered the speech reception threshold or SRT (Miller 1947).

Masking can be said to occur at any site throughout the entire auditory pathway from lower levels, at the periphery (e.g. the cochlea), to higher levels centrally, within the brain (Durlach 2006). Therefore masking has often been discussed as occurring peripherally or centrally. Interest in central masking is growing to further establish masking effects at higher levels within the auditory system to understand cognitive (top-down) processing of speech. Moreover, it is of great importance to better determine effects of background sounds that occur in every day realistic environments, with complex backgrounds of multiple talkers (Durlach 2006). Durlach (2006) explains however, that the definitions for the mechanisms of masking effects are somewhat obscure and not clearly classified. Pollack (1975) coined the terms "energetic" and "informational" masking to describe the mechanisms of masking in terms of peripheral and central masking respectively. Durlach (2006) claims, that the mechanisms of these masking effects however, need stronger clarification, to understand the reasons behind why this masking may occur. Our ability to segregate realistic target and masker sounds will now be considered in more detail to further understand the possible interference effects of background sounds and explore the mechanisms involved.

2.2 Effect of background on speech intelligibility

2.2.1 Auditory Scene Analysis

The mechanisms behind our ability to segregate sounds are not fully established (Wightman & Kistler 2005), but theories of how we are able to listen to one sound in the presence of others will now be discussed.

Attending to a particular sound source in the auditory scene can be difficult as incoming sounds combine together to form one complex acoustical mix (Ihlefeld & Shinn-Cunningham 2008). A theory put forward by Bregman (1994) describes auditory scene analysis, and suggests that the brain and auditory system analyse the complex waveform mixture in the acoustical scene by segregating sounds into their different sources to determine what and where they have come from. The theory suggests that listeners segregate sounds in two stages. The first is a primitive stage, whereby incoming sounds are grouped based on their acoustic patterns and regularities (bottom-up influences), and the second is a schema based stage which enables the listener to build up structures over time, upon which they can advance sound source segregation owing to their experience with sounds (top-down influences). Sounds can be segregated into streams based on factors including frequency differences, spectral profile, spatial location, temporal offsets and onsets, and gender of target and masker voices (Bregman 1994). Such acoustic qualities that are common amongst incoming waveforms will tend to be grouped and streamed together as one sound source (Bregman 1994).

To study how we segregate sounds, auditory streaming paradigms can be used (Leibold 2012). An example of an auditory streaming experiment involves presenting two pure tones to a listener which are separated in frequency. When the frequency difference between these two tones is small, listeners may report hearing one stream of audio. When this frequency difference is increased, two distinct streams will be perceived (Leibold 2012). When listening to one talker in background noise, segregation can be aided as the listener can attend to this one talker based on their voice characteristics (Bonino et al. 2013). Furthermore, if the talker remains fixed throughout a speech perception test, performance will be better than if the target talker changes (Ryalls & Pisoni 1997) as the listener can maintain attention on the object to improve performance (Maddox & Shinn-Cunningham 2012).

This ability of sound source segregation has not been widely researched in children (Wightman & Kistler 2005). If we can understand how sound source segregation develops we can better understand the ways in which children may learn speech and language in noisy real world environments (Leibold et al. 2007). Using the same auditory streaming paradigm as described above, Sussman et al. (2007) found that children aged 5-10 years required frequencies to be separated more greatly than adults in order to perceive two separate streams, suggesting that sound source segregation may develop throughout childhood, although data with children is limited.

Shinn-Cunningham (2008) expands further on Bregman's theory (1994) of sound source segregation and explains that it is attention which enables us to focus on certain aspects within a signal mix. Since theories of attention on auditory perception are not fully established, ShinnCunningham (2008) uses theories from visual attention to suggest that the control of attention may be applied across different sensory modalities and thus may help to explain auditory attention.

Theories of visual attention state that objects within a visual scene are what we attend to (e.g. a person, a table, a door). In an auditory scene we can name objects as sounds we perceive to originate from one physical source, whether we are correct or not (e.g. a doorbell, a dog bark, the radio). In this way, objects may therefore be what we attend to in an auditory scene but the identification of an auditory object as opposed to a visual object is less clear. Within a visual scene short term object formation occurs based on local structures (e.g. edges, borders, contours) (Feldman 2003). This short term object formation may also occur in a similar way within an auditory scene, with the grouping together of sounds that have similar spectral and temporal characteristics (e.g. onsets, offsets, dynamic and static frequencies). Higher order perceptual effects can also contribute to the formation of whole objects in the visual scene (e.g. from knowledge of colour and texture) and in the auditory scene (e.g. by streaming together similar sounds from knowledge of pitch, location and semantics). This formation from local structures (bottom-up) and higher order effects (top-down) does not occur in a particular order but can interact going both ways (Shinn-Cunningham 2008).

The process of object formation may be disrupted if the object is amongst a similar background, causing both the object and background sounds to group together. Also masking at the periphery may occur as background sounds overlap with the target object, rendering the object indistinguishable. Shinn-Cunningham (2008) uses models of visual attention (Desimone & Duncan 1995) to explain that there is competition occurring between objects within an auditory scene. Like in a visual scene, Shinn-Cunningham (2008) proposes that in the auditory scene we are only able to perceive one object at a time, and the process of which object is attended to, i.e. object selection, depends on the salience of bottom-up and top-down influences. Using a visual analogy Shinn-Cunningham (2008) demonstrates how bottom-up salience can interfere with top-down processing. Imagine a page full of separated words equally spaced all in one colour (pale pink) with one word presented in black. Objects form, as the boundaries of words are clear (letters within words are closer together than spaces between words), but object selection is determined from bottom up salience, as the eye is drawn to the black word regardless of any directed voluntary (top-down) control. So in absence of any top-down direction (such as being told where to look, what to look at etc.) bottom-up salience will prevail, but if told to attend to a particular feature of the visual scene (e.g. to look at a particular area on the page) top-down effects can override bottom-up salience, as the word most prevalent now will not be the word in black but the word within the attended visual vicinity. So when voluntarily selecting features of a sound to listen to, topdown effects can succeed, yet object selection may fail if two objects have indistinguishable features causing the listener confusion in how best to focus attention causing the wrong object to be attended to.

Shinn-Cunningham (2008) uses another theory of visual attention which may be used to explain how we can listen to one talker in the presence of many other talkers. Change blindness explains that in a visual scene, a change is not detected due to focus being directed somewhere specific within that visual scene (Simons & Rensink 2005). In this same way, we are able to focus on one particular talker and supress others. Shinn-Cunningham (2008) proposes that we are only able to perceive one object at a time and that we must shift attention between objects (e.g. two simultaneous talkers at a party), and fill in the gaps of what we have missed using our short term memory (Warren 1970). Over time, focus on one object may be improved as auditory streaming builds up, yet in shifting attention to another object, streaming must be reset so keeping track of multiple talkers may become more challenging.

Shinn-Cunningham (2008) argues that the reasons why central masking may occur are not simply due to a similarity of the target and maskers and the uncertainty in picking out the target stimuli, but underlying these reasons why central (i.e. informational) masking may affect a listener could be due to failures in object formation and object selection. An example of when object formation may occur but object selection may fail is demonstrated in a study (Brungart & Simpson 2004) measuring speech intelligibility using the coordinate response measure (CRM) corpus whereby the target sentence and speech masker sentence follow the same structural pattern of "Ready (call sign) go to (colour) (number) now". There are 8 different call signs, 4 different colours and numbers 1-8. The listener's task was to report the correct colour and number spoken by the talker identified by a specific call sign. Looking at the errors made by the listeners it was found that many wrong words repeated by listeners consisted of words that came from the masker voice instead of the target so this shows that listeners may have been able to accurately achieve object formation but not accurately select the correct object to attend to. Such activities may require greater use of top-down processes so cognitive issues may be influential and thus significant. The point about only being able to attend to one speech sound at a time and so switching attention to help us manage in complex environments may be displayed by one study in particular (Best et al. 2006). Best et al. (2006) found, that when listeners were asked to report words from two different simultaneous speech streams when both were spatially coincident, words were able to be reported but words from each stream became intertwined and mingled with each other. When the spatial separation was increased, streams were more correctly repeated but listeners recalled fewer words overall. So when the streams are spatially coincident we may be attending to what seems more like one signal, because two sounds are not well segregated and objects are not well formed, but by providing a spatial cue performance increased as if attention was switched between the two (Shinn-Cunningham 2008). During the attention shift however, it is possible we may miss some parts that we later fill in the through use of short term working memory to replay the parts we missed to then process them (Shinn-Cunningham 2008).

Working memory may play an important role in speech intelligibility in speech backgrounds due to the necessary processing of language requiring the storing and integrating of incoming speech signals to form words and understandable speech streams (Baddeley 2012). Working memory is often described as a top down cognitive process which is limited in its capacity and resources (Francis 2010). Listening in background speech places greater demand on working memory and so it may become difficult to filter out irrelevant sound (Francis, 2010). The listener may have to inhibit irrelevant speech or process more than one talker, the former will mean inhibition skills are important and the latter will mean that cognitive capacity may be stretched and so increase the risk of limited understanding (Schneider et al. 2013). It may be thought that when listening to one talker amongst several other interfering talkers the listener can only listen to one at a time and uses buffers to fill in missed information about other sources from working memory (Conway et al. 2001). Inter-subject variability in informational masking is often large and sometimes this is attributed to differences in working memory capacity (Schneider et al. 2013). Pichora-Fuller & Singh (2006) suggest that older adults with hearing loss have to focus more on top-down processing leaving limited resources to cope with memory load when attempting to ignore background speech. Children are found to have a limited working memory capacity compared to adults (Gathercole et al. 2004) and this may relate to one aspect of why children perform more poorly than adults when it comes to perceiving speech in noise.

So it seems listening to one talker in the presence of other interfering talkers may be challenging especially if both sounds overlap at the level of the periphery causing the target to imperceptible. If both sounds are perceptible however the auditory scene must be parsed and sounds must be segregated and objects formed. The formation of target objects as separate from background objects may be made more difficult when they are not distinctive from each other. Furthermore, success in accurately attending to the target talker may also depend on accurate object selection which may be made more difficult when there are uncertainties as to which sound is the target and which sound is the background. It appears clear that both bottom-up and top-down factors may contribute to successful communication although investigating such factors independently may not be straightforward.

2.2.2 Energetic masking

The term energetic masking is often used to describe a mechanism of masking when two sounds overlap in frequency and time activating similar excitation patterns along the basilar membrane in the cochlea (Schneider et al. 2013). At the periphery, the cochlea contains an array of overlapping band-pass filters and hence the cochlea and auditory nerve code sound in terms of time and frequency. When two incoming sounds overlap spectrally and temporally, they may be passed through the same or overlapping auditory filters, thus representation of the target sound may become degraded at the peripheral level and energetic masking may then occur (Shinn-Cunningham 2008). This energetic masking can be compared to the way that a visual object obscures another visual object placed behind it in a visual scene so it becomes imperceptible (Shinn-Cunningham 2008).

Models of these auditory band-pass filters in the periphery can be used to predict the extent of energetic masking by calculating the amount of energy and thus overlap in each frequency band (Moore & Glasberg 1987). These models can predict successfully the masking effect with maskers which may overlap speech like steady-state noise maskers (Wightman et al. 2010) to show that with decreasing signal to noise ratios and greater spectral and temporal overlap between the target and masker, performance on speech intelligibility tests decreases monotonically (Brungart 2001). The long term average speech spectrum (LTASS) can be used to describe the energy within the speech signal (Miller 1947). Most of the energy within the LTASS lies in the low frequencies below 1000 Hz, although the instantaneous spectrum is constantly changing (Miller 1947). Optimal masking will thus occur with a noise that has most energy in the lower frequency range from 100 to 4000 Hz, similar to the LTASS (Miller 1947). Steady speech-shaped noise could be considered an effective masker since it contains continuous spectral and temporal qualities and covers a wide range of frequencies (Miller 1947). In realistic conditions, talkers may change their voice quality and timing according to the background to help improve signal to noise ratio and reduce effects of energetic masking (Mattys et al. 2012). The slope measured in psychometric functions for performance against signal to noise ratio in laboratory conditions is usually quite steep as the overlap of the masker eventually renders the signal inaudible (Freyman et al. 2004).

Whilst models may predict how much energetic masking interferes with target identification, in instances when performance is worse than predicted, or in cases where spectral and temporal overlap is restricted, energetic masking becomes no longer the limiting factor in performance and other mechanisms may be responsible (Shinn-Cunningham 2008).

2.2.3 Informational masking

Masking with speech backgrounds encompass mechanisms usually known as informational masking but also sometimes referred to as central masking or perceptual masking (Schneider et al. 2013). The term informational masking has come to be used to encompass all masking effects which cannot be attributed to be energetic masking, however this definition is very broad and explains what informational masking is not, rather than what it actually is (Durlach 2006). It is not clear from such definitions what mechanisms informational masking is actually referring to and whether there is need for such a term at any rate. Durlach (2006) explains that when considering a definition of energetic masking which refers to the overlap of sounds producing degradation in the target, informational masking may also just be energetic masking which occurs at higher auditory sites along the auditory pathway, where auditory neurons are busy with one stimulus so that they cannot accurately represent another (Kidd et al. 2007).

Nevertheless, informational masking has been referred to in the literature as masking occurring at higher levels in the auditory system in addition to energetic masking, due to similarities between the target and masker and uncertainties about the masker, making it challenging for the listener to hear out the target (Durlach et al. 2003). This term has been used broadly; both in paradigms where tones are the targets and maskers, and where speech sounds are the targets and maskers. For example, elevation in threshold detection of a target tone is sometimes found when a tonal masker does not contain overlapping spectral and temporal qualities (limited energetic masking), but varies in frequency over time from trial to trial, causing the listener uncertainty in determining the masker from the target (Neff & Green 1987).

Another example is when speech maskers are used instead of steady-state noise maskers to mask speech. Speech maskers may reduce effects of energetic masking, due to their spectral and temporal fluctuations and sparseness owing to low energy consonant sounds and pauses between words enabling the listener to hear “glimpses” of the target and improve speech intelligibility (Miller & Licklider 1948; Howard-Jones & Rosen 1993; Freyman et al. 2004). Using top-down cognitive influences like buffers may keep representations in working memory to restore the missing parts (Warren 1970). Despite this however, in some cases, more masking may be shown with speech maskers compared to steady-state noise maskers (Brungart 2001; Hall et al. 2002). When changing the background sound from a single-talker speech masker to a two-talker speech masker Carhart et al. (1969) found there to be an 11 dB difference in speech reception thresholds (SRTs) which the authors suggested could not be predicted based on the additive energetic summation of the energies of two single-talker maskers, reflecting a masking mechanisms in addition to energetic masking possibly related to difficulties in picking out the target in a similar background .

Like with a frequency varying tonal masker, speech maskers are thought to be associated with informational masking, due to the uncertainty caused from the unpredictability of speech maskers in spectral and temporal configurations compared to steady-state noise maskers, meaning that the listener may not be able to build up any kind of knowledge about the masker which may make segregating the two sounds difficult (Stickney et al. 2004). Also, similarity between a speech target and speech masker may make it difficult to direct attention towards the target (Carhart et al. 1969). Furthermore there could be interference at higher cognitive levels owing to the linguistic content of speech backgrounds (Schneider et al. 2013).

In order to achieve successful communication, speech must be mapped from acoustic input to mental representations of meaning which can be comprehended (Ryalls & Pisoni 1997). Acoustic speech patterns are analysed as phonemes, which access the mental representations of words and language (the lexicon) to unlock further linguistic processing (Ryalls & Pisoni 1997). Speech backgrounds may have a particular interfering effect due to their linguistic content which seems to have some obligatory access capturing our attention (Banbury et al. 2001) and may interfere with the accurate linguistic processing of the target speech.

The difficulty with the description of informational masking is that there are no measures of similarity or uncertainty, or models which take both these factors into account (Durlach 2006). Furthermore causes of this informational masking are likely to encompass many stages of auditory processing anywhere from the periphery onwards, so may comprise effects of sound source segregation, attention, memory and general cognitive processes where contributions from each effect may be hard to pin down (Kidd et al. 2007). Informational masking studies have also shown that when segregation cues are present, e.g. when the target and masker are spatially separated, improvement is more likely to be seen with informational maskers than energetic maskers (Arbogast et al. 2002; Freyman et al. 2004; Bernstein & Brungart 2011).

In contrast to the contributions of energetic masking, informational masking (i.e. masking with speech backgrounds) often shows measured psychometric functions displaying shallow slopes compared to the steeper slopes seen with noise maskers (Brungart 2001; Arbogast et al. 2002; Freyman et al. 1999; Freyman et al. 2004; Oxenham & Simonson 2009). Therefore a small improvement in signal to noise ratio yields only a small improvement in performance. Often psychometric functions may be shallower for sparse maskers (i.e. speech) where glimpses can occur, as such glimpses may be even more important at lower SNRs causing a flattening of the slope (MacPherson 2013). Shallower slopes may also occur owing to the confusion between the target and masker limiting the use of top down influences. Small improvements in bottomup influences (i.e. increased signal to noise ratio) may not be able to combine with top-down influences (i.e. contextual constraints) to improve performance and steepen the slope, thus the slopes may be shallower (MacPherson 2013).

Other factors which characterise informational masking are that larger individual differences are often observed compared to those observed with energetic maskers (Lutfi et al. 2003; Wightman et al. 2003; Kidd et al. 2007). In addition, studies investigating informational masking with both tonal stimuli and speech stimuli appear to show that children are more susceptible to the effects of informational masking than adults (Wightman et al. 2003; Hall et al. 2002). Due to the broad definition of informational masking it is however, hard to know what exactly is responsible for such trends. Moreover, it is difficult to isolate the effects of informational masking from energetic masking, especially in experiments capturing realistic listening conditions, since both effects are likely to occur together (Brungart 2001).

2.2.4 **Release from masking**

A release from masking refers to any situation whereby the effects of masking can be reduced. Investigating what factors (i.e. acoustical/linguistic manipulations) can lessen the masking effects from different background sounds, may provide information about the mechanisms of different types of masking and the origin of whereabouts such interference is taking place (Freyman et al. 2004).

One manipulation of background sounds that can cause a release from masking (as discussed previously) are spectral and temporal fluctuations. This has sometimes been referred to as the fluctuating masker benefit (FMB) and is often apparent when a stationary noise masker is changed to a single-talker speech masker. Festen & Plomp (1990) found an improvement in speech reception thresholds of 4-6 dB when a stationary noise masker was changed to a modulated noise masker (with amplitude modulations) and 6-8 dB improvement when a stationary noise was changed to a single-talker speech masker. Similarly Duquesnoy (1983) and Peters et al. (1998) found a 7 dB and 8 dB difference between noise and speech maskers respectively. It could be said that such improvement is primarily due to a reduction in peripheral energetic masking because the modulated noise and single-talker speech maskers are less sparse and so contain less overlapping acoustical energies. When the numbers of talkers in the masker mix increase however, this fluctuating masker benefit appears to become reduced as energetic masking is thought to increase (Bronkhorst 2000).

When the spectral separation between target speech and masker speech is increased there is less physical overlap and this can help the auditory system segregate the sound, as opposed to perceiving a fusion of these sounds (Bregman 1994). Such a release from masking may be primarily a release from peripheral level energetic masking, but there could also be a release from cognitive level informational masking as the two speech sounds become less alike (Schneider et al. 2013). For example, in speech intelligibility tasks where the target and masker talker consisted of different voices or different genders, ability to repeat back the target sentence was improved compared to conditions where the voices were the same gender or the same talker (Brungart 2001; Brungart et al. 2001).

Spatially separating two sounds is another manipulation that can provide a release from masking (Ihlefeld & Shinn-Cunningham 2008). When sounds are spatially separated listeners can make use of binaural factors like inter-aural time and level differences within the target and masker to achieve some benefit and improve segregation of sounds (Moore & Gockel 2012). This may be due to a release from energetic masking, whereby speech can be attended to at the ear with the better signal to noise ratio owing to the head shadow effect (Schneider et al. 2013), but such a manipulation however, does not allow us to look at the effects of informational masking alone. In order to measure the effects of spatial separation on informational masking alone Freyman et al. (1999) developed a paradigm to induce a perceived spatial separation and ensure there would be no actual release from energetic masking, so any improvements in performance could be thought of as a release from informational masking. In experiments by Freyman et al. (1999; 2001) speech intelligibility of nonsense sentences was investigated in the presence of what is thought a primarily energetic masker (speech-shaped noise) and a primarily informational masker (single and two-talker maskers). In order to examine any release from masking a perceived spatial separation was induced, by making use of the precedence effect to present the masker at a speaker 60 degrees to the right leading the target onset by 4 ms, and compared to a coincidental location.

A benefit in perceived spatial separation was seen with the single and two-talker masker backgrounds but not for the noise masker. The authors suggest that there was no release in energetic masking but informational masking may be diminished where factors better improve attention on the target. Thus this can be said to tell us something about the mechanism of informational masking, that it may originate from problems in directing attention (Freyman et al. 2004), which fits in with Shinn-Cunningham's (2008) theories on how auditory attention acts on auditory objects. A further experiment in the study by Freyman et al. (2004) using similar methods showed that as the number of background talkers increased from two to ten, the improvements in perceived spatial separation and thus release from informational masking decreased, which the authors attributed to the background masking becoming less informational and more energetic.

In a further study investigating factors which provide a release from masking, speech intelligibility was investigated to examine the beneficial effect of visual speech cues by comparing performance in speech-shaped noise backgrounds (the energetic masker) and in two-talker speech backgrounds (the informational masker) (Helfer & Freyman 2005). Release from masking was examined by comparing performance in audio only conditions with an audio-visual condition (consisting of a video recording of the talkers face speaking the target sentence). It was found that compared to the audio only condition the audio-visual condition improved performance by 3 dB with the noise masker but by 9 dB with the speech masker. Therefore it can be said that the visual speech cues directed attention towards the target and since the improvement was largest in the informational masking condition, it tells us again that top-down cognitive factors of attention perhaps may be contributing to informational masking (Schneider et al. 2013).

Another factor which may be thought to provide a release from masking is a priori knowledge of the target, including the target voice, target context and target location (Brungart 2001; Kidd et al. 2005; Schneider et al. 2013). Prior knowledge of the talkers voice is thought to aid segregation and also prior knowledge of the topic which may be because top-down contextual cues enable missing or degraded parts of speech to be filled in (Schneider et al. 2013). In another experiment in the study by Freyman et al. (2004) they investigated speech intelligibility of nonsense sentences in either a speech-shaped noise or a two-talker masker. The listener's task was to repeat back the final word in the sentence. In some conditions the sentences were presented without any prior prime, but in other conditions the sentences were presented with a prime beforehand. The prime consisted of a presentation of the subsequent target sentence presented in quiet with the final word omitted and replaced with noise. Because the sentences were nonsense sentences it was thought that the final word could not be predicted by contextual cues. It was found that the prime increased speech intelligibility by 4 dB with the two-talker masker but only 1.3 dB with the speech-shaped noise masker. This finding was replicated even when the prime consisted of the sentence spoken by a talker different to the one in the target sentence or even if the prime was read. Therefore it is suggested that knowledge of the words help the listener to focus on the target, and informational masking (thus occurring from speech maskers and not noise maskers) can be reduced. So cues which can help focus attention are beneficial in informational masking conditions and less so in energetic masking conditions (Schneider et al. 2013), suggesting again that interference from informational masking may be occurring at higher cognitive levels.

From studies looking at the release from masking it can be seen that unlike energetic masking, the interfering effects of informational masking may be overcome by factors that increase attention to the target (Mattys et al. 2012). Therefore in listening conditions with greater perceptual demands and increased informational masking, cues which can improve attention towards the target may facilitate the use of top-down processing (Freyman et al. 2004).

2.3 Effect of background on speech intelligibility: child/adult differences

2.3.1 Noise backgrounds

Many studies have shown that children require more favourable signal to noise ratios (SNR) than adults in speech intelligibility tasks with steady-state noise maskers (Elliott 1979; Papso & Blood 1989; Nittrouer & Boothroyd 1990; Fallon et al. 2000; Hall et al. 2002; Litovsky 2005; Wightman & Kistler 2005; Johnstone & Litovsky 2006; Hall et al. 2012; Bonino et al. 2013; Leibold & Buss 2013). This effect is shown to be developmental in children as preschool children are more affected by noise than school aged children (Wightman et al. 2003). Furthermore, many studies agree that by around 8 or 9 years of age children perform at adultlike levels with noise maskers (Elliott 1979; Papso & Blood 1989; Bonino et al. 2013).

In a study by Nittrouer & Boothroyd (1990) children (4-6 years) needed a 3 dB increase in signal to noise ratio in order to perform at the same level as adults when recognising phonemes in speech-shaped noise. The findings of this study agree with those from Blandy & Lutman (2005) who also found children (7-8 years) needed a 3 dB improvement in SNR to achieve intelligibility levels comparable to adults. Blandy & Lutman (2005) carried out a large study investigating the effect of speech-shaped noise on the speech intelligibility of target sentences (male talker) in an open set task. The results from 189 children were compared with results from 17 adults (22-29 years) from a previous identical study by Cattermole (2003).

Speech reception threshold was measured in an adaptive procedure to yield 70% correct performance. A significant difference in performance was found with mean SRTs (dB SNR) of -3.9 dB in children and -6.9 dB in adults. The authors explained that the significant difference was seen between children and adults despite children having equal or better hearing threshold levels. Therefore it was suggested that poorer speech intelligibility in children in the noise background may not be attributable to any poorer peripheral auditory processing but possible immature central auditory processing. Similarly Hall et al. (2012) also found that 4-6 year old children (n=10) performed at comparable speech intelligibility levels (50% correct performance) as adults (n=10) only when the SNR was significantly higher for children by 3 dB (at -2.3 dB SNR) than that for adults (at -5.3 dB SNR) with target sentences presented in speech-shaped noise conditions.

Vaillancourt et al. (2008) investigated child/adult differences using a group of children across a slightly wider age range, in order to observe the developmental trajectory. Speech intelligibility of target sentences was investigated in 56 children (6-12 years) and 14 adults (18-30 years) in the presence of speech-shaped noise. In an adaptive procedure it was found that performance improved with increasing age with significant differences between the youngest (6 years) and oldest (12 years) child groups. No significant difference was found between adults and the oldest group of children suggesting that by age 12 children perform at adult like levels in speech-shaped noise. The authors ascribe their findings to developmental effects in auditory processing as well as possible developments in linguistic knowledge.

Some studies investigating child/adult differences have looked at speech intelligibility in the presence of steady broadband noise as well as an amplitude modulated noise (Stuart 2005; Stuart 2008). In a study by Stuart (2005) 80 children (aged 6-15 years) and 16 adults were required to identify words in an open set task presented over earphones. The stimuli were presented over 5 fixed SNRs of -20, -10, 0, and 10 dB. As expected, performance was found to be better in the amplitude modulated condition which likely reflects a release from energetic masking owing to the dips within the masker enabling “glimpses” of the target (Cooke 2006). Like Vaillancourt et al. (2008), Stuart (2005) found that performance was shown to improve significantly with increasing age and children performed adult like after 11 years of age. Whilst there was a trend for children to perform disproportionately poorer than adults in the amplitude modulated noise condition compared to the steady noise condition, there was no significant interaction suggesting that both backgrounds affected children and adults to the same extent. In a later study similar study by Stuart (2008) comparable findings were observed using sentence stimuli presented in an adaptive procedure where adult like performance was achieved after 14-15 years of age.

Whilst child/adult differences are seen for speech perception performance in the (mostly energetic) noise maskers, child/adult differences become more pronounced in more complex listening conditions, particularly where aspects of informational masking come into play (Hall et al. 2002; Wightman et al. 2003). In realistic listening conditions, masking effects from energetic and informational masking are likely to occur together, so it is difficult to test contribution from each masker type separately. Studies investigating informational masking have therefore used paradigms where the target and masker are tonal stimuli which vary randomly on a trial to trial basis but are well separated in frequency, making it less likely that representation of these sounds will overlap in the cochlea thus reducing effects of energetic masking. In this way, it is thought that informational masking can be addressed as a separate entity (Leibold & Bonino 2009). Although such research does not describe speech intelligibility, it is noteworthy to report that several studies have investigated this topic in children and have found that school aged children display a greater susceptibility to informational masking than adults (Allen & Wightman 1995; Oh et al. 2001; Lutfi et al. 2003; Wightman et al. 2003; Hall et al. 2005; Leibold & Bonino 2009).

2.3.2 **Multi-talker babble**

Some studies have investigated child/adult differences in speech intelligibility against a background of multi-talker babble and found that younger children perform more poorly than adults. In an early study by Elliot (1979) speech intelligibility of 24 children (aged 11, 13, 15 and 17 years) for the recognition of monosyllabic words (male talker) was investigated in a quiet condition and in the presence of a multi-talker babble (consisting of 12 male talkers) presented over earphones at three fixed signal to noise ratios of -5, 0 and +5 dB. There were no significant differences in percent correct across all age ranges in the quiet condition. It was found that performance in the babble condition improved with increasing age. The authors suggest that differences between children and adults may occur due to greater effects of masking from babble noise in children and their limited knowledge of language.

In a later study by Papso and Blood (1989) speech intelligibility was compared between 30 younger children (aged 4-5 years) and 30 adults (19-28 years) in the presence of a multi-talker babble (consisting of 20 talkers) as well as a speech-shaped noise background. Masker*age interactions were investigated to determine if children were affected by speech babble to the same extent as adults. Target speech consisted of words within a carrier phrase and participants were required in a 6 alternative forced choice (AFC) task to select the correct picture matching the target word. The target-masker complex was presented in the sound field at 6 dB SNR and percentage correct was recorded. A significant masker*age interaction was found, where children displayed poorer performance compared to adults which was disproportionately poorer with the multi-talker compared to the noise masker condition. Therefore children were found to be more detrimentally affected than adults by a speech masker compared to a noise masker. It is possible however, that ceiling effects were present in this study since adults performed very highly in all three conditions possibly yielding a false masker*age interaction. Furthermore, since only one SNR (and only one part of the psychometric function) was tested and since results are dependent on SNR (Bernstein 2012), it may be likely that the differences found between children and adults in the noise and speech conditions were confounded due to the differences between children and adults in the noise (baseline) conditions (see Chapter 4 section 4.1.1 for more details). So the masker*age interaction and the conclusion that children are more detrimentally affected by the speech babble masker than adults should perhaps be interpreted with caution.

In a study by Fallon (2000) speech intelligibility of 24 children in each of three different age groups (5, 9, and 11 years) and 24 adults (19-28 years) was investigated. Target words within a carrier phrase were presented in the sound field in the presence of a multi-talker babble (consisting of 8 talkers both male and female). In a 4AFC task participants were required to match the correct picture to the target word. For each background noise condition the target was presented at chosen fixed SNRs expected to give 85% correct performance in what was called a low noise condition, tailored to the age group (-28, -30, -31, -33 dB SNR for 5, 9, 11 year olds and adults respectively). A high noise condition was also presented whereby the SNR decreased from that presented in the low noise condition by 7 dB. Percentage correct was as expected suggesting that children aged 5 required a 5 dB higher SNR than adults to achieve comparable intelligibility levels. It was also found that performance in the low noise and high noise conditions with children and adults were comparable. The authors suggest that since children were not more greatly affected by the high noise condition than adults, age related differences may largely be due to differences in hearing sensitivity as opposed to cognitive differences.

In one experiment from a study by Bonino et al. (2013) 8 younger children (5-7 years), 8 older children (8-10 years) and 10 adults (18-30 years) were tested on their ability to identify words in a multi-talker babble consisting of 20 talkers. It was found that younger children performed more poorly than older children and adults and the performance between adults and older children was not significantly different. This suggests that younger children may be more detrimentally affected by the babble masker but by 8-10 years of age performance becomes adult like. This experiment again however, only considered one fixed SNR so therefore only one part of the psychometric function and the result is dependent on SNR (Bernstein 2012). So there could be here possible confounds relating to the SNRs tested (see Chapter 4 section 4.1.1 for more details).

2.3.3 Single-talker/two-talkers

When a stationary noise masker is switched to a fluctuating masker (such as speech itself), performance on speech intelligibility tasks are shown to improve which may be thought to be due to the spectral and temporal dips and sparseness within the masker (Festen & Plomp 1990). As previously discussed, this can provide a release from energetic masking and enable the listener to listen in the gaps to "glimpse" parts of the signal (Festen & Plomp 1990). Whilst a single-talker speech masker (associated with informational masking) with a gender different to the speech target can yield improved speech intelligibility results (e.g. Brungart 2001), compared to those in stationary noise (associated with energetic masking), studies with children show interesting results with speech backgrounds.

Few studies have actually investigated speech intelligibility with speech targets in speech backgrounds (i.e. single-talker, or two-talker backgrounds) in children, and those that have, demonstrate some inconsistencies across results. A study by Hall et al. (2002) investigated the ability of nineteen children (5-10 years) and fourteen adults (19-48 years) to identify target spondee words (two monosyllabic words which together form one two-syllable word, spoken by a male talker) in the presence of both steady state speech-shaped noise and a two-talker masker (consisting of two simultaneously presented sentences spoken by male talkers). In a 4AFC task where participants had to choose the correct target word picture, SRTs were recorded using an adaptive procedure targeted to measure the 79.4% correct level.

A main effect of age was found with children performing more poorly than adults. A main effect of masker was also found with performance being poorest in the two-talker condition compared to the noise. The deterioration in performance found when switching the speechshaped noise to the two-talker masker was considered to quantify the amount of informational masking. The authors explained that when results from the two-talker masker were compared to those from the speech-shaped noise masker there was a significant masker*age interaction whereby the adverse effects of the speech maskers were even greater in children compared to adults, with children showing a larger effect from informational masking (6.7. dB), compared to the adults (2.3. dB). Performance significantly increased in children with increasing age in the noise masker condition and although there was a trend for improved performance with increasing age in children with the speech masker this did not reach significance, and showed large individual differences. The authors suggested the results showed children to be more detrimentally affected by speech backgrounds than adults and they also suggested a possible developmental delay in ability to process speech with the speech background. The results however were compared at the same intelligibility level (79.4% correct) and again therefore only one part of the psychometric function was examined. So the deterioration in results with the two-talker background compared to the speech-shaped noise masker could be due to differences in the baseline (noise) conditions between children and adults, and that children may be poorer than adults on speech intelligibility tasks in general and not specifically more affected by speech backgrounds (see Chapter 4 section 4.1.1 for more details).

A similar study by Litovsky (2005) described a method suggested to be used in paediatric settings to assess children’s functional hearing due to its simplicity and swift delivery. Litovsky (2005), like Hall et al (2002) investigated spondee recognition. Background noise conditions included a quiet condition, a speech modulated noise (which fluctuated in amplitude in accordance with either a female single-talker or female two-talker speech masker) and a speech masker (consisting of either a female single-talker or female two-talker masker). Litovsky (2005) investigated speech intelligibility of spondee words (male talker) in noise in 36 children (aged 4.5-7.5 years) and in 9 adults. In a 4AFC task where participants had to choose the correct target word picture, SRTs were recorded using an adaptive procedure targeted to measure the 79.4% correct level. Adults underwent all masker conditions, but children were split into 4 groups and each group (n=9), underwent one of the 4 masker conditions.

It was found that noise maskers were more effective maskers than speech maskers, which contrasts with the results of Hall et al (2002) who found the two-talker masker to be more effective than the speech-shaped noise. This result may be because informational masking with the single-talker speech masker is not the limiting factor. The fact that Litovsky (2005), unlike Hall et al. (2002) used a speech masker of a different gender to the speech target, may have provided a strong enough segregation cue whereby listeners could use the fundamental frequency differences in the voice quality to separate the two sounds which agrees with results from Brungart et al. (2001).

Considering child/adult differences, SRTs were better overall in adults than in children. There were found to be no significant difference in SRTs between noise and speech maskers in children, so it may not be the speech content of the background that is particularly interfering but could just be the background modulation. Interestingly, Litovsky (2005) found that the amount of masking (as calculated by the difference between performance in quiet and in the presence of a masker) in children and adults was comparable in both noise and speech masker types, however the authors state that this could be due to the simplicity of the 4AFC used, producing ceiling effects in the adults' performance in the quiet condition. Moreover, since a repeated measures design was not used, and not all children underwent all conditions, direct masker*age interactions were therefore not reported and could not be analysed. Therefore, it would be of interest to examine the speech intelligibility performance of children and adults, in a similar procedure to that of Litovsky (2005) by adapting a standard clinical test to incorporate masking from a speech masker in order to determine directly any masker*age interactions.

A further study by Johnstone and Litovsky (2006), leading on from Litovsky (2005), investigated the ability of 20 children (5-7 years) and adults to discriminate spondees (male talker) in both single-talker speech (female talker), speech modulated noise (amplitude modulated based on a single female talker), and speech which was reversed in time. The same methods as Litovsky (2005) were used except with a 25AFC task for the adults and a 4AFC task for the children in attempts to prevent ceiling effects in adults. It was found that like Litovsky

(2005) performance was poorer in speech modulated noise than either speech masker in adults. Similarly again to Litovsky (2005), in children there was no difference in performance with either masker type. Unlike Litovsky (2005) the amount of masking was greater in children with the speech maskers compared to adults suggesting a greater susceptibility to informational masking in children. Moreover, it was found that children performed worse with the time reversed speech masker which the authors attribute to the notion that the time reversed speech was somewhat of a novelty for the children and so was more distracting to them than adults who may not have been as distracted. The authors propose that this may reflect a development of central auditory processing and auditory attention in children. Results for the different masker types were however compared at the same intelligibility level, therefore only one part of the psychometric function was examined so it seems again here that there could have been a possible confound in the results relating to differences between children and adults in the noise conditions (see Chapter 4 section 4.1.1 for more details).

Litovsky (2005) states that the reasons why children may have found it more difficult than adults with the speech maskers is not likely to be due to difficulties segregating fundamental frequency differences between the target and masker voices, since abilities in children's frequency resolution are found to be comparable to adults by 4 years of age (Hall & Grose 1994). Johnstone and Litovksy (2006) claim that these difficulties may be due to the notion that a child's central auditory processes needed for the segregation of sound sources is still developing.

Whilst some studies have shown child/adult differences appear larger with a speech masker compared to a speech-shaped noise masker, it also appears that research investigating performance in different age groups has observed a prolonged course of development when listening with the speech masker compared to the noise masker (Leech et al. 2007). Hall et al. (2002), for example observed a trend for spondee word recognition in the presence of a speech-shaped noise masker to significantly improve across a 5 year age range in a group of 19 children spanning from 5-10 years of age. In contrast, although there was a trend with the same group of children for better performance in a two-talker speech masker with increasing age, significant correlation here was not found. The authors explain that this may be because children experience a delay in their development to process speech in the more challenging listening condition of a speech background compared to a speech-shaped noise background. Larger individual differences are found with speech maskers and in some cases it can take up until the teenage years before performance is adult like (Wightman & Kistler 2005) with some young children performing adult like at 5 years of age and some still performing child like at 14 years of age (Wightman et al. 2010).

In a series of studies carried out by Wightman and colleagues (2005; 2006; 2010) the CRM corpus was used to investigate differences in how children and adults may be affected by informational masking, as separate from energetic masking, by using a dichotic listening task with target and maskers presented to separate ears. As previously discussed the CRM corpus consists of sentences used as both the target and masker sentences which have the same sentence structure but consist of a different call sign, colour and number which must be identified in a closed set task. Results from such studies have shown when target and masker are presented to the same ear, children perform more poorly than adults, needing higher SNRs, with even children as old as 16 years not reaching adult like performance levels and children displaying large individual differences (Wightman & Kistler 2005; Wightman et al. 2006). When presenting a speech-shaped noise masker to the contralateral ear, performance remained unaffected in children and adults (Wightman & Kistler 2005). When the contralateral masker was another CRM sentence however, performance was reduced in the youngest group of children (4-5 years). An analysis of errors showed most incorrect responses to come from the message in the contralateral ear. The authors suggest therefore that since error patterns occur mostly from the distractor and not at random, effects of informational masking are present as the listener is able to hear out the words (so they are not being masked energetically) but cannot accurately select the correct talker. It is interpreted thus that young children may be more detrimentally affected by informational masking and lack attentional strategies.

In a later study by the same authors (Wightman et al. 2010) a baseline condition was established to prevent floor and ceiling effects possibly seen in previous experiments. A target sentence was presented monaurally together with an amplitude modulated speech-shaped noise at a fixed SNR chosen to yield 51% correct baseline performance. Three background noise conditions (either an amplitude modulated speech-shaped noise, a same sex CRM sentence or a different sex CRM sentence) were then presented to the contralateral ear and a shift in baseline performance was recorded. It was found that speech modulated noise in the contralateral ear did not affect performance in children or adults. The single-talker speech maskers had a detrimental effect on performance by up to 20 dB in the youngest group of children (5-8 years) and by up to 4 dB in adults, with comparable results between male and female maskers. Large individual differences were observed, particularly in children between 7-12 years. From the results of these studies combined, the authors speculate that the larger detrimental effect of the single-talker maskers in children may be due to development of selective attention, possibly developing at different rates.

Bonino et al. (2013) investigated performance on an open set word recognition task whereby 3 participant groups aged either 5-7 years (n=9), 8-10 years (n=9) and adults (n=16) were required to identify words presented in either a speech-shaped noise or a two-talker speech masker. A significant masker*age interaction was found, with children performing disproportionately poorer than adults with the speech masker compared to the noise masker, suggesting that children may be more susceptible to informational masking. A study by Leibold and Buss (2013), using the same masker types also found a similar significant masker*age interaction with 62 children (5-13 years) and 28 adults. It is suggested that children may show immaturity in their ability to segregate speech sounds and maybe more detrimentally affected by speech backgrounds than adults. In both studies however, as in the other previous studies (e.g. Hall et al. 2002; Johnstone & Litovsky 2006) performance was considered only at one SNR, therefore only one part of the psychometric function was examined so differences found between children and adults with speech backgrounds could occur due to differences relating to the SNR and differences in the baseline speech-shaped noise conditions (see Chapter 4 section 4.1.1 for more details).

Concerning the developmental trajectory, whilst Bonino et al. (2013) found a significant difference in performance between 5-7 year olds and 8-10 year olds with a speech-shaped noise masker, no significant difference was found between 8-10 year olds and adults. This result suggests that at 8-10 years of age children are able to process speech in noise at adult like levels, a finding which agrees with other previous studies (Elliott 1979; Papso & Blood 1989; Leibold & Buss 2013). Child/adult differences in the two-talker masker condition, on the other hand were found to be significant, yet there was no significant difference between the two child age groups, suggesting in agreement with Hall et al. (2002) and also more recently Leibold & Buss (2013), a slower development in children’s ability to process speech in speech maskers. Such cross-sectional studies have provided valuable insight into the developmental trajectory of speech intelligibility in noise and speech maskers. Previous research however has not yet examined development within the same sample of children. Owing particularly to large individual differences seen within speech intelligibility studies amongst children often of the same age group (e.g. Wightman & Kistler 2005), it is of interest to characterise development with both masker types within the same group of children.

There are some conflicting results across studies and comparisons become difficult considering differences in methodologies. In particular complex speech intelligibility tasks have often been used and/or complex masker types. Moreover, cognitive skills of children have not been taken into account to ensure they are age appropriate. Suggestions are made that children's central auditory processing is developing and so may account for the differences between children and adults with regard to informational masking effects (Johnstone and Litovsky 2006). Investigating how cognitive factors in children and adults may contribute to performance would thus be of interest. Further studies are also needed to determine robust results to understand if there are real child/adult differences and also to investigate the reasons for individual differences. Table 2.1 compares the results from several studies investigating child/adult differences with single-talker and/or two-talker speech maskers and noise maskers.

Table 2.1: A comparison of normal hearing child/adult differences across studies incorporating child and adult data in the same study with various noise and single-talker or two-talker speech masker types.

Author/s (year)	Sample size	Target stimuli	Masker stimuli	Findings	Comments
Hall et al. (2002)	19 children (5-10 years) 14 adults	Words (male talker)	Speech-shaped noise Two-talker speech (male)	Main effect of age: children performed poorer than adults. Main effect of masker: performance poorest in two-talker speech than noise. Significant masker*age interaction: adverse effect of speech masker even greater in children than adults. Significant improvement with increasing age amongst children in noise masker but not speech masker (although trend for improvement present).	Authors suggest possible developmental delay in ability to process speech in speech masker. Results compared at same intelligibility level, therefore only one part of psychometric function examined so possible SNR confound.
Litovsky (2005)	36 children (4-7 years) 9 adults	Words (male talker)	Quiet Speech modulated noise Single-talker and two-talker speech (female)	Amount of masking (as compared to a quiet condition) was larger in speech modulated noise compared to speech maskers. SRTs were better in adults compared to children in all conditions. No significant difference between SRT with any masker type in children. No significant difference in amount of masking between children and adults.	The author suggests amount of masking may be similar in children and adults due to ceiling effects in quiet conditions with adults. The author suggests the child/adult differences may be due to developing central auditory skills and language differences. Masker*age SRT interactions could not be directly evaluated (due to independent measures design).
Wightman and Kistler (2005)	38 children (4-16 years) 8 adults	Monaural CRM sentences (male)	Speech-shaped noise Single-talker CRM sentence (male or female)	With an ipsilateral speech masker, performance gets poorer with decreasing age. Addition of noise distractor to contralateral ear had no effect on performance in children and adults but contralateral speech masker detrimentally affected performance monotonically in children and adults Errors occurred from the maskers mostly in the youngest group of children.	The authors suggest that the youngest children were not able to direct attention to a single ear. Children require higher SNRs to achieve adult like performance even in the oldest group of children 13-16 year olds.

Wightman et al. (2006)	23 children (6-16 years) 10 adults	Monaural CRM sentences (male)	Single-talker CRM sentence (male)	Children required higher SNRs than adults to reach the same intelligibility levels. Even the oldest age group (12-16 years) did not achieve adult like levels. Large individual differences, particularly in the intermediate age groups.	Attentional strategies for discriminating speech in speech backgrounds may develop at different rates in children.

Author/s (year)	**Sample size**	**Target stimuli**	**Masker stimuli**	**Findings**	**Comments**
Johnstone and Litovsky (2006)	20 children (5-7 years) 20 adults	Words (male talker)	Quiet Speech modulated noise Single-talker speech (female) Time reversed single-talker speech (female)	Amount of masking (as compared to a quiet condition) was larger in speech modulated noise compared to speech maskers in adults. There was no significant difference between SRT with either masker type in children. When increasing set size (25AFC in adults, 4AFC in children), children experienced larger amounts of masking than adults with reversed speech, but less amounts of masking than adults with speech modulated noise, same amount of masking with speech.	Children did not perform better with speech compared to speech modulated noise (as adults did). The authors propose this may reflect development of central auditory processes and auditory attention. Results compared at same intelligibility level, therefore only one part of psychometric function examined so possible SNR confound.
Wightman et al. (2010)	36 Children (5-16 years) 24 adults	Monaural CRM sentences (male)	Speech modulated noise (ipsilateral and contralateral presentation) Single-talker CRM sentence (male or female, in contralateral ear)	Speech modulated noise in the contralateral ear did not affect performance in children or adults. Single-talker speech maskers in contralateral ear detrimentally effected performance by up to 20 dB in the youngest children (5-8 years) and 4 dB in adults. Male and female contralateral speech maskers produced comparable results. Large individual differences were observed, particular in children between 7-12 years.	The authors speculate that the larger detrimental effect of the single talker maskers in children may be due to development of selective attention, possibly developing at different rates.

Bonino et al. (2013) **Experiment 2**	9 children (5-7 years) 9 children (810 years) 16 adults	Words (male)	Speech shaped noise Two-talker speech (males)	Children's performance was poorer than adults' in all masker conditions. Significant masker*age interaction: larger child/adult differences for two-talker speech masker compared to speech-shaped noise. In speech-shaped noise condition, younger children performed more poorly than older children. In two-talker speech masker no significant difference in performance between younger and older children and larger individual differences.	Development of speech perception more prolonged with two-talker masker compared to speech shaped noise. Only tested at one SNR therefore only one part of the psychometric function, so possible SNR confound.
Leibold and Buss (2013)	62 children (5-13 years) (in 3 age groups) 28 adults (19-34 years)	Consonantvowel (CV) tokens (female)	Speech-shaped noise Two-talker speech (female)	Main effect of age: Adults better than youngest children for both maskers. Main effect of masker type: performance poorest with two-talker speech. Masker*age interaction: larger child/adult differences for two-talker speech masker compared to speech-shaped noise. Children's scores in speech-shaped noise were adult like by 11-13 years of age but remained poorer than adults even at 13 years with two-talker masker.	Children may show immaturity in their ability to segregate speech sounds from running speech, and development may be prolonged. Results from children and adults in both masker types compared only at one SNR, therefore only one part of psychometric function examined so possible SNR confound.

2.3.4 Reasons for child/adult differences

Whilst results from many studies have revealed child/adults differences, the reasons for such differences are not well understood. Many suggestions have been put forward by investigators which will now be discussed. It is often proposed that the differences are not likely to be due to developmental effects in the peripheral auditory system since this is said to be fully developed by birth (Eggermont et al. 1996; Eisenberg et al. 2000). Furthermore, the ability of a child to hear before they are born is reflected in the behaviour of the new-born when they orient their heads towards the sound of their mothers' voice over other voices (DeCasper & Fifer 1980). Fine aspects of hearing such as coding of intensity, frequency and temporal qualities of sound develop over the first 6 months and are thought to be adult-like around this time (Leibold et al. 2007). Child/adult differences are therefore often said to be due to differences within their central auditory systems (Mlot et al. 2010). It is found from electrophysiological tests that maturation of neurons in the auditory cortex continues throughout childhood until around 12 years of age (Moore & Linthicum 2007). Cadaveric anatomical investigations have also shown that central processes are not fully developed until adolescence (Moore & Linthicum 2007).

Skills of sound source segregation are said to develop over a prolonged time course and general cognitive developments are also said to develop throughout childhood, including selective attention and working memory (Gomes et al. 2000; Werner 2007). It seems that adults may appear to listen more selectively than children. For example, Dai et al. (1991) found that when adults were faced with the task of listening to one particular frequency in the presence of background noise, they were able to focus on that frequency in order to improve their performance, yet their ability to detect unexpected frequencies was reduced. In a later contrasting study (Bargones & Werner 1994), children (9 months old) performed equally well for the detection of unexpected frequencies (played in 25% of trials) and expected frequencies (played in 75% of trials) in noise suggesting that children may listen using a "broadband" strategy (Werner 2007). Werner (2007) suggests that whilst this may make children more susceptible to informational masking, this "broadband" listening may be needed to help them learn the important cues and features of speech.

As previously discussed, large individual differences are seen in the majority of studies investigating speech intelligibility in noise (particularly in speech maskers) in children. Some studies suggest that these large individual differences could reflect the fact that children may develop the ability to utilise cues for sound source segregation at different rates (Leibold & Bonino 2009), and this could relate to differences in their exposure to various listening environments (Wightman & Kistler 2005). It is also suggested that in complex backgrounds, children may be less able to choose the optimal cues needed for separating out speech from noise (Hazan & Barrett 2000), which may improve with experience.

Another aspect which may influence the results of speech intelligibility tests showing child/adult differences is that adults may have more experience using language and so be able to use linguistic knowledge to help identify speech when degraded by maskers (Fallon et al. 2000; Mlot et al. 2010). Children's language skills develop over time which includes increasing vocabularies and the learning of grammatical structures (Leech et al. 2007) and may account for child/adult differences, yet even when vocabulary is age appropriate in speech intelligibility tests, differences are still apparent (Elliott 1979).

It has been investigated whether reading ability may influence speech intelligibility in noise in children (Lewis et al. 2010), since phonological awareness is shown to predict reading ability in children (Hogan et al. 2005), so phonological awareness may help identify the phonological structure of sounds within noise (Nittrouer 2002). Snowling et al. (1986) found that in children (9-12 years) with and without reading disabilities, there was no correlation with speech intelligibility in noise. Brady et al. (1983) however, found that 8 year olds who were poor readers performed more poorly in speech intelligibility tasks in noise than those with normal reading abilities. Lewis et al. (2010) also investigated if children's (5-7 years) phonological awareness could predict their speech intelligibility in noise, but found this was not the case. Older children had better phonological awareness performance but performance on this task did not account for the wide variability observed in the speech intelligibility test. Lewis et al. (2010) explains that differences in these studies may be due to different tasks used and so evidence on this topic is inconclusive, and may benefit from further research.

Despite the many suggested reasons for child/adult differences, it is important to note that some differences seen in children and adults with the speech maskers may be due to an SNR confound occurring from differences in baseline noise conditions with which differences in speech masker conditions are compared. Bernstein & Brungart (2011) have recently suggested that when adaptive procedures are used in speech intelligibility tests (e.g. Hall et al. 2002; Bonino et al. 2013; Leibold & Buss 2013) different signal to noise ratios (SNRs) are tested for each individual participant depending on their performance, so the results are SNR dependent. Since children are generally found to perform worse than adults with a stationary masker (Elliott 1979; Allen & Wightman 1995; Litovsky 2005; Wightman & Kistler 2005), comparisons of performance with the speech masker and the stationary noise masker (or comparisons of masking release) may not represent actual performance. Bernstein and Brungart (2011) propose that when higher SNRs are tested (such as those which may be tested with children who generally perform worse than adults) low masking release is found and when lower SNRs are tested (such as those which may be tested with adults who generally perform better than children), high masking release is found. Bernstein and Brungart (2011) suggest that this is due to differences between the slopes of the psychometric functions (often shallow with speech maskers and steeper with stationary noise maskers) which display a larger masking release at a certain per cent correct level which may not be representative. It may therefore be that differences in the baseline condition where performance is worse with the stationary noise, reflects differences with the stationary noise masker and not differences in the way that the speech masker in particular affects performance. Thus conclusions of previous studies may be flawed. It is therefore of great importance to examine psychometric function slopes measured in this way to determine the authenticity of such results (see Chapter 4, section 4.1.1 for more details).

Understanding the differences between children and adults is necessary for Audiologist to make clinical decisions about what constitutes normal hearing in children so to appreciate their needs. Determining whether a difference of speech intelligibility in speech maskers between children and adults is just due to differences in baseline noise conditions is therefore a prominent issue which needs resolving. Nonetheless, when listening conditions are poor, top-down skills may be required when bottom-up factors are compromised (Lewis et al. 2010). For children with hearing loss, listening could thus be particularly difficult if a degraded sound reaches the brain to be perceived by an immature central auditory system (Werner, 2007) so understanding differences in central auditory systems and thus cognitive factors is therefore too, important.

2.4 Effect of background on reading

2.4.1 Mechanisms of reading

In situations where energetic masking is limited, speech intelligibility can still be degraded through informational masking. Whilst such sounds may not overlap acoustically with each other it appears possible that the target and the masker may be confused at higher cognitive levels. Speech backgrounds have often been suggested to produce effects of informational masking affecting speech intelligibility. Whilst some speech intelligibility studies have tried to isolate effects of informational masking as separate from effects of energetic masking (e.g. Wightman & Kistler, 2005; Wightman et al., 2010;), by presenting targets and maskers to separate ears, this still does not separate out contributing factors of parsing the auditory scene and so is not specific in explaining the origins of informational masking. In order to investigate interference from speech backgrounds specifically at higher cognitive levels, a reading paradigm can be employed investigating interfering effects of speech backgrounds on the process of reading. Because written language and spoken language are presented across different sensory modalities, any interference may be attributed to occurring at a cognitive level beyond peripheral sensory levels (Banbury et al. 2001).

Listening to speech in speech backgrounds is a common everyday occurrence, yet we are also required to cope with speech backgrounds when carrying out tasks other than listening. At school and in many workplaces, reading is a necessity and is prone to distractions (Cauchard et al. 2012). It is interesting to study the effects that background sounds have, particularly those associated with informational masking, on reading. It is commonly reported that when we read we experience hearing a voice inside our heads, known as the 'inner voice' (Baddeley & Lewis 1981). Whilst energetic masking of this inner voice is not possible, it is not unreasonable to assume that informational masking may have an effect on reading, causing confusion at higher levels between the inner voice and the auditory masker. It is thus interesting to determine which background maskers, for example stationary noise maskers (associated with energetic masking), or speech maskers (associated with informational masking) cause most disruption in reading. This following section will provide some further background into language comprehension in general, the mechanisms of reading and explore the ways in which different background sounds affect reading.

2.4.1.1 *Language comprehension*

Considering the literature discussed so far, speech intelligibility in the presence of speech backgrounds has been shown to incorporate additional effects of masking deemed informational masking. Although a definition of what informational masking entails is not clear, it has been described as incorporating components of object formation/sound source segregation and object selection (Shinn-Cunningham 2008). It is important to also consider that interference from speech backgrounds, when listening to speech or reading text, may interfere with the linguistic processing of a target (Schneider et al. 2013). Therefore, it is necessary to consider how language (spoken or read) is processed to deliberate how speech backgrounds may interfere.

Processing language (whether spoken or read) involves the formation of phonemes (units of sound), combined into morphemes (units of meaning) which then form words able to access the mental lexicon; lexical processing (Cutler & Clifton 2000). The mental lexicon refers to where we store in the mind, information about word properties and where we retrieve information to match with linguistic input to understand language (Treiman et al. 2003). Language input must rapidly map onto similar patterns in the mental lexicon and as words are accessed and lexical processing is achieved, semantic processing (meaning of language) and syntactic processing (structure of sentences) become available to thus comprehend language (Treiman et al. 2003). Working memory is an aspect which is argued to be involved in language comprehension of both spoken and written language (Gathercole & Baddeley 1993). Despite this however, much research has been carried out in the audiological field on speech intelligibility, without looking at its interplay with cognitive resources (Mattys et al. 2012). It is not entirely clear whether higher level linguistic knowledge effects both pre-lexical processing or just post lexical processing (and this is beyond the scope of this thesis), but it is certain however that such factors do play a part in the eventual perception of language spoken or read (Treiman et al. 2003; Heald & Nusbaum 2014).

Without evidence proposing otherwise, it is assumed beyond the periphery that auditory language comprehension follows that same process as written language comprehension (Cutler & Clifton 2000). Early models describing the processing of spoken language in particular state that candidate words compatible with (portions of) an incoming speech signal are simultaneously activated in the mental lexicon and long term memory, with all words then competing for recognition. Many candidates may be activated and are tested against a hypothesis then rejected until fewer and fewer candidates exist until a decision is made on the best match that remains (Heald & Nusbaum 2014). This type of mapping is similar in the early models like the Cohort model (Marslen-Wilson & Welsh 1978), the TRACE model (McClelland & Elman 1986) and the Shortlist model (Norris 1994).

More recently, the developing field of Cognitive Hearing Science puts an emphasis on the role of cognition in communication, which has previously not received a great deal of attention (Rönnberg et al. 2013). It is suggested that a phonological representation of language is stored in working memory to enable this matching process with long term memory (Gathercole & Baddeley 1993). Working memory may be thought of as having a limited capacity whereby background sound may compete for resources needed for the focal task (Baddeley 2012). The Ease of Language Understanding (ELU) model is similar to these earlier models in that it refers to the matching and retrieval of linguistic input with representations in long term memory to access the lexicon and comprehend language (Rönnberg et al. 2013). It is largely based on Baddeley's (2012) research on working memory and it differs from earlier models in the fact that it includes aspects of working memory which come into play when mismatches between language input and representations in long term memory occur (Rönnberg et al. 2013). A model of working memory proposed by Baddeley (2012) consists of three component parts, a phonological loop which stores verbal information, a visuospatial sketchpad which stores visual information and the central executive component which controls both these things. The phonological loop component is separated further into two separate parts; the store where information is held and the sub-vocal rehearsal mechanisms to refresh its contents (Chenoweth & Hayes 2003). In conditions which may interfere with this matching process (i.e. with energetic or informational masking), obtaining a match between language input and long term memory may be challenging. Bottom-up implicit and top-down explicit processing will interact differently, changing according to demands throughout the discourse (Rönnberg et al. 2013).

The Ease of Language Understanding model from Rönnberg et al. (2013) will now be explained in more detail. Regarding long term memory it is separated into two parts. The episodic memory is personal to the language user of events they have experienced and the semantic memory is non personal and considers vocabulary and phonology (Rönnberg et al. 2013). The Ease of Language Understanding (ELU) model proposes that language input (whether spoken or read) is Rapidly, Automatically, Multi-modally, Bound into a Phonological representation in the episodic buffer (RAMBPHO) of working memory. If the sub-lexical language input matches with a representation of phonemes in semantic long term memory then lexical access will have been achieved and understanding will be rapid and implicit without the occurrence of top down processing (Rönnberg et al. 2013). If the information in the RAMBPHO does not match with semantic long term memory and there is a mismatch, lexical access will be

delayed. Thus for successful understanding to occur the explicit involvement of top down processing will be required, specifically working memory to help compensate for any mismatch. So the time taken for processing will slow down. When there is a RAMPHO induced mismatch an explicit processing loop is set off feeding back to pre-set the RAMBPHO and promote the tuning of attention to the subsequent input (Rönnberg et al. 2013). There is a phonological and semantic influence on RAMBPHO whereby the expectations of what may come next in the language input (i.e. contextual cues) constrain possibilities in working memory (Rönnberg et al. 2013) and modulate the RAMBPHO. Semantic long term memory may help to explicitly fill in the gaps or infer meaning of missing or degraded information or even help in priming of implicit processes for expected input (Rönnberg et al. 2013). An example of when a mismatch may occur may be where attention is switched between two things and where irrelevant information must be inhibited (Rönnberg et al. 2013). The slow explicit process and the rapid implicit process can occur in parallel and the implicit process is always running. As a mismatch occurs working memory is invoked to keep track and retrieve phonological and semantic information from long term memory to help compensate for the degraded input so comprehension is accomplished and the next loop can include some contextual constraints.
Information is then committed to episodic long term memory (Rönnberg et al. 2013).

When listening or reading, language input will initiate phonetic, semantic and syntactic processing. The processing of speech is thought to be automatic and obligatory (Hawley et al. 2004), therefore in the presence of background speech, if both target and interferer contain linguistic properties they will both activate linguistic and cognitive systems and the phonologic and semantic processing of both messages may interfere with the processing of the target language input (Francis 2010). So when lexical access is not supported, top-down influences may be employed to aid comprehension (Mattys et al. 2012). When the language input from both target and interferer are from the auditory modality this may be especially difficult if the auditory scene has not been adequately parsed (Schneider et al. 2013). It is thought that it may not just be linguistic knowledge of lexicon and syntax which contribute to the restoration of degraded speech but also non-linguistic factors such as cognitive faculties like attention (Mattys et al. 2012).

The ELU model predicts that those with high working memory will be able to better infer missing information and so experience less effort (Pichora-Fuller & Singh 2006). In some studies investigating subjective listening effort this has been shown to be true, higher working memory capacity lowers effort (Rudner et al. 2012). In some studies investigating objective listening effort however, through measures of pupil dilation, this has been shown to be the other way around; higher working memory capacity is shown to increase pupil dilations thus inferring higher listening effort (Zekveld et al. 2011; Koelewijn et al. 2012a). Although this sounds counterintuitive it may be thought that those with high working memory capacity have a greater engagement with the task so allocate more cognitive resources to it (Koelewijn et al. 2012a).

2.4.1.2 *Processing the text*

During reading, the reader must extract information which requires analyses from both low to high levels (Schotter et al. 2012). Beginning with low level features, visual information such as orthographic (letters), phonologic (sounds) and morphologic (units of meaning) information is coded for. Next, higher level semantic (meaning) and syntactic (grammar) properties are integrated to enable understanding of the text (Schotter et al., 2012). Since a lot of information is processed at once, readers may use up cognitive resources (Schotter et al., 2012) leaving little resources available to cope with any interference from background noise (Cauchard et al., 2012).

The process of reading has been widely studied by measuring reading comprehension, and over the past thirty years, the study of reading has progressed through eye movement measures (Rayner 1998). Studying eye movements can reveal the ways in which words and sentences are processed and how different variables can affect language processing (Rayner, 1998). Eye movement studies can provide insights into how we read moment to moment and reveal information about cognitive processes (Rayner et al. 1978). Measures of eye movements include fixation durations (time spent fixating one part of the text), saccade size (size of jump from one fixation to the next) and regressions (backward eye movements to reread parts of the text) (Johansson et al. 2011). The time fixated on one word is usually about 200-250 ms (Rayner, 1998). Fixation durations may be longer for words which are uncommon (low frequency words) and shorter for regularly occurring words (high frequency words) (Inhoff & Rayner 1986). Also, when words are present which do not make sense semantically within a sentence (implausible words), longer fixations may be made, and the reader may be more likely to make regressions (Warren & McConnell 2007). Regressions may also occur if too long a saccade has been made, if the reader finds words difficult to process or if the reader does not understand the text (Rayner, 1998).

2.4.1.3 *The inner voice*

Humans are well adapted to process speech and language (Chomsky 1959). One may think however, that since spoken language is transient and cannot be re-heard, comprehending spoken language may be more challenging than comprehending written language as the listener has to work out the boundaries and segments of words across a continuous acoustical signal over time (Treiman et al. 2003). Spoken words cannot be re-heard in the same way that text may be re-read as eyes make regressive (backward) movements. It is thought nonetheless that auditory input stays longer in working memory than visual input, lasting several seconds (Cowen 2005). This may be why orthographic (spelling of words) representations are transformed into phonological representations encompassing what is known as our "inner voice" in silent reading (Slowiaczek & Clifton 1980).

It seems universal that when we read silently, we experience an inner voice (or subvocalisation) inside our heads (Leinenger 2014), and that people actually have a sensation of "hearing" what they read (MacKay 1992). The function of this subvocalisation is not entirely clear (Slowiaczek & Clifton 1980). During reading, since words must be combined together to form the understanding of sentences and paragraphs, the reader needs to be able to keep track of previously read words (Baddeley & Lewis 1981). Subvocalisation is thought of as an articulatory rehearsal process to help keep words in memory to combine them with other words in order to reflect the meaning of sentences (Baddeley et al. 1981). The inner voice is thought to translate letters on a page into their phonological codes (sounds) which are thought to stay longer in working memory than the visual codes (Baddeley 1979). It is also said that the prosody (rhythm, stress and intonation) within spoken speech may help group words into phrases and sections and enable the processing of syntactically ambiguous sentences which may facilitate the reading process (Cutler & Clifton 2000).

Experiments which have looked at the role of the inner voice have tried to supress subvocalisation by asking participants to repeat aloud words or numbers while reading ((Kleiman 1975; Baddeley et al. 1981). In this way, a study by Kleiman (1975) asked participants to detect rhyme in sentences (e.g. does cream rhyme with: He awoke from his dream? does soul rhyme with: the referee called a foul?). It was found that this ability was impaired when carrying out a subvocalisation suppression task which involved listening and repeating back strings of random numbers. Baddeley et al. (1981) suggest however that the suppression task used could have been too demanding cognitively and so overloaded short term memory. A further similar study by Baddeley et al. (1981) found that supressing subvocalisation by repeating a word aloud caused detrimental effects to reading as participants were not able to detect sentences which were semantically incorrect. Baddeley et al. (1981) explain that it is unlikely that their results are due to the overload of the subvocalisation task, directing away attention, since when they repeated their experiment allowing subvocalisation, but with participants tapping at the same rate as they repeated the word to reduce subvocalisation, performance was not affected. The effect of subvocalisation suppression on reading comprehension was investigated by (Levy 1978). It was found that reading comprehension was possible but verbatim (word for word) recognition was affected; a finding which is also observed by Slowiaczek & Clifton (1980).

It seems that the inner voice plays an important role during reading but it is not clear which processes of reading (e.g. from low to high levels) it facilitates. If background maskers like speech affect reading ability, they may interfere with this articulatory rehearsal process by mapping onto phonological codes (Slowiaczek & Clifton, 1980), and interfering with phonological formation of read words (Martin et al. 1988).

2.4.2 **Noise backgrounds**

Whilst stationary noise backgrounds are often shown to cause greater detrimental effects to speech intelligibility compared to speech backgrounds, in reading this effect seems to be the other way around. Stationary noise backgrounds may affect reading less than speech backgrounds because they may not initiate any linguistic processing which could be likely to occur with speech backgrounds (Francis 2010) and thus interfere with the linguistic processing of the text. One study however (Zimmer & Brachulis-Raymond 1978) found that the presence of industrial noise caused more of a disruption to reading than a speech background yet this noise was intermittent and thus became more dynamic than the speech and so could have directed attention away more easily owing to its changing state (Martin et al. 1988).

That being said, it may not just be the changing characteristics of an auditory background, diverting attention, that affects reading. Martin et al. (1988) investigated the disruption to reading comprehension from a white noise background, instrumental music and speech. Since there was no intermittence in the backgrounds and all background sounds played continuously, Martin et al. (1988) expected that if reading comprehension was affected by changing backgrounds, the speech and music may produce similar effects. It was found that performance in the white noise and instrumental music conditions was not significantly different and did not affect reading comprehension, but the speech background did. Since the music background is constantly changing, this result suggests that it is not just meaningfully organised and changing backgrounds that will affect reading but the properties of speech in particular could interfere with the phonological and/or semantic processes of reading. This will be discussed further in the next sections. A more recent study by Johansson et al. (2011) investigated the effects of café noise on reading comprehension, which the authors described as a consistent buzz. Like Martin et al. (1988), it was also found that reading comprehension was not affected in café noise as performance was no different to reading in silent conditions.

2.4.3 Speech backgrounds

It is said that background speech can interfere with reading by reducing reading comprehension. Speech is automatically encoded for, even involuntarily (Beaman et al. 2007), so the conflict in semantic processing of the text and the automatic semantic processing of the speech (Marsh et al. 2009) may affect reading. Whilst speech may interfere with such post lexical levels of semantic processes engaging with the text and the masker, it may not be wholly due to semantics within the speech but it could also be due to the speech masker interfering at pre-lexical levels by disrupting storage of phonological codes (Boyle & Coltheart 1996).

Although few studies have directly investigated the effects of background sounds on reading, it seems noteworthy to consider the large body of research which has investigated effects of background sounds on serial recall performance. It is of relevance to explain how such studies can relate to reading as some of the processes may be similar. Often referred to as the irrelevant speech effect, interfering effects of background speech on the ability to recall lists of visually presented items has been widely measured (Banbury et al. 2001). It has been thought that background speech gains obligatory access to the phonological store (Banbury et al. 2001) which is used to maintain rehearsal of to-be-remembered items so competes for resources (Baddeley 1979). Prohibiting subvocalisation is found to interfere with this rehearsal process in serial recall tasks (Slowiaczek & Clifton 1980). Background sound is also shown to interfere with serial recall tasks and because the detrimental effects disappear when subvocalisation is also restricted, background sound is also thought to interfere by way of disrupting the rehearsal process (Salamé & Baddeley 1982). Since effects of speech backgrounds cause most disruption to serial recall tasks compared to white noise and tones (Salamé & Baddeley 1982), and since it does not matter whether this speech contains semantically correct or semantically anomalous material (Colle & Welsh 1976; Jones et al. 1990), it is suggested that the speech sounds gain access to the phonological store in working memory and this is where the interference occurs by preventing the rehearsal process, thus preventing phonological processing (Baddeley 2012).

For tasks involving reading comprehension however, it seems semantic processing of background speech does play a role (Martin et al. 1988; Jones et al. 1990). Meaning within background speech may affect reading because the task of reading itself involves deciphering meaning. In contrast, the serial recall tasks do not require the processing of meaning and so this may be why background speech does not have to be meaningful to interfere with serial recall (Banbury et al. 2001). Interference to reading from background speech could be occurring owing to the shared semantic processing between the background speech and the read text (Marsh et al. 2009).

In a study by Martin et al. (1988) reading comprehension was investigated in the presence of speech with and without meaning. Participants were required to read passages of text whilst ignoring background speech, and were afterwards asked a comprehension question about what they had read. The background interference was either meaningful speech, foreign speech, random words, non-words or white noise. Reading comprehension was found to be poorest with the meaningful speech and random words compared to the non-words and foreign speech. The non-words and foreign speech were also no different in their effect on comprehension to white noise. This shows interference is less likely to be phonological but more semantic, since white noise does not contain phonological information whilst the foreign speech does contain phonological information, albeit meaningless (Martin et al., 1988). Martin et al. (1988) do however point out that the phonological code of the foreign (Russian) speech may have been different enough from English to reduce any phonological interference. Considering the results from this reading comprehension study and the serial recall studies (e.g. Colle & Welsh, 1976; Salamé & Baddeley, 1982) however, speech backgrounds appear to affect different tasks differently. Speech backgrounds may prevent verbatim (word for word) recognition of visual stimuli through phonological interference whilst speech may affect the comprehension of texts through semantic interference (Martin et al., 1988). In the serial recall task, the subjects may not need semantic processing to relay verbatim lists and so meaningful speech, which automatically activates semantic processing (Beaman et al. 2007), may not interfere in this way (Martin et al. 1988). Martin et al (1988) suggest that interference occurs if both irrelevant material and read text require the same analyses, like semantic processing.

Similarly, Oswald et al (2000) also found meaningful speech to interfere with reading more so than meaningless speech. Participants were required to read sentences and answer questions after reading in both meaningful speech backgrounds (news recordings), meaningless speech backgrounds (the same news recordings time reversed), as well as a silent condition. It was found that both speech backgrounds differed compared to quiet with meaningful speech being more interfering. The authors support the conclusions from the study by Martin et al. (1988) that the background speech is being semantically processed and thus may interfere with the semantic processing of the read material (Oswald et al. 2000).

Sörqvist et al. (2010) argued however that studies using comprehension questions after texts have been read (e.g. Martin et al. 1988) are largely examining long term memory and less so the present processing of the text. Therefore, in their study they tried to minimise the interval between reading the text and answering questions by having participants read shorter texts and presenting the question and the text on the same screen all at once. In the presence of background speech (fictitious stories), participants made more errors in reading comprehension than compared with reading in silence. They also found that performance on the reading comprehension task was associated with working memory and the ability to suppress irrelevant material as individuals with better reading comprehension also performed better in a test of working memory.

The effects of background speech on the efficiency of proof reading has been investigated (Jones et al. 1990; Halin et al. 2014). Participants had to identify errors in grammar and spelling and implausible words when speech and reversed speech played in the background. It was found that speech, but not reversed speech, caused a detrimental effect to the task and interfered most with low level spelling errors and not syntactic or semantic errors. The reasons for interference at this level are not clear but and it could be thought that the background speech is not interfering at a semantic level. Boyle & Coltheart (1996) explain however, that the proof reading task may not have shown any detrimental effects to semantic processing because the task involved just skimming and not comprehension as such, so the text may not have involved any actual semantic processing.

In contrast to Jones et al (1990), Halin et al. (2014) found in a proof reading task that ability to identify semantic errors within text was reduced by the presence of background speech. The discrepancy between these studies is not clear but supports findings of other similar studies (e.g. Venetjoki et al. 2006) . Interestingly however, it was found that when the visual material in the proof reading task was made more difficult to see (due to obscured text), the interfering effect of the speech background disappeared (and this was not a floor effect).The authors explained that results may be because attention towards the text is improved as the participant must become more engaged with the text. This resonates with findings on speech intelligibility tasks in speech backgrounds where effects of informational masking are overcome when attention towards the target is improved (Freyman et al. 2004).

Few studies have explored the effects of background sounds on reading, but those that have, have mostly measured any disruption using reading comprehension tasks. This may be considered an “offline” measure since it records comprehension accuracy after comprehension may have occurred. Therefore such a measure does not tell us anything about how background sound interferes with the “online” processing of reading as it is happening.

The use of eye movement measures through eye tracking technology can provide moment to moment information about online cognitive processing (Rayner 2009) and is considered a benchmark in reading research (Rayner 2009). The process of reading involves phonological mapping to achieve lexical access of words and integration across sentences involving post lexical processing. Studies of reading have shown eye movements can tell us both about lexical and post-lexical processing (Rayner 2009). At the lexical level some studies have shown low frequency (low occurring) words can cause longer fixation durations (time spent with eyes fixed on one part of the text) (Inhoff & Rayner 1986), and at the post lexical level words which do not fit in semantically within sentence context can also cause longer fixations (Warren & McConnell 2007). It is thought that difficulties in lexical processing mainly cause longer fixation durations but difficulties in post lexical processing mainly cause both longer fixations and regressive eye movements going back over the text to re-read parts (Reichle 2011).

Using eye movement measures to look at the effects of acoustic backgrounds on reading is an area which has received little research. To our knowledge, only one study to date by Cauchard et al. (2012) has made use of this technique to examine effects of background speech on reading. Thirty-two adult participants were required to read short paragraphs silently (14 lines of text on one screen) whilst background speech played over speakers (at 60-70 dB SPL) which they were instructed to ignore. The speech background consisted of speech recorded from a radio talk show about books which included conversations between the host of the show and the book writers lasting 29 minutes (Cauchard et al., 2012). After each paragraph, participants answered a question about the text and also gave a rating of their perceived comprehension difficulty. During reading, participants had their right eye tracked and recorded by an Eyelink 1000 eye tracker. They were required to place their head in a head and chin rest to prevent them from moving their head so their eyes could be accurately tracked. The eye movements measured were fixation durations (time spent with eyes fixed on one part of the text), number of fixations, total reading time and proportion of regressions (backward eye movements to reread parts of the text). It was found that compared to silent reading, reading time increased with speech backgrounds and this was largely due to longer fixation durations and more regressions. This type of eye movement suggests that reanalysis or reprocessing of the text is needed showing difficulties perhaps at higher processing levels of later post lexical integration (Rayner & Liversedge 2011; Reichle 2011). However Cauchard et al. (2012) explain that their study is not in a position to tease apart exactly at which level of linguistic processing the interference is occurring. It would therefore be interesting to use eye movement measures and further manipulations of acoustic stimuli with different speech background and reading text combinations to investigate how different levels of language processing may be affected during reading.

The results from the Cauchard et al. (2012) study agree with those using reading comprehension measures (Martin et al. 1988; Oswald et al. 2000; Sörqvist et al. 2010) in that the speech background interferes with reading. In contrast however to these studies, Cauchard et al. (2012) failed to find any effect of background speech on reading comprehension. The authors explain this may be likely due to possible ceiling effects since the comprehension scores were very high. They also state however, that since their experiment allowed readers to read with no time constraints, as these other reading comprehension studies have, readers slowed down and re-read parts to compensate for any detrimental effects to comprehension. So comprehension was not effected but at the cost of efficiency. The results of the subjective questionnaires also showed that readers perceived reading comprehension to be more difficult with the speech background than the music and silent background conditions (Cauchard et al. 2012).

2.4.3.1 *Music*

It seems noteworthy to discuss the effects that background music has on reading since it can consist of lyrical or just instrumental music and the effects on reading are shown to be somewhat complex (Cauchard et al. 2012). Studies on this topic have shown results which can vary widely depending on factors such as music type (Kiger 1989), and also individual personality (Furnham & Bradley 1997). In some cases people may often play music when studying to help them concentrate and block out further distractions (Cauchard et al. 2012).

In a second experiment by Martin et al (1988) the effect of background music both with and without lyrics was investigated and it was found that reading comprehension was better in music without lyrics than with lyrics. Henderson et al. (1945) also found a similar result and showed that music with lyrics disrupted performance on a reading test, but the disruption of classical instrumental music was no more disruptive than the quiet condition. It is unclear whether this was due to the presence or absence of lyrics or the two different kinds of music (Martin et al. 1988). The study by Cauchard et al. (2012) also included a music condition and found no effect on reading efficiency with instrumental music.

Another study which did use eye tracking measures similar to that of Cauchard et al. (2012) was a study by Johansson et al. (2011) whereby eye movements were tracked whilst 24 participants read large bodies of text but not in the presence of speech, only music or café noise and in a quiet condition. There were found to be however, no significant differences within the eye movement data suggesting that music and café noise (described as babble) did not disrupt the normal process of reading. It is not clear why there were no effects of eye movements seen but it could be that these background noise conditions were not distracting enough. It is difficult to interpret the findings of this study however as each participant read in different background music that they had chosen themselves, so there may have been differences in acoustical and linguistic properties between all the music backgrounds.

Eysenck (1967) states that introverts and extroverts are affected by background sounds differently. Eysenck (1967) suggests that people who are introverted require lower levels of external stimulation to reach optimal arousal than extroverts, and this arousal is inhibited by higher levels of external stimulation (Eysenck 1967). This theory is supported by studies which have found that introverts are more affected by background sounds than extroverts on tasks which are cognitively demanding (Furnham & Bradley 1997; Cassidy & MacDonald 2007).

2.4.3.2 *Level*

To our knowledge, no studies exist which have manipulated the level of background sounds during reading. The level of background sound on serial recall tasks however, has been investigated and most studies have found that increasing the level of background sound does not increase disruption to serial recall (Banbury et al. 2001). It has been found in particular that when speech backgrounds are whispered (e.g. at 48 dB A) or shouted (e.g. at 76 dB A) there is no change in its disruption to a serial recall task (Colle 1980; Salamé & Baddeley 1987). Furthermore the disruption to serial recall has been shown to remain unchanged when the level of background is manipulated across trials (Tremblay & Jones 1999). Therefore it would be of interest, especially using online measures to examine whether sound level moderates any disruptive effects of background sound during reading to see if such effects are replicated here. Such findings may have implications for background noise levels in schools and offices.

2.5 Summary and Research Aims

In summary it appears that background sounds with different properties may affect speech intelligibility and reading tasks differently. This alludes to different mechanisms associated with the way these background sounds interfere, and those mechanisms associated with informational masking (i.e. with speech backgrounds) are not fully understood. It is thought that cognitive processes are involved with regards to interference from speech backgrounds and it seems that children may be more susceptible to the masking effects of speech backgrounds than adults. Such findings have been interpreted to suggest poorer central auditory or cognitive development in children. It is important however to resolve important issues relating to the methods and analyses used to investigate such child/adult differences, to determine whether or not they are subject to SNR confounds or are in fact genuine. By looking at the effects speech backgrounds have on reading may then provide a further insight into mechanisms involved in informational masking and too provide further evidence for how we are able to cope with daily reading tasks in the presence of background sounds.

Therefore the main research aims of this thesis are:

1) To identify how noise and speech backgrounds may affect speech intelligibility in children and adults using a cognitively undemanding speech intelligibility task, with simple masker backgrounds and taking into account vocabulary scores of children.

   How child/age differences may change during development will also be investigated.

2) To determine if the background sound type*age interaction is authentic by tackling the possible methodological anomaly.

3) To further understanding how certain background sounds contribute to cognitive interferences and so identify how noise and speech backgrounds can affect reading, and relate these findings to better understand the mechanisms of informational masking in speech intelligibility tasks.

Chapter 3: **Effect of background on speech intelligibility in children and adults**

3.1 Introduction

This chapter consists of an experiment to investigate child/adult differences using a simple clinical speech intelligibility task. The vocabulary skills of children are also considered to ensure children with age appropriate vocabulary scores were tested. The speech intelligibility and vocabulary tasks utilised will firstly be detailed.

3.1.1 Speech intelligibility tests with children

3.1.1.1 *Adaptive procedures*

When aiming to determine a threshold of speech intelligibility (i.e. speech reception thresholds, SRTs) often we are just interested in one point, e.g. the signal to noise ratio whereby the listener achieves 50% correct. This could be found using a method of constant stimuli to determine the shape of the psychometric function, as it is difficult to know where along the x axis this 50% correct point will lie (Leek 2001). However, the method of constant stimuli would take a great deal of time and effort simply to determine one point (Leek 2001). Furthermore, when testing speech intelligibility in children, it is of great importance that the test be quickly and easily implemented to avoid fatigue and to prevent loss of attention from the child. Therefore adaptive procedures are made use of to greatly reduce test time which assumes the psychometric functions are monotonic (Leek 2001).

In an up-down staircase procedure the signal to noise ratio is varied according to the responses which have gone before. In a 1-down-1-up adaptive procedure the signal to noise ratio is decreased after a participant responds once correctly and increased after a participant responds once incorrectly. In this way the adaptive procedure targets the 50% intelligibility level. In order to seek higher performance levels, a transformed staircase method can be used whereby more trials are required to be correct before the down rule comes into play i.e. a 3down-1-up procedure targeting 79.4% correct on the psychometric function. This may be more highly desired since a percent correct level around 80% is a level considered more satisfying for conversation and relevant communication (Brand & Kollmeier 2002).

3.1.1.2 *The McCormick Automated Toy test*

The McCormick Automated toy test is one speech intelligibility test which is currently used clinically to determine the quietest level whereby children are able to identify words using an adaptive procedure. It is normally presented via loudspeakers and in quiet (Summerfield et al. 1994). Word discrimination thresholds are determined and are then often used to infer hearing threshold level in children who are too young to perform more demanding behavioural hearing testing (Summerfield et al. 1994).

In order to examine test-retest reliability of the McCormick Automated toy test and speech intelligibility tests in general there are various methods which can be implemented. Summerfield et al (1994) described specific methods used to examine test-retest reliability of the Automated Toy test.

The first method is replicability which involves determining any correlation between repeats to establish if the results are replicable, whereby a high correlation suggests that SRTs from one repeat can predict those from the next. This method however, does have limitations as Summerfield et al. (1994) points out. A high correlation may be found in the case where the results are not similar from one repeat to the next, but shifted one way or the other consistently, reflecting perhaps not replicability but learning or fatigue effects. In addition, there may be a case where poor correlation is found when results are replicable, but where the range of the results is very small.

A second method of test-retest reliability described by Summerfield et al. (1994) which tries to overcome these issues is the within subjects standard deviation of scores, i.e. the variability of thresholds. Variability determines the within subject standard deviation across repeats for each participant and thus calculates the within subject standard deviation of scores for a given group of participants by taking into account the number of tested participants and the number of repeats. From this within subjects standard deviation score the 95% confidence intervals can be determined.

In a study by Summerfield et al. (1994) using 8 adults, test-retest variability of the McCormick automated toy test was determined when averaging points across 6 reversals presented in steady state speech-shaped noise. The 95% confidence intervals were found to be ± 2.5 dB and test-retest repeatability improved with a greater number of reversals. Also the test-retest reliability was improved with the noise masker compared to when tested in quiet. However, more reversals increases test time which must not be too long with children as it is important to prevent fatigue and maintain attention.

Concerning children, Summerfield et al. (1994) also investigated test-retest reliability in 127 children (aged 2-13 years) tested in quiet, and the 95% confidence interval was found to be ± 4.9 dB. Lovett et al. (2012) since examined reliability with 13 children aged either 3 or 7 years tested in speech-shaped noise, and the 95% confidence intervals were ± 6.27 dB. Lovett et al. (2012) highlights that this result is less reliable than that found when tested in quiet in the Summerfield et al. (1994) study, but explains this may be due to a lack of statistical power from using a small sample size. In a later study it was confirmed that reliability was greater with a sample size of 53 children (aged 3-6 years), (± 4.8 dB) and found it to be slightly higher with a foreign speech babble masker in 42 children (± 5.8 dB).

3.1.2 **Vocabulary tests**

Where comparisons in speech intelligibility scores between children and adults are made, it is important to consider their language differences. Child/adult differences could become apparent simply if the language used in the speech intelligibility test lies outside the child's vocabulary range. Therefore it is important to make use of speech material that is appropriate for a child's vocabulary. It may then be necessary to implement a measure of vocabulary to ensure that the participant sample displays no vocabulary deficits and has a vocabulary range that is age appropriate.

One vocabulary scale which could be used is the British Picture Vocabulary Scale (BPVS). The BPVS is a Standard English vocabulary assessment test which is a well-established tool used widely for clinical, educational and research purposes. It is a measure of receptive (hearing) vocabulary and can be used to identify those children who may have language impairment and those performing below their age range. It consists of 14 sets of 12 words spoken orally by the examiner which increase in difficulty from one set to the next. For each word, the participant is required to point to the picture which best represents the meaning of that word in a 4 alternative forced choice (4AFC) task. Scores are converted to standardised scores to allow comparisons with well-defined age related norms and enable individual scores to be compared against large groups of people of the same age. The standardisation of scores is based on the normal distribution of scores expected for a particular age range and the raw scores are transformed into standardised scores with a mean of 100 and a standard deviation of 15, where 68% of people score between 85-115 (Dunn et al. 1997).

3.1.2.1 *Development, reliability and validity of the BPVS*

The BPVS was developed from an American vocabulary test, the Peabody Picture Vocabulary Test (PPVT) which was established in 1959 in the USA and is there an accepted vocabulary test. In order to determine which words and pictures to use, words together with black and white artwork was tested on large groups of children to determine which age groups could correctly identify which words and whether the pictures were adequate. Items were included covering a wide range of categories involving concepts likely to be encountered by children. Words were selected if they showed an increase in percent correct across successive age groups, displaying a linear steep growth curve. Standardisation of the BPVS was carried out across a range of schools in England and Wales. A total of 2571 children took part from 152 schools which were found to be largely representative of the national population.

The internal consistency of the BPVS was evaluated using split-half reliability calculations whereby the performance on odd and even numbered items was compared and correlated. Corrections were calculated on the raw scores of both sets to estimate the reliability of a test which was twice as long (using the Spearman-Brown formula). The reliability coefficients for each age group are shown in table 3.1 and are shown to be high.

Table 3.1: Reliability of the BPVS scores using the split-half reliability coefficient in each range adapted from Dunn et al. (1997)

School year	Correction Split-half reliability coefficient
Pre-School (3-5 years)	0.89
Year 1 (5-6 years)	0.86
Year 3 (7-8 years)	0.81
Year 5 (9-10 years)	0.95
Year 6 (10-11 years)	0.89
Year 7 (11-12 years)	0.86
Year 10 (14-15 years)	0.86
Median	0.86

For each standardised score there is also a standard error of measurement which has been calculated as 5.6 standard score points, therefore the standardised score falls within a confidence band of ±5.6 points for a 68% level of confidence. For a 95% level of confidence the standardised score falls within ±11.2 points.

An experiment comparing the results between native English participants and those with English as an additional language (Whetton 1997) found that, across 410 pupils (spread evenly across pre-school, year 1 and year 3), native English participants scores were superior, which is expected and thus substantiates the validity of the BPVS.

Tests of vocabulary can be used to measure verbal intelligence since this measure been found to be the best single predictor of success at school (Dale & Reichert 1957) and the PPVT, on which the BPVS was based, has been reported to best predict cognitive ability in young children (McCormick et al. 1994). Used in native English speakers only, the BPVS can be utilised to assess scholastic aptitude but does not measure other important aspects of general intelligence.

3.1.3 SNR confounds

As mentioned previously in Chapter 2 section 2.3.4, It is important to consider the fact that differences found in previous research comparing SRTs across both a steady state speechshaped noise masker and a single-talker speech masker between children and adults, may be due to an SNR confound (Bernstein & Brungart 2011). A SNR confound may be present when using adaptive procedures because the slope of the psychometric functions are not taken into account when comparing one intelligibility level and comparing performance across masker types and age groups. The presence of a possible SNR confound will be discussed in this first experiment and measures will be taken to determine if child/adult differences with speech backgrounds are eliminated when this confound is taken into account. For more details see Chapter 2 section 4.1.1.

3.2 Aims of experiment 1

1) **To determine the differences between speech intelligibility in children and adults in different masker types (steady state speech-shaped noise and a single-talker speech masker) to establish any masker*age interactions.**

   Rationale: Whilst speech intelligibility in background noise in children and adults has previously been considered, focus on the difference between children and adults' speech intelligibility particularly with a speech masker has received little consideration. Furthermore, the studies that do exist have often used cognitively *demanding speech tasks and complex maskers without taking into account the cognitive skills of children.*

Predictions: (i) A steady state speech-shaped noise masker will be more detrimental to speech perception than a single-talker masker (Festen and Plomp 1990); (ii) compared to the stationary noise masker, single-talker maskers will affect children more greatly than adults (Hall et al. 2002; Bonino et al. 2013; Leibold and Buss 2013)

2) To determine how the differences between speech intelligibility in children and adults in these two masker types changes over time.

Rationale: Few studies have investigated development over time within the same sample. It is of interest to determine how performance with children changes over time with both masker types. This could provide a greater insight into how different age groups cope with background sound, with regards to the specific features of each masker type, to further understand the course of development.

Predictions: (i) Children aged 5-6 years will reach near adult-like levels with the noise masker when tested one year later aged 6-7 years (Elliot 1979; Papso and Blood 1989; Leibold and Buss 2013; Bonino et al. 2013) (ii) Children aged 5-6 years will not reach near adult-like levels with the speech masker when tested one year later aged 6-7 years (Hall et al. 2002; Wightman and Kistler 2005; Leibold and Buss 2013; Bonino et al. 2013).

3.3 Methods

3.3.1 Overview and rationale

Participants were tested on their ability to listen to and identify words in the presence of two different noise maskers (see stimuli section below for details) presented diotically over headphones. The masker types were chosen to encompass features of both 'energetic' and 'informational' masking as previously defined. A modified version of the clinical speech test, the McCormick Automated Toy Discrimination test (Summerfield et al. 1994) was used to estimate the speech reception threshold, SRT (dB SNR) required for participants to achieve

79.4% of words correct in a 14AFC task. The McCormick Automated Toy Discrimination test was chosen as this is test is widely used with children in current clinical settings, and is not particularly cognitively demanding. A repeated measures design was used in which each participant underwent all of the conditions being tested. A vocabulary test was also carried out in both children and adults to ensure that their vocabulary range was at age appropriate levels. The speech intelligibility test was carried out with both adults and children, and then a sample of the children tested were re-tested one year later to investigate any developmental effects.

3.3.2 **Participants**

Ethics and research governance approval was obtained from the Faculty of Engineering and the Environment Human Experimentation Safety and Ethics Committee before commencing this experiment (see Appendix A for safety and ethics approval emails).

3.3.2.1 *Children*

Normal hearing, native English children (aged 5-6 years), with no known special educational needs were recruited from a participating school. Invitation letters consisting of consent forms and otological health questionnaires (see Appendix B) were sent home to parents of all children in year 1 (n=90). Year 1 children (aged 5-6 years) were chosen to enable comparisons with previous studies (Hall et al. 2002; Litovsky 2005; Johnstone & Litovsky 2006; Bonino et al. 2013; Leibold & Buss 2013). This age range was also chosen since this is an age where children are beginning to enter noisy environments at school and so an age where background sound may begin to have more of an impact on them. Finally, this age range was chosen for the practical reason that this decision best suited the participating school.

Ninety invitation letters were distributed of which 69 signed consent forms were returned. Of those 69 children, 12 children were excluded from this study because they were either nonnative English speakers, were not deemed otologically normal, or had special educational needs determined from the answers to the paper questionnaires filled out by parents. One further child did not complete testing due to absence during allocated testing time. Fifty-six children went on to complete the hearing screening test which was carried out in a quiet room within the school. Pure tone audiometry with circumaural earphones (ISO_389-8: 2004) was carried out with fifty-six children. Only if the child reliably responded to 20 dB HL at 1-4 kHz and 30 dB HL at 500 Hz in both ears did he or she carry on with the study. Five children did not reliably respond and were sent a letter home to their parents detailing the result. One child displayed poor attention to the main experimental task, thus a total of 50 children (23 males, 27 females) aged from 5 years and 10 months to 6 years and 8 months (mean 6 years 3.5 months), participated in this study.

All 50 children were tested from year 1 in the first year of testing. One year later the same 50 children were recruited again in the same way as before, and consent forms were obtained from 34 of those 50 children. In the second year of testing all 34 children's audiometric, otological and educational status was rechecked. All 34 children (16 males, 18 females) now aged from 6 years 11 months to 7 years 10 months (mean 7 years 9 months) participated in the follow up experiment. Only the 34 children in the follow up study underwent the vocabulary test as this test was decided upon only in the second year of testing.

3.3.2.2 *Adults*

Fifty normal, native English speaking adults (21 males, 29 females) aged from 18 years 10 months to 43 years 3 months (mean 24 years 9 months), with no known otological disorders or special educational needs were recruited from students and friends of the University of Southampton (see Appendix B for consent forms and questionnaires). Hearing threshold levels satisfied the same criteria as with the children. Thirty of these fifty adults were tested alongside the fifty children in the first year of testing so did not undergo the vocabulary tests. A further twenty adults were recruited in the second year of testing and so did undergo the vocabulary tests. Attempts were made to re-recruit the previously tested thirty adults from the first year of testing so to carry out the vocabulary checks, and fifteen of these adults were available to be retested. Therefore a total of 35 of the 50 adults recruited were assessed on their vocabulary range.

3.3.3 **Equipment**

A laptop was connected to a Creative Extigy external sound card. A custom written MATLAB code (MATLAB 2012) controlled the entire procedure and generated and controlled the stimuli for both the experiment and the hearing screen. Sennheiser HDA 200 cicumaural headphones were used to deliver sound diotically. A touch-screen monitor was used for the participants to provide their responses to the stimuli. The level of all stimuli through the headphones was calibrated using a sound level meter attached to an artificial ear. This objective calibration took place at the beginning of the experimental period and once every three weeks thereafter. Subjective listening checks occurred at the start of every experimental session by the primary researcher.

3.3.4 Stimuli

The target stimuli consisted of target words presented in a carrier phrase which was either "point to the", "where's the" or "show me the", spoken by an English female. The target words were from the clinical McCormick Automated Toy Discrimination test (Summerfield et al., 1994) which consists of seven pairs of words with similar sounds (duck/cup, house/cow, horse/fork, spoon/shoe, tree/key, man/lamb and plate/plane). Target words were selected at random.

Two masker conditions were used. In one condition the speech-shaped noise associated with the toy test (Summerfield et al. 1994) was used (i.e. the spectrum had been adjusted to match the LTASS of the speech targets). The second masker condition consisted of single-talker speech sentences used as the masker. The sentences were taken from the corpus of the Institute of Hearing Research (IHR) sentence lists (Macleod & Summerfield 1990) spoken by an English male. Each sentence by itself contained meaningful grammatical speech. For this study, two randomly selected sentences were concatenated and a random segment of this was then used as the masker. So the two sentences may have been "she ironed her skirt" and "the floor was quite slippery" but the final masker used may have included "her skirt the floor was quite". This was done to enable control of the duration of the speech masker. The segment was checked to be equivalent to the long term average rms level of the noise. The masker duration was chosen to be longer than the target phrase by 0.5 s to 1 s, the exact value being varied randomly from presentation to presentation. The target was temporally centred within the masker. This was to add unpredictability to the start of the target presentation to reduce the likelihood that the participant could use an onset cue to facilitate sound source segregation. The maskers were presented with a raised-cosine onset and offset ramp of 100 ms in duration. All stimuli were generated digitally and played out with a 44100 Hz sampling rate via a 16 bit Creative Extigy external sound card. When the SNR was set at 0 dB SPL, the level of the target and masker individually was at 55 dB A. The levels of the both the target and masker were then varied to generate required SNRs.

3.3.5 Vocabulary test

The vocabulary test carried out was the British Picture Vocabulary Scale (BPVS) (Dunn et al. 1997). The BPVS is a widely used and well established vocabulary assessment tool designed to assess the range of vocabulary in children (primarily) and adults. It involves the participant being shown four pictures at a time whereby they will be asked to point to one picture from those four options which best describes a word (spoken by the experimenter). This was carried out with several different words and in total lasted approximately 15 minutes. The BPVS was carried out only in the second year of testing. Therefore only the 34 repeated children in the second year of testing underwent this test alongside the 20 adults in the second year of testing, as well as a further 15 adults who were re-recruited from the first year of tested adults, producing a total of 35 adults who completed the BPVS.

3.3.6 **Procedure**

3.3.6.1 *Overall Structure*

Testing was carried out on two separate days and a single testing session lasted approximately 25-40 minutes each for the children and 20-35 minutes each for the adults. Two SRT (dB SNR) estimations were obtained in each session, on each day, for each masker condition. Therefore, four SRT (dB SNR) estimations were obtained in total for each participant. Each testing session was identical to enable the repeatability of the results to be examined. Pure tone audiometry was however carried out in session one and the vocabulary test was carried in session two with the 34 follow up children only, and for previously tested adults the vocabulary test was carried out in in a standalone session.

Figure 3.1: The experimental arrangement in a room within the school

Child participants were seated in a quiet room within the school; adult participants were seated in a quiet room within the university. Ambient noise levels were less than 40 dB A. Participants sat at a table facing a touch-screen monitor (see Figure 3.1). Fourteen pictures were displayed corresponding to all fourteen words which make up the targets within the McCormick Automated Toy test. On all trials, participants listened to the stimulus and then touched the appropriate picture on the touch screen. This made up one trial.

Participants were instructed to ignore 'the man' or 'the hissing noise' (i.e. the masker) and to listen carefully to 'the lady' (i.e. the target). Feedback was not given except positive encouragement. Children were given a short rest after each run where the headphones were removed and the child and the experimenter engaged in conversation to prevent the child becoming bored and inattentive before commencing the next adaptive run. Adults generally kept the headphones on throughout the duration of testing as inattention was not an issue. The orders of the conditions were counterbalanced between participants using a Latin square.

In the follow up experiment with children 1 year later, the structure of the experiment was identical to the first year except with the addition of the vocabulary test.

3.3.6.2 *Familiarisation*

At the start of the first session (after the hearing check in both ears) each participant was familiarised with the picture-word combinations of the toy test. If the participant was not familiar with the picture-word combination, they were told, and the experiment did not proceed until it was clear that the participant associated that picture and word (as per Summerfield et al. 1994). Most children knew all of the picture-word combinations straight away.

Participants began one run of the adaptive procedure where the SNR was generally set to + 10 dB. The run was then stopped after 10 familiarisation trials.

The starting SNR for the main experiment was decided according to the SNR of the familiarisation trials where the listener gave the first incorrect response. The starting SNR was set approximately 10 dB above this level. This starting SNR often began at approximately 0 dB for most participants and with both masker types. The starting SNR was chosen in this way for each participant for each masker type so as to focus as many trials as possible close to the participants' SRT (dB SNR) and to prevent wandering attention, which was especially important with the children. These 10 familiarisation trials were conducted in the same way at the start of Session 1 and Session 2.

3.3.6.3 *Adaptive procedure*

In an adaptive tracking procedure the level of both the targets and the maskers varied in order to generate required signal to noise ratios (SNR) for the test, with a decrease in SNR resulting in a reduction in intensity level of the signal together with an increase in intensity level of the masker to meet the required SNR and vice versa. Both the target and the masker were varied in this way to prevent delivered sound becoming too loud for the participants. It has been found that varying the signal level and fixing the masker level, or varying the masker level and fixing the signal level results in differences of less than 1 dB with speech in noise measurements which are not considered noteworthy (Wilson & McArdle 2012).

The adaptive procedure for the McCormick Automated Toy test was modified to a 3-down-1up adaptive tracking procedure (Litovsky, 2005), with a theoretical asymptote of 79.4% on the psychometric function. This was in an attempt to improve the reliability of the test whilst using fewer trials, and also to allow comparison to previous studies (Litovsky 2005; Johnstone & Litovsky 2006). At the start of the adaptive procedure, the SNR changed in 8 dB steps using a 1down rule until the first incorrect response. If the very first trial was incorrect this was ignored and another trial was given at the same level.

Thereafter the increments followed a 3-down-1-up rule whereby the participant had to obtain three correct responses before the SNR decreased and one incorrect response, followed by a null trial at that same level, before the SNR increased. Figure 3.2 shows an example of an adaptive track for a single participant. For every incorrect response a null trial was presented at that same SNR which was not included in the analysis. Null trials were included in an attempt to provide the listener with more practise at the lower SNRs so in an attempt to get a truer estimate of their SRT (dB SNR) (Trahiotis et al., 1990). For the first trials encompassing the 3-down-1-up rule the signal to noise ratio changed by step sizes of 4 dB, which then halved to 2dB after the first reversal. If a step size was used twice consecutively in the same direction then the step size was doubled. Signal to noise ratio was set never to exceed +20 dB and testing was terminated after 5 reversals. Results from each adaptive track were calculated to determine SRTs (dB SNR) using two different analyses.

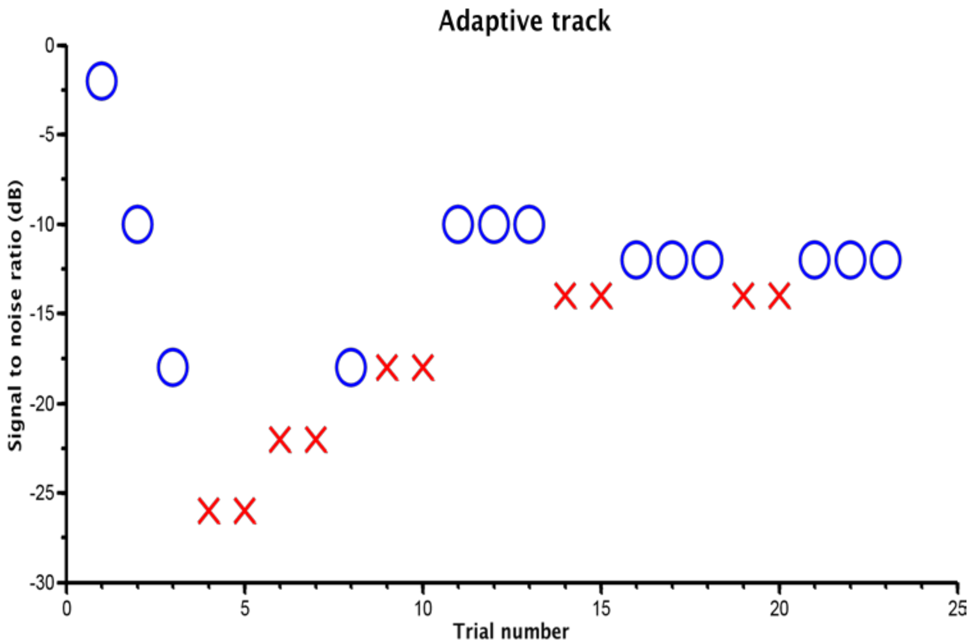


Figure 3.2: An example of one adaptive track (or run). The circles mean that the participant responded correctly, and the crosses mean that participant scored incorrectly. After an incorrect response a null trial was given which was not included in the analysis or adaptive tracking procedure.

The traditional approach of Levitt (1971) was used, whereby the points at which reversals occur were averaged over a certain number of reversals. In this case an average of these points over the last four reversals was taken. Another approach used by Litovsky (2005) was implemented using the maximum likelihood estimate (MLE) (Wichmann & Hill 2001a; Wichmann & Hill 2001b) to fit a logistic function to all of the points in the adaptive track so to estimate the psychometric function to deduce SRT (dB SNR). By looking at the number of trials amassed at each tested signal to noise ratio and taking the performance level, the psychometric function is estimated from which SRT (dB SNR) is deduced. Average SRTs (dB SNR) were then taken across 4 adaptive runs (four repeats, two within session 1, and two within session 2) for each masker type and for each method.

Test-retest data was collected to calculate the reliability of both methods by measuring the variability of SRT (dB SNR) measurements.

3.4 Results

3.4.1 Overview of results

The BPVS was carried out to determine the vocabulary range of both children and adults. Each participant's score was compared to a standardised score calculated according to their age. The results from the BPVS will be presented first.

SRTs (dB SNR) were estimated at the SNR required to correctly identify 79.4% of words within a noise masker or a speech masker. SRT (dB SNR) was estimated using the two methods previously described; averaging the last 4 reversals of the adaptive track and fitting a psychometric function to all the points within the adaptive track to determine the 79.4% correct level. Four separate measurements (repeats) were obtained for each masker type and each SRT (dB SNR) estimation method was applied to each measurement. Averages of these four measurements were then taken.

A comparison of the two SRT (dB SNR) estimation methods was carried out to establish the test-retest reliability of each method across all four repeats and will be presented next. Results and further analyses will then be presented in terms of the method which yielded better reliability. The SRT results will then be displayed in box plot diagrams and further analyses will be carried out. The error bars of the box plots represent the range of participants' maximum and minimum SRTs (dB SNR). The centre horizontal line within each box indicates the median of the results for the labelled conditions and the upper and lower horizontal lines of the boxes denote the upper and lower quartiles. The discussions of results will not be considered here, but will be in the discussions section 3.5.

3.4.2 Vocabulary test

Table 3.2: The mean, range and standard deviation of the standardised scores from the British Picture Vocabulary scale in children and adults.

	Adults (n=35)	**Children aged 6-7 years (n=34)**
Mean	134	109
Range	115 – 160	90 – 124
Standard deviation	15	9

The results of the British picture vocabulary scale (BPVS) are displayed in table 3.2. Results are shown for the 34 children taken when they were in year two (aged 6-7 years) and for 35 of the total 50 adults.

The BPVS calculates standardised scores based on age population means, only up to 15.8 years of age. The mean standardised score of both children and adults are displayed in table 3.3. A standardised score of 100 would suggest that a child had an average vocabulary range for their age. The standard deviation for standardised scores is 15, so 68% of people will score between 85 and 115 (Dunn et al. 1997). For adults, the standardised scores have been based on a child aged 15.8 years.

As can be seen from the table, the mean child standardised score in this sample was above the population mean. The mean adults standardisation scores is also above the population mean and is significantly higher than that of the child standardisation scores t (67) = 4.57,p<.001. The adults mean scores ranged from 115 to 160 and thus were all above the population mean. This is not surprising, but it is however reassuring that the adults did not display any vocabulary deficits with this test that may have interfered with the speech perception test.

The children's mean scores ranged from 90 to 124 and thus span above and below the population mean. The BPVS calculates age equivalent scores and percentile ranks together with their confidence bands. Table 3.3 shows details of individual standardised scores, age equivalent scores and the percentile rank for each child, in order of increasing standardisation score. Seven children's vocabulary scores fell within their age range (participants 1, 2, 14, 25, 28, 31, 33) and 24 children's vocabulary scores fell above their age range and for three of the children, scores fell below their age range (participants 3, 10, 32).

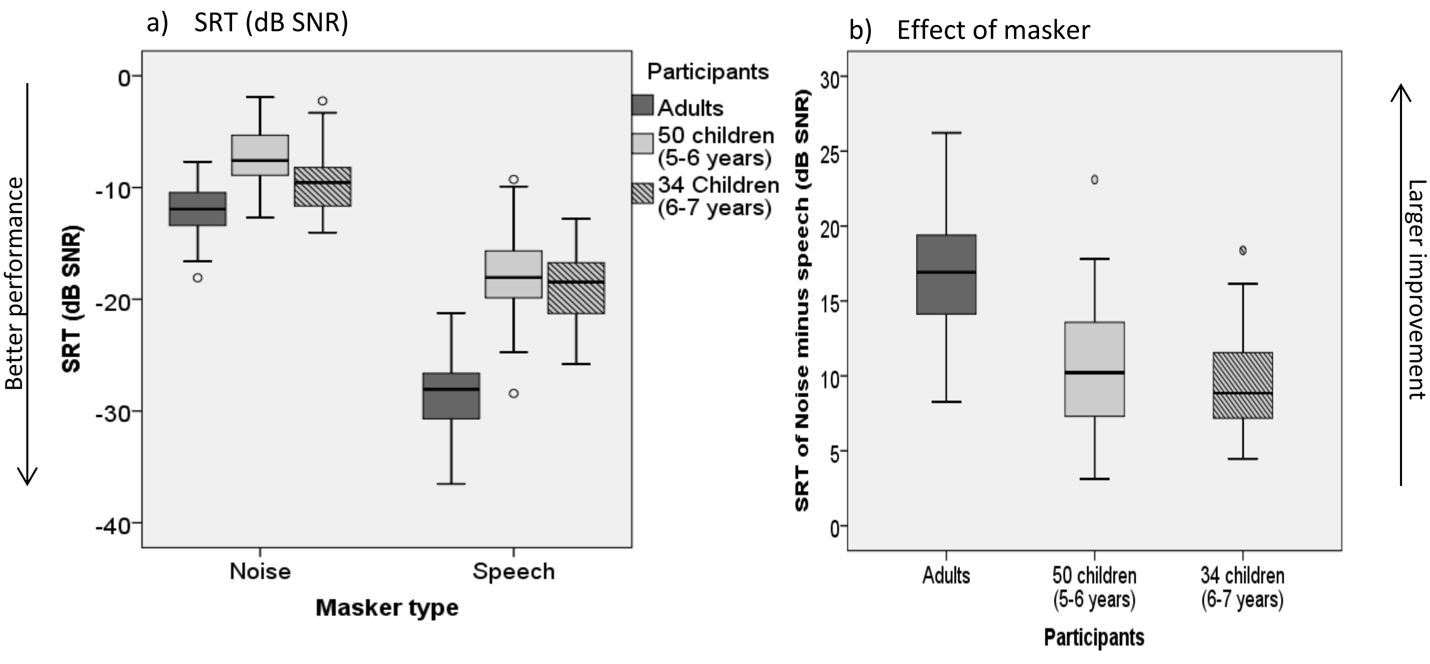
Table 3.3: The standardised scores, age equivalent scores and percentile ranks calculated from the BPVS for children (6-7 years) together with their upper and lower 68% confidence bands. An age of 7:01 denotes 7 years 1 month and an age of 7:10 denotes 7 years 10 months. A percentile rank of 77 indicates that 77 out of 100 children tested of the same age scored equal or below the standardisation score of this participant.

ID No.	Standardised score	Actual age (years: months)	Age equivalent (confidence bands)	Percentile rank (confidence bands)
10	90	7:06	6:05 (5:10-7:01)	26 (14-40)
3	90	7:10	6:08 (6:01-7:04)	26 (14-40)
32	92	7:10	7:00 (6:05-7:08)	30 (18-45)
33	96	7:00	6:07 (6:00-7:03)	40 (26-55)
25	100	7:01	7:01 (6:06-7:09)	50 (34-66)
1	100	7:08	7:06 (6:11-8:02)	50 (34-66)
34	102	7:03	7:06 (6:11-8:02)	55 (40-70)

28	102	7:07	7:10 (7:03-8:06)	55 (40-70)
14	104	7:06	8:00 (7:05-8:09)	60 (45-74)
31	105	7:08	8:01 (7:06-8:10)	63 (48-77)
2	106	6:11	7:05 (6:10-8:01)	66 (50-78)
26	106	7:09	8:07 (7:10-9:04)	66 (50-78)
9	108	7:08	8:05 (7:09-9:02)	70 (55-82)
18	109	7:02	7:11 (7:04-8:08)	72 (58-84)
8	109	7:06	8:07 (7:10-9:04	72 (58-84)
27	109	7:08	8:07 (7:10-9:04)	72 (58-84)
4	110	7:00	8:00 (7:05-8:09)	74 (60-86)
20	110	7:06	8:08 (8:00-9:06)	74 (60-86)
17	111	7:70	8:09 (8:01-9:08)	77 (63-87)
16	112	7:00	8:02 (7:07-8:11)	78 (66-89)
13	112	7:08	8:10 (8:02-9:10)	78 (66-89)
12	113	7:09	9:04 (8:06-10:03)	80 (68-90)
7	115	7:05	8:10 (8:02-9:10)	84 (72-92)
23	115	7:06	9:04 (8:06-10:03)	84 (72-92)
15	115	7:09	9:08 (8:09-10:06)	84 (72-92)
21	115	7:09	9:08 (8:09-10:06)	84 (72-92)
24	117	7:01	8:09 (8:01-9:08)	87 (77-94)
5	117	7:01	8:10 (8:02-9:10)	87 (77-94)
22	118	7:07	9:10 (8:10-10:09)	89 (78-94)
11	120	7:04	9:08 (8:09-10:06)	91 (82-96)
6	121	7:01	9:07 (8:07-10:04)	92 (84-96)
29	121	7:04	9:11 (9:00-10:10)	92 (84-96)
19	124	7:00	9:11 (9:00-10:10)	94 (89-98)
30	124	7:01	9:11 (9:00-10:10)	94 (89-98)

A standardised score of 85 and a standardised score of 115 represent ±1 standard deviation of the population mean respectively where the results of 68% of people from a certain age range will fall (Dunn et al. 1997). According to Dunn et al. (1997), standardised scores between 85 and 100 are classed as “low average scores” and those between 100 and 115 are classed as “high average scores”. Scores between 115 and 130 are classed as “moderately high scores”. The three poorer scoring children do not fall below “low average scores” or 1 standard deviation below the population mean. These three participants were therefore were included in all analyses but were excluded in an extra child/adult comparison in order to see if their poorer vocabulary range made any difference to the observed differences. For all the other children, it can be suggested that they have a vocabulary range which is appropriate for their age. All participants tested with the vocabulary test, apart from three of those children, did not display any vocabulary deficits which may be said to make them particularly poor with the speech perception test. Furthermore, the words within the speech test all have an age of acquisition of 4.95 years or under according to the age of acquisition ratings database

3.4.4 Speech Reception Thresholds (SRTs)



(Kuperman et al. 2012). There was found to be no significant correlation between vocabulary score and SRT (dB SNR) in children when in year one with the noise ($r = .18$, $p > .05$) or with the speech masker ($r = .04$, $p > .05$). There was also no significant correlation when children were in year 2 with the noise ($r = .02$, $p > .05$) or with the speech maskers ($r = -.28$, $p > .05$).

3.4.3 Comparisons of test-retest repeatability between methods of SRT (dB SNR) estimation

Table 3.4: Measurement error (95% confidence intervals) for each method of SRT (dB SNR) estimation in each child group and adults with both noise and speech maskers. These confidence intervals were calculated from the variability measure of within subjects standard deviation of scores as detailed by Summerfield et al. (1994) and described in section 3.1.1.2.

	Analysis 1: Average reversals		Analysis 2: MLE	
	Noise (dB)	**Speech (dB)**	**Noise (dB)**	**Speech (dB)**
Year 1 - 50 children (5-6 years)	±8.2	±9.6	±10.1	±10.0
Year 1 - 34 children (5-6 years)	±7.6	±10.5	±9.8	±10.8
Year 2 – 34 children (6-7 years)	±7.1	±10.1	±7.6	±12.3
Adults	±6.7	±9.1	±7.2	±10.7

A comparison of the two SRT (dB SNR) estimation methods was carried out to establish the reliability of each method across all four repeats. For both SRT estimation methods, reliability was investigated by calculating measurement error across four measurements (repeats) as detailed in (Summerfield et al. 1994) and explained in section 3.1.1. It can be seen from table 3.4 that measurement error is consistently larger with the MLE method than the average reversals method across both age groups and masker types. Therefore the average reversals method shows greater reliability. The reasons for such findings are unclear and will be considered in the discussions section. The main results will thus be displayed in terms of the average reversals method.

Figure 3.3: a) Box plots to show the effect of masker type on SRT (dB SNR) in each child group and in adults with both masker types, b) Box plot to show the difference in SRTs (dB SNR) between masker types as calculated by subtracting SRT (dB SNR) with speech masker from that with the noise masker. In each graph the circles represent the outliers.

Figure 3.3 (a) shows the spread of participants' SRT (dB SNR) results (calculated from the average reversals method) for all participants' average scores (averaged across four repeats, both sessions). The trends in the graph show that performance is better with the single-talker speech masker compared to the noise masker, for all participant groups. It can be seen for the noise masker that adult's performance is better than the children's, and substantially better than children's with the speech masker. Children's performance appears to improve from one year to the next and this improvement looks to be larger with the noise masker compared to the improvement seen with the speech masker.

Figure 3.3 (b) shows the spread of participants' SRT (dB SNR) results for all participants' average scores in terms of the difference between the speech masker and noise masker's SRTs. This shows the improvement observed when the masker is switched from the noise masker to the speech masker. It can be seen that adults experience a larger improvement than children, and the improvement for children appears to be similar in both age groups.

3.4.5 **Statistical analyses**

To examine any significant effects and interactions from these results firstly the repeated measures variables will be defined. These are repeat number (repeat 1 vs. repeat 2), session number (session 1 vs. session 2) and masker type (noise vs. speech). The three participant groups cannot be treated as independent variables, since the same child participants are used between years 1 and 2. Therefore three separate ANOVA's were conducted to determine any statistically significant differences within the results. A fourth ANOVA was also included to examine any possible effects that removing the results from those three children with vocabulary deficits (from the group of 34 children aged 6-7 years) may have on any masker*age interactions. See Appendix C for ANOVA tables.

These four separate ANOVA's were:

1) Four-way mixed ANOVA to compare one between subjects variable of age (all 50 year 1 children vs. adults) and three within subjects variables of repeat number (repeat 1 vs. repeat 2), session number (session 1 vs. session 2) and masker type (noise vs. speech).

2) Four-way mixed ANOVA to compare one between subjects variable of age (only 34 year 2 children 6-7 years vs. adults) and three within subjects variables of repeat number (repeat 1 vs. repeat 2), session number (session 1 vs. session 2) and masker type (noise vs. speech).

3) Four-way repeated measures ANOVA to compare four within subjects variables of repeat number (repeat 1 vs. repeat 2), session number (session 1 vs. session 2), year group, comparing children aged 5-6 years with children aged 6-7 years (year 1 vs. year 2), and masker type (noise vs. speech).

4) Four-way repeated measures ANOVA to compare one between subjects variable of age (only 31 children with age appropriate vocabulary in year 2, 6-7 years vs. adults)

   and three within subjects variables of repeat number (repeat 1 vs. repeat 2), session number (session 1 vs. session 2) and masker type (noise vs. speech).

3.4.5.1 *Stability of SRTs*

In order to determine the stability of any main effects or interactions across repeats and across sessions, firstly the results from the four ANOVAs with regard to the effects of repeat number and session number will be detailed.

The results of the first ANOVA (comparing children aged 5-6 years with adults) show no significant main effects of repeat number $F(1, 98) = 3.72$, $p=.06$ but a significant main effect of session number F (1, 98) = 27.97, p<.001. There was a significant interaction between session and age F (1, 98) = 5.14. p =.03, meaning that there was improvement seen from session 1 to session 2 which was larger in children than adults. A significant interaction was also found between masker type and session number F (1, 98) = 5.33, p =.02 and masker type and repeat number F (1, 98) = 5.06, p=.03 suggesting that participants appeared to improve to a greater extent with the speech masker than the noise masker both within and across sessions. Although there were learning effects within these results, there was no significant interaction between masker, session and age F (1, 98) = 1.65, p =.20, or masker, repeat and age F (1, 98) = .63, p =.43, suggesting that the presence of any masker age interaction would be repeatable across and within sessions.

The results of the second ANOVA (comparing children aged 6-7 years and adults) showed no significant main effects or interactions including repeat number. The results were therefore pooled and averaged across both repeats and a three-way mixed ANOVA was carried out (comparing session number, masker type and age). There was found to be a significant main effect of session number F (1, 82) = 8.86, p<.01 where similarly to the first ANOVA, overall SRTs were lower (better) in session 2 compared to session 1. This shows again that overall participants experienced a learning effect which occurred across but not within sessions. Unlike with the younger children in ANOVA 1, there was no significant interaction between session and age F (1, 82) = .17. p =.68, meaning that the improvement seen from session 1 to session 2 was the same in children and adults. Also dissimilarly to the younger children, there was no significant interaction between masker type and session number F (1, 82) = 3.74, p =.06 suggesting that improvement across sessions was the same with both masker types. Finally, there was no significant interaction between masker, session and age F (1, 82) = 1.37, p =.25, which again shows that despite learning effects, any masker age interaction with the children when one year older would be repeatable across and within sessions.

Like ANOVA 2 the results of the third four-way repeated measures ANOVA showed no significant interaction which included repeat number. The results were therefore again pooled and averaged across both repeats and a three-way mixed ANOVA was carried out. There was found to be a significant main effect of session number F (1, 33) = 19.18, p<.001. This shows that overall SRTs were lower (better) in session 2 compared to session 1, showing the same learning effect again across but not within sessions. No significant interactions were found between year and session F (1, 33) = 2.47, p =.13, suggesting that children, when in both years, showed a similar amount of improvement across sessions. Finally, there were no significant interactions between masker type and session number F (1, 33) = .86, p =.36, showing that the improvement across sessions was to the same extent in both masker types.

The results of the fourth ANOVA (comparing only those children without any vocabulary deficits and adults) showed no significant main effects or interactions including repeat number. The results were therefore pooled and averaged across both repeats and a three-way mixed ANOVA was carried out (comparing session number, masker type and age). There was found to be a significant main effect of session number F (1, 79) = 8.98, p<.01 where similarly to other ANOVAs, overall SRTs were lower (better) in session 2 compared to session 1. This shows yet again that overall participants experienced a learning effect which occurred across but not within sessions. As with ANOVA 2 there was no significant interaction between session and age F (1, 79) = .27. p =.61, meaning that the improvement seen from session 1 to session 2 was the same in children and adults. Also as with ANOVA 2 there was no significant interaction between masker type and session number F (1, 79) = 2.75, p =.10 suggesting that improvement across sessions was the same with both masker types. Finally, there was no significant interaction between masker, session and age F (1, 79) = 1.78, p =.19, which again shows that despite learning effects, any masker age interaction with the children when one year older would be repeatable across and within sessions.

3.4.5.2 *Effect of masker type in children and adults*

Considering the effects of masker type in children and adults, the results of the first ANOVA (comparing children aged 5-6 years and adults) show a significant main effect of age F (1, 98) = 292.08, p <.001, and a significant main effect masker type F (1, 98) = 1160.92, p <.001. This indicates that overall SRTs were lower (better) for adults than children and lower (better) with the speech masker than the noise masker. A significant interaction was found between masker type and age F (1, 98) = 64.62, p <.001 suggesting that the masker type effected children and adults' SRTs (dB SNR) differently in that children's SRTs (dB SNR) were disproportionately higher (poorer) than adults' with the speech masker compared to the noise masker.

The results of the second pooled three way ANOVA (comparing children aged 6-7 years and adults) showed a significant main effect of age F (1, 82) = 141.97, p <.001, masker type F (1, 82) = 959.15, p <.001. Similarly to the first ANOVA comparing adults with the younger children, overall SRTs (dB SNR) were lower (better) for adults than children and lower (better) with the speech masker than the noise masker. In the same way as with the younger children, a significant interaction was found between masker type and age F (1, 82) = 71.38, p <.001 suggesting that the masker type effected the older children and adults SRTs (dB SNR) differently in that children's SRTs (dB SNR) were again disproportionately higher (worse) than adults' with the speech masker compared to the noise masker.

Considering the third ANOVA (comparing children aged 6-7 years with children aged 5-6 years), there was found to be a significant main effect of year F (1, 33) = 15.48, p<.001 and masker F (1, 33) = 457.70, p<.001. This shows that overall SRTs were lower (better) when children were in year 2 than when in year 1 and lower (better) with the speech masker than the noise masker. No significant interactions were found between year and masker F (1, 33) = .22, p=.65, implying that the both masker types affected the SRTs (dB SNR) of children in year 1 and year 2 in the same way.

The results of the fourth pooled three way ANOVA (comparing only those children without vocabulary deficits aged 6-7 years and adults) showed a significant main effect of age F (1, 79) = 134.77, p <.001 and masker type F (1, 79) = 885.69, p <.001. Similarly to the first and second ANOVA comparing adults with the younger children, overall SRTs (dB SNR) were lower (better) for adults than children and lower (better) with the speech masker than the noise masker. In the same way as with all children aged 6-7 years, a significant interaction was found between masker type and age F (1, 79) = 64.44, p <.001 suggesting that the masker type effected the older children and adults SRTs (dB SNR) differently in that children's SRTs (dB SNR) were again disproportionately higher (worse) than adults with the speech masker compared to the noise masker. Therefore the masker*age interaction was still present despite removing those children with vocabulary deficits. Moreover, the 3 children who had below average vocabulary scores achieved an SRT (dB SNR) of-7.3, -6.6, and -5.0 dB SNR in the speech-shaped noise masker with a collective average of -6.3 dB SNR. The average SRT in the noise masker for the rest of the group was -7.2 dB SNR. With the speech masker they scored -16.9, -14.6 and -20.2 dB SNR with a collective average of -17.2 dB SNR whilst the average for the rest of the group was -17.6 dB SNR. Therefore, the SRT results of these participants were not too dissimilar from the rest of the children.

Figure 3.3 (b) shows the spread of participants' SRT (dB SNR) result for all participants' average scores displayed as the difference between the speech masker and the noise masker. This shows the improvement observed when the masker is switched from the noise masker to the speech masker. To examine any significant differences from these results independent samples t-tests were carried out to compare adults results with all 50 year 1 children (aged 5-6 years) and year 2 children (aged 6-7 years). Adults (M=16.85 dB, SE=.57) showed a significantly higher improvement than the children when in year 1 (M=10.36 dB, SE=.57), t (98) = 8.05, p<.001 and a significantly higher improvement than children when in year 2 (M=9.63 dB, SE=.61), t (82) = 8.44, p<.001. A paired samples t-test showed no significant difference between the improvement observed by the same 34 children when in year 1 (M=9.98 dB, SE=.58) and children when in year 2 (M=9.63 dB, SE=.61), t (33) =.46, p>.05.

3.4.5.3 *Stability of individual differences*

The results from the ANOVAs in section 3.4.5.1 indicate that there is an effect of session, with participants showing a learning effect from one session to the next. It is of interest to explore the individual differences in participants' results across sessions to explore the stability of individual results.

Figure 3.4 shows the individual average data points of SRT (dB SNR) from session 1 and 2 plotted against each other with all adults, all 50 year 1 children and 34 year 2 children. Pearson's correlation coefficient was determined to investigate the stability of individual differences from one session to the next, to test the directional (one tailed) hypothesis that those participants who do well in one session also do well in the next session. The results are displayed in table 3.5.

Table 3.5: Results from Pearson's correlation coefficient with participant groups in each masker type across sessions

	Masker	Pearson's correlation coefficient	Significant Correlation?
Adults	Noise	r = .12, p>.05	No
	Speech	r = .42, p<.01	Yes
All 50 year 1 children	Noise	r = .23, p=.06	No
	Speech	r = .43, p <.01	Yes
34 year 2 children	Noise	r = .34, p=.03	Yes
	Speech	r = .16, p=.18	No

To examine any significant differences between the group results from session 1 and session 2, paired samples t-tests were also carried out. The results are displayed in table 3.6.

Table 3.6: Results from t tests with participants groups in each masker type across sessions.

	Masker	t test statistic.	Significant difference?
Adults	Noise	t (49) = -.08, p>.05	No
	Speech	t (49) = 3.12, p<.01	Yes
All 50 year 1 children	Noise	t (49) = 3.74, p<.001	Yes
	Speech	t (49) = 3.92, p <.001	Yes
34 year 2 children	Noise	t (33) = 1.67, p>.05	No
	Speech	t (33) = 1.47, p>.05	No

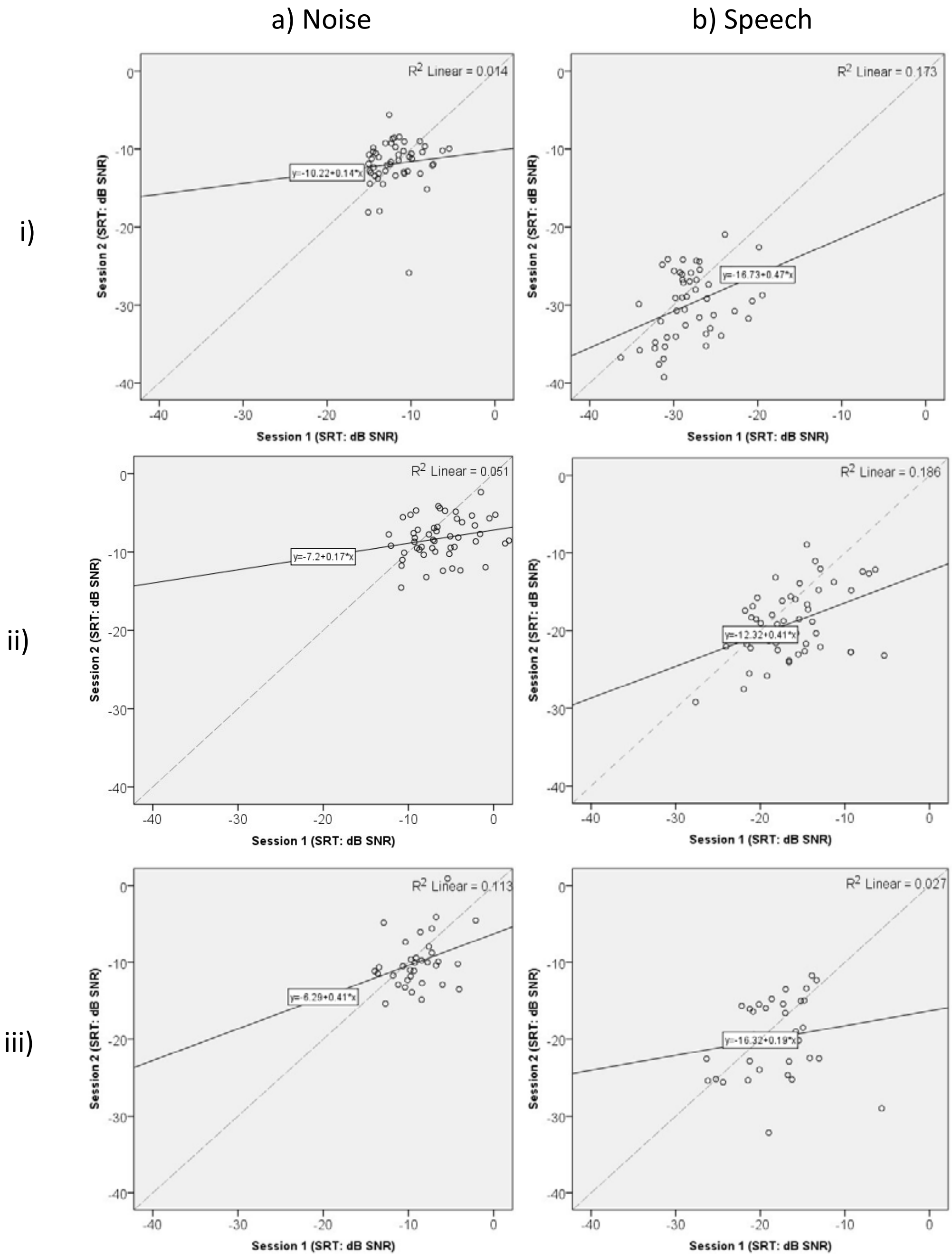


Figure 3.4: Scatterplot of individual participants' SRT (dB SNR) in session 1 plotted against SRT (dB SNR) in session 2. (i) Adults, (ii) all 50 year 1 children, (iii) 34 year 2 children. R^2 indicates the coefficient of determination showing the amount of variability in session 1 shared by session 2.
Data points below the dotted line indicate that SRTs were lower in session 2 compared to session 1. Data points above the dotted line indicate that SRTs were higher in session 2 compared to session 1.

Owing to the large variability within the results, correlation analysis shows that the methods may not be sufficient to determine individual differences. Therefore the experiment may only best capture groups differences between and not within cohorts.

3.4.6 Differences between the two methods of SRT (dB SNR) estimation

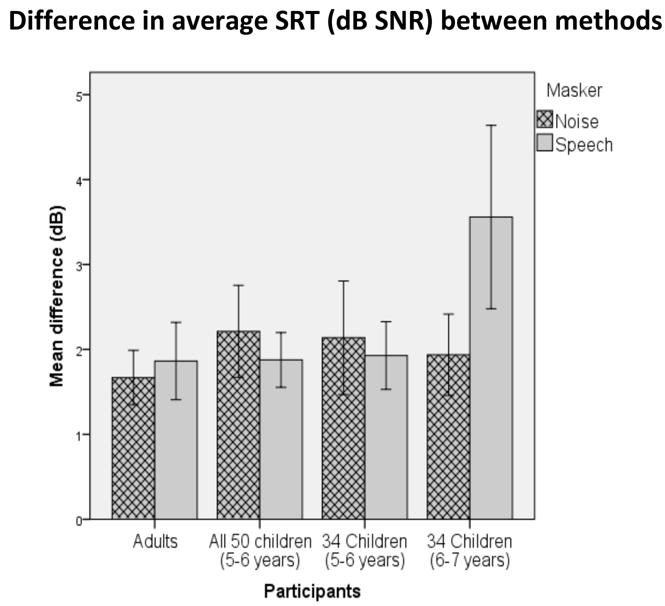


Figure 3.5: Differences in average (across 4 repeats) SRT (dB SNR) estimation methods for all test groups and both masker types. Error bars show ±95% confidence intervals.

Whilst variability is improved with the average reversals method it is of interest to portray just how different the results from both methods are. Figure 3.5 shows the differences in average SRT (dB SNR) estimation in all child groups and adults with both masker types. It can be seen from this graph that the differences between methods are small and are similar in all groups. The difference with the speech masker in the year 2 group (34 children 6-7 years) however appears larger than the rest; the reasons for this are unclear but will be considered in the discussion section.

To further investigate the impact of the difference between the two SRT (dB SNR) estimation methods, a three-way mixed ANOVA was carried out. As previously mentioned, since the children in year 1 and year 2 consist of the same children, each age group cannot be treated as an independent measures, therefore two separate three-way mixed ANOVAs were carried out to compare one between subjects' variable of age (ANOVA 1: all 50 year 1 children vs. adults/ANOVA 2: 34 year 2 children vs. adults), with two within subjects' variables of masker type (noise vs. speech maskers) and method type (MLE vs. average reversals) to do determine if the two method types produced the same interactions. See Appendix C for ANOVA tables.

In both ANOVA's a significant main effect of method was found (ANOVA 1: $F (1, 98) = 305.37$, $p<.001$/ANOVA 2: $F (1, 82) = 137.40$, $p<.001$), showing that the average reversals method yielded lower SRTs (dB SNR) than the MLE method. There was no significant interaction between method and age (ANOVA 1: $F(1,98) = 1.42$, $p>.05$/ANOVA 2: $F(1,82) = 2.12$, $p>.05$) meaning that the average reversals method yielded better results compared to the MLE method to the same extent in both children and adults. There was no significant interaction between masker and method type (ANOVA 1: $F (1, 98) = .24$, $p>.05$/ANOVA 2: $F (1, 82) = 1.63$, $p>.05$) suggesting that masker type had the same effect on SRT (dB SNR) regardless of which method was used. Finally, there was no significant interaction between masker, method and age (ANOVA 1: $F (1, 98) = 1.90$, $p>.05$/ANOVA 2: $F (1, 82) = .53$, $p>.05$). This suggests that the masker age interaction was not significantly different between method types and so the interaction was repeatable in both methods.

To further investigate the effect of method type in only 34 year 1 children and 34 year 2 children a final three-way repeated measures ANOVA was carried out to compare three within subjects' variables of year group (year 1 vs. year 2) masker type (noise vs. speech maskers) and method type (MLE vs. average reversals) to do determine if with this comparison the two method types produced the same interactions. A significant main effect of method was found $F (1, 33) = 115.00$, $p<.001$ showing that SRT (dB SNR) estimations were lower with the average reversals method than the MLE method. There was no significant interaction between method and masker $F (1, 33) = .21$, $p<.05$ indicating that the effect of masker was the same across both methods. Finally, there were no significant interactions between year and method $F (1, 33) = .41$, $p>.05$ meaning that the effect of method was repeatable from one year to the next.

Although both methods yield significantly different results, the main interaction between masker and age was repeatable in both methods suggesting that this finding is present regardless of which method is used.

3.4.7 Exploration into reasons for the masker*age interaction

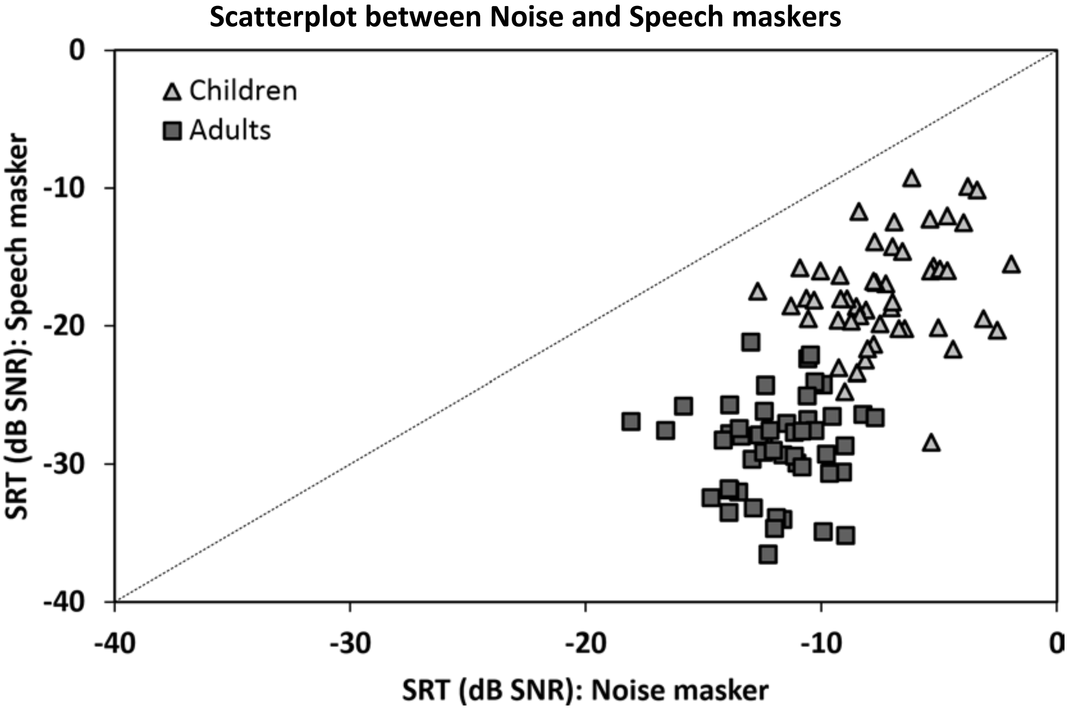


Figure 3.6: Scatterplot of SRT (dB SNR) with the speech masker against SRT (dB SNR) with the noise masker in all 50 adults and all 50 children (year 1, aged 5-6 years). In all cases data points below the dotted line indicate lower SRTs (dB SNR) with the speech masker compared to the noise masker. Regression analysis were attempted to establish if the regression lines with children and adults had the same slope and intercept. If so, Bernstein and Brungart's (2011) proposition would apply here, and larger SRT differences between children and adults with the speech masker would be due simply to the differences in the baseline condition. Regression analyses proved inconclusive.

In order to explore the reasons behind the masker*age interaction, attempts were made to investigate possible SNR confounds. Results from children (all 50 children year 1, aged 5-6 years) and adults with the noise masker were correlated with results with the speech masker (Figure 3.6). This was to determine whether, as highlighted by Bernstein and Brungart (2011), the larger observed differences between children and adults with the speech masker are due to the differences in the baseline conditions (i.e. with the noise masker).

Regression analyses were attempted to establish if the regression lines with children and adults had the same slope and intercept, if so Bernstein and Brungart’s (2011) proposition would apply here, and larger SRT differences between children and adults with the speech masker would be due simply to the differences in the baseline condition. So the theory was that if regression lines overlapped then if there had been children who performed as well as adults with the noise masker, they also would have performed just as well as adults with the speech masker. In this same way if there had been adults who performed as poorly as children with the noise masker, they too would have performed just as poorly as children with the speech masker. So differences in SRTs (dB SNR) with the speech masker would then be due to baseline differences with the noise masker. If the lines did not overlap however then if there had been children who had performed just as well as adults with the noise masker they may have still remained poorer with the speech masker. In this same way if there had been adults who had performed as poorly as children with the noise masker, their performance still would have remained better with the speech masker.

Linear regression was considered where both the slope and intercept did not overlap, however it was realised this was not a viable statistical analysis. Owing to errors in both the noise masker and speech masker variables linear regression was not an accurate measure so orthogonal regression was attempted, which does take into account the errors in both variables. Such analyses however remained inconclusive regarding whether the slopes and intercepts were the same due to the large variability between measurements thus regression lines are not presented.

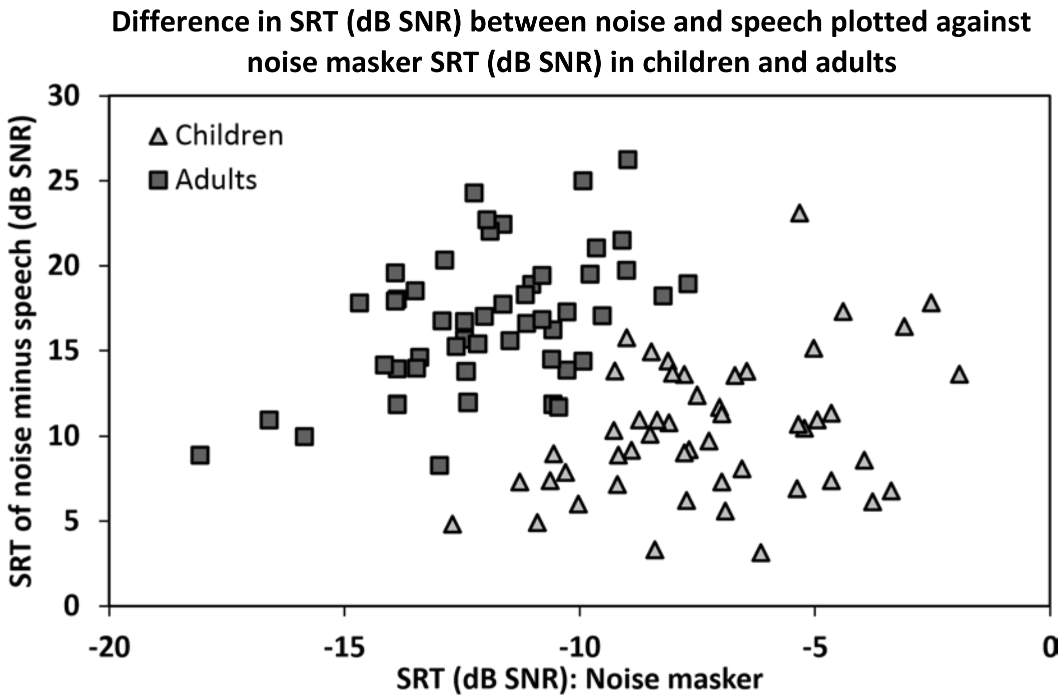


Figure 3.7: Graph to show the differences in SRT (dB SNR) between masker types plotted against SRT (dB SNR) in the baseline noise conditions.

To look again more closely at the presence of possible SNR confounds a method of analysis proposed by Bernstein and Grant (2009) was further employed. The improvement afforded when switching the masker from a noise background to a single-talker speech background (i.e. the SRT (dB SNR) in noise minus speech) was plotted against the baseline noise SRT (dB SNR). This is shown in Figure 3.7. Such an analysis was thought to take into account differences between children and adults in the baseline noise conditions (see Chapter 4, section 4.1.2.1 for more details). In the same way as above, linear regression lines are not presented owing to the large errors in both variables. Despite this however, it is possible to look at the results where children and adults overlap and have the same baseline noise SRT (dB SNR), for example in the approximate region of – 13 dB SNR to – 7 dB SNR in the noise masker. In this region it can be seen that children still seem to get less improvement with the speech masker compared to the noise masker in comparison to adults. If the differences between listener groups mainly reflected differences in the baseline noise conditions then it may have been that children adult differences would be reduced when comparing them at the same baseline noise condition. Here however, this does not seem to be the case. Analyses using more reliable methods may provide more informative results to establish the relevance of this possible SNR confound and are addressed in Chapter 4.

3.5 Discussion

3.5.1 Overview of results

The results of this study suggest that children need higher (better) signal to noise ratios than adults to achieve the same speech intelligibility levels in a noise masker, a finding which was expected and one that coincides with previous research (Wightman & Kistler 2005; Hall et al. 2002; Fallon et al. 2000; Nittrouer & Boothroyd 1990; Papso & Blood 1989; Elliott 1979; Litovsky 2005; Johnstone & Litovsky 2006; Bonino et al. 2013; Hall et al. 2012; Leibold & Buss 2013). Children also require a disproportionately higher (better) signal to noise ratio than adults with a speech masker, which also mirrors previous research (Hall et al. 2002; Bonino et al. 2013; Leibold & Buss 2013). These findings will now be discussed with reference to reliability, previous research and their implications.

3.5.2 Effect of masker type in children and adults

To fulfil the first aim of this experiment, masker*age interactions were explored with children aged 5-6 years and adults and with children (1 year on) aged 6-7 years and adults. Masker*age interactions were found between both ages of children and adults indicating that with the speech masker, children across an age range of 5-7 years had disproportionately higher (poorer) SRT (dB SNR) results than adults compared to the results with the noise masker. Such results were as expected. Considering the speech-shaped noise masker, the average child/adult difference in this study was 4.7 dB for children when in year 1 decreasing to 2.5 dB for children when in year 2. This finding fits in with previous research as Nittrouer & Boothroyd (1990) found that children aged 4-6 years displayed a 3 dB child-adult difference when assessing ability to recognise phonemes in speech-shaped noise at the same intelligibility level. Hall et al. (2002) who also used a similar adaptive procedure to the present study with children aged 5-10 years, found average child/adult differences of 3 dB with speech-shaped noise. The child/adult differences with the two-talker speech masker in the Hall et al. (2002) study were larger than the child/adult differences with the noise masker at 7 dB. In the present study the child/adult difference with the single-talker masker was found to be 11.1 dB for children when in year 1 and 9.7 dB for children when in year 2. Although such differences are slightly larger with the present study, this could be due to variations in the methods, since Hall et al. (2002) used a two-talker masker and not a single-talker masker and slightly older children overall.

To fulfil the second aim of this experiment, children were tested again 1 year later to gain insight into their ability to cope with background noise in order to understand more about how the energetic and informational features of a masker affect them as they develop. Whilst there was a significant effect of year, suggesting an improvement in children's results, there was no significant masker*age interaction. This shows that the children 1 year on displayed improvements in the results with the noise masker and speech masker which were proportionate. Such results suggest that children aged 5-6 years and children aged 6-7 years still remain poorer than adults when listening to speech within a speech masker. In the first year children's (5-6 years) mean SRTs were –17.6 dB SNR in the speech masker and -7.2 dB SNR in the noise masker. In the second year children's (6-7 years) mean SRT were -19.1 dB SNR in the speech masker and -9.4 dB SNR in the noise masker. The mean SRT for adults in the noise masker was -11.9 dB SNR and in the speech masker was -28.8 dB SNR. These results tie in with previous research that shows development of speech intelligibility in children to be particularly prolonged with speech maskers compared to stationary noise maskers (Wightman et al. 2010; Bonino et al. 2013; Leibold & Buss 2013). Such findings may be interpreted as an increased susceptibility in children to 'informational masking'; however it is important to consider that the findings could be the result of the SNR confound as highlighted by Bernstein and Brungart (2011), (see Chapter 4 section 4.1.1 for more details).

3.5.3 Learning effects

Results from each participant group showed an effect of learning, in that SRT (dB SNR) results were improved across sessions but not within sessions.

Such a learning effect within the results suggests that reliability may have been improved if there was an extended training period at the start of testing. Despite this however, the masker*age interaction was found to be present and repeatable across both sessions, thus the learning effects within this experiment do not affect the interpretation of the results.

When the children were in year one (aged 5-6 years) this across session improvement was found to be larger in children than adults and showed that participants appeared to improve to a greater extent with the speech masker than with the noise masker. Such findings could be employed to explain that the masker*age interaction would not be present had children received an extended period of training, and that perhaps more practice with the task would have dispelled these child/adult differences. However, when the children were in year 2, improvement from session 1 to 2 was found to be the same between year 2 children and adults and between children in both years, and was the same in each comparison with both masker types. This would suggest that improvements at this age 6-7 years may be proportionate in children and adults and with both masker types, thus it could be thought that further training would not dismiss the interaction. Children when in year 1 may have presented with a greater learning effect due to their age and due to the larger measurement error within the results.

It is important to recognise that the issue of longer and extended periods of training with children has not been examined here. The reason for the masker*age interaction therefore could well be that children require more practice, and with this practice could perform at adult like levels, or at least perform proportionately worse than adults with both masker types.

An unpublished study carried out at the Institute of Sound and Vibration Research (Thrift, 2014) aimed to address this issue of training. Using exactly the same experimental design as in the present study, single adaptive tracks were obtained with only the single-talker speech masker in 19 children (6-7 years) and 12 adults across 5 consecutive days. Comparing SRTs (dB SNR) from the first day with the final day children's results were shown to significantly improve by 4.2 dB, whilst adults improved by 2.8 dB which was not statistically significant. Whilst an improvement post training was observed in children, performance still did not reach even pretrained adult like levels (-27.1 dB SNR) and remained 4.9 dB poorer (-22.2 dB SNR). Whilst this improvement narrows the performance gap between children and adults with the speech masker, it still remains that this child adult difference of 4.9 dB is larger than the child/adult difference when compared with the same aged children from the present study with the speech-shaped noise masker (which was 2.5 dB). Further more rigorous testing in this area is however needed as it still could be the case that more training may lessen the child/adult differences. Such questions must be addressed with future research carefully, since too long a training session could cause fatigue which also could produce unreliable results. In the absence of extensive training periods in the present study at least, there still remains a masker*age interaction which still shows that children are more detrimentally affected when listening to speech in speech maskers than adults.

3.5.4 **Vocabulary test**

Another reason which could be put forward to explain the large child/adults differences in SRT (dB SNR)'s within the speech masker are differences between the educational backgrounds of both child and adult cohorts. The children were recruited from a single school within one region of the UK whilst the adults were recruited largely from the University of Southampton which comprises people from various regions and with privileged educational foundations.

Since the experiment consisted of a speech intelligibility test, to tackle this issue the BPVS vocabulary test was carried out to examine vocabulary differences between the child and adult groups. As expected, vocabulary scores with adults were larger than scores with children; however, the mean children's standardised score on the BPVS was shown to be above the population mean. When investigating the masker*age interaction, when those children performing below the level expected for their age were removed, this did not make any difference to the results and the significant interaction remained. This suggests that the interaction was not present owing to poor vocabulary in those children. Moreover, the words within the speech test had an age of acquisition of 4.95 years or under (Kuperman et al. 2012) so would have been appropriate for the age group tested.

Whilst differences in vocabulary are thus not likely to be the reason for the masker*age interactions between children and adults, it cannot be ruled out that experience with language may be the cause. Adults clearly have more experience with language than children and so could better recognise aspects of the speech using their linguistic knowledge in order to identify them (Mlot et al. 2010; Fallon et al. 2000). Children's language skills develop over time which includes increasing vocabularies and the learning of grammatical structures (Leech et al., 2007) and so could account for child/adult differences. For a closed set word identification task however, knowledge of grammatical structures is likely unnecessary. Furthermore, in previous research, even when vocabulary is age appropriate in speech intelligibility tests, differences are still apparent (Elliott, 1979). Such a difference between children and adults cannot be easily controlled for (since adults simply have years more experience) however further extensive training with children may help to further understand this issue.

3.5.5 Comparisons of SRT (dB SNR) estimation methods

In the speech intelligibility study using a rapid adaptive procedure the SRT (dB SNR) estimation method which gave the most reliable results was the average reversals method (taking an average of the points at last four reversals in the adaptive track) as opposed to the MLE method (using all the points within the adaptive track to fit a psychometric function to estimate the SRT (dB SNR). This result is surprising since the MLE method averages more trials from the whole adaptive track, whilst the average reversals method only takes an average here of the last 4 points of reversal. The MLE method, fitting the psychometric function, is very sensitive to lapses in attention during the adaptive procedure which can cause biases when estimating the psychometric function (Wichmann & Hill 2001a). In order to overcome biases from lapses the upper bound of the psychometric function was constrained (as described by Wichmann & Hill 2001a). It is possible however that lack of feedback provided to the participants created lapses and the upper bound may not have been tightly constrained enough and perhaps set lower (Wichmann & Hill 2001a). Furthermore, psychometric functions estimated this way from adaptive procedures can cause slope biases especially if learning is occurring throughout the experiment (Leek 2001). This may be because speech stimuli are not homogenous, possibly leading to variability in the adaptive track which can cause slope biases (Leek 2001). This is possibly why the MLE method shows poorer reliability since a psychometric function was fitted in this way to each of the four repeats (adaptive track runs) for each masker type.

The measurement error within the test was quite large (see table 3.4) and found to be higher than that of Summerfield et al. (1994) and less than that of Pure Tone Audiometry (Schmuziger et al. 2004). Owing to this high measurement error it may not be useful to examine individual differences within the data as the results from one repeat to the next may be too variable and lead to only speculative conclusions. This may have implications for clinicians who use this test clinically to monitor a child's progress with hearing aids and cochlear implants.

Compared to previous research, Litovsky (2005) does not talk about test-retest variability measures between the two methods but states that a t-test was used to compare the MLE method with the average last 3 (in their case) reversals and found there to be no significant differences. The present study used the same methods as Litovsky (2005) and found there to be significant differences between the MLE approach and the averaging of the last four reversals so the reasons for discrepancies between the findings are unclear. Litovsky (2005) did however carry out the t-test on results when the speech was presented in a quiet condition, which was not included in the present study. The results in quiet may have been more reliable than the results in noise and this could be one reason why the results from the

MLE and average 4 reversals in the present study differ. Furthermore, it could be that the present study had greater statistical power with larger number of participants and repeats, so a statistical difference emerged.

Despite the differences between SRT (dB SNR) estimation methods the masker*age interaction of interest was found to still be present with either method, so it can be construed that using either method to present the results does not make a difference to our conclusions.

3.5.6 **Individual differences**

Looking at the stability of individual results within the data, the correlation from one session to the next appears to show no clear patterns. With the adult group and the youngest child group (all 50 year 1 children) there is shown to be no correlation between session 1 and 2 with the noise masker but correlation with the speech masker. Considering the spread of the results (being larger with the speech masker than the noise masker) the lack of correlation with the noise masker could be interpreted in adults as there being not much variation in the results, which is backed up by the fact that the t-tests shows no significant differences between session 1 and 2. However, a t-test with children in this condition shows there to be a significant difference between the results. With the speech masker, correlation is seen yet the results are more greatly spread and t-tests also show a significant difference suggesting that there is more variation in this condition. With the results from the older child group (34 year 2 children), there is shown to be correlation with the noise masker, but no correlation with the speech masker, and no statistically significant differences across sessions with both masker types. Since the reliability of the data is poor, individual differences may not be stable from one repeat to the next so cannot accurately be explored.

Regarding the spread of the individual data it can be seen in the results that the data for children are more greatly spread than adults with a wider range of results seen with the speech masker compared to the noise masker. This could be because there are larger differences between individuals in children particularly with the speech masker. Similarly to the present study previous research has also shown large individual differences in children with speech maskers (Hall et al. 2002) that may take up until the teenage years before performance becomes adult-like (Wightman & Kistler 2005) with some young children performing adult-like at 5 years of age and some still performing child-like at 14 years of age (Wightman et al. 2010).

3.5.7 Explorations into the reasons for the masker*age interaction and implications

Considering Bernstein & Brungart's (2011) suggestion, this reason for the masker*age interaction could be due to differences in the baseline conditions, and children could be disproportionately worse than adults with the speech masker compared to the noise masker simply because they are poorer with the noise (baseline) condition. Attempted orthogonal regression was inconclusive and clear interpretations cannot be made from this data.

Looking however, at the results obtained from children 1 year later there is a significant improvement with the noise masker showing that children have become more adult-like in this condition. They appear to remain however, far from adult-like levels with the speech masker. If Bernstein and Brungart's (2011) suggestion holds true then a small change within the results from the noise masker would surely yield a larger change within the results from the speech masker but this is not the case and the significant masker*age interaction with adults is still present with the year 2 children. Such observations imply that the reason for this interaction may not be due to the differences between children and adults in the noise (baseline conditions) and that there may be something specific about the speech masker which causes children to be more greatly affected than adults by the speech masker.

The notion that children may be more susceptible to informational masking cannot be denied as the children's performance with the speech masker is not comparable to adults. The widespread individual differences in the performance of children with the speech masker may suggest that some children are more susceptible to informational masking than others which could support the view that children develop attentional strategies to overcome informational masking effects at different rates (Wightman et al., 2010). Owing to such large child/adult differences with the speech masker and the widespread results in children with the speech masker, the implications of such results could mean that children (some more so than others) may require more attention at school and need to sit nearer the teacher for improved signal to noise ratios. This may be especially relevant since children spend a vast majority of their day listening to one talker with competing talkers all around them.

Further research needs to be carried out to determine more reliable results in order to explore conclusively whether the masker*age interaction is or is not due to differences within the noise (baseline) conditions, but regardless of this children still require higher SNRs than adults to perform at comparable intelligibility levels.

3.6 Conclusions

- Children across an age range of 5-7 years performed disproportionately poorer than adults with a single-talker speech masker compared to a steady speech-shaped noise masker.

- Over a 1 year period children improved with the both the noise masker and speech masker to the same extent, but still fell short of adult like levels at 7 years of age, more so with the speech masker.

Chapter 4: **Are child/adult differences subject to SNR confounds?**

4.1 Introduction

This chapter consists of an experiment to investigate the validity of previously found child/adult SRT differences with speech backgrounds. This is to determine if children perform more poorly than adults owing to particular properties of the speech masker or whether children perform more poorly than adults owing to a general processing deficit displayed with all masker backgrounds.

4.1.1 Possible SNR confounds

Recent research and reports by Bernstein & Grant (2009), Bernstein & Brungart (2011) and Bernstein (2012) have highlighted that when listener groups differ in performance with stationary noise conditions (as adults and children differed in experiment 1), comparing SRTs between stationary noise and fluctuating maskers (i.e. speech maskers) may be subject to an SNR confound and so conclusions without taking this into account could be flawed. The exact nature of this SNR confound will now be explained in further detail beginning with descriptions of normal hearing and hearing impaired group comparisons, such concepts will then be applied to explain how this might similarly affect adult and child comparisons.

Comparing speech intelligibility can be especially problematic when large performance differences exist within different masker types and across different listener groups. It can be difficult to choose an appropriate SNR at which to measure performance that does not simultaneously yield floor effects for one listener group/masker type, and ceiling effects for the other (Bernstein 2012); comparisons here would be meaningless. Adaptive procedures are therefore often adopted to determine the SNR required to achieve a common intelligibility level (e.g. SRT), at which to make masker type and listener group comparisons (Levitt 1970). One specific example where adaptive procedures can prove extremely useful is when comparing speech intelligibility between maskers which are stationary (e.g. stationary speechshaped noise) and those which are fluctuating (e.g. amplitude modulated noise or a speech masker). Such masker types can yield very different results when measured at the same SNR since the fluctuating masker may consist of spectral and temporal dips that enable the listener to 'glimpse' parts of the target speech, improving performance greatly (Festen & Plomp 1990). Therefore comparing SRTs across such listening conditions provides a metric to allow for meaningful comparisons.

As explained previously, making use of the dips within a fluctuating masker is sometimes referred to in the literature as dip listening ability or fluctuating masker benefit (FMB) (Christiansen & Dau 2012). FMB is often measured by calculating the difference between SRTs (the SNR required to achieve a certain intelligibility level) with a baseline stationary noise masker and a fluctuating masker (Bernstein et al. 2012). Many previous studies investigating FMB have shown that hearing impaired listeners display a reduced FMB compared to normal hearing listeners (Festen & Plomp 1990; Eisenberg et al. 1995; Bacon et al. 1998; Gustafsson & Arlinger 1994; Peters et al. 1998; Dubno et al. 2003; George et al. 2006; Jin & Nelson 2006; Lorenzi et al. 2006; Wilson et al. 2007; Bernstein & Grant 2009; Strelcyk & Dau 2009). Such results have previously been interpreted as a reduced dip listening ability amongst the hearing impaired owing possibly to poorer supra-threshold acuity such as poorer spectral and temporal resolution (Bernstein et al. 2012). This interpretation has also been reinforced by studies using normal hearing listeners listening to speech processed to remove temporal fine structure and spectral details (Bernstein & Brungart 2011). Under these conditions the normal hearing listeners have also shown a reduced FMB similar to those who are hearing impaired suggesting that the reduced FMB in hearing impaired individuals is related to a reduction in temporal and spectral resolution, disrupting dip listening ability (Bernstein & Brungart 2011).

Recent studies by Bernstein and Grant (2009) and Bernstein and Brungart (2011) however, have recognised that previous interpretations could be flawed. Inspired by the findings of Oxenham and Simonson (2009) Bernstein and Grant (2009) and Bernstein and Brungart (2011) realised that the amount of FMB depends on the baseline stationary noise SNR at which you calculate it. Oxenham and Simonson (2009) investigated the discrimination of speech sentences (male talker) in the presence of a steady speech-shaped noise masker and a single talker (male) masker. Their results showed that the slope of the psychometric functions differ between masker types, being steeper for the steady speech-shaped noise and shallower for the single-talker masker. Therefore, the measured FMB decreases with increasing baseline stationary noise SNR. So if the FMB was to be compared at higher baseline stationary noise SNRs, the FMB would be smaller than if compared at lower baseline stationary noise SNRs where the curves deviate most greatly (Bernstein et al. 2012). The reason why a greater FMB is seen at lower SNRs is often because the fluctuating masker has a wide dynamic range where it spans from low to high levels. Therefore, there will be a larger SNR range where the speech is audible compared to the stationary noise which does not change in level and where sound will go very quickly from inaudible to audible (Rhebergen & Versfeld 2005).

Because of this, when a listener has a speech processing deficit, like a hearing impaired listener for example or a normal hearing listener listening to processed speech, the tested SNRs will be higher with the baseline stationary noise SNR and so yield a smaller FMB (Bernstein 2012). Even normal hearing listeners would similarly show a reduced FMB if they were compared at higher baseline stationary noise SNRs (Bernstein 2012). Therefore, the reduced FMB in hearing impaired listeners may not be due not to any reduced dip listening ability at all but due instead to higher SNRs tested overall as a result of the adaptive tracking procedure. So it may not be that the hearing impaired listener shows a deficit with the speech masker owing to any particular properties of the masker (e.g. its fluctuating properties), but they may just show a speech processing deficit which affects speech intelligibility equally in all types of background noise (Bernstein 2012).

Applying this theory to conclusions that have previously been made about normal hearing and hearing impaired listeners, it is possible that false conclusions have been drawn too when comparing speech intelligibility differences between children and adults. Where previous research has found differing SRTs between children and adults in different masker types, and concluded that larger child/adult differences arise with speech maskers in particular because children are more susceptible to informational maskers than adults, this may be a false conclusion. It is possible from the explanations above that children are not especially affected by any specific properties of the speech masker (e.g. its linguistic content) but just experience a deficit affecting their speech processing in all types of background noise.

So whilst the SRT metric measured via adaptive procedures is a time efficient way of measuring speech intelligibility, tested SNRs are allowed to vary according to the listener's performance. Where a general speech processing deficit of some sort is present (e.g. in individuals with hearing loss or perhaps even in normal hearing children), the tested SNRs will be higher and so will likely show a smaller FMB owing simply to an SNR confound (Bernstein 2012).

4.1.2 Methods for tackling SNR confounds

A recent report by Bernstein (2012) has explained three separate methods which can be employed to control for such SNR confounds. These methods take into account differences in baseline stationary noise conditions when comparing performance in noise and speech maskers across two population groups. Such methods will now be discussed.

4.1.2.1 *Measuring FMB at fixed baseline stationary noise SNRs*

One method involves measuring FMB (i.e. the difference between SRT (dB SNR) for baseline noise and speech maskers) at the same baseline noise SNR but at different percent-correct levels for the two population groups (Bernstein 2012). This reflects how children for example, at the same baseline stationary noise SNRs, have poorer intelligibility levels than adults for example as two different percent correct points will be identified, thought to represent the same amount of audible speech between both population groups (Bernstein 2012). Next the SNR required to reach the same intelligibility level as found in the baseline stationary noise condition in the fluctuating masker condition (i.e. with the speech masker) is estimated and FMB is calculated as the difference here between the SNRs with the fluctuating masker and the selected SNR with the noise masker (Bernstein 2012). If the measured FMB now becomes the same for the two population groups then this would suggest that the previously measured FMB (using traditional methods comparing at the same intelligibility levels with adaptive procedures for example) incurred an SNR confound. Therefore it could be concluded that there may be no extra difficulties relating to one population group with regard to the specific properties of the fluctuating/speech masker, but just an overall speech processing deficit applying equally to all types of backgrounds (Bernstein 2012). Figure 4.1 illustrates this concept further with stationary and fluctuating maskers across two population groups. Looking at the graph the two population groups, for example a hearing impaired group (green lines) and a normal hearing group (magenta lines), are shown to have different SRTs (dB SNR), with the hearing impaired group achieving a 50% intelligibility level at higher baseline stationary noise SNRs than the normal hearing group. Comparing performance at the same intelligibility level reveals a smaller FMB for the hearing impaired group (e.g. 3 dB) compared to the FMB for the normal hearing group (e.g. 9 dB). When taking into account differences between the baseline stationary noise conditions and comparing FMB at the same baseline noise SNRs however it is revealed that both groups have the same FMB (e.g. 9 dB). Therefore in this illustration the results would be subject to an SNR confound. It must be made clear that this data is not real and is for illustration purposes only and based on the work by Bernstein (2012).

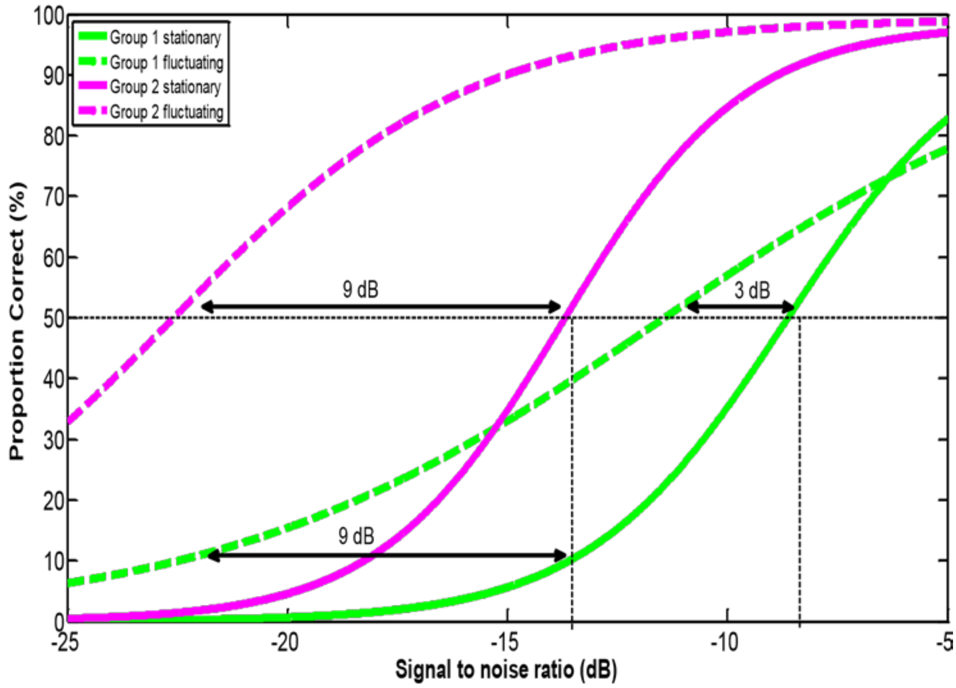


Figure 4.1 Graph to illustrate how the magnitude of the fluctuating masker benefit may change when compared at the same intelligibility level and when compared at the same stationary noise SNR across two listener groups. The solid lines represent a stationary background and the dotted lines represent a fluctuating background. The green lines represent one population group (for example 'hearing impaired') and the magenta lines represent another population group (for example 'normal hearing'). This graph is based on the work from Bernstein (2012) and is for illustration purposes only.

Bernstein and Grant (2009) therefore employed this proposed method to further investigate the differences in FMB between normal hearing and hearing impaired listeners. They measured the proportion of keywords amongst sentence stimuli correctly identified across various SNRs in the presence of a stationary speech-shaped noise, a modulated speech-shaped noise or a single-talker speech masker. They then fitted logistic functions to their data to estimate the shapes of the psychometric functions. It was found that, when comparing FMB traditionally, at the same 50% intelligibility level, the hearing impaired group showed a FMB of -2 to -1 dB (actually a deficit) compared to the normal hearing group who showed a FMB of 7 to 10 dB. They did note however that the baseline stationary noise conditions for a 50% intelligibility level was at a higher SNR for the hearing impaired group (at 0 dB) compared to the normal hearing group (at -6 dB) which they suggested could be the reason for observed reduction in FMB for hearing impaired group.

They thus employed the above proposed method and compared FMB across population groups at the same baseline stationary noise SNR however, for various SNRs and plotted FMB for each population group on a graph against increasing baseline noise SNR (with separate graphs for each masker type). It was found that FMB reduced with increasing baseline SNR, suggesting that listeners were receiving more benefit from the fluctuating masker at more negative SNRs. Furthermore it was found that with the speech modulated noise masker, the curves for each population group overlapped (i.e. the same FMB was found in both listener groups at the same baseline SNRs). This findings suggested that there were no dip listening deficits for the hearing impaired group and the previously reduced FMB was present due to the SNR confound. In contrast to this result however, it was found that with the single-talker speech masker the curves did not overlap and FMB whilst reduced, remained poorer for the hearing impaired listeners compared to the normal hearing group. The FMB with the speech modulated noise decreased from 7 dB (at 50% intelligibility) to 1 dB and with the single-talker speech masker decreased from 11 dB (at 50% intelligibility) to 5 dB. So while some of the reduced FMB in hearing impaired listeners could be accounted for by differences in the baseline conditions a difference between population groups did remain (Jensen & Bernstein 2013). This can be interpreted as some real deficit when listening to speech in those specific maskers (Bernstein & Grant 2009).

This method could be employed with children and adults by estimating the psychometric functions based on proportion correct at fixed SNRs to determine FMB at different baseline stationary noise SNRs. One possible limitation of this method however is that psychometric functions must overlap and there may be larger amounts of uncertainty present in regions where the psychometric functions are flat (Bernstein 2012).

4.1.2.2 *Plotting raw percentage correct data for two population groups against each other*

If there is no overlap with the psychometric functions across population groups another method can be used plotting the raw percentage correct from one population group against the other for each common tested SNR point for each masker (Bernstein 2012). This method factors out differences in supra-threshold distortion between the two groups that will apply equally to each masker type. Therefore, if the curves plotted overlap between masker types then the differences in FMB can be said to be due to baseline stationary noise differences as any distortion specific to the single-talker speech masker would actually in this case have the same effect on performance with both maskers. If the curves differ however, then there may be something specific relating to the speech masker that is causing differences in FMB across the two groups (Bernstein 2012).

Bernstein and Brungart (2011) applied this method when comparing FMB between normal hearing listeners and normal hearing listeners listening to speech processed to removed spectral and temporal cues (to stimulate factors associated with hearing loss). FMB was calculated with a speech modulated noise masker and a single-talker masker compared to a baseline stationary noise masker. They found that plotting raw percentage correct points of one group against the other gave curves that did overlap with both maskers and so suggested that traditionally reduced FMB found in hearing impaired listeners was not due to any reduced dip listening deficits but due to the SNR confound reflecting overall processing difficulty.

4.1.2.3 *Adjusting the task to equate population group performance*

If floor and ceiling effects are an issue and common SNRs between population groups cannot be found, the third method proposed by Bernstein (2012) is to adjust the task difficulty to equalize performance in baseline stationary noise conditions. This method involves decreasing the response set size for the hearing impaired listener (or child in this case) to equate performance at the same SNR. This is achieved through making the task “easier” by limiting the choices in an alternative forced choice task, providing the same % correct level in the baseline stationary noise condition.

If the difference between performance with the baseline stationary noise masker can be equated (by reducing response set size) then the SNR confound can be accounted for as differences in supra-threshold distortion will be offset (Bernstein 2012). Bernstein and Brungart (2011) again employed this method to show how performance between normal hearing listeners (with a 1000 word response set size, without the viewing of this list) and normal hearing listeners listening to speech processed to mimic supra-threshold deficits associated with hearing loss (with a 72 response set size) was equated here in the baseline stationary noise condition. When the fluctuating masker conditions were tested FMB appeared the same across population groups, and so it was concluded that these distortions were not the reason for reduced FMB but were in fact due to the differences in baseline stationary noise conditions.

Using set size adjustments to equalise stationary noise baseline performance with sentence material may be difficult however as lists of sentences or keywords would need to be remembered or displayed in lists which may become confusing for the participants. Moreover reducing set size could result in decreased external validity if the set is not changing all the time. Therefore it was thought that using the above two methods may be more straightforward and further using children slightly older (aged 7-8 years) may serve to equalise performance between children and adults in the baseline stationary noise condition since SRTs in this conditions have previously been found to be adult-like by around 8 years of age (e.g. Bonino et al., 2013; Elliott, 1979).

4.2 Aims of experiment 2

1) **To determine the speech intelligibility differences between children and adults in different masker types (a steady state speech-shaped noise and a single-talker speech masker) using sentence material to establish masker*age interactions.**

   Rationale: Whilst speech intelligibility in background noise in children and adults has previously been considered, using sentence material and slightly older children (aged 7-8 years) should provide more reliable results to base conclusions.

   Predictions: (i) a stationary noise masker will be more detrimental to speech intelligibility than a single-talker masker; (ii) compared to the stationary noise masker, single-talker maskers will affect children more strongly than adults.

2) **To explore the extent to which a masker*age interaction should be interpreted by investigating whether the expected smaller improvement from a stationary noise masker with the speech masker in children compared to adults is due to an SNR confound (Bernstein 2012).**

   Rationale: It has recently been shown that much previous research into the interaction between populations is methodologically flawed and so could have drawn erroneous conclusions (e.g. Bernstein & Brungart 2011; Bernstein & Grant 2009). Future research with speech intelligibility measures needs to implement improved methodology. Interactions between children and adults have not previously, in this way, been considered and must be if any interactions are to be accurately explained.

   Predictions: (i) Previous studies regarding this issue have not been carried out before. It is therefore not known whether or not the difference will be due solely to a SNR confound.

4.3 Methods

4.3.1 Overview and rationale for methods

Participants were tested on their ability to listen to and identify sentences in the presence of two different maskers (see stimuli section below for details) presented diotically over headphones. As per experiment 1 the masker types remained the same, chosen to encompass features of both 'energetic' and 'informational' masking as previously defined. Target sentences to be identified were taken from the standard clinical speech intelligibility tests: the Bamford, Kowel and Bench (BKB) sentence test (Bench et al. 1979) and from the corpus of the Institute of Hearing Research (IHR) sentence lists (Macleod & Summerfield 1990). In an open set task a method of constant stimuli was used to determine the proportion of correctly identified sentences at different signal to noise ratios. This was so to estimate the shape of the psychometric functions. A test encompassing target sentences instead of target words (as in experiment 1) was chosen since such a test yields greater reliability (Summerfield et al. 1994). The signal-to-noise-ratios chosen to be tested, differed between children and adults, and were chosen based on previous research and pilot testing and to provide common SNRs across population groups where possible. A repeated measures design was then used whereby each participant underwent all of the conditions being tested. As with experiment 1, a vocabulary test was also carried out with both children and adults to ensure that their vocabulary scores were at age appropriate levels.

4.3.2 **Participants**

Ethics and research governance approval was obtained from the Institute of Sound & Vibration Research Human Experimentation Safety and Ethics Committee before commencing this experiment (see Appendix A for safety and ethics approval emails).

4.3.2.1 *Children*

Normal hearing, native English children (aged 7-8 years), with no known special educational needs were recruited from a participating school. Invitation letters consisting of consent forms and otological health questionnaires (see Appendix B) were sent home to parents of all children in year 3 (n=120). Year 3 children (aged 7-8 years) were chosen for a number of reasons. One reason was to enable comparisons with other studies (e.g. Hall et al. 2012; Hall et al. 2002; Litovsky 2005). Another reason was to suit the more complex sentence discrimination task as opposed to the word discrimination task examined with slightly younger children (aged 5-6/6-7 years) in experiment 1. The sentence task required more involvement from children as they had to repeat back whole sentences instead of pointing to pictures of words on a touch screen, so it was thought this age group would produce more reliable results than younger participants. A final reason for the chosen age range was because it was expected that this age range would perform closer to adult-like levels in the baseline stationary speech-shaped noise condition (e.g. Bonino et al. 2013; Papso & Blood 1989; Elliott 1979) and so one step taken to control for the possible SNR confound.

One-hundred-and-twenty invitation letters were distributed of which forty-two consent forms were returned. Of those forty-two children, three were excluded from the study because they were either non-native English speakers, or were not deemed otologically normal determined from the answers to the otological health questionnaire filled out by parents. Of those thirtynine children, fifteen were not tested due to timing issues when the children broke up from school for their summer holidays. Twenty-four children therefore went on the complete a basic hearing screening test which was carried out in a quiet room within the school. A basic hearing screen (pure tone audiometry with circumaural earphones in accordance with ISO 3898:2004) was carried out with twenty-four children. Only if the child passed the screen (and reliably responded to 20 dB HL at 1-4 kHz and 30 dB HL at 500 Hz) in both ears did he or she carry on with the study. All twenty-four children passed the hearing screen. Four of these children took part in some pilot testing to establish the final conditions and signal to noise ratios of the experiment. Thus a total of 20 children (10 males, 10 females) aged from 7 years 10 months to 8 years 9 months (average age 8 years 4 months) participated in the main study.

4.3.2.2 *Adults*

Twenty normal, native English speaking adults (7 males, 13 females) aged 18 years 2 months to 29 years 10 months (average age 23 years 3 months), with no known special educational needs were recruited from students and friends of the University of Southampton (see Appendix B for consent forms and questionnaires). All 20 adults were deemed otologically normal, determined from otological health questionnaires filled out by participating adults, so all twenty adults went on to complete and pass the basic hearing screen (as detailed above) and participate in this study.

4.3.3 **Equipment**

A laptop was connected to a Creative Extigy external sound card. A custom written Matlab code (MATLAB 2012) controlled the entire procedure and generated and controlled the stimuli for both the experiment and the hearing screen. Sennheiser HDA 200 cicumaural headphones were used to deliver sound diotically. The level of all stimuli through the headphones was calibrated using a sound level meter attached to an artificial ear. This objective calibration took place at the beginning of the experimental period and once every three weeks thereafter. Subjective listening checks occurred at the start of every experimental session by the primary researcher.

4.3.4 **Stimuli**

4.3.4.1 *Target stimuli*

The target stimuli consisted of target sentences. The target sentences were obtained from two separate corpora; the BKB sentence lists (Bench et al. 1979) and the IHR sentence lists (Macleod & Summerfield 1990) both spoken by the same male voice. The BKB sentence lists consisted of 21 lists with each list containing 16 sentences. The IHR sentence lists consisted of 18 lists with each list containing 15 sentences. The reason two different corpora were combined was because alone there would not have been enough target speech material to test all the conditions, and using sentences more than once may have given a possible false improvement of scores through familiarity. In the interest of maximising available target speech material a male target voice was used instead of a female voice (as used in experiment 1). This was because the IHR sentences were only recorded with a male voice; the same male voice from the BKB sentences, yielding more target speech material. It has been shown that a male target voice is, on average, less intelligible than a female voice but this effect of voice gender is found to be the same in adults and children (Markham & Hazan 2004). Thus changing the target gender should have no effect on child/adult comparisons. Both the BKB and IHR sentences have been shown to be of equal difficulty provided the rms level is equalised across all sentences (Parfect 2002) which they were for this experiment. The BKB sentence lists were developed based on the expressive language of hearing impaired children aged 8-15 years (Bench et al. 1979) and are used widely in clinical settings to assess the speech perceptual abilities of young children and adults. The sentences are thus deemed appropriate for testing normal hearing 7 year olds (Blandy & Lutman 2005).

For each listening condition one list was presented to the participant and lists were chosen at random but never repeated. Because there were an unequal number of sentences in each of the BKB (16 sentences) and IHR (15 sentences) lists, only the first 15 sentences from the BKB sentence lists were used. Therefore for each listening condition, 15 sentences were presented. Correct identification of sentences was based on a loose scoring method. Each sentence contained three or four keywords and the participant was required to correctly identify two or more of these keywords. Keywords were scored correct if the root of the word was repeated correctly irrespective of tense or order (Blandy & Lutman 2005). Two examples of the sentences used are 'the little boy was tired' and 'the lorry drove up the road', where the underlined words signify the keywords.

4.3.4.2 *Masker stimuli*

As per experiment 1, two masker conditions were used. In one condition a steady-state noise was used whereby the spectrum had been adjusted to match the long term average speech spectrum of the speech targets.

The second masker condition consisted of speech sentences used as the masker. The sentences were taken from the EUROM sentence database (Chan et al. 1995) consisting of different speakers reading different meaningful and grammatically correct sentences and passages. Female talkers were chosen in order to be distinct from the male target speech so that children could be easily directed as to which voice to listen to. In order to have enough unique speech material from which to create the speech maskers, two specific female talkers were chosen because they provided the greatest number of unique passages with similar speaking rate. Each talker had 15 unique recorded passages 21 seconds in duration. The rms level of each passage was equalised to that of the speech-shaped noise and then for each distinct talker the passages were concatenated end to end in MATLAB, and pauses between words were deleted if more than 100 ms. Owing to the fact there was not enough time to do so, the spectrum was not equalised between the target sentences and masker sentences. A continuous stream of female talker A and female talker B was thus obtained and random segments of these were selected and used as the masker, whereby a segment selected at random was used for each target sentence. The masker duration was set so that the target sentence began 0.5 s after masker onset. This was not varied as it was thought that children could find the varying onset too difficult to follow. Both maskers were ramped on and off slowly using a raised-cosine ramp with 100 ms duration. All stimuli were obtained digitally and converted to analogue form using a 44100 Hz sampling rate via a 16 bit digital to analogue converter. Half the participants were tested with female talker A and half with female talker B.

For each of the two masker conditions five SNRs were tested, chosen to estimate the shape of the psychometric functions. The SNRs chosen were specific to each masker type and each age group, and were based on both previous research and pilot testing (see section 4.3.7 below).

The level of the both the target and masker was then varied to generate required SNRs. A decrease in SNR resulted in a reduction in intensity level of the signal together with an increase in intensity level of the masker to meet the required SNR and vice versa. Both the target and masker were varied in this way to prevent the delivered sound becoming too loud for the participants, as per experiment 1. When the SNR was set at 0 dB SPL the level of the target and masker was at 60 dB SPL.

4.3.5 **Vocabulary test**

The vocabulary test carried out was the same as in experiment 1; the British Picture Vocabulary Scale (BPVS), and followed the same methods.

4.3.6 **Structure**

4.3.6.1 *Overall Structure*

Testing was carried out on two separate days and a single testing session lasted approximately 40-50 minutes each for the children and 20-30 minutes each for the adults. Each testing session was identical to enable the repeatability of the results to be examined. The proportion of sentences scored correctly (out of 15) was measured for each of the two masker conditions at each of the five SNRs. The vocabulary test was carried out on the second testing session with all participants.

Child participants were seated in a quiet room within the school; adult participants were seated in a quiet room within the university. All participants were seated on a chair at a table. Over headphones, participants listened to the target sentence presented within a masker after which the participant was required repeat back as much of the sentence they heard, which was then scored correct or incorrect, depending on how many keywords were repeated correctly. This made up one trial.

The experiment consisted of two masker conditions: the speech target presented within the noise masker and the speech target presented within the speech masker. The order of the

SNRs were presented so that, after some practise with higher SNRs (see familiarisation section 4.3.6.2 below), the first SNR tested was the second highest, the second SNR tested was then the highest, the third SNR tested was the second lowest, the fourth SNR tested was the middle SNR level and the last SNR tested was the lowest. The SNRs were tested in this order in an attempt to maintain the attention of the children. It was thought that, if the order began with the “easier” SNRs and progressed down towards the “harder” SNRs successively; children would lose interest in the task and not try. It was of great worth to be able to tell child participants in particular that the next SNR would be a little bit easier as opposed to just telling them it was to get harder and harder. For accurate comparisons adults were tested in the same way. Fifteen sentences were presented for each of the two listening conditions for five SNRs, therefore a total of 150 sentences were tested.

Participants were instructed to ignore the lady or the noise and to listen carefully to the man’s voice, and to repeat back as many of the words they heard the man say, even it is didn’t make any sense. During the main test feedback was not given except occasionally if the participant needed to be prompted as to which voice to listen to. Children were given a very short rest after each SNR was tested and a sticker was placed on a record card that became filled up with stickers which they could take home once the whole experiment was completed (see Appendix D for reward cards). This was used to help maintain the child’s interest and keep them on task. Children were given a short break after each masker condition where the headphones were removed and the child and the experimenter engaged in conversation to prevent the child becoming bored and inattentive before commencing the masker condition. Adults generally kept the headphones on throughout the duration of testing as inattention was not an issue. For each masker condition all SNRs tested were completed (in the specific order as noted above) before moving onto the next masker condition. The order of the masker conditions was counterbalanced between participants in attempts to counter any order effects.

4.3.6.2 *Familiarisation*

At the start of each tested masker condition an additional list of 15 sentences (chosen at random from the IHR and BKB sentence lists, and not a list tested in the main experiment) was presented to the participant. The SNRs were varied, starting at +15 dB, and were moved progressively down to lower SNRs until the chosen starting SNR (for the masker type and age group) was reached, whereby the last 5 trials were tested at this level. The participants then began the test trials and were instructed that the first 15 sentences would be at the level that was just tested. For each sentence that the participant got wrong the sentence was repeated verbally by the experimenter and then the sentence was played back to the participant again so they could focus again on the target talker. These 15 practise trials were conducted in the same way at the start of session 1 and session 2 and results were not recorded.

Additionally, two familiarisation trials were added to the start of each tested SNR in the main experiment which were taken from BKB sentence list 1 and 2 at random and not included in the analysis. Because of this, BKB sentence lists 1 and 2 were not used in the main experiment and it is noteworthy to state that at the start of each tested SNR, as the experiment went on participant would have been presented with the familiarisation trials they had heard before. This was to give them a little bit of practise at the start of each newly tested SNR.

4.3.6.3 *Statistical analysis*

Psychometric functions were fitted to the results across all five SNRs for both masker conditions for session 1 and session 2 independently. The results from session 1 and 2 were also pooled (by adding up scores out of a possible 30 for each tested SNR) and were too fitted with psychometric functions. This enabled the test-retest reliability of each session to be examined. Goodness of fit calculations were applied to all psychometric function fits for each participant. The parameters for the fitted psychometric functions from each individual participant's pooled data were then averaged to produce a single psychometric function for each masker type. The methods proposed by Bernstein (2012) were then applied to investigate the impact of possible SNR confounds.

4.3.7 **Selecting the SNRs**

4.3.7.1 *Selecting the SNRs based on previous research*

In order to choose an array of SNRs which would best capture the psychometric functions for both children and adults with both masker types, both previous research and some pilot testing with children was taken into account.

Table 4.1: Signal to noise ratios and proportion correct recorded from previous studies which have used target sentence material and a speech-shaped noise masker with children.

Children: Speech-shaped noise				
Proportion correct (%)	**SNR (dB)**	**Study**	**Sample**	**Methods**
70.7%	-4.7	Unpublished data from the ISVR (2011)	53 children (9-10 years)	BKB sentences presented in an adaptive procedure
	-3.8	Blandy and Lutman (2005)	84 children (7 years)	
50%	-3.0	Hall et al. (2012)	11 children (7-11 years)	

Table 4.2: Signal to noise ratios and proportion correct recorded from previous studies which have used target sentence material and speech-shaped noise masker with adults.

Adults: Speech-shaped noise				
Proportion correct (%)	**SNR (dB)**	**Study**	**Sample**	**Methods**
99%	+3	Oxenham and Simonson (2009)	24 adults	HINT sentences, fixed SNRs
95% (approx.)	0 (approx.)	Bernstein and Grant (2009)	5 adults	IEEE sentences, adaptive procedure
96%	0	Oxenham and Simonson (2009)	24 adults	HINT sentences, fixed SNRs
91%	0	Blyth (2013)	2 adults	BKB sentences, fixed SNRs

97%	-2	Blyth (2013)	2 adults	BKB sentences, fixed SNRs
80% (approx.)	-3 (approx.)	Bernstein and Grant (2009)	5 adults	IEEE sentences, adaptive procedure
78%	-3	Oxenham and Simonson (2009)	24 adults	HINT sentences, fixed SNRs
59%	-4	Blyth (2013)	2 adults	BKB sentences, fixed SNRs
70.7%	-5	Unpublished data from the ISVR (2011)	30 adults	BKB sentences, adaptive procedure
45% (approx.)	-6 (approx.)	Bernstein and Grant (2009)	5 adults	IEEE sentences, adaptive procedure
47%	-6	Blyth (2013)	2 adults	BKB sentences, fixed SNRs
43%	-6	Oxenham and Simonson (2009)	24 adults	HINT sentences, fixed SNRs
28%	-8	Blyth (2013)	2 adults	BKB sentences, fixed SNRs
15%	-9	Oxenham and Simonson (2009)	24 adults	HINT sentences, fixed SNRs
6%	-10	Blyth (2013)	2 adults	BKB sentences, fixed SNRs
3%	-12	Blyth (2013)	2 adults	BKB sentences, fixed SNRs
2% (approx.)	-12 (approx.)	Bernstein and Grant (2009)	5 adults	IEEE sentences, adaptive procedure
0.4%	-12 dB	Oxenham and Simonson (2009)	24 adults	HINT sentences, fixed SNRs

Table 4.3: Signal to noise ratios and proportion correct recorded from previous studies which have used target sentence material and a single-talker speech masker with adults

Adults – Single talker				
Proportion correct (%)	**SNR (dB)**	**Study**	**Sample**	**Methods**
97%	+6	Oxenham and Simonson (2009)	24 adults	HINT sentences, adaptive procedure, male target and male masker

90%	0	Oxenham and Simonson (2009)	24 adults	HINT sentences, adaptive procedure, male target and male masker
94%	-6	Oxenham and Simonson (2009)	24 adults	HINT sentences, adaptive procedure, male target and male masker
85% (approx.)	-9 (approx.)	Bernstein and Grant (2009)	5 adults	IEEE sentences, adaptive procedure, female target male masker
84%	-12	Oxenham and Simonson (2009)	24 adults	HINT sentences, adaptive procedure, male target and male masker
75% (approx.)	-12 (approx.)	Bernstein and Grant (2009)	5 adults	IEEE sentences, adaptive procedure, female target male masker
50% (approx.)	-15 (approx.)	Bernstein and Grant (2009)	5 adults	IEEE sentences, adaptive procedure, female target male masker
58%	-18	Oxenham and Simonson (2009)	24 adults	HINT sentences, adaptive procedure, male target and male masker
40% (approx.)	-18 (approx.)	Bernstein and Grant (2009)	5 adults	IEEE sentences, adaptive procedure, female target male masker
20% (approx.)	-21 (approx.)	Bernstein and Grant (2009)	5 adults	IEEE sentences, adaptive procedure, female target male masker
55%	-22	Blyth (2013)	2 adults	BKB sentences, fixed SNRs, female target male masker
15% (approx.)	-24 (approx.)	Bernstein and Grant (2009)	5 adults	IEEE sentences, adaptive procedure, female target male masker
40%	-26	Blyth (2013)	2 adults	BKB sentences, fixed SNRs, female target male masker
7%	-30	Blyth (2013)	2 adults	BKB sentences, fixed SNRs, female target male masker
26%	-24	Oxenham and Simonson (2009)	24 adults	HINT sentences, adaptive procedure, male target and male masker

Previous research was only considered if sentences were used as the target material and if a speech shaped noise and/or a single-talker was used as the masker. Regarding children, there was little previous research which fulfilled these criteria. Table 4.1 shows the SNRs and their corresponding intelligibility levels from three previous studies which have used the same target sentences and speech-shaped noise masker as the present study. These studies did however implement an adaptive procedure so do differ from this present study. There were no studies, using sentence target material, which had looked at results from children using a single-talker speech masker. Therefore pilot testing with children in particular was needed to ensure appropriate SNRs were chosen. For adults previous research in table 4.2 and table 4.3 gave a better insight into which SNRs would be suitable for both the speech-shaped noise masker and single-talker masker respectively.

4.3.7.2 *Selecting the SNRs based on Pilot testing*

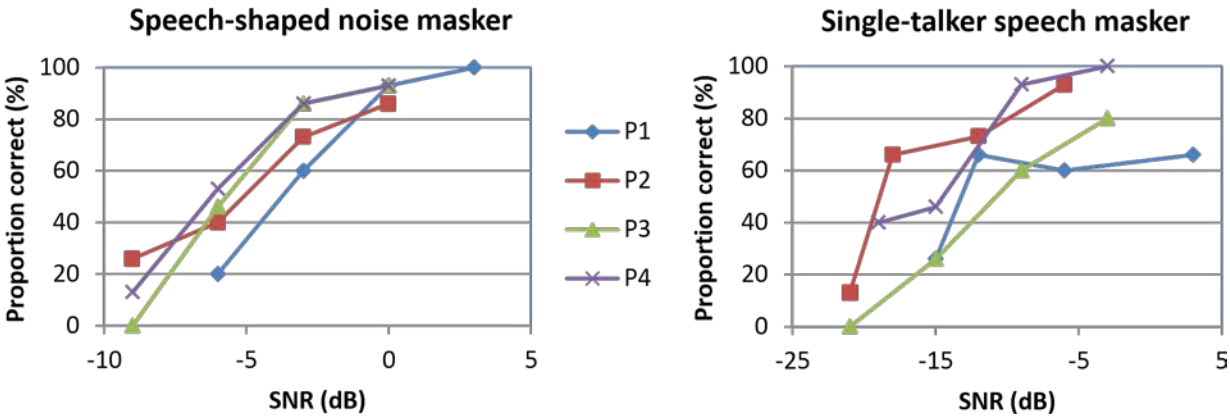


Figure 4.2: Proportion correct recorded at an array of different SNRs for 4 different child participants. Different coloured lines represent different participants labelled P1 to P4.

Considering the previous research from children and adults with target sentence material in both masker types, the pilot SNRs were chosen. The first four children to be recruited took part in the pilot testing in order to determine which SNRs would be appropriate to use within the main experimental sample that would best capture the shape of the psychometric function. From one pilot participant to the next the SNRs were changed slightly to examine the result and these results are displayed in figure 4.2.

Taking into account previous research and pilot testing the final SNRs for children and adults in both masker types were chosen as detailed in table 4.4. SNRs were selected so that they had the same decibel interval between consecutive SNRs per masker type and had common tested SNRs between children and adults where possible.

Table 4.4: The chosen signal to noise ratios for children and adults for each masker type.

Masker	Participants	SNRs (dB) from low to high				
Speech-shaped noise	Adults	-14	-11	-8	-5	-2

	Children	-11	-8	-5	-2	1
Single-talker speech	Adults	-28	-23	-18	-13	-8
	Children	-23	-18	-13	-8	-3

4.4 Results

4.4.1 Overview of results

The BPVS was carried out to determine the vocabulary of both children and adults. Each participants score was compared to a standardised score calculated according to their age, the results of which will be presented first.

The proportion of correctly identified sentences was determined within a speech-shaped noise masker and a single-talker speech masker at five various fixed SNRs in both children and adults. For each fixed SNR and each masker condition measurements were obtained across two sessions. Logistic functions were then fitted to individual measurements from each session and each masker condition to show individual psychometric functions in session 1 and session 2 to evaluate reliability. Scores from session 1 and session 2 were then added together and logistic functions were fitted to this pooled calculation for each individual participant. The goodness of fit for each of the individual psychometric fits was then calculated and the psychometric function parameters from those deemed good fits were averaged to show an average result across participants.

SRTs (dB SNR) corresponding to 79.4% correct (as measured in experiment 1) were then calculated and results were compared between masker types between children and adults. Using methods proposed by Bernstein & Grant (2009) SRTs (dB SNR) were calculated and compared between masker types and between children and adults at various fixed signal to noise ratios, in order to take into account the differences in the baseline speech-shaped noise masker conditions to see if any child/adult differences disappear. Methods proposed by Bernstein and Brungart (2011) were also investigated by plotting the raw percent correct data from children against adults with each masker type to again explore the extent of a possible SNR confound. The discussion of results will not be considered here, but will be in the discussion section 4.5.

4.4.2 Vocabulary test

Table 4.5: The mean, range and standard deviation of the standardised scores from the British Picture Vocabulary scale in children and adults.

	Adults (n=20)	Children aged 7-8 years (n=20)
Mean	127	109
Range	111-160	94-121
Standard deviation	14	8

The results of the British picture vocabulary scale (BPVS) are displayed in table 4.5. Results are shown from the 20 children in year 3 (aged 7-8 years) and the 20 adults.

To recap, the BPVS calculates standardised scores based on age population means, only up to 15.8 years of age. The mean standardised score of both children and adults are displayed in Table 4.5. A standardised score of 100 would suggest that a child had an average vocabulary range for their age. The standard deviation for standardised scores is 15, so 68% of people will score between 85 and 115 (Dunn et al. 1997). For adults, the standardised scores have been based on a child aged 15.8 years.

As can be seen from the table, the mean child standardised score in this sample was above the population mean. The mean adults standardisation scores is also above the population mean and is significantly higher than that of the child standardisation scores $t(30) = 4.96, p<.001$. The adults mean scores ranged from 111 to 160 and thus were all above the population mean. This is not surprising, but it is however reassuring that the adults did not display any vocabulary deficits with this test that may have interfered with the speech intelligibility test.

The children’s mean scores ranged from 94 to 121 and thus span above and below the population mean. The BPVS calculates age equivalent scores and percentile ranks together with their confidence bands. Table 4.6 shows details of individual standardised scores, age equivalent scores and the percentile rank for each child. Nine children’s vocabulary scores fell within their age range and eleven children’s vocabulary scores fell above their age range.

Table 4.6: The standardised scores, age equivalent scores and percentile ranks calculated from the BPVS for children (7-8 years) together with their upper and lower 68% confidence bands. An age of 8:02 denotes 8 years 2 months and an age of 7:11 denotes 7 years 11 months. A

percentile rank of 77 indicates that 77 out of 100 children tested of the same age scored equal or below the standardisation score of this participant

ID No.	**Standardised score**	**Actual age (years: months)**	**Age equivalent (confidence bands)**	**Percentile rank (confidence bands)**
2	94	8:03	7:09 (7:02-8:04)	34 (22-50)
15	96	8:05	7:11 (7:04-8:08)	40 (26-55)
11	96	8:08	8:01 (7:06-8:10)	40 (26-55)
4	103	7:10	8:02 (7:07-8:11)	74 (60-86)
18	103	8:02	8:09 (8:01-9:08)	58 (42-72)
19	103	8:03	8:09 (8:01-9:08)	58 (42-72)
14	103	8:08	9:00 (8:03-9:11)	58 (42-72)
16	103	8:05	8:09 (8:01-9:08)	58 (42-72)
9	106	8:04	9:01 (8:04-10:01)	66 (50-78)
20	106	8:09	9:10 (8:10-10:09)	66 (50-78)
17	111	7:11	9:00 (8:03-9:11)	77 (63-87)
3	111	8:01	9:07 (8:07-10:04)	77 (63-87)
13	111	8:08	10:02 (9:03-11:01)	77 (63-87)
12	112	8:07	10:04 (9:05-11:03)	78 (66-89)
1	115	8:06	10:10 (9:10-11:07)	84 (72-92)
5	117	8:07	11:01 (10:02-11:11)	87 (77-94)
8	118	7:11	10:04 (9:05-11:03)	89 (78-94)
6	120	8:02	10:11 (10:00-11:09)	91 (82-96)
7	121	8:00	11:01 (10:02-11:11)	92 (84-96)
10	121	8:02	11:01 (10:02-11:11)	92 (84-96)

A standardised score of 85 and a standardised score of 115 represent ±1 standard deviation of the population mean respectively where the results 68% of people from a certain age range will fall (Dunn et al. 1997) According to Dunn et al. (1997), standardised scores between 85 and 100 are classed as “low average scores” and those between 100 and 115 are classed as “high average scores”. Scores between 115 and 130 are classed as “moderately high scores”. All children scored within or above their age range and thus it can be suggested that they have a vocabulary range appropriate for their age. All participants therefore did not display any vocabulary deficits which may be said to make them particularly poor with the speech intelligibility test. There was found to be no significant relationship between vocabulary score and SRT (dB SNR), as derived from the location parameter in the psychometric function fits, in children with the noise ($r = -.18$, $p > .05$) and speech masker ($r = -.01$, $p > .05$).

4.4.3 **Goodness of fit**

Logistic functions were fitted to the scores obtained at each SNR using “Palamedes” code written in MATLAB (Prins & Kingdom 2009). This fitting formula estimates the participants’ psychometric function and produces a range of parameters. The alpha parameter denotes the position of the curve along the abscissa whereby the proportion correct is 0.5. The beta parameter denotes the slope of the curve and the Lambda denotes the upper asymptote.

Before any further analyses were carried out, the goodness of fit for each psychometric function was calculated to ensure only good fits were included in the analysis. This measured how well the fitted psychometric function describes the recorded data points using a bootstrap analysis (Prins & Kingdom 2009). The goodness of fit is denoted by the deviance value whereby a corresponding p-value of less than 0.05 would indicate the fit was unacceptably poor (Kingdom & Prins 2009). Most of the fits were classed as acceptable (see Appendix E for tables all the goodness of fit deviance and p-values). Table 4.7 details which participants, for which masker and session displayed poor fits (with p-values less than 0.05). Participants who displayed poor fits were not completely excluded from the analysis, but were excluded only when analysing data which involved those sessions in which the fits were unacceptably poor. Figure 4.3 gives three examples of fitted psychometric functions and their data points.

Table 4.7: Fitted logistic functions where fits were deemed poor according to Kingdom & Prins (2009) so were not included in analyses involving the corresponding session.

Participant group	Masker	Session	Poor fits
Adults	Noise masker	Session 1	Participant 9 Participant 11
		Session 2	N/A
		Both sessions	Participant 7 Participant 15
	Speech masker	Session 1	N/A
		Session 2	N/A
		Both sessions	N/A
Children	Noise masker	Session 1	Participant 6
		Session 2	N/A
		Both sessions	Participant 1 Participant 11
	Speech masker	Session 1	N/A
		Session 2	Participant 16

Both sessions	N/A

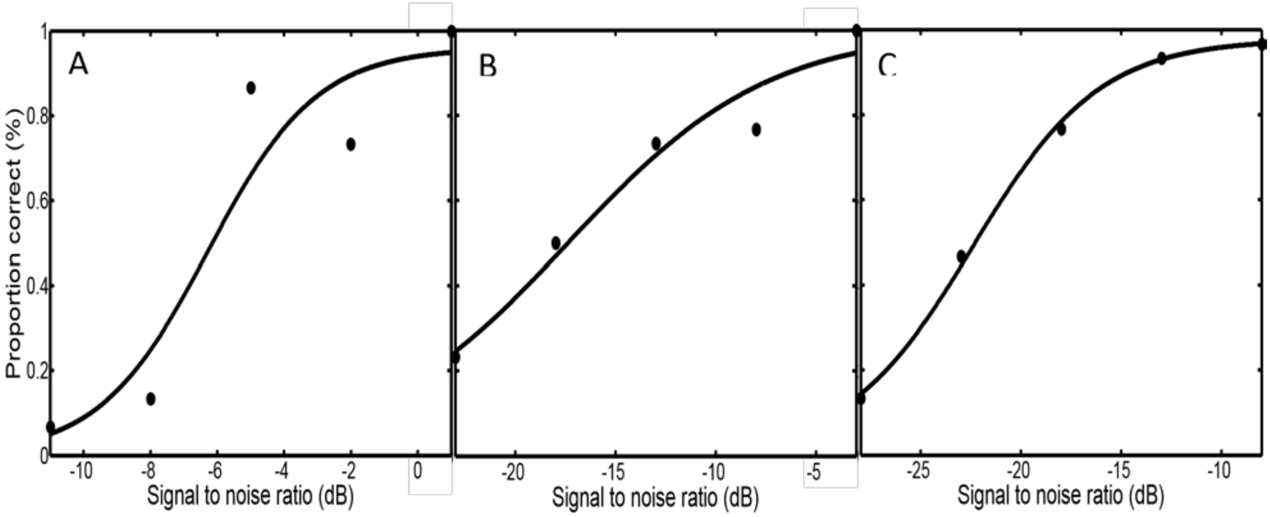


Figure 4.3: Three examples of fitted psychometric functions. The solid curve shows the fitted psychometric function and the black dots show the actual recorded data points. Panel A shows a fitted psychometric function where the deviance had a p-value of 0.02 (thus deemed unacceptably poor). Panel B shows a fitted psychometric function where the deviance had a pvalue of 0.05 and panel C shows a fitted psychometric function where the deviance had a pvalue of 0.95.

4.4.4 **Reliability**

The individual fitted functions from all 20 adults in noise and speech maskers and all 20 children in noise and speech maskers are displayed in figure 4.4, 4.5, 4.6, and 4.7 respectively. The green line displays the results from the first session whereby 15 sentences were presented at each SNR in each masker condition. The magenta line displays the results from the second session, where a further 15 sentences were presented, as in the first session. The black dotted line shows the results from session 1 and 2 added together, so shows the results from 30 sentences in each SNR. The steepness of the curves shows that results with the noise masker display a steeper gradient whilst the speech masker results in a much shallower slope. This pattern occurs with both children and adults.

It can be visualised that these lines across sessions appear very similar with the noise masker in both children and adults suggesting good repeatability within the results from one session to the next. The lines with the speech masker are less similar in both children and adults, yet don’t appear too different.

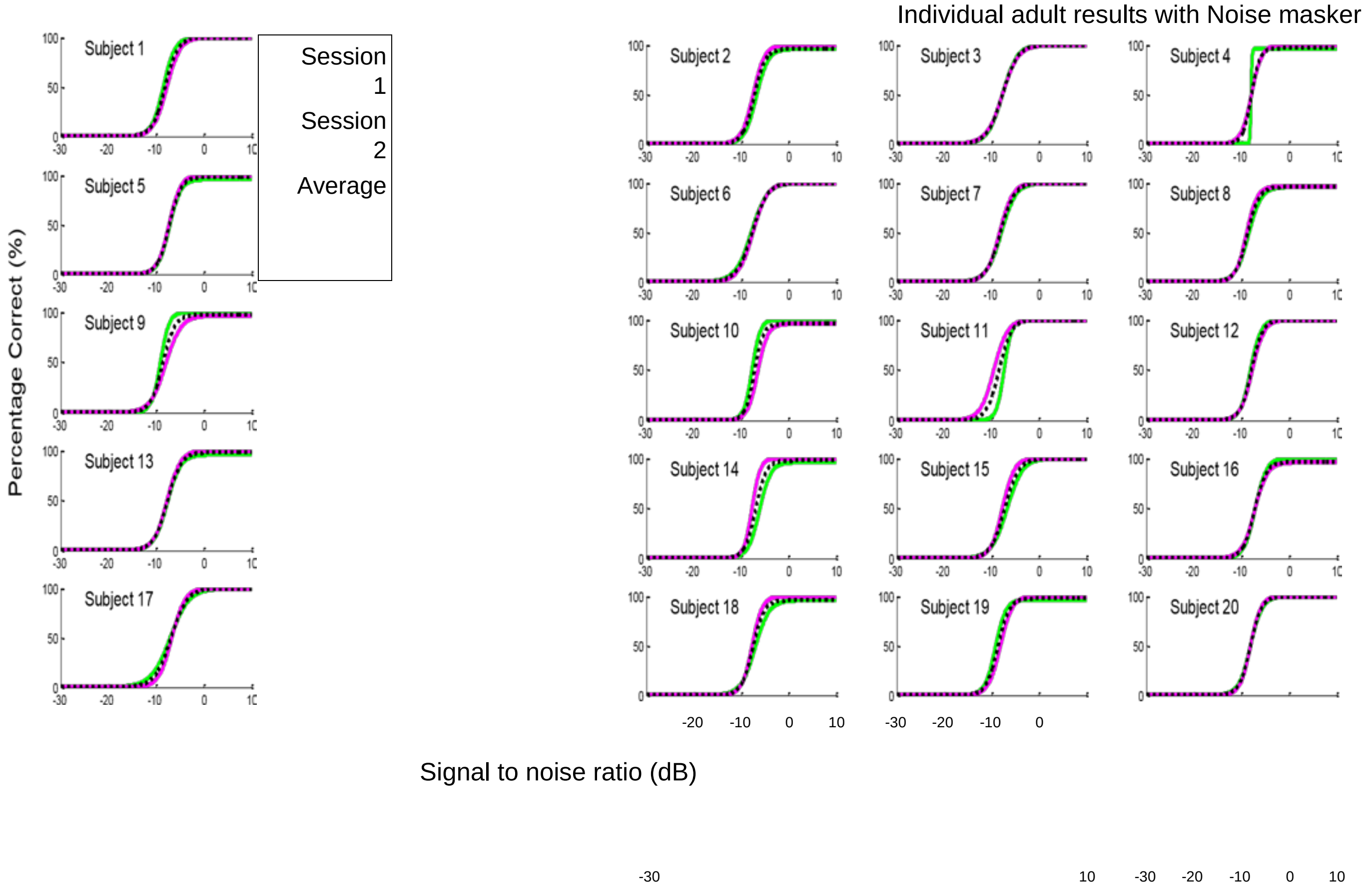
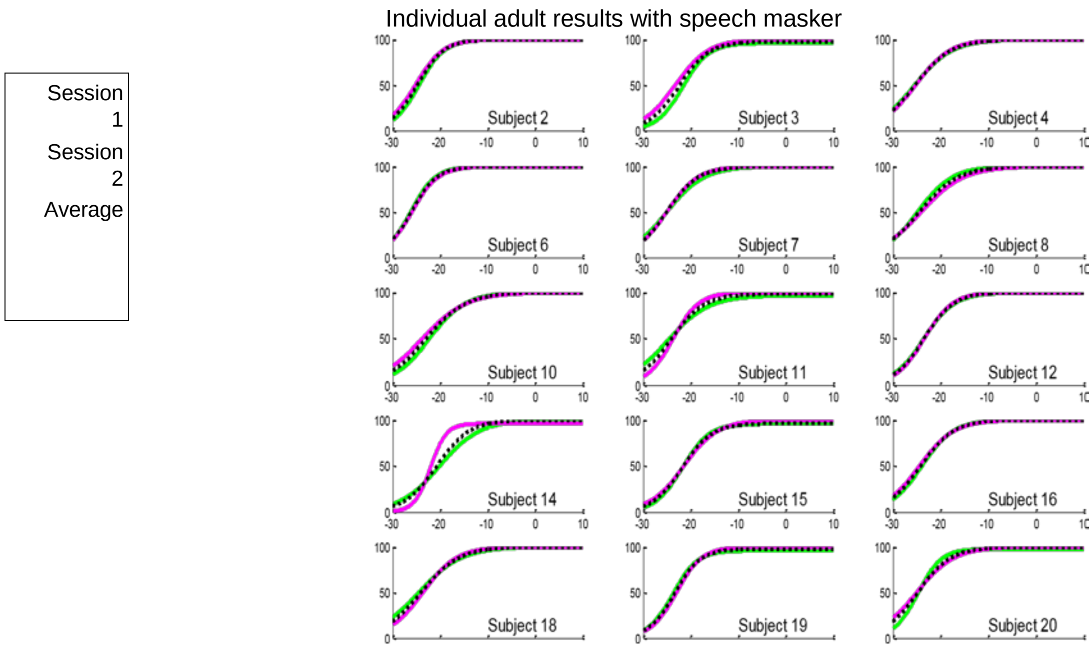


Figure 4.4: Fitted logistic functions of results from adults with the noise masker in session 1 (green line), session 2 (magenta line) and with both sessions added together (black dotted line).



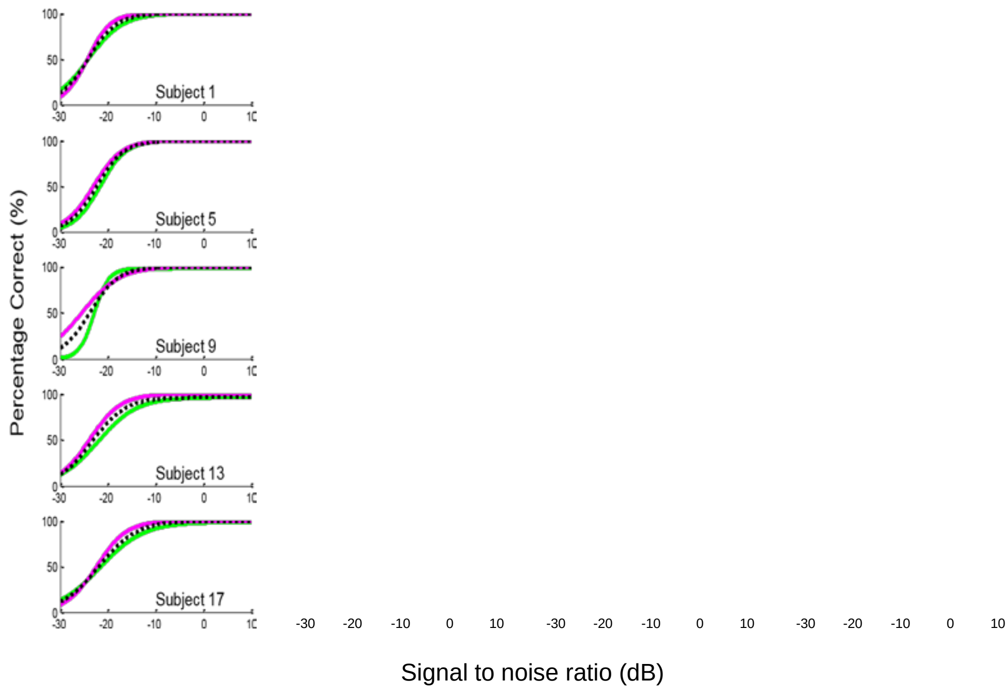


Figure 4.5: Fitted logistic functions of results from adults with the speech masker in session 1 (green line), session 2 (magenta line) and with both sessions added together (black dotted line).

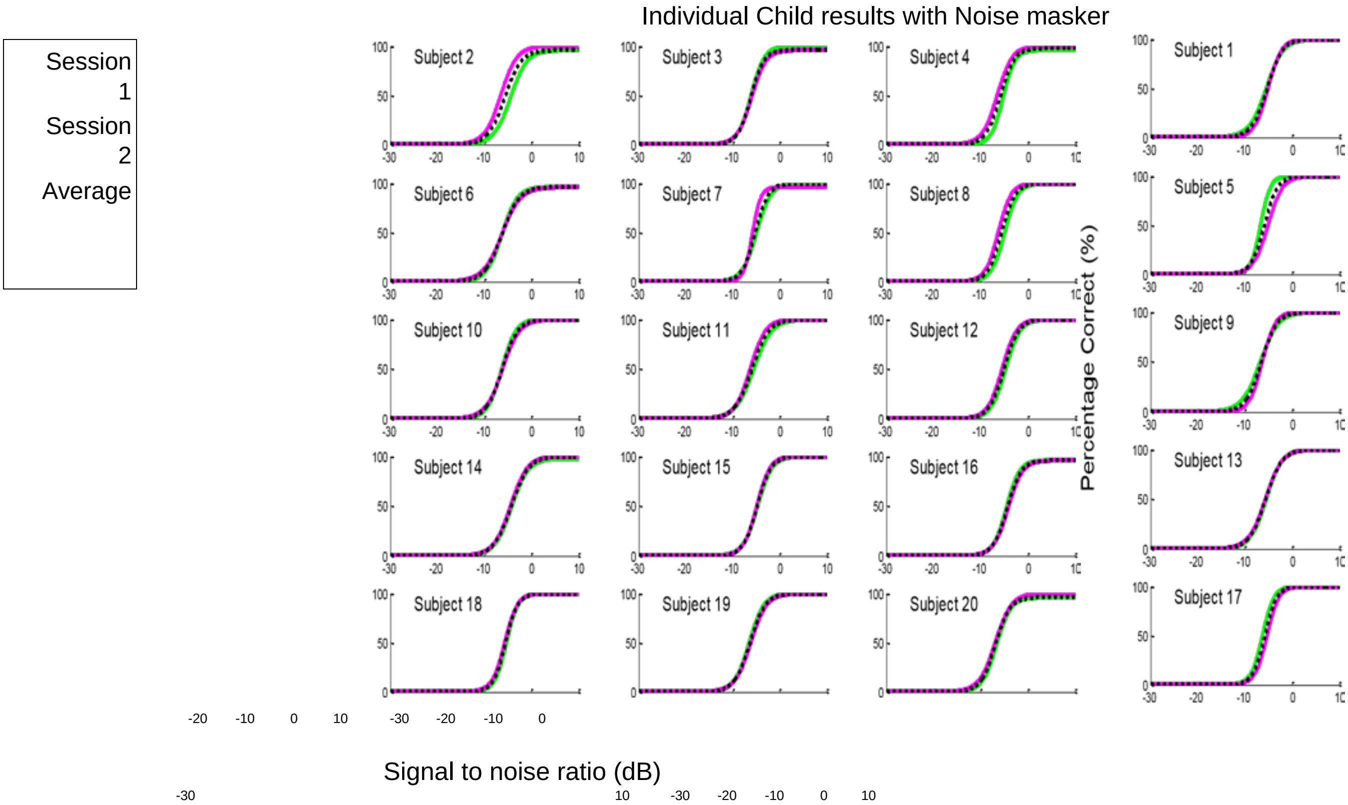


Figure 4.6: Fitted logistic functions of results from children with the noise masker in session 1 (green line), session 2 (magenta line) and with both sessions added together (black dotted line)

Individual Child results with Speech masker

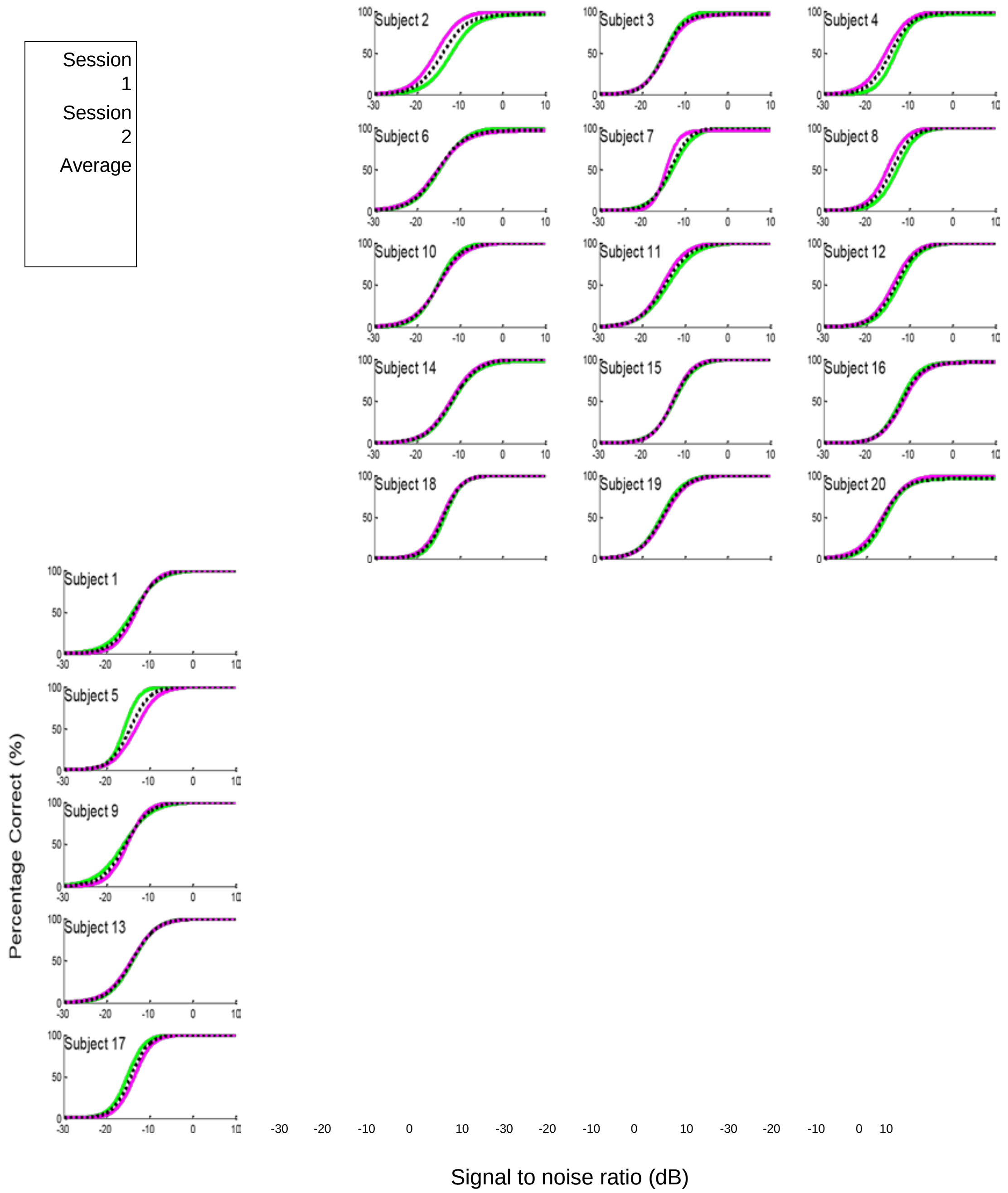
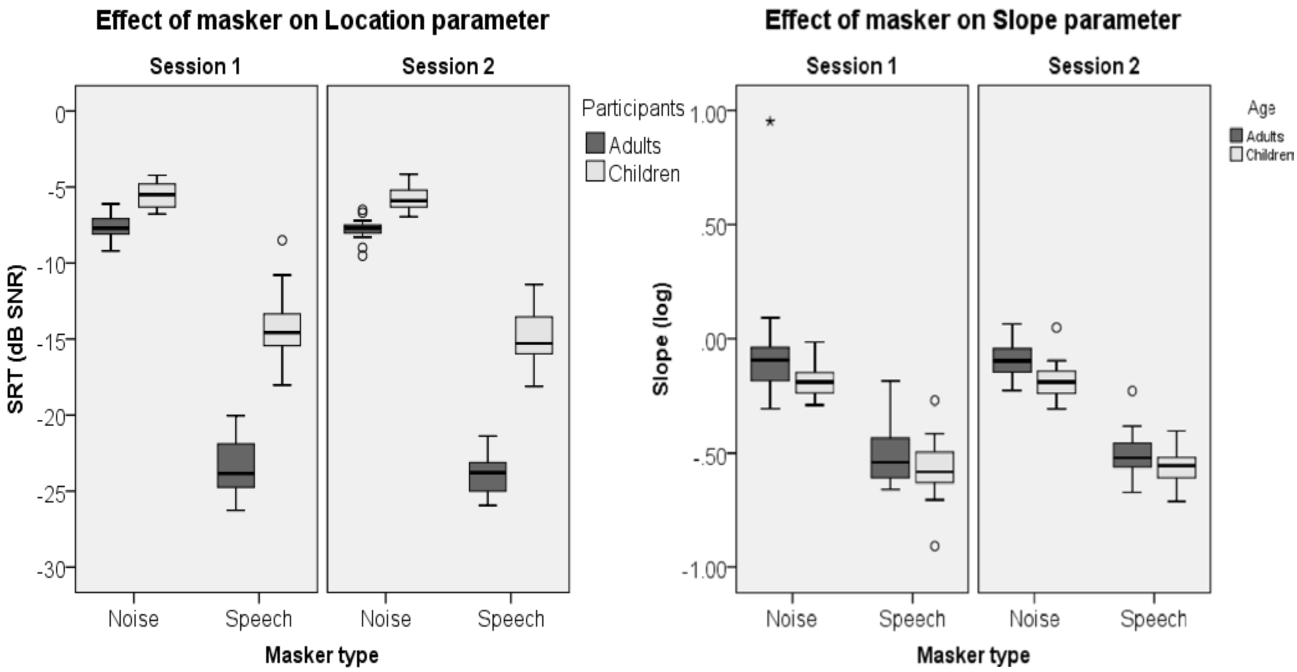


Figure 4.7: Fitted logistic functions of results from children with the speech masker in session 1 (green line), session 2 (magenta line) and with both sessions added together (black dotted line).

Figure 4.8: Box plots to show the effect of masker type on Location and slope parameters across session 1 and 2.



In order to investigate the reliability of individual results more thoroughly, the fitted alpha (location at the 50% correct level) and beta (the slope) parameters for individual psychometric functions in both masker types in children and adults were explored. The results for which are displayed in figure 4.8. Paired samples t-tests were carried out from only the acceptable fits (chosen according to goodness of fit calculations). For the slope parameters the data was logged because distribution of the slopes was skewed and the t-test was carried out on this logged data. Whilst there may have been a slight trend for improvement from one session to the next, there were no significant differences between any of the comparisons with adults (noise location: $t\ (17) = 0.28$, $p = .79$; noise slope: $t\ (17) = 0.46$, $p = .62$; single-talker location: $t\ (19) = 1.92$ $p = .07$; single-talker slope: $t\ (19) = -0.16$, $p = .88$) or with children (noise location: $t\ (18) = 1.07$, $p = .30$; noise slope: $t\ (18) = -0.07$, $p = .94$; single-talker location: $t\ (18) = 1.95$, $p = .07$; single-talker slope: $t\ (18) = -0.014$, $p = .98$). This suggests reliable results. It is interesting however that both children and adults showed an effect near significance for the location parameter with the single-talker masker ($p = .07$). In adults the mean location parameter in session 1 was -23.4 dB SNR and in session 2 was -23.9 dB SNR and although the difference was non-significant it did represent a medium effect size $r = .04$. In children the mean location parameter in session 1 was -14.1 dB SNR and in session 2 was -14.9 dB SNR and again the difference was non-significant but did represent a medium effect size $r = .42$.

Variability was also investigated by calculating measurement error across the two alpha (location at the 50% correct level) measurements (session repeats) as detailed in Summerfield et al. (1994) and as used in Chapter 3 (see section 3.1.1.2 for more details). It can be seen from Table 4.8 that measurement error is small suggesting reliable results.

Table 4.8: Measurement error (95% confidence intervals) for SRT (dB SNR) at the 50% correct level in children and adults with both noise and speech maskers. Measurement error was calculated as detailed in Summerfield et al. (1994).

	Noise (dB)	Speech (dB)
Year 3 - 20 children (7-8 years)	±1.2	±2.8
Adults	±1.0	±1.6

4.4.5 Effect of masker type in children and adults

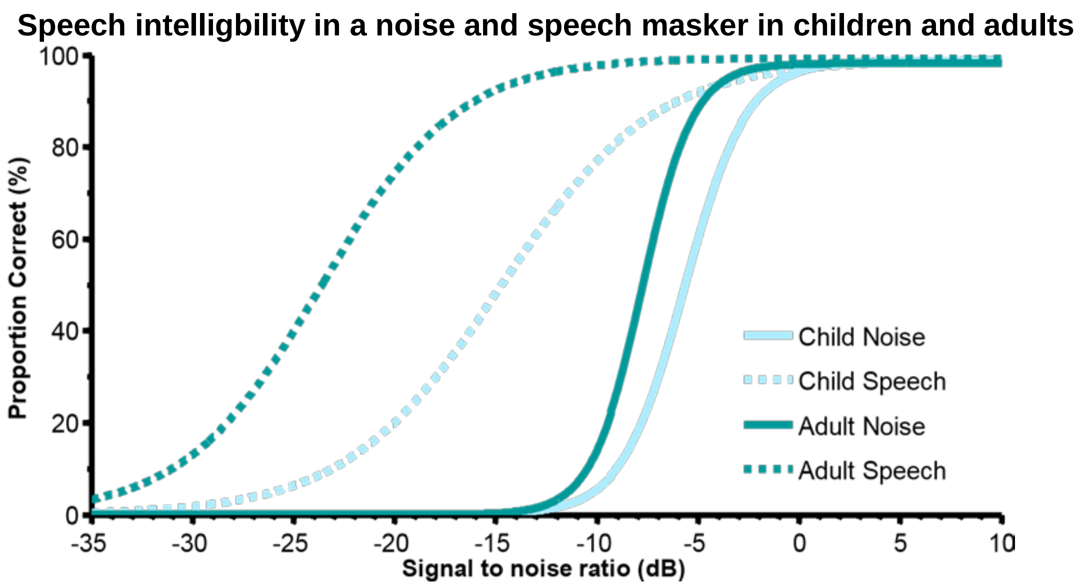


Figure 4.9: Mean psychometric function parameter values across good fits (n = 18 children, n= 18 adults) obtained from adding scores in session 1 and 2, so from both sessions.

Figure 4.9 displays the mean results from 18 children and 18 adults. These psychometric functions were calculated by adding each participant's results from session 1 and 2 to provide a combined session's score out of thirty. The scores at each tested SNR out of thirty were then fitted with individual psychometric functions and location and slope parameters were found.

The mean of these location and slope parameters was then obtained and this is displayed in the figure. The data from child participants 1 and 11 and adult participants 7 and 15 were excluded from this analysis owing to their poor fits. The solid lines display the results from the noise masker whilst the dotted lines display the results from the speech masker. The dark blue lines present the results from adults and the light blue lines present the results from children.

With the results from the noise and speech maskers presented on the same graph it can be seen that slopes are shallower with the speech masker and much steeper with the noise masker. It can also be seen that the results with the noise masker in adults are slightly steeper with the adults than the children and that children require slightly better signal to noise ratios than adults to achieve similar intelligibility levels. With the speech masker however, the curves are again slightly shallower with children but here children need a much greater signal to noise ratio than adults to achieve the same intelligibility level.

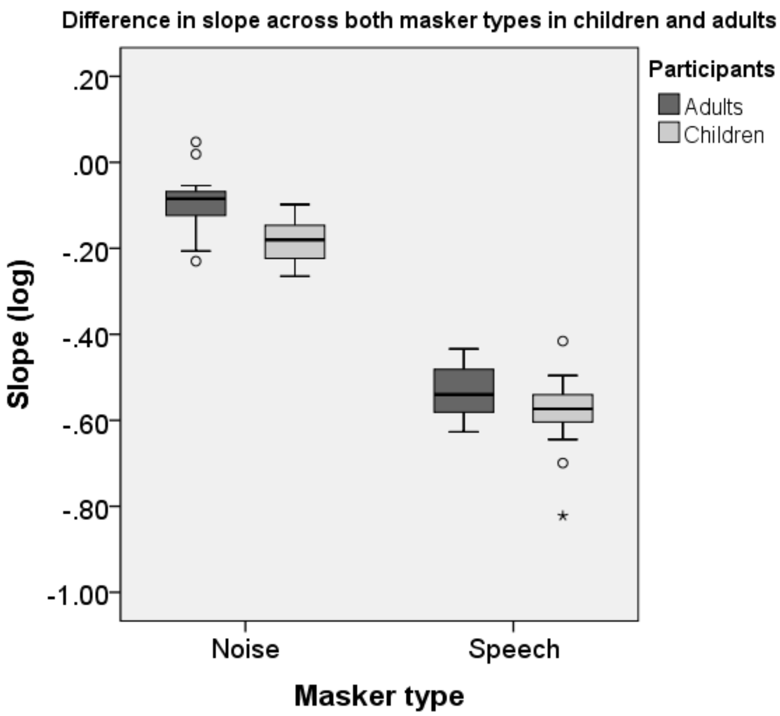


Figure 4.10: Box plot to show the effect of masker type on slope in children and adults.

Figure 4.10 plots the differences in slope parameters across masker types between children and adults. The data was logged because the distribution of the slopes was skewed. It can be seen from the figure that slopes appear to be shallower with the speech masker than the noise masker and shallower in children than in adults. A two-way mixed ANOVA was carried out to investigate this further. There was found a significant main effect of age $F(1, 34) = 1744.65$, $p < .001$ and masker type $F(1, 34) = 699.78$, $p < .001$. This indicates that overall, the slope was steeper in adults compared to children and in noise compared to speech. There was no significant interaction between masker type and age F (1, 34) = 1.17, p = .29 suggesting that the slopes were shallower in children compared to adults to the same extent in both masker types.

The SRTs (dB SNR) for achieving 79.4% correct intelligibility level (as measured in experiment 1) were determined via reading off each individuals psychometric function (combined from both sessions). A two-way mixed ANOVA was carried out to investigate the effects of masker type and age on this SRT (dB SNR). There was a significant main effect of age F (1, 34) = 275.59, p < .001, and masker type F (1, 34) = 881.67, p < .001. This indicates that overall SRTs were lower (better) for adults than children and lower (better) with the single talker speech masker than the noise masker. A significant interaction was found between masker type and age F (1, 34) = 125.29, p < .001, suggesting that the masker type affected children and adults SRT (dB SNR)'s differently in that children's SRT (dB SNR)'s were disproportionately higher (poorer) than adults with the speech masker compared to the noise masker, a finding which agrees with that found in experiment 1. See Appendix C for ANOVA tables.

4.4.6 Effect of masker type comparing at fixed SNRs

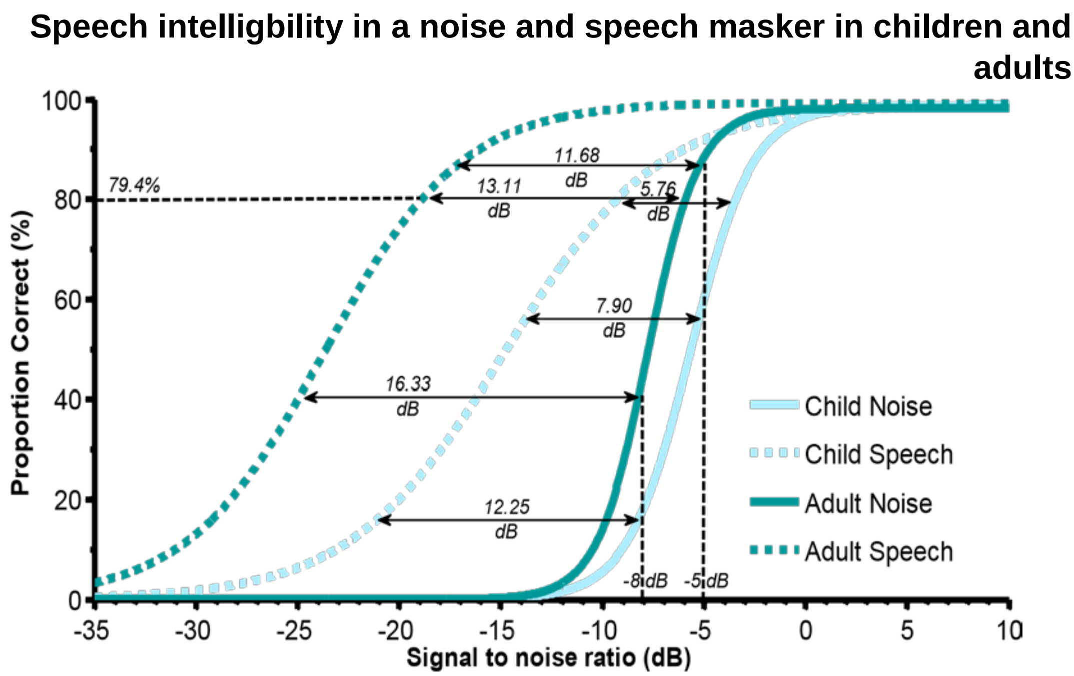


Figure 4.11: Mean psychometric function parameter values (as in figure 4.9) with labels denoting masking release at the same intelligibility level (79.4% correct) and at a few example fixed SNRs of -5 and -8 dB.

Figure 4.11 shows the same graph again but labelled with the decibel difference found from the horizontal differences between psychometric functions for the noise and speech masker (i.e. SRT dB SNR noise minus speech). On the graph these differences have been labelled when comparing SRTs (dB SNR) at the same intelligibility level of 79.4% correct (as would be measured in a 3-down-1-up adaptive procedure and as measured in experiment 1). Also these differences have been labelled when comparing performance at fixed SNRs (as recommended by Bernstein and Grant, 2009) in order to take into account differences in the baseline (noise) conditions. SNRs of -5 dB and – 8 dB have been chosen to be labelled and displayed on the graph as examples. Table 4.9 shows the exact child/adult differences at each comparison point. It can be seen from the graph and table that the larger child adult difference, found when comparing SRT (dB SNR) at the same intelligibility level is reduced when comparing performance at fixed stationary noise SNRs, when taking into account differences in the baseline condition. Whilst this child/adult difference is reduced, it is not however eliminated, therefore a child/adult difference may exist despite this possible SNR confound.

Table 4.9: Child/adult differences in dB between FMB scores as calculated from the average psychometric function parameters at 79.4% correct, and then at 2 stationary noise SNR comparison points to be shown as examples

Comparison point	Child/adult difference (dB)
79.4%	7.4
-5 dB SNR	3.8
-8 dB SNR	4.1

In order to look into the child/adult differences further, this SRT (dB SNR) difference between masker types within children's and adult's results was plotted on a graph to show the difference as different points on the psychometric functions are compared (i.e. different stationary noise SNRs). Since measured fluctuating masker benefit (FMB), as explained previously by Bernstein and Brungart (2011), decreases with increasing SNR due to the slopes of the psychometric functions, it was possible that since children perform at increased SNRs compared to adults in the stationary noise baseline condition the previously observed reduced FMB (or poorer performance in children with the speech masker) could have been found solely due to the poorer baseline performance and the point at which comparison took place. It could be because children were simply poorer overall and that there is not anything particular about the speech masker which children find particularly difficult.

Figure 4.12 shows the FMB (or SRT noise minus speech dB SNR) plotted against the stationary noise baseline SNR. The results for this graph were obtained from each of the 18 children's and 18 adults' individual good fit psychometric functions calculating the FMB at a variety of fixed stationary noise SNRs and also the results calculated at the same intelligibility level of 79.4% correct. The confidence intervals on this data were obtained and are plotted in the graph.

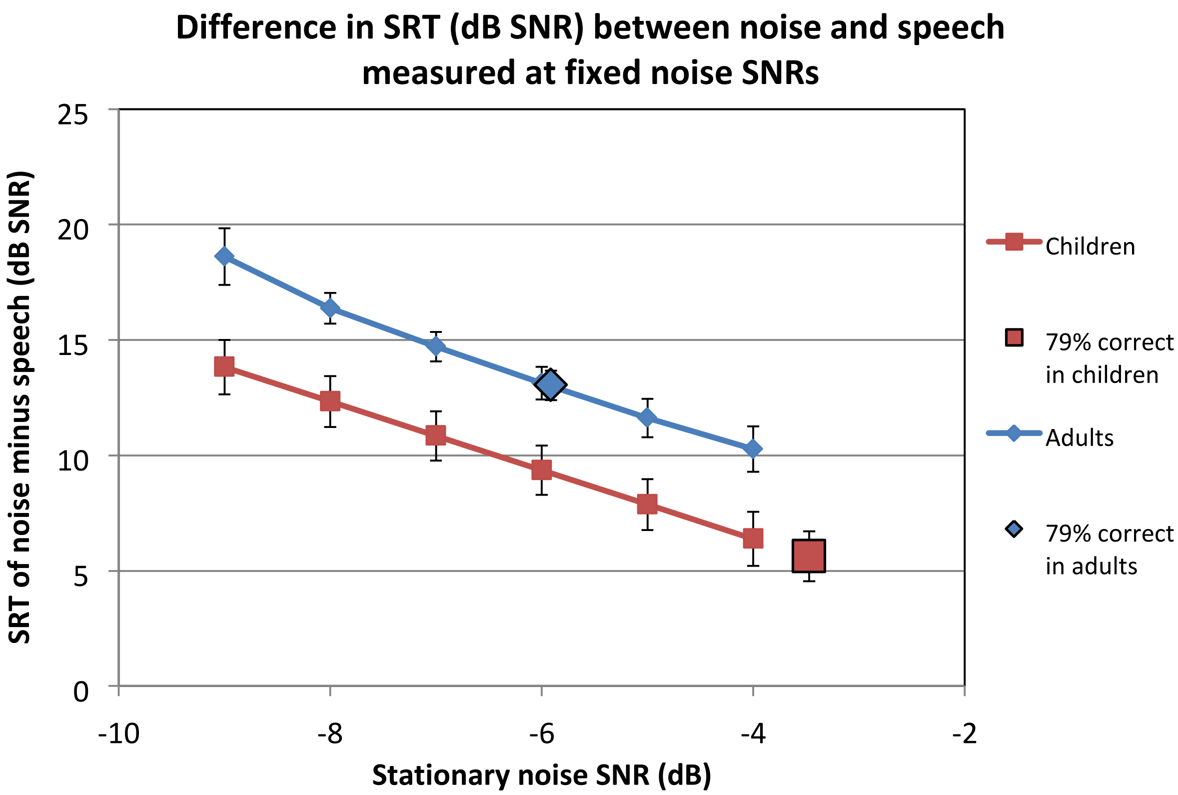


Figure 4.12: Graph to show the difference in SRT (dB SNR) between masker types when compared at fixed stationary noise baseline SNRs. The error bars show the 95% confidence intervals as calculated taking into account each individual participants' difference in SRT (dB SNR)

As can be been in the figure, FMB decreases as baseline stationary noise SNR increases. The FMB measured at 79.4% correct was larger for adults compared to children. However, the stationary noise SNR for 79.4% correct was higher for the children (at -3.5 dB) than for the adults (at -5.9 dB) which may have contributed to the reduced FMB in children. When comparing FMB at fixed stationary noise baseline SNRs it can be seen that children have a consistently lower FMB compared to adults and the confidence intervals do not overlap. This results suggests that there is a reduction in FMB from comparing at the same intelligibility level to comparing at fixed stationary noise SNRs so some of the observed child/adult difference can be said to be due to the SNR confound. Since a difference still remains however, not all of the child/adult difference can be attributed to this SNR confound. Table 4.10 details the mean child/adult differences at each comparison point together with the differences based on the upper and lower confidence intervals.

Table 4.10: Child/adult differences in dB between FMB scores as calculated from the average FMB scores across individual participants at 79.4% correct, and then at various stationary noise SNR comparison points. The smallest child/adult differences (calculated using children's +95% CI and the adults -95% CI and displyed as well as the largest child/adult differences (calculated using children's -95% CI and the adults +95% CI). The child/adult difference at 79% correct, -5 dB and -8 dB is different to that seen in table 4.9 because these results are based on FMB for each individual participant and those in table 4.9 were based on the average psychometric functions obtained from averaging all the good fit parameter values.

Comparison point	**Mean child/adult difference (dB)**	**Smallest child/adult difference (children's +95% CI and adults -95% CI) (dB)**	**Largest child/adult difference (children's -95% CI and adults +95% CI) (dB)**
79.40%	7.2	5.5	8.8
-4 dB SNR	3.7	1.5	5.8
-5 dB SNR	3.5	1.5	5.4
-6 dB SNR	3.5	1.7	5.3
-7 dB SNR	3.6	1.9	5.3
-8 dB SNR	3.8	2.0	5.6
-9 dB SNR	4.6	2.2	7.0

It can be noted from table 4.10, that no matter where along the x axis for the stationary noise masker FMB is calculated, a child/adult difference still remains. Taking the smallest difference based on the children's -95% confidence intervals and adults +95% confidence intervals, a 1.5 dB difference still remains and such a difference could be as larger as 7.0 dB even when taking into account the possible SNR confounds.

4.4.7 **Effect of masker type comparing percentage correct data for**

children against adults

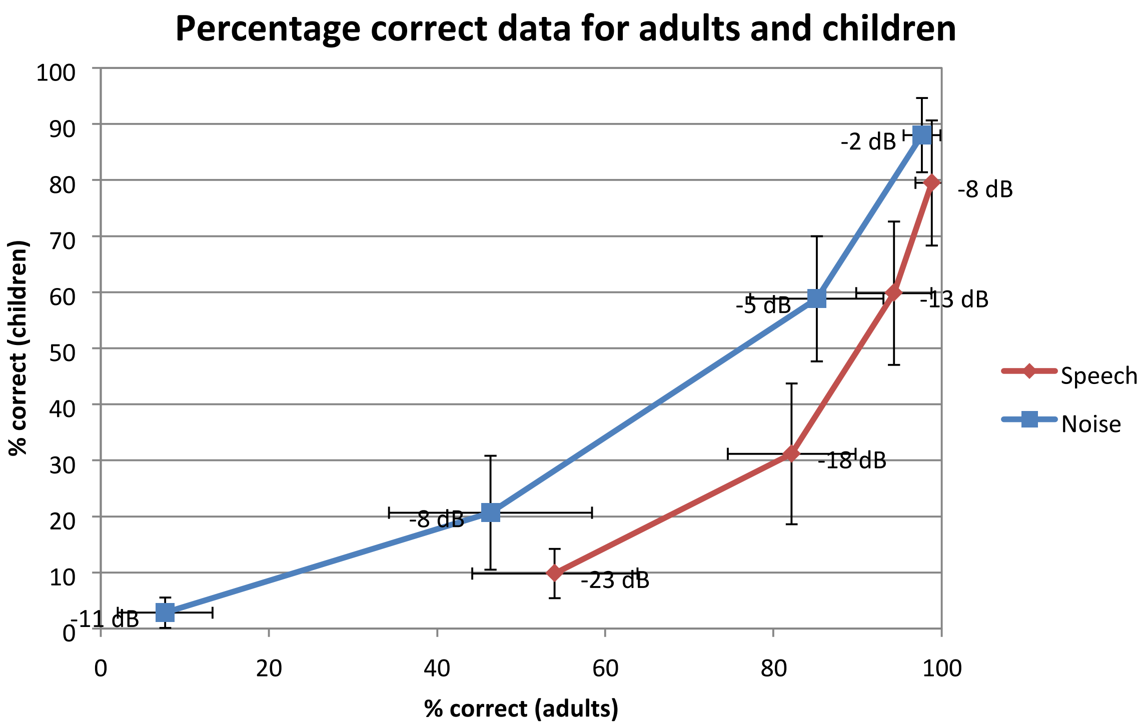


Figure 4.13: Percent correct data for children plotted against those for adults at the same tested SNRs. Error bars denote 95% confidence intervals for % correct data.

A further method proposed by Bernstein (2012) for controlling for possible SNR confounds when comparing performance with noise and speech maskers, is to plot the raw percentage correct data from one participant group against the other. If there is a real difference in performance between children and adults with the speech masker which is not dependant on the difference with the noise masker (i.e. if poorer performance in children with the speech masker is found due to only a difference between children and adults with the baseline stationary noise condition), then the curves should differ across masker type. As can be seen from figure 4.13, the curves do not overlap and so it can be suggested again that a real child adult difference remains which cannot be solely contributed to an SNR confound.

4.5 Discussion

4.5.1 Overview of results

The results of this study agree with results from experiment 1 that children require disproportionately higher (better) signal to noise ratios than adults with a speech masker compared to a noise masker to achieve comparable intelligibility levels with sentence test stimuli. This finding also ties in with previous findings (Hall et al. 2002; Bonino et al. 2013; Leibold & Buss 2013). It can be suggested from the analyses of this study's results, that previous interpretations suggesting children may be more susceptible to informational masking compared to adults however could have been over exaggerated. Taking into account differences in the baseline stationary noise conditions suggests that child/adult differences do exist regarding their improved performance with speech compared to noise makers, but may not be as large as previously thought. A difference however does still remain so not all child/adult differences with the speech masker can be explained by an SNR confound and in terms of differences in the baseline stationary noise conditions. These findings will now be discussed.

4.5.2 Vocabulary test

The results from the BPVS suggest that all the children within this experiment had a vocabulary range which was appropriate for their age. This provides confidence that the child/adult differences existing within this study are not down to the fact that children have an inappropriate vocabulary range. It cannot be ruled out however, as with experiment 1, that better performance displayed with adults is simply due to greater experience with language. This point applies even more so with open set sentence identification tasks because there are more contextual cues available (Wilson et al. 2007) but this factor cannot be easily controlled for.

4.5.3 Reliability

The results carried out for both children and adults with both masker types were found to be reliable from one session repeat to the next. This finding may reflect the fact that sentence targets in comparison to word targets are found to yield steeper psychometric function slopes so are thus suggested to be more reliable since they are more sensitive in measuring speech reception threshold (Neumann et al. 2012). Furthermore using sentences stimuli has greater external validity reflecting more greatly than word stimuli how listeners are able to integrate aspects of speech for speech intelligibility in realistic conditions. Moreover, the standard deviation of the SRTs obtained with BKB sentences has been found to be 2.4 dB (Bench & Bamford 1979) and test re-rest reliability within this experiment was very low.

4.5.4 Effect of masker type in children and adults

Regarding the psychometric functions derived from the results with children and adults in both masker types the slopes were shallower with the speech masker compared to the noise masker. This finding ties in with previous research (Oxenham & Simonson 2009; Bernstein & Brungart 2011; Arbogast et al. 2002) and suggests that since the speech masker has a larger dynamic range the local SNR is increased enabling more opportunities for “glimpsing” at lower SNRs (MacPherson & Akeroyd 2014). This also reflects the fact the speech maskers are simply more variable from trial to trial than noise maskers.

A study by MacPherson & Akeroyd (2014) investigated data in the literature on psychometric function slopes with various masker types from 139 different studies. It was found that with an open set task (as used in the current study) slopes are generally shallower than those found with closed set tasks and suggested thus that cognitive factors can influence slope gradient. Slopes between children and adults were also compared and it was found that slopes tended to be shallower in results from children compared to adults in speech maskers but not with stationary noise maskers. This was not quite the case within the current study as slopes were shallower in results from children compared to adults in speech maskers and also in noise maskers to the same extent. The reasons for such differences are not entirely clear. MacPherson and Akeroyd (2014) subsequently discuss that the identification of the sentences or keywords within a sentence may be based on both bottom-up and top-down influences. Top-down influences can lead the listener to guess the target word if the rest of the sentence constrains word options and so in combination with small increased in bottom up information the slope may be steeper. Therefore there is more top-down influence in closed set tasks because the target options are already constrained. So, small increases in SNR (bottom-up salience) may facilitate the use of top-down information with a combination of these two things steepening the slope. A lack of top-down information may provide then greater reliance on bottom-up information and slopes may be shallower (Dirks 1986; Dubno et al. 2000; Elliott 1979; Kalikow et al. 1977; Pichora-Fuller et al. 1995). MacPherson and Akeroyd (2014) discuss this further and state that slope steepness may thus be able to provide insights into how much a listener is depending on bottom-up processes as if steeper then there may be a greater reliance on top-down information to combine with bottom up information constraining options for the target word. MacPherson and Akeroyd (2014) explain that it was found in the survey that elderly listeners with hearing loss (and thus distorted bottom-up information) were found to have steeper slopes and so maybe more greatly relied on top-down information to overcome their hearing difficulties. In this same way it could be suggested that children's slopes are shallower than adults because children may perhaps not be relying on top-down processes because it could be these cognitive factors are underdeveloped in children compared to adults, but they may have adequate bottom-up processes.

4.5.5 Tackling SNR confounds

Results from experiment 1 showed a masker*age interaction suggesting that children were more susceptible than adults to masking by a speech masker; this was found using a speech test which employs a relatively low cognitive load. Results from the current experiment also show this to be the case using a speech test which is more realistic and engaging of a higher cognitive load. Although children were slightly older than previously tested in experiment 1, they still remained poorer than adults in the baseline stationary noise conditions. When taking into account these baseline differences it was clear that FMB decreased with increasing SNR, owing to the shapes of the psychometric functions which agreed with Bernstein and Grant (2009). Whilst FMB decreased, a child adult difference in FMB was not eliminated so this difference could not be said to be solely attributable to the SNR confound. Therefore, whilst previous studies have shown large child/adult differences with speech maskers (e.g. Bonino et al., 2013; Leibold & Buss, 2013; Hall et al., 2002) it could be that some of this difference may be due to confounds of the measured SNR, although perhaps not all of it.

In experiments with similar methods to the current study, the SNR confound was found to change previous conclusions, that hearing impaired people cannot take advantage of the dips within the masker due to poorer temporal resolution (Bernstein & Grant 2009; Bernstein & Brungart 2011). These studies actually showed that there was nothing specific about the properties of the fluctuating masker which affected hearing impaired people more than normal hearing people, but just that they had a speech intelligibility difficulty in general. In contrast, the findings of the current experiment suggest that there may actually be real differences in the way speech backgrounds affect children and adults' speech intelligibility which relate specifically to the speech background and so not just a processing deficit in general.

One study which does show some child/adult speech intelligibility differences may be present owing to an SNR confound, is a study by Hall et al. (2012). There was found a significant masker*age interaction where adults showed a larger release from masking when a speechshaped noise background was switched to a spectrally and temporally modulated noise background compared to a group of younger children (4-6 years) and older children (7-11 years). The authors suggested that ability to piece together speech segments across spectral and temporal gaps may not be developed in children. Hall et al. (2012) did however, add in a supplementary condition aiming to tackle any possible SNR confounds by aiming to adjust baseline performance. A further 10 adults were recruited and speech intelligibility was measured in an adaptive procedure to determine SRTs corresponding to a 70% correct performance level (as opposed to a 50% correct performance level in children) in order to attempt to take into account differences in the baseline conditions by attempting to compare performance between children and adults at the same SNR for the steady noise conditions. It was found that adults in the supplementary condition showed poorer results (of +0.1 dB) compared to adults in the main experiment (-5.3. dB), and the masker*age interaction become no longer significant. Therefore the authors suggested the interaction may have been the result of SNR confounds and that differences were not likely to be due to any possible underdeveloped temporal processing in particular in children just poorer speech intelligibility in general.

Whilst the study by Hall et al. (2012) used different noise maskers, the current study used speech maskers and found that the results cannot be wholly attributed to SNR confounds. This study is not in a position to explain why children and adults differ in this respect but reasons for such differences may be speculated. Since speech backgrounds are suggested to produce effects of informational masking, and effects of informational masking are found to relate to higher level cognitive factors, it may be that child/adult differences exist due to cognitive developmental factors in children. It may be that children are not able to segregate target and masker speech in the same way that adults can, so differences could relate to immature auditory scene analyses. It could also be the case that children are particularly poorer than adults when it comes to ignoring the interfering speech masker and that attention is drawn away from the target talker. This would tie in with previous research that showed young children could not selectively attend in a dichotic listening task compared to adults (Wightman & Kistler 2005). Furthermore, it is found that selective attention develops across childhood (Bargones & Werner 1994) so child/adult differences could reflect this development. It is also possible that reduced linguistic experience (Calandruccio et al. 2014) or working memory (Gomes et al. 2000) in children prevents them from using top-down processes to understand the target talker and inhibit the interferer. It may therefore be of great interest to further investigate cognitive factors in children and see how these relate to interference from speech backgrounds.

4.5.6 Implications

The implications of this study together with findings of previous research suggest that children may be at a disadvantage compared to adults when faced with the task of listening to speech in speech backgrounds and advocate cause for quiet classroom environments. Noisy classrooms could make it harder for children to discriminate speech invoking the greater involvement of top-down processes which in turn could increase cognitive load leading to fatigue (Edwards 2007).

Cognitive load is said to have a limited capacity and resources are divided between tasks (Moray 1967). The allocation of resources may change when one task is more cognitively demanding (Kahneman 1973) and if children are allocating resources to filtering out speech backgrounds they could be left with fewer resources to cope with learning. The current findings may highlight a greater need for quiet learning environments. Although hearing impaired children were not tested in this study it could be speculated that they may be even more greatly disadvantaged. Hearing impaired children who are at a disadvantage with reduced bottom-up salience may be requiring a greater engagement of top-down processes already which seems likely to be made even more difficult when the background sounds consist of speech. Relating to these discussions therefore, such children could be said to benefit from the use of FM systems in combination with their hearing aids which would enable them to hear the teachers voice directly into their hearing aids whilst blocking out surrounding sounds. Such systems would improve SNR, when the teacher's voice is the target of interest, and may help overcome the detrimental effects speech backgrounds have on speech intelligibility and to facilitate improved learning conditions.

It is important to consider however the reasons behind why the masker*age interaction exists and why it seems to persist up until adolescence (Wightman et al. 2010). Werner (2007) explains that children listen in a less focussed and broader fashion than adults and this may be necessary for them to learn the important cues and features of speech. Concerning this idea it could also be that many other aspects of developmental learning within children are dependent on this broad listening strategy and so could be the reason for this reduced selective listening. Overhearing may then be critical to enable incidental learning (Akhtar 2005) which may be important for a child's academic and social development. If it is true that this broad listening strategy is necessary for learning then it is possible the benefit associated with the use of FM systems in children with hearing loss could actually summate to a disadvantage. Limiting the use of overhearing may even hinder a child's development possibly leading to delayed development of language and other social skills. Further research would thus be of interest to investigate this issue in more detail.

4.6 Conclusions

- Children across an age range of 7-8 years performed disproportionately poorer than adults with a single-talker speech masker compared to a steady speech-shaped noise masker.

- This interaction can be cannot be solely attributed to an SNR confound, therefore real child/adult differences exist regarding speech intelligibility in speech maskers.

Chapter 5: **Effect of background on speech intelligibility and reading**

5.1 Introduction

To recap on the discussions from the literature review, different background maskers are said to contribute in differing amounts to the two main mechanisms of masking in speech intelligibility. Those background maskers which overlap spectrally and temporally with target speech at the peripheral level are thought to encompass energetic masking. A steady state stationary noise may be thought of as predominately contributing energetic masking consisting of a high acoustical energy (Wightman et al. 2010). Predictions of speech intelligibility based on energetic masking can be made considering the listener's hearing threshold levels and the measured amount of overlap in frequency and time between target and masker materials (Moore & Glasberg 1987).

Where the effect of a background masker on speech intelligibility cannot be wholly explained in terms of energetic masking, a second mechanism is described. Informational masking, often considered a catch-all phrase (Shinn-Cunningham 2013), is used to describe such interference (Durlach 2006). This term describes the presence of masking owing to similarities between the target and masker and uncertainties in distinguishing the target from the masker. Informational masking refers to interference of maskers at more central as opposed to peripheral levels, specifically at higher cognitive levels (Durlach et al. 2003). A speech background may be thought of as predominately contributing informational masking since it consists of a lower concentration of acoustical energy (so limited energetic masking), but contains similarities to the target speech which may be confused at higher cognitive levels (Stickney et al. 2004). The underlying properties of informational masking however are not fully understood (Kidd et al. 2007).

5.1.1 Properties of different types of background

5.1.1.1 *Number of talkers*

Energetic masking and informational masking mechanisms in speech intelligibility can be present within one single background noise simultaneously. Depending on the properties of the background noise, contributing amounts from each masking mechanism can be manipulated. For example, it seems a single-talker masker is thought of as contributing mostly informational masking due to its sparse acoustical structure (so limited energetic masking) and similar qualities to a speech target. Performance with a single-talker masker compared to a stationary noise masker however is often found to be improved and this is largely attributed to a release from energetic masking (Freyman et al. 2004; Litovsky 2005). The informational masking effect therefore of a single-talker masker is not especially detrimental to speech intelligibility; listeners can still achieve high intelligibility levels even in very adverse listening conditions (i.e. low SNRs).

A two-talker masker of the same sex as the target speech is thought of as contributing mostly informational masking. Whilst energetic masking is increased as compared to a single-talker masker, performance on speech intelligibility tasks are often found to be worse than would be predicted based on energetic masking alone (Freyman et al. 2004; Rosen et al. 2013). Performance with a two talker masker compared to a stationary noise masker are found to be worse despite there still being a release from energetic masking with its sparser structure. Because of this, a two-talker background masker is thought to comprise strong informational masking.

In general previous research shows that a two-talker masker is a particularly effective masker contributing large effects of informational masking more so than a single-talker masker (Brungart 2001; Hall et al. 2002; Freyman et al. 2004). As the number of talkers is increased beyond two or three, performance has been shown to improve (Brungart 2001; Freyman et al. 2004; Rosen et al. 2013). This is thought to be because, whilst the addition of talkers fill in the spectral and temporal dips increasing energetic masking and becoming more like stationary noise, the effects of informational masking are decreasing as the acoustic mix becomes less and less similar to the target speech so speech intelligibility improves (Rosen et al. 2013).

5.1.1.2 *Understanding the cognitive demands of informational masking*

Since speech maskers are thought to indicate a greater cognitive involvement compared to noise maskers, attempts have been made to investigate the cognitive demands of such maskers using pupil dilation measures. Pupil dilation measurements have previously been used to determine cognitive load, and Beatty (1982) explains that several studies have shown pupil dilations to be observed in various mental tasks where larger pupil dilations are found with tasks expected to be ‘difficult’ (Koelewijn et al. 2012b). Koelewijn et al. (2012b) investigated participants’ speech intelligibility in a stationary noise masker, a single-talker speech masker and a fluctuating noise masker (a noise masker with amplitude fluctuations in sync with the speech masker). Not surprisingly, it was found that speech intelligibility improved with the speech masker and the fluctuating noise masker compared with the stationary noise masker, likely to be due to some reduction in energetic masking. What was most interesting however, was the observed pupillary responses that were recorded during the task. When the task was set so that in each of the three masker conditions participants achieved the same intelligibility levels (either 50% correct, or 84% correct), larger pupil dilations were actually observed when listening to the speech in the speech masker, compared with both the stationary and fluctuating noise maskers. This result reveals that although speech intelligibility measures showed listeners were able to achieve the same intelligibility level at lower SNRs with the speech masker compared to the stationary noise masker, more effort is exerted when perceiving speech in the presence of another talker than in the presence of stationary and fluctuating noise suggesting a specific interfering effect from speech itself. The authors discuss this as an effect from informational masking. Such findings therefore highlight the importance for measurements of cognitive load over and above purely speech intelligibility measures. Current measures of speech intelligibility used to evaluate hearing difficulties clinically do not tell us anything about the cognitive demands of listening in speech backgrounds, the interference of informational masking, or how we are able to cope with communication more generally, in realistic settings.

It was thought that by looking at the effect these background maskers have on the normal process of reading, we may be able to further understand how speech maskers interfere at higher cognitive levels because any physical masking at the peripheral level (i.e. the cochlea) would be removed. Furthermore measuring eye movements when reading is said to provide insight into processing difficulties (Rayner 2009) so could tell us about interference when processing language which may also relate to processing difficulties during speech intelligibility tasks. By manipulating properties of the background maskers we will be able to see if those maskers associated with most informational masking in speech intelligibility tasks are also those associated with most interference during reading to understand further this interference at higher cognitive levels.

5.1.2 Assessing the process of reading

To recap, most studies investigating the effect that background sound has on the reading process have used measures of reading comprehension to show that speech backgrounds produce larger detrimental effects to comprehension compared to stationary noise backgrounds (e.g. Sörqvist et al. 2010; Martin et al. 1988). This finding is of course in contrast to the findings with speech intelligibility tests but suggests that the informational masking associated with speech causes cognitive interference more so than a noise background. By looking at eye movements in reading however, we can not only look just at the detrimental effects of background noise when reading (which may have implications for working in open plan offices and for children working in noisy classrooms), but eye movements also tell us about online moment to moment cognitive processes and so online processing difficulties (Rayner 2009). For example, when readers read a low frequency occurring word compared to a high frequency occurring word their fixation durations increase (Inhoff & Rayner 1986), and when readers read a word which is implausible within the sentence fixation durations increase as do the proportion of regressions (Warren & McConnell 2007). To our knowledge there has only been one study which has investigated the effects of background speech on reading, measuring eye movements and that is the study by Cauchard et al. (2012). They found that compared to silent reading, reading rate was slowed with a single-talker speech background as number of fixations and regressions increased, suggesting that participants had to re-read the text to comprehend it. By including different manipulations of background sound we can extend this study to further understand the effect realistic background sounds have on the reading process and use these findings to additionally aid explanations of informational masking in speech intelligibility tasks, which have not been investigated in this way previously.

5.1.3 Test of word reading efficiency (TOWRE)

When carrying out tests on reading it is important to ensure that the participant sample has efficient reading ability, as this could interfere with interpretation of the results. The TOWRE tests word reading skill with regards to the recognition of familiar printed words and the accurate pronunciation of printed words. It includes two subtests of ability to pronounce both real words and non-words. Both accuracy of pronunciation and the speed of these processes are measured (Torgesen et al. 1999).

The subtest of sight word efficiency (SWE) measures ability to identify real printed words and records the number of words pronounced accurately in 45 seconds. This subtest evaluates sight word vocabulary and critical word reading skills out of context based solely on the words' visual printed appearance.

The subtest of phonemic decoding efficiency (PDE) is carried out in the same way but measures ability to identify printed non-words. With the inclusion of non-word identification, the reader is required to apply graphophonemic knowledge in decoding words (Siegel 1989) where any contextual cues are removed, demanding full analysis of each word.

The number of accurately pronounced words/or non-words in the 45 seconds are converted to standardised scores to allow comparisons with well-defined age related norms and enable individual scores to be compared against large groups of people of the same age. As with the BPVS, the standardisation of scores is based on the normal distribution of scores expected for a particular age range and the raw scores are transformed into standardised scores with a mean of 100 and a standard deviation of 15, where 68% of people score between 85-115.

The standardised scores are based on a sample of 1507 participants aged 6-24 years across thirty US states. The TOWRE is designed to decipher differences amongst children primarily and the words for the SWE subtest were chosen based on how frequent words occur in text for beginning school level. As the test progresses words become less frequent. For the PDE subtest, difficulty increases as the test progresses with increases number of phonemes and syllables (Torgesen et al. 1999).

For a test such as the TOWRE to be considered reliable it has been said that reliability coefficients of .80 or above, ideally .90 are required (Aiken 1994; Nunnally & Bernstein 1994). The internal consistency of the TOWRE was investigated using alternate-form reliability (because split half coefficients are not appropriate when a test of speed is included). It was found that coefficients exceed .94 magnitude. Such a test is thus a good indicator of reading efficiency to be used experimentally (Torgesen et al. 1999).

5.2 Aims of experiment 3

1) **To determine what type of background sound is most disruptive to a listening task and which is most disruptive to a reading task.**

Rationale: It has previously been found in Audiological speech intelligibility tests that intelligibility is improved with a single-talker masker compared to a steady state noise masker (e.g Festen & Plomp 1990) thought to be due to the "glimpsing" of the target within the dips of the masker. It is apparent however, that these tests may not provide any information concerning the cognitive demands of such a task as Koelewijn et al. (2012b) have recently found increased pupil dilation (linking this to increased cognitive load) with a singletalker masker compared to a noise masker despite improved intelligibility.

It is thought that investigating the effect these background sounds have on eye movements during a reading task will not only provide insight into how written language processing (i.e. reading) in everyday noisy situations may be affected, but also advance our understanding of how speech maskers (informational maskers) may interfere at higher cognitive levels in communication tasks more generally.

Predictions: (i) a two-talker masker will be most detrimental to a speech intelligibility task because this masker is associated with most informational masking. (ii) A stationary noise masker will be more detrimental to a speech intelligibility task than a single-talker masker; (iii) those maskers associated with the most informational masking (i.e. a two-talker masker) masker will be more detrimental to a reading task than those associated with the least informational masking (i.e. a stationary noise masker).

2) To determine the extent to which level of background noise interacts with the effects of background noise interference in a speech intelligibility task and a reading task.

Rationale: In poorer signal to noise ratios speech intelligibility declines. This is due to the increased concentration of acoustical energy causing a physical masking of the target speech. With different background maskers, measured psychometric function slopes are found to differ and are often shallower with speech maskers. It is therefore of interest to determine the effect of background noise level on speech intelligibility to see how these slopes differ with maskers thought to encompass both energetic and informational masking in differing amounts.

This physical masking cannot occur in the reading task owing to the presentation of background sound and reading material across different sensory modalities. Previous findings have shown that disruption to serial recall tasks do not depend on the level of background noise (Salamé & Baddeley 1987; Colle & Welsh 1976), but the effect of level on a reading task using eye movements has not previously been investigated and so it is of interest to determine whether disruption depends on level.

Predictions: (i) the slope of the psychometric functions for speech masker will be shallower than the slope of the psychometric function for the noise masker. (ii) Based purely on the lack of an effect of background level on serial recall tasks (Salamé & Baddeley 1987; Colle & Welsh 1976), it may be predicted that the interference of background noise on a reading task will not depend on level.

5.3 Methods

5.3.1 Overview and rationale

Participants took part in two separate tasks; a listening task and a reading task. The listening task involved similar methods as per experiment 2 and tested participants' ability to listen to and identify sentences but in the presence of four different backgrounds instead of two (as in experiment 2) presented diotically over headphones (see stimuli section 5.3.4 below for details). These four backgrounds were chosen to encompass features of both 'energetic' and 'informational' masking as previously defined, but in suspected differing amounts. Target sentences were the same as in experiment 2 and as per experiment 2 the proportion of correctly identified sentences was measured at different chosen signal to noise ratios (SNRs) to estimate the shape of the psychometric functions.

The reading task involved participants reading silently sentences displayed visually on a computer monitor. Sentences were read in the presence of the same four different backgrounds presented diotically over headphones, and additionally in the presence of a no noise condition. During reading, eye movements were tracked and various measures were recorded to determine reading fluency in each masker condition. Participants were also required to answer simple yes/no comprehension questions which occurred randomly throughout the reading task in order to keep participants on task. It was originally planned to match the two tasks using the same response requirements whereby participants repeated back what they read (as with the listening task), but owing to the sensitivity of the eye tracking equipment to movements (that would be incurred through speaking aloud) this idea was abandoned.

As with experiment 1 and 2, the same vocabulary test was also carried out to ensure appropriate vocabulary range. Additionally a test of word reading efficiency was also carried out to ensure that reading ability was appropriate for the reading stimuli.

5.3.2 **Participants**

Ethics and research governance approval was obtained from the Institute of Sound & Vibration Research Human Experimentation Safety and Ethics Committee before commencing this experiment (see Appendix A for safety and ethics approval emails).

Thirty normal hearing, native English speaking adults (15 males, 15 females) aged 19 years 4 months to 35 years 7 months (average age 25 years 3 months), with normal or corrected-tonormal vision with no known special educational needs and no known reading disabilities were recruited from students and friends of the University of Southampton (see Appendix B for consent forms and questionnaires). All 30 adults were deemed otologically normal, determined from otological health questionnaires filled out by participating adults, so all thirty adults went on to complete and pass the basic hearing screen. A basic hearing screen (pure tone audiometry with circumaural earphones in accordance with ISO 389-8:2004) was carried out. Only if the participant passed the screen (and reliably responded to 20 dB HL at 1-4 kHz and 30 dB HL at 500 Hz) in both ears did he or she carry on with the study. All 30 adults passed the hearing screen and so participated in this study. Adult participants were paid £10 for their participation.

5.3.3 **Equipment**

The equipment for the listening task was exactly the same as for experiment 2. See Chapter 4 section 4.3.3 for details.

For the reading task, an SR-Research 1000 eye tracker, with 1000 Hz sampling rate was used to record eye movements. A 21” ViewSonic P227F CRT monitor with a resolution of 1024x768 pixels and a refresh rate of 1000 Hz was used to present the reading stimuli. A custom written code in SR-Research Experiment Builder controlled the entire procedure for the reading task and controlled the presentation of the masker stimuli. A head and chin rest, placed 71 cm from the monitor, ensured participants had restricted movement of their head throughout testing. Calibration of eye movements was carried out using a three-point calibration procedure which was only classed as acceptable if the mean calibration error was less than 0.5 degrees of visual angle. Also a drift correct occurred at the start of each trial. Participants responded to comprehension questions via a handheld button box connected via the USB port. As with the listening task, the same Sennheiser HDA 200 cicumaural headphones were used to deliver sound diotically. The level of all stimuli through the headphones was calibrated using a sound level meter attached to an artificial ear. This objective calibration took place at the beginning of the experimental period and once every three weeks thereafter. Subjective listening checks occurred at the start of every experimental session by the primary researcher.

5.3.4 **Stimuli**

5.3.4.1 *Target stimuli*

For the listening task, target stimuli were identical to the target stimuli in experiment 2; BKB and IHR sentences spoken by the same male talker. See Chapter 4 section 4.3.4 for details.

For the reading task, target stimuli were semantically and syntactically correct sentences 9 to 17 words long (13 words average) see Appendix F for lists of sentences. Sentences were displayed one at a time on the monitor in front of the participant. The font used was Courier font and all characters were in lower case.

5.3.4.2 *Masker stimuli*

Four masker conditions were used. These masker conditions were the same in both the listening and reading task but differed regarding the way the segments were selected to be the masker for each task (which will be detailed below). The reading task alone also included an additional no-noise condition. A no-noise condition was not included in the listening task as this would have led to ceiling effects and thus unnecessarily increased test time.

In condition one a steady-state noise was used (as in experiment 1 and 2) whereby the spectrum had been adjusted to match the long term average speech spectrum of the speech targets from the listening task.

Condition two consisted of single-talker speech sentences used as the masker. The sentences were taken from the EUROM sentence database (Chan et al. 1995) which consists of different speakers reading different meaningful and grammatically correct passages. Differing to experiment 2, male instead of female talkers were chosen to investigate the masking effects of same sex targets and maskers (applying only to the listening task). In order to have enough unique speech material from which to create the speech maskers, two specific male talkers were chosen because they provided the greatest number of unique passages. Each talker had 10 unique recorded passages approximately 21 seconds in duration. The rms level of each passage was then equalised to that of the speech-shaped noise and then for each distinct talker the passages were concatenated end to end in MATLAB, and pauses between words were deleted if more than 100 ms. This produced one continuous stream of male talker A and B. For the listening task half the participants were tested with male talker A and half with male talker B for the single-talker condition. For the reading experiment however, only male talker A was used for the single-talker condition for consistency.

Condition three consisted of two-talker speech sentences used as the masker. These sentences were a combination of male talker A and B chosen for condition one. Again the rms level of each passage was equalised to that of the speech-shaped noise and the passages were concatenated and pauses longer than 100 ms were removed. Each stream was then laid down on top of each other, and then truncated so the longest stream was the same length as the shortest one.

Condition four consisted of multi-talker babble used as the masker. These sentences included those from male talker A and B (from condition 1 and condition 2) but also added in 14 additional male talkers speaking sentences. The multi-talker condition thus consisted of 16 male talkers. In the same way as for condition three, each of the sentences for the each of the 16 talkers were concatenated end to end in MATLAB and pauses longer than 100 ms were removed. Then each of the 16 streams were laid down on top of each other and streams were truncated so the longest one was the same length as the shortest one.

For all four maskers the spectrum was then equalised to that of the 16 talker babble and the rms level of each masker was again equalised to that of the speech-shaped noise. For the listening task, for each target sentence, a segment of the masker stream was selected at random and set so the target sentence began 0.5 s after masker onset (as in experiment 2). Both maskers were ramped on and off slowly using a raised-cosine ramp with 100 ms duration. All stimuli were obtained digitally and converted to analogue form using a 44100 Hz sampling rate via a 16 bit digital to analogue converter.

For the listening task, for each of the four masker conditions, five SNRs were tested chosen to estimate the shape of the psychometric functions. Therefore there were 20 conditions. The SNRs chosen were specific to each masker type, and were based on previous research and pilot testing (see section 5.3.8 below). As in experiment 2, the level of the both the target and masker was then varied to generate required SNRs. A decrease in SNR resulted in a reduction in intensity level of the signal together with an increase in intensity level of the masker to meet the required SNR and vice versa. Both the target and masker were varied in this way to prevent the delivered sound becoming too loud for the participants, as per experiment 1 and 2. When the SNR was set at 0 dB SPL the level of the target and masker was at 60 dB SPL.

For the reading task, the required number of segments of the masker stream were selected in advance prior to testing and saved in wav files ready to be selected at random and presented together with a target sentence also chosen at random. This was because it was not possible for the software to select a segment at random for each new trial. Each segment was set to be 15 s long. This length was chosen based on brief piloting and so was set in attempts to prevent the participant from running out of background noise whilst they were reading a single sentence. Because it was not possible to know how long each participant would take when reading the target sentences in the background maskers, segments were selected to be as unique as possible, to prevent participants from hearing the same background maskers which could affect the way they ignore/or may be distracted by the sound. There was however some overlap, owing to the limited availability of background male talker A's speech material, but it was thought participants would not need the whole 15 seconds to read the sentence. It was decided to only include one male talker for the single talker speech condition so to not include too many variables which may complicate things when interpreting and analysing the results. The onset of the background sound and the sentence stimuli was set to be presented as the same time.

For the reading task the level of each of the maskers was presented either at 55 dB (A) or 75 dB (A) to investigate the effect of level of background noise on the reading process. These levels were chosen to represent quiet speech and loud speech. Therefore the reading task consisted of 9 conditions, four background maskers presented at two different levels and a no noise condition.

5.3.5 Vocabulary test

The vocabulary test carried out was the same as in experiment 1 and 2; the British Picture Vocabulary Scale (BPVS), and followed exactly the same methods.

5.3.6 Test of word reading efficiency

In order to ensure that participants had an age appropriate reading ability, the test of word reading efficiency (TOWRE) was also carried out. It evaluates both sight word reading efficiency and phoneme decoding efficiency in children (primarily) and adults. It involves the participant being shown lists of words and non-words whereby they were asked to read down each list of words aloud and as fast as they could. The number of accurately pronounced words was recorded over 45 seconds. This test lasted less than 5 minutes.

5.3.7 Structure

5.3.7.1 *Overall Structure*

Testing was carried out in a single session which lasted approximately 90 minutes in total. Half the participants underwent the listening task first and half underwent the reading task.

At the start of the session participants underwent the basic hearing screening test, the BPVS and the TOWRE which together lasted approximately 20 minutes. For the listening task, testing lasted approximately 40 minutes. The proportion of sentences scored correctly (out of 15) was measured for each of the four masker conditions at each of the five SNRs. Participants were seated in a chair at a table in a quiet room within the University. Participants were instructed to ignore the background sounds and listen to the target male, then repeat back as much of what they heard. Sentences were scored correctly depending on how many keywords were repeated correctly. The structure of the listening task was the same as for experiment 2 and the order of the SNRs were presented in the same way alternating “easier” then “harder” SNRs as opposed to successively decreasing SNRs (see chapter 4 section 4.3.6.1 for details).

Fifteen sentences were presented for each of the four listening conditions for five SNRs, therefore a total of 300 sentences were tested. The orders of the background noise conditions were counterbalanced in attempts to counter any order effects.

For the reading task testing lasted approximately 20-30 minutes. Eye movements were recorded during each of the four masker conditions at 55 dB (A) and 75 dB (A) and a no noise condition. Participants were seated in a chair at a table in a dimly lit room to capture a clear visualisation of the pupil. Participants were seated facing a monitor with their head stabilised via a head and chin rest at a distance of 71 cm from the monitor. This was necessary to restrict head movements that could interfere with calibration of the eye movements. Participants were instructed to read each sentence silently for comprehension and to ignore the background sound. They were told they would be asked a comprehension question which would occur randomly, in order to keep them on task. Comprehension questions occurred after 25% of the trials whereby the participant had to answer a yes/no questions about the sentence they just read. Because the movement involved with talking would interfere with the calibration of eye movements, participants responded via a handheld button box, pressing the left hand button to answer “yes” and the right hand button to answer “no”. Participants were not provided with any feedback. Initial calibration of the eye movements required the participant to follow a dot on the screen with their eyes; this took approximately 1 minute and occurred at the start and midway through the experiment, and any time throughout testing where necessary. After each sentence the participant pressed any button on the button box to move onto the next sentence. For each masker condition, eight sentences were presented. Because there were 8 sentences per condition and 9 conditions, there were 72 trials in total. The background conditions were completely randomised throughout the experiment from one trial to the next, so the listener would not know which condition to expect next. Participants viewed the sentences with both eyes but recordings were only measured from the right eye.

5.3.7.2 *Familiarisation*

For the listening task, the same familiarisation procedure was followed as for the experiment 2 (see Chapter 4 section 4.3.6.2) for each of the four maskers.

For the reading task, 5 practice sentences were given at the start which included each of the four background noise conditions presented at either 55 dB (A) or 75 dB (A), and a no noise trial. Comprehension questions occurred after 3 of these trials. This practice period was given in order to enable the participant to become familiar with the procedure regarding the target and background stimuli, sitting in the head and chin rest and using the button box to respond and move on to the next trials. Results from these familiarisation trials were not analysed.

5.3.8 Selecting the SNRs for the speech intelligibility task

5.3.8.1 *Selecting the SNRs based on previous research*

For the listening task, in order to choose an array of SNRs which would best capture the psychometric functions for adults with all four masker types, previous research, findings from experiment 2 and some pilot testing with adults was taken into account. Since the speechshaped noise masker and the single-talker speech masker (although an opposite sex masker) had been used in experiment 2, this previous experiment informed the chosen SNRs for those conditions.

Table 5.1: Signal to noise ratios and proportion correct recorded from previous studies which have used target sentence material and two-talker speech maskers with adults.

Adults – Two-talker				
Proportion correct (%)	**SNR (dB)**	**Study**	**Sample**	**Methods**
50% (approx.)	0 dB (approx.)	Freyman et al. (2004)	8 adults	Nonsense sentences, fixed SNRs, female target female masker
56%	-2 dB	Rosen et al. (2013)	16 adults	IEEE sentences, fixed SNRs, male target male masker
14% (approx.)	-4 dB (approx.)	Freyman et al. (2004)	8 adults	Nonsense sentences, fixed SNRs, female target female masker
8%	-6 dB	Rosen et al. (2013)	16 adults	IEEE sentences, fixed SNRs, male target male masker
4% (approx.)	-8 dB (approx.)	Freyman et al. (2004)	8 adults	Nonsense sentences, fixed SNRs, female target female masker
0% (approx.)	-12 dB (approx.)	Freyman et al. (2004)	8 adults	Nonsense sentences, fixed SNRs, female target female masker

Table 5.2: Signal to noise ratios and proportion correct recorded from previous studies which have used target sentence material and multi-talker speech masker with adults.

Adults – Multi-talker babble				
Proportion correct (%)	**SNR (dB)**	**Study**	**Sample**	**Methods**
70% (approx.)	0 dB (approx.)	Freyman et al. (2004)	8 adults	Nonsense sentences, fixed SNRs, female target female masker (10 talkers)
62%	-2 dB	Rosen et al. (2013)	16 adults	IEEE sentences, fixed SNRs, male target male masker (16 talkers)
55% (approx.)	-2 dB (approx.)	Freyman et al. (2004)	8 adults	Nonsense sentences, fixed SNRs, female target female masker (10 talkers)
34% (approx.)	-4 dB (approx.)	Freyman et al. (2004)	8 adults	Nonsense sentences, fixed SNRs, female target female masker (10 talkers)
13%	-6 dB	Rosen et al. (2013)	16 adults	IEEE sentences, fixed SNRs, male target male masker (16 talkers)
2% (approx.)	-8 dB (approx.)	Freyman et al. (2004)	8 adults	Nonsense sentences, fixed SNRs, female target female masker (10 talkers)

Previous research was only considered if the target material consisted of sentences, and if the masker material consisted of a two-talker of multi-talker babble of the same gender as the target sentence material. Table 5.1 and table 5.2 show the SNRs and their corresponding intelligibility levels from previous studies which have used a two-talker and multi-talker masker respectively. Whilst these previous results provide insight into which SNRs to choose, the test materials do differ from the current study. Therefore the result from pilot testing with these maskers was taken into consideration.

5.3.8.2 *Selecting the SNRs based on pilot results*

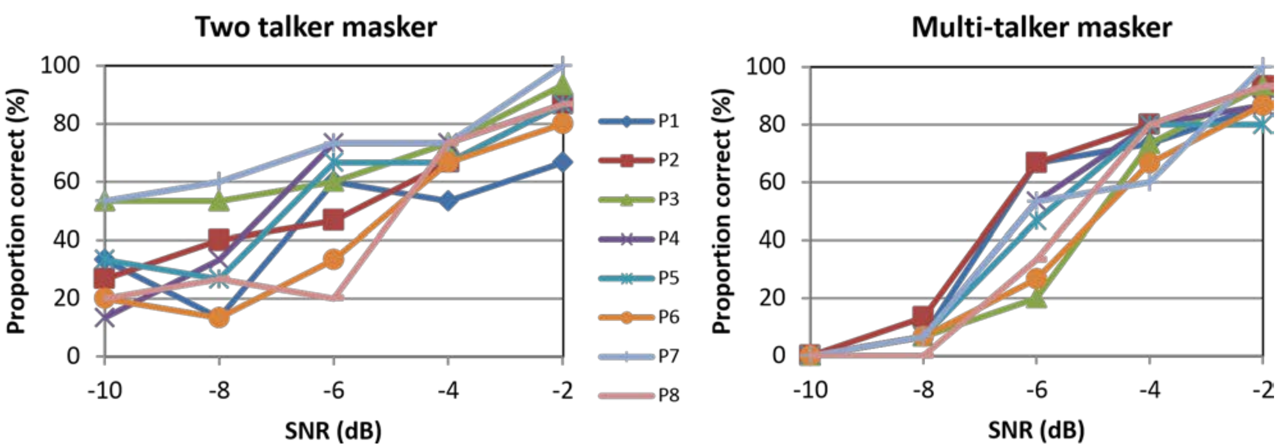


Figure 5.1: Proportion correct recorded at 5 fixed SNRs for each individual adult pilot participant. Different coloured lines represent results from different participants labelled P1-P8.

Considering the previous research with target sentence material in the two-talker and multitalker babble masker types the pilot SNRs were chosen. Eight normal hearing adults took part in the pilot testing in order to determine which SNRs would be appropriate to use within the main experimental sample that would best capture the shape of the psychometric function. The tested SNRs were set to -2, -4, -6, -8 and -10 for both masker types and results are displayed in figure 5.1. The results for the multi-talker masker appear to show a fair representation of the psychometric function however -10 dB SNR displays floor effects. The results for the two talker masker show much greater variation in results from person to person, however the tested SNRs appear to capture the psychometric function slope.

Taking into account the previous research, pilot testing and results from experiment 2, the final SNRs for adults in each masker type were chosen as detailed in table 5.2. The SNRs for the speech shaped noise masker were changed slightly from those tested in experiment 2 since poor performance was observed at -14 dB, so this SNR was removed and the interval between consecutive SNRs was decreased from 3 dB to 2 dB in attempts to provide improved psychometric function fits. For the single-talker masker condition, the SNRs were kept the same as those used in experiment 2, despite now being a male target male masker mix (as opposed to the male target female masker mix in experiment 2). The reason behind keeping the SNRs the same was because even at -28 dB in experiment 2, intelligibility was well above floor level so it was thought that there would be room for performance to get worse with a same sex target and masker and that those SNRs would still be able to capture the full psychometric function. SNRs were selected so that they had the same dB interval between consecutive SNRs per masker type.

Table 5.3: The chosen signal to noise ratios for each masker type

Masker	**SNRs (dB) from low to high**				
Single-talker speech	-28	-23	-18	-13	-8
Two-talker speech	-10	-8	-6	-4	-2
Multi-talker speech	-8	-6	-4	-2	0
Speech-shaped noise	-12	-10	-8	-6	-4

5.4 Results

5.4.1 Overview of results

The BPVS and the TOWRE were carried out to determine the vocabulary range and word reading efficiency of participants. Each participant's score was compared to a standardised score; the results of these tests will be presented first.

For the listening task, the proportion of correctly identified sentences was determined within a speech-shaped noise masker, a single-talker speech masker, a two-talker speech masker and a 16-talker babble masker. Speech intelligibility was measured in each of these conditions at five various fixed SNRs. Logistic functions were then fitted to individual measurements from each masker to check goodness of fit before psychometric function parameters were averaged to show average results across participants.

For the reading task, eye movements were measured during sentence reading within the same four background conditions at two different levels, plus a no noise condition. Eye movement data was then cleaned before being analysed.

5.4.2 Vocabulary test

Table 5.4: The mean, range and standard deviation of the standardised scores from the British Picture Vocabulary scale in adult participants.

	Adults (n=30)
Mean	142
Range	107-160
Standard deviation	16

The results of the British picture vocabulary scale (BPVS) are displayed in table 5.4. Results are shown from the all thirty adults. To recap, the BPVS calculates standardised scores based on age population means, only up to 15.8 years of age. The mean standardised score of adults are displayed in Table 5.4. For adults, the standardised scores have been based on a child aged 15.8 years.

As can be seen from the table, the mean adult's standardisation score is well above the population mean for a 15 year old. This result is not surprising, but it is however reassuring that the adults did not display any obvious vocabulary deficits with this test that may have interfered with the speech perception or reading tests.

5.4.3 Test of word reading efficiency

Table 5.5: The mean, range and standard deviation of the Total Word Reading efficiency Standardised Scores from the Test of Word Reading Efficiency in adult participants.

	Adults (n=30)
Mean	106
Range	90-120
Standard deviation	9

The results of the TOWRE are displayed in table 5.5. Results are shown from the all thirty adults. To recap the TOWRE is standardised based on samples of participants aged between 624 years. For all participants in the present study the standardised score was calculated based on the age bracket of 17 years to 24 years 11 months. Most of the participants within this study were within this age range yet some were slightly older than this. As can be seen from the table, the means adult's standardisation score lies above the population mean and all individual results were either deemed "average" or "above average" according to the examiner's manual (Torgesen et al. 1999). This again is reassuring that the adults did not display any obvious reading difficulties that may have interfered with the reading test.

5.4.4 Effect of background on speech intelligibility

5.4.4.1 *Goodness of fit*

Psychometric functions were fitted to the individual results from all 30 participants across all five SNRs for each of the four masker conditions. Unlike in experiment 2, a repeat was not included in this experiment so reliability of identifying the BKB/IHR sentences within the four background maskers was not investigated. This was largely because reliability had been examined in experiment 2 and also due to time constraints for data collection. In the same way as for experiment 2, goodness of fit calculations were applied to all psychometric function fits for each participant to ensure only good fits were included in further analysis (see Chapter 4 section 4.4.3 for more details). Results from each acceptable psychometric function fit parameters were then averaged to produce a single psychometric function for each masker type. All participants' fitted psychometric functions for each masker type were deemed acceptable fits as the p-value on the deviance measured for each fit was more than 0.05. The poorest fit deviance had a p-value of 0.06 (see Appendix E for tables of all the goodness of fit deviance and p-values). Therefore results from all participants were included in further analyses.

5.4.4.2 *Psychometric functions*

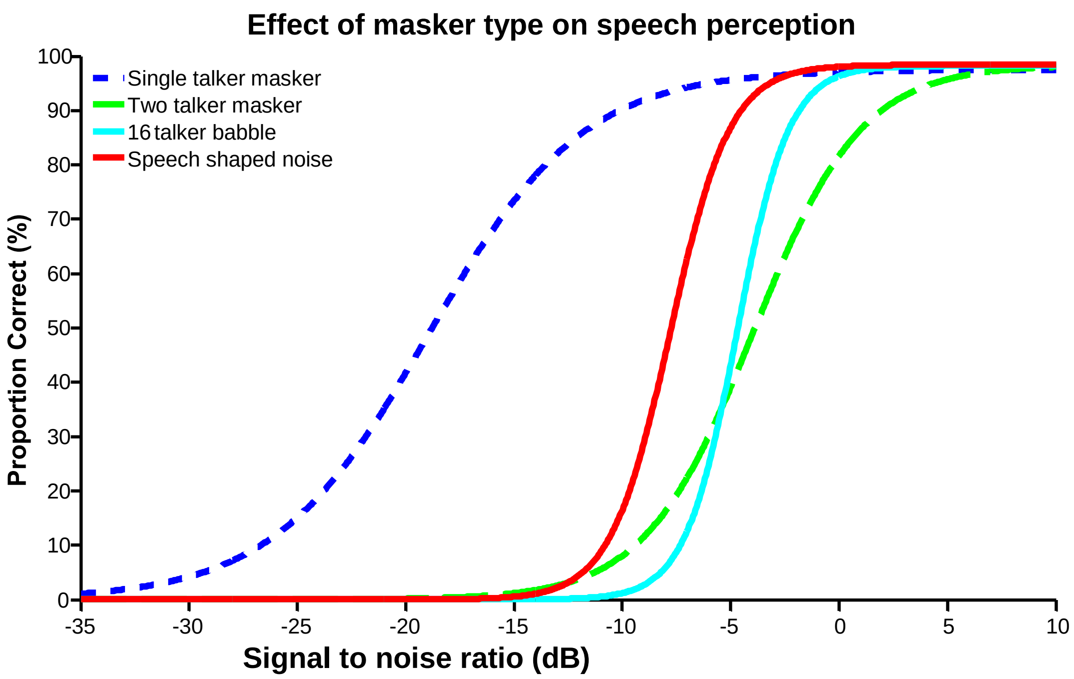


Figure 5.2: Mean fitted psychometric function parameters for speech intelligibility in four masker types in adults.

Figure 5.2 displays the average results from all 30 adults who participated in this experiment. The four lines on the graph show the mean psychometric functions as calculated by averaging individually fit parameters across all 30 participants. The dark blue line shows the results with the single-talker speech masker. The green line shows the results with the two-talker speech masker. The turquoise line shows the results with the multi-talker masker and the red line shows the results with the speech-shaped noise masker. With the results for each masker type displayed on the same graph, it can be seen that the slopes are shallower with the single and two-talker maskers and steeper with the babble and noise maskers.

The results with the single-talker masker are shifted to the far left showing that speech can be identified in this condition in very low (poor) signal to noise ratios. The signal to noise ratios must be increased in the speech-shaped noise conditions to reach comparable intelligibility levels as the single-talker masker, and must be increased further still with the multi-talker babble masker. The two-talker masker appears to require the highest (best) SNRs to achieve comparable intelligibility levels; however owing to some overlap with the babble masker, this depends on which intelligibility level you are interested in.

To further investigate the effects of masker type on speech perception a one-way repeated measures ANOVA was carried out comparing the location parameter threshold (50% intelligibility level) for the four background masker conditions. Mauchly's test indicated that the assumption of sphericity had been violated for the main effects of masker type $\chi^2(5) = 29.36$, $p < .001$. Therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .72$).There was a significant main effect of masker on the location parameter $F(2.15, 62.33) = 821.06$, $p < .001$. Post hoc Bonferroni adjusted pairwise comparisons, revealed that the location parameter threshold was significantly lower (improved) with the single-talker masker (M = -18.97, SE = .42) compared to the two-talker masker (M = -3.97, SE = .32) ($p < .001$), multi-talker masker (M = -4.72, SE = .17) and noise masker (M = -7.79, SE = .18). The location parameter threshold with the noise masker was significantly lower (improved) than that with the multi talker masker ($p < .001$) and the two talker masker ($p < .001$). There was however, no significant difference between the location parameter threshold with the two-talker masker and the multi-talker masker ($p = .06$) suggesting similar SNRs are required to reach comparable intelligibility levels. See Appendix C for ANOVA tables.

A one way repeated measures ANOVA was also carried out on the slope parameter of the psychometric functions for the four background masker conditions. The ANOVA was carried out on the log transformed slope parameter in order to reduce the skew in the data. There was a significant main effect of masker on the slope parameter $F(3, 87) = 49.73$, $p < .001$. Post hoc Bonferroni adjusted pairwise comparisons, revealed that the slope parameter was significantly shallower with the single-talker masker (M = 0.28, SE = .02) compared to the twotalker masker (M = 0.40, SE = .03) ($p < .01$), multi-talker masker (M = 0.83, SE = .04) ($p < .001$) and noise masker (M = 0.73, SE = .02) ($p < .001$), and the trend seemed to show an increase in steepness of slope as more talkers were added to the speech masker mix. The slope parameter with the noise masker was not significantly different from the slope with the multi-talker masker ($p = 1.00$) but was significantly steeper than the slope with the two-talker masker (p < .001). Lastly, the slope with the two-talker and multi-talker maskers was significantly different (p < .001) being shallower with the two-talker masker compared to the multi-talker masker. This finding suggests that the two-talker and multi-talker maskers do affect intelligibility differently but above and below the location parameter.

5.4.4.3 *Comparing results to previous results*

In order to place these results from the listening task in the context of previous research, it was decided to compare the results from the current study with those from the Rosen et al. (2013) study on the same graph. To be specific, Rosen et al. (2013) investigated the intelligibility of IEEE sentences spoken by a male talker (which are longer compared to the BKB sentences used in the current study) amongst a background of EUROM sentences also spoken by a male talker (which were taken from the same database of speech maskers used in the present study) at two fixed SNRs (-2 and -6 dB). These findings are plotted in boxplots in figure 5.3.

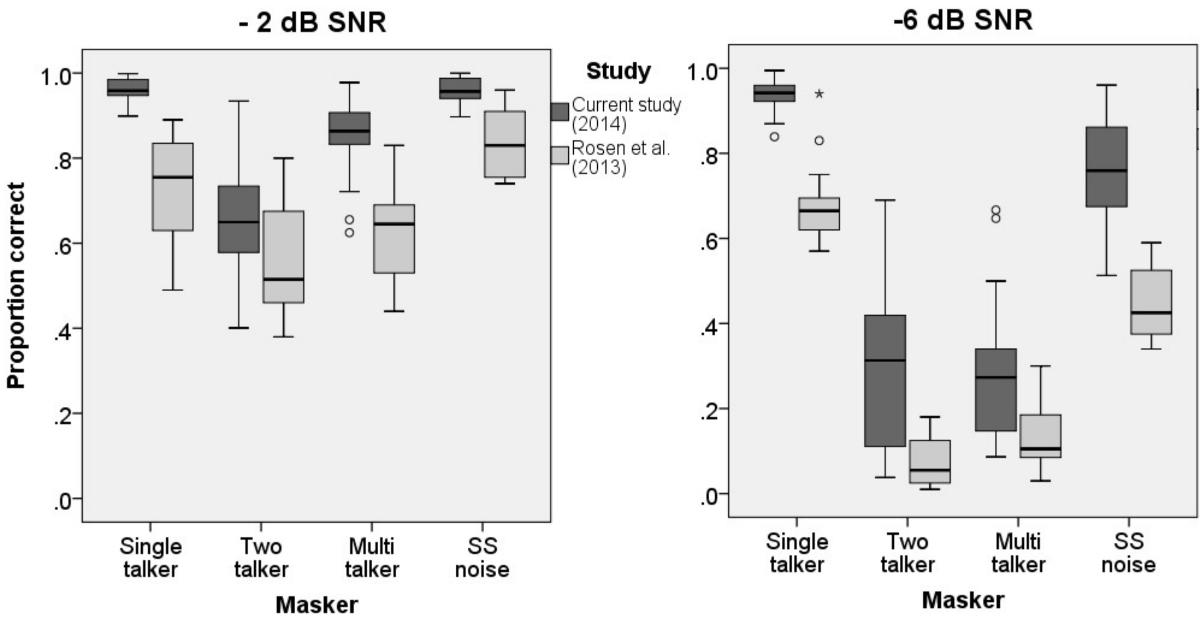


Figure 5.3: Boxplots to show the range of results from the current study and from Rosen et al. (2013) at the two SNRs tested in the Rosen et al. (2013) study with permission. The dark grey boxes denote the results from the current study and the pale grey boxed denote the results from the Rosen et al. (2012) study.

Since Rosen et al. (2013) examined intelligibility at two SNRs (-2 and -6 dB), and these SNRs were not tested in every masker condition in the current study, proportion correct was determined at these SNRs by reading off each fitted psychometric function. With the data from Rosen et al. (2013) at -2 dB SNR it can be noticed that performance seems poorest with the two-talker masker compared to the other maskers, which is also the trend within the current study. Performance looks to improve as you move from a two-talker to multi-talker masker and performance with the single-talker and noise maskers are the best. At -6 dB SNR it can be seen that performance appears poorest with the two-talker and multi-talker maskers and best with the single-talker masker compared to the noise masker.

It is apparent that the results within the current study appear to show better performance overall compared to the findings from Rosen et al. (2013). The reasons for this are not clear but it is possible these differences occur due to the different target speech material and perhaps differences in fundamental frequency between the target speech and their chosen masker voices. Also the target sentences used in the Rosen et al. (2013) study are more complex than the simple BKB and IHR sentences used in the current study. Furthermore differences may exist because the proportion correct at each SNR in the current study is extrapolated from the psychometric functions whereas in the Rosen et al. (2013) study they are measured at fixed SNRs.

5.4.5 Effect of masker type on reading

Eye movement data from the reading task were analysed in terms of global measures. These measures investigate the fixations and saccades that occur throughout each trial as a whole, i.e. across the whole sentence as opposed to specific words within the sentence to investigate how background noise affects the reading process in general.

The five global eye movement measures investigated were number of fixations (the number of times eyes stopped on parts of the text), mean fixation durations (the average length of time spent with eyes stopped on parts of the text), total fixation time (the total time spent with eyes stopped on parts of the text), proportion of regressions (backward eye movements to reread parts of the test) and saccade amplitude (the length of separation between one fixation to the next). Such eye movement measures can reveal processing difficulties (Rayner 2009). The number of fixations may increase as readers may have to look at several parts of the text more than once and without skipping portions. The duration of recorded fixations may increase as readers may have to look longer at parts of the text. Total fixation time may thus increase. The proportion of regressions may increase as readers may have to re-read parts of the text. Saccade amplitude may decrease as the length of separation between one fixation to the next may be reduced so not to skip any portions of the text (Rayner, 2009). Reading comprehension was also investigated.

Two-way repeated measures ANOVAs were carried out to examine the main effects of background masker (single-talker, two-talker, multi-talker and speech-shaped noise) and background level (55 dB (A) and 75 dB (A)). For each measure there were no significant main effects or interactions including background level. Therefore, the results across levels were pooled and the no noise condition was included in subsequent one-way repeated measured ANOVAs to investigate effect of background (single-talker, two-talker, multi-talker, speechshaped noise and no noise). Significant main effects were explored through t-tests Bonferroni adjusted for multiple comparisons.

5.4.5.1 *Data preparation*

After the eye tracker was calibrated for each participant, the position of the corneal reflection (the reflection of the infrared light beamed from the camera) was tracked. This was to investigate whereabouts within the sentence the participant looked (fixated) at any given time and for each if individual fixation, how long they fixated for. Fixation and saccadic data was recorded for each trial (i.e. for each sentence read). Average results were then calculated across sentences within the same background condition. Fixations that coincided with the display onset were removed as standard practice. Fixations which coincided with a button press were also removed. Fixations <80 ms or >1200 ms were classed as outliers and too removed. This resulted in 2% of the fixations being removed. The final data set consisted of 22646 fixations in total.

5.4.5.2 *Reading comprehension*

The mean proportion of correct responses to comprehension questions in each of the five background conditions are displayed in table 5.6. The mean accuracy of answering the comprehension questions across each background was above 85% which is deemed an acceptable level. There was found to be no significant main effect of background on mean accuracy, F (4, 116) = .40, p = .81. These results indicate that there were no differences in reading comprehension across each of the five background conditions.

Table 5.6: The mean proprtion of comprehension questions answered correctly and the standard deviation in each of the five different backgrounds.

All participants

BACKGROUND	MEAN	STANDARD DEVIATION
Single-talker	0.89	0.22
Two-talker	0.90	0.13
Multi-talker	0.93	0.19
Speech-shaped noise	0.89	0.13
No noise	0.86	0.25

5.4.5.3 *Number of fixations*

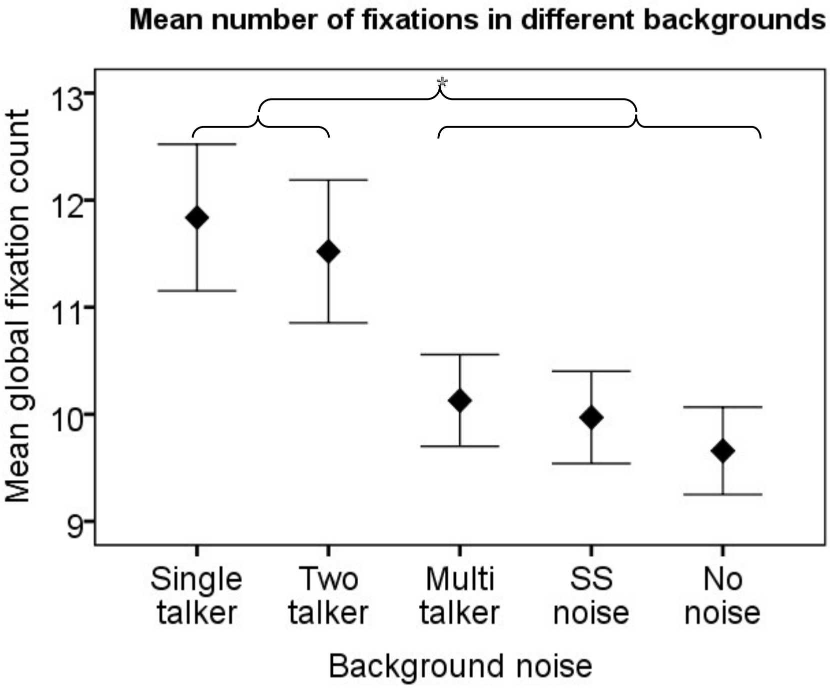


Figure 5.4: The mean number of fixations across both levels when reading sentences in five different backgrounds. The error bars denote ± 1 standard error. The brackets and asterisk represent significant differences.

Table 5.7: The mean number of fixations and standard deviation in each of the five different backgrounds.

All participants

BACKGROUND	MEAN (n)	STANDARD DEVIATION
Single-talker	11.84	3.75

Two-talker	11.52	3.65
Multi-talker	10.13	2.35
Speech-shaped noise	9.97	2.36
No noise	9.66	2.23

The global fixation count is the mean number of fixations across participants in each of the five background conditions, the results of which are displayed in figure 5.4 and table 5.7. Mauchly's test indicated that the assumption of sphericity had been violated for the main effects of background $\chi^2(9) = 70.41$, $p < .001$. Therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .45$). There was a significant main effect of background on global fixation count, $F(1.80, 52.07) = 12.38$, $p < .001$. Compared to the no noise condition (M = 9.66, SE = .41), the single-talker condition (M = 11.84, SE = .69) and twotalker condition (M = 11.52, SE = .67) yielded a significantly higher global fixation count ($p = .01$, $p = .02$ respectively). There was no significant difference found between the single-talker and two-talker conditions ($p = 1.00$). Compared to the no noise condition, the multi-talker condition (M = 10.13, SE = .43) and speech-shaped noise condition (M = 9.97, SE = .43) were not significantly different ($p = .79$, $p = 1.00$ respectively). There was also no significant difference found between the multi-talker and speech-shaped noise conditions ($p = 1.00$). A significant difference was found between the single-talker and multi-talker conditions ($p = .01$) and single-talker and speech-shaped noise conditions ($p < .01$). A significant difference was also found between the two-talker and multi-talker conditions ($p = .04$) and two-talker and speech-shaped noise conditions ($p < .01$). See Appendix C for ANOVA tables.

These results indicate that the number of fixations during reading significantly increased from the baseline no noise condition with single and two-talker backgrounds but not with multitalker and speech-shaped noise backgrounds. Across all four backgrounds the single-talker and two-talker backgrounds were the most disruptive to the reading task and there were no significant differences in how a multi-talker and speech-shaped noise background affected the reading process.

5.4.5.4 *Fixation durations*

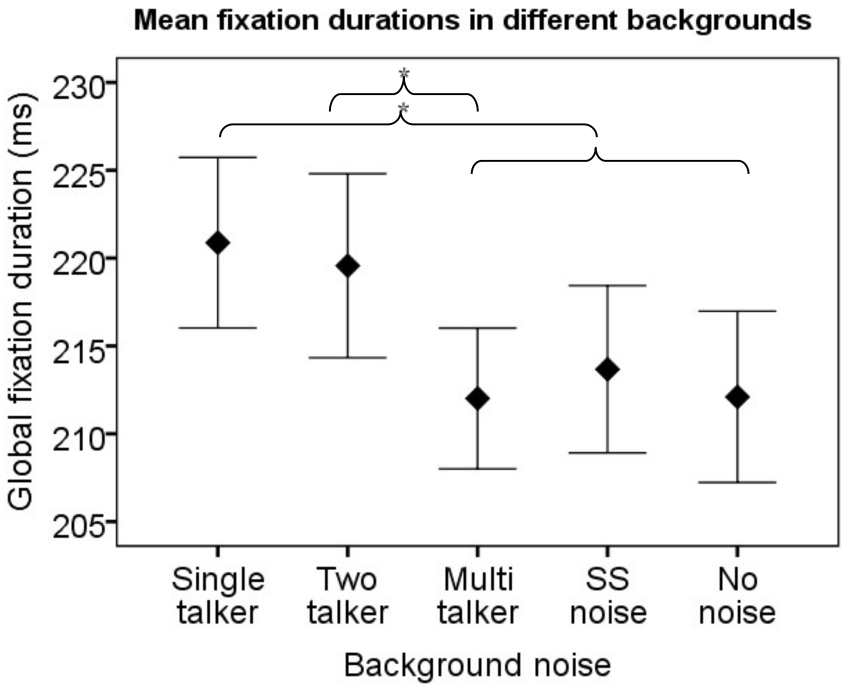


Figure 5.5: The mean global fixation durations across both levels when reading sentences in five different backgrounds. The error bars denote ± 1 standard error. The brackets and asterisks represent significant differences.

Table 5.8: The mean global fixation duration and standard deviation in each of the five different backgrounds.

All participants		
BACKGROUND	MEAN (ms)	STANDARD DEVIATION
Single-talker	220.88	26.59
Two-talker	219.56	26.69
Multi-talker	212.01	21.95
Speech-shaped noise	213.67	26.07
No noise	212.10	26.70

The global fixation durations are the mean fixations durations across participants in each of the five background conditions, the results of which are displayed in figure 5.5 and table 5.8. Mauchly's test indicated that the assumption of sphericity had been violated for the main effects of background $\chi^2(9) = 20.33$, $p = .02$. Therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .74$).There was a significant main effect of background on global fixation duration, $F(2.98, 86.33) = 8.23$, $p < .001$. Compared to the no noise condition (M = 212.10, SE = 4.87), the single-talker condition (M = 220.88, SE = 4.85) yielded a significantly higher global fixation duration ($p = .01$). Although the two-talker condition (M = 219.56, SE = 5.24), produced a higher mean global fixation duration compared to the no noise condition this difference was found not to be statistically significant ($p = .10$). The single-talker and two-talker conditions were not significantly different ($p = 1.00$). The multi-talker condition (M = 212.01, SE = 4.76) and the speech-shaped noise condition (M = 213.67, SE = 4.76) also were not significantly different from the no noise condition ($p = 1.00$ for both comparisons) or each other ($p = 1.00$). The two-talker condition gave significantly longer fixation durations compared to the multi-talker condition ($p = .02$) but was not significantly different from the speech-shaped noise condition ($p = .14$). The single-talker condition gave significantly longer fixation durations compared to the multi-talker ($p < .01$) and speechshaped noise conditions ($p < .01$). See Appendix C for ANOVA tables.

These results indicate that the mean fixations durations significantly increased from the baseline no noise condition with only a single-talker background but not with any other backgrounds. There were however increased mean fixation durations with the two-talker masker but this increase was not significant. Across all four background maskers the singletalker was significantly more disruptive to the reading task than all other maskers except the two-talker masker, and the two talker-masker was more disruptive than the multi-talker masker but not the speech-shaped noise masker.

5.4.5.5 *Total fixation time*

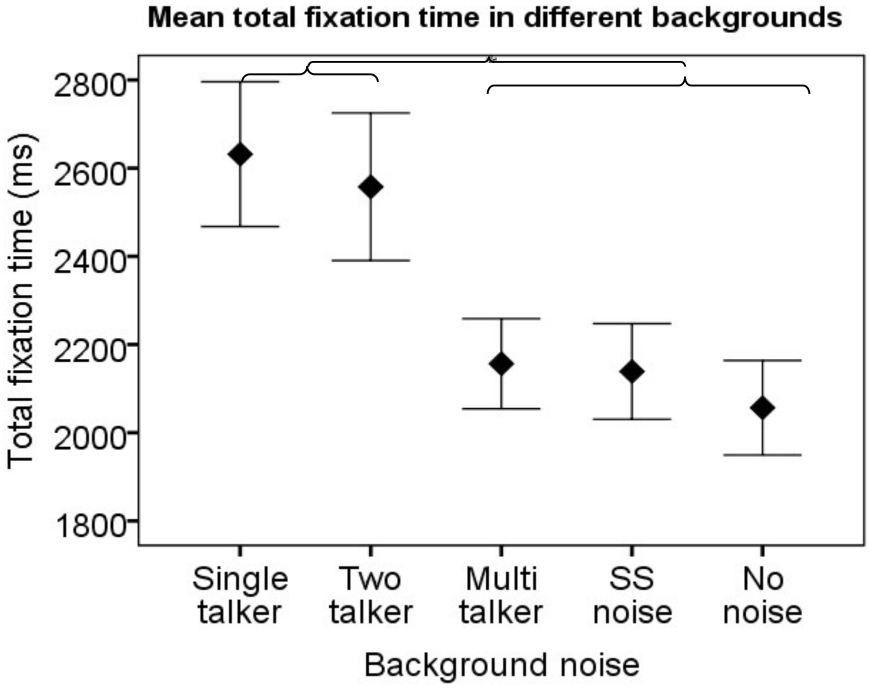


Figure 5.6: The mean total fixations time across both levels when reading sentences in five different backgrounds. The error bars denote ± 1 standard error. The brackets and asterisks represent significant differences.

Table 5.9: The mean total fixation times and standard deviations in of the five different backgrounds.

All participants		
BACKGROUND	MEAN	STANDARD DEVIATION
Single-talker	2631.76	898.79
Two-talker	2557.72	915.61
Multi-talker	2156.50	559.37
Speech-shaped noise	2138.90	593.87
No noise	2056.63	587.79

The mean total fixation time is the total time spent fixating across participants in each of the five background conditions, the results of which are displayed in figure 5.6 and table 5.9. Thus mean total fixation time is the number of fixations multiplied by the mean fixation durations and explains the total time spent reading (as information is only acquired during fixations). Mauchly's test indicated that the assumption of sphericity had been violated for the main effects of background $\chi^2(9) = 65.11$, $p < .001$. Therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .46$).There was a significant main effect of background on total fixation time, $F(1.83, 53.11) = 15.37$, $p < .001$. Compared to the no noise condition (M = 2056.63, SE = 101.32), the single-talker condition (M = 2631.76, SE = 164.10) and the two-talker condition (M = 2557.717, SE = 167.35) yielded a significantly longer total fixation time ($p < .01$, $p = .01$ respectively).There was no significant difference found between the single-talker and two-talker conditions ($p = 1.00$). Compared to the no noise condition, the multi-talker condition (M = 2156.50, SE = 102.13) and speech-shaped noise condition (M = 2138.90, SE = 108.43) were not significantly difference ($p = 1.00$ for both comparisons). There was no significant difference found between the multi-talker and speechshaped noise conditions ($p = 1.00$). A significant difference was found between the singletalker and multi-talker conditions ($p < .01$) and single-talker and speech-shaped noise conditions ($p < .001$). A significant different was also found between the two-talker and multitalker conditions ($p = .01$) and two-talker and speech-shaped noise conditions ($p < .01$). See Appendix C for ANOVA tables.

These results indicate that the total time spent reading significantly increased from the baseline no noise condition with single and two-talker backgrounds but not with multi-talker and speech-shaped noise backgrounds. Across all four backgrounds the single-talker and twotalker backgrounds were the most disruptive to the reading task and there were no significant differences in how a multi-talker and speech-shaped noise background affected the reading process.

5.4.5.6 *Proportion of regressions*

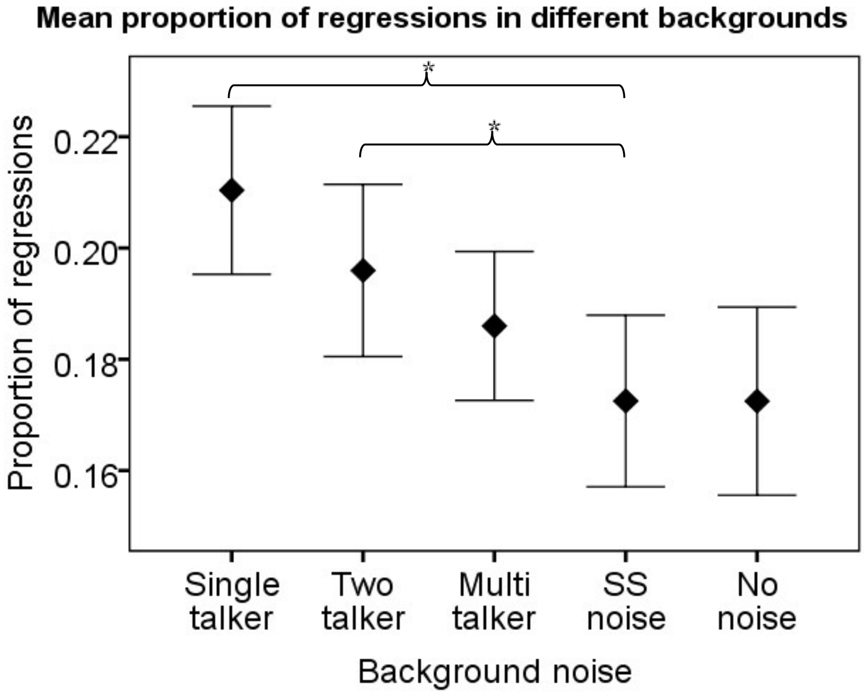


Figure 5.7: The mean proportion of regressions across levels when reading sentences in five different backgrounds. The error bars denote ± 1 standard error. The brackets and asterisks represent significant differences.

Table 5.10: The mean proportion of regressions and standard deviations in each of the five different backgrounds.

All participants		
BACKGROUND	MEAN	STANDARD DEVIATION
Single-talker	0.21	0.08
Two-talker	0.20	0.08
Multi-talker	0.19	0.07
Speech-shaped noise	0.17	0.08
No noise	0.17	0.09

The proportion of regressions describes the proportion of fixations made that occurred after backwards (right to left) eye movements across participants in each of the five background conditions, the results of which are displayed in figure 5.7 and table 5.10. Therefore this measure details how often participants went back to re-read already read text. Mauchly's test indicated that the assumption of sphericity had been violated for the main effects of background $\chi^2(9) = 37.58$, $p < .001$. Therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .57$).There was a significant main effect of masker on the proportion of regressions made, $F(2.29, 66.53) = 5.81$, $p < .01$. Compared to the no noise condition (M = .17, SE = .02), although the mean proportion of regressions were higher with the single-talker condition (M = .21, SE = .02), the two-talker condition (M = .20, SE = .02) and the multi-talker condition (M = .19, SE = .01), none of these differences were statistically significantly higher ($p = .13$, $p = .54$, $p = 1.00$ respectively). The proportion of regressions with the speech-shaped noise condition (M = .17, SE = .02) were also not significantly different from the no noise condition ($p = 1.00$). The proportion of regressions when compared to the speech-shaped noise condition however, were found to be significantly higher with the single-talker condition ($p < .01$) and higher with the two-talker condition ($p = .01$). The single-talker condition and two-talker condition did not differ significantly ($p = .68$) and the speech-shaped noise and multi-talker conditions also did not differ significantly ($p = .77$). There were no significant differences between the single-talker and multi-talker conditions ($p = .06$) and the two-talker and multi-talker conditions ($p = 1.00$). See Appendix C for ANOVA tables.

These results indicate that although there was a trend for the regressions to increase from the baseline no noise condition with a single-talker, two-talker and multi-talker backgrounds there were no significant differences. Across all four background maskers the single-talker and twotalker backgrounds were significantly more disruptive to the reading task than a multi-talker and speech-shaped noise background as regressions were significantly higher in these two conditions.

5.4.5.7 *Saccade amplitude*

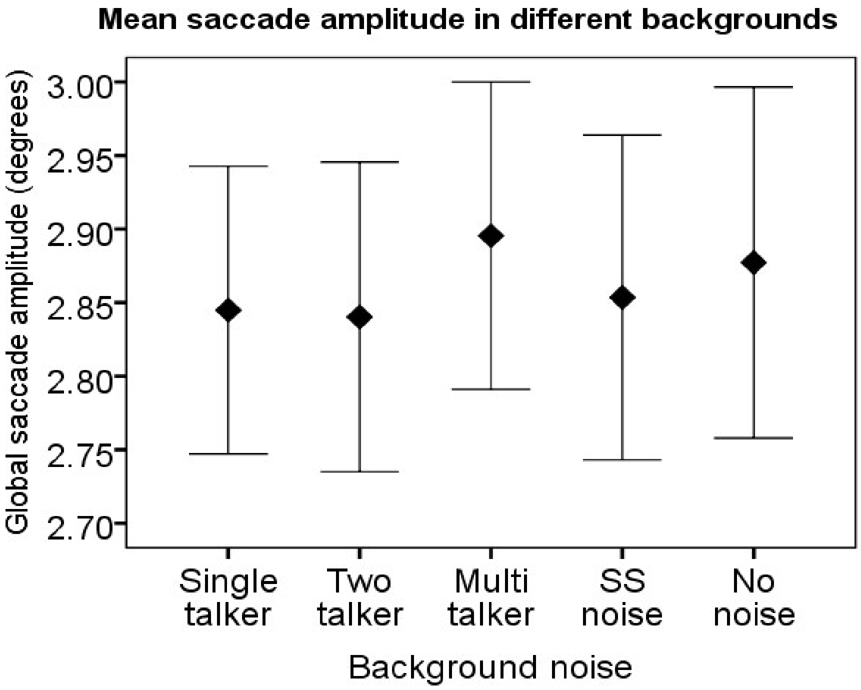


Figure 5.8: The mean global saccade amplitude across levels when reading sentences in five different backgrounds. The error bars denote ± 1 standard error.

Table 5.11: The mean saccade amplitude and standard deviations in each of the five different backgrounds

All participants

BACKGROUND	MEAN	STANDARD DEVIATION
Single-talker	2.84	0.54
Two-talker	2.84	0.58
Multi-talker	2.90	0.57
Speech-shaped noise	2.85	0.60
No noise	2.88	0.65

The global saccade amplitude is the mean degrees of visual angle within a saccade, across participants in each of the five background conditions, the results of which are displayed in figure 5.8 and table 5.11. Mauchly's test indicated that the assumption of sphericity had been violated for the main effects of background $\chi^2(9) = 30.85$, $p < .001$. Therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .64$).There was no significant main effect of background on global saccade amplitude, $F(2.54, 73.66) = .65$, $p = .56$. See Appendix C for ANOVA tables.

5.5 Discussion

5.5.1 Overview of results

The results of this study suggest that the two-talker background is the most detrimental to the speech intelligibility task. The multi-talker background does however become most disruptive at low intelligibility levels. Intelligibility is generally improved as the number of talkers is increased from two to sixteen talkers and improves further with a steady state speech-shaped noise. The single-talker background is least detrimental to the speech intelligibility task, particularly at poor SNRs. These findings were as expected and echo previous research which show two-talker maskers to be particularly detrimental to speech intelligibility tasks (Hall et al. 2002; Freyman et al. 2004; Brungart 2001; Rosen et al. 2013), that show improvements in speech intelligibility when more talkers are added to become more like stationary noise (Freyman et al. 2004; Brungart 2001; Rosen et al. 2013), and that show single-talker backgrounds to be least the effective masker (e.g. Litovsky 2005).

In comparison, the eye movement results suggest that the same single-talker background is most detrimental to the reading task and the same multi-talker and stationary noise backgrounds do not appear to disrupt the normal reading process. The same two-talker background whilst disruptive is no more so than the single-talker background, and the trend in the results suggest it may even be less distracting. This finding is in stark contrast to the speech intelligibility results which suggest the single-talker background is least disruptive and the twotalker background most disruptive. Such a result was not expected since it was hypothesised those maskers known to provide most informational masking in speech intelligibility studies (i.e. a two-talker masker) would interfere the most with the reading process. This finding coincides with previous research (Cauchard et al. 2012) that shows a single-talker background to be more disruptive to a reading task than a background of music in an eye tracking study, and also mirrors previous findings that suggest speech backgrounds have detrimental effects on reading comprehension (Sörqvist et al. 2010; Martin et al. 1988). These findings will now be discussed in detail with reference to previous research and their implications.

5.5.2 Effect of background on speech intelligibility

The results from the speech intelligibility task show that intelligibility is best in the presence of a single-talker background and worst in the presence of a two-talker background. These findings agree with those previously found in the literature that suggest a two-talker background carries the most informational masking, since performance is worse than would be expected based on its energetic content (Freyman et al. 2004; Rosen et al. 2013).

It is considered that speech maskers contain elements of informational masking (due to their similarities to the target speech) over noise maskers, and that finding is also reflected in the slopes of the psychometric functions in this experiment, which agree with previous findings and are shallower with the speech maskers compared to noise maskers (Freyman et al. 2004; Brungart 2001; Arbogast et al. 2002). Slopes have also been shown to increase as the number of talkers is increased contributing to greater energetic masking and less opportunities for the listener to “glimpse” parts of the target speech at low SNRs (MacPherson & Akeroyd 2014) which is the case with the results from the current study.

Overall, the findings from the speech intelligibility task alone suggest poorest speech intelligibility in a two-talker masker when informational masking is expected to be at its greatest. This has implications again for listening to speech amongst a background of other talkers, particularly two. Whilst this effect of masker type on speech intelligibility has been shown in previous research, it was important to confirm that the backgrounds used in the current study gave the same effect as previously found so the properties of the backgrounds (with regards to speech intelligibility) could be used and compared with the reading task.

5.5.3 Effect of background on reading

The results from the reading task show that the total reading time significantly increased with a single and two-talker background in comparison to a no noise background. This shows that there is a significant disruption to the normal reading process. When this background is changed however to a speech-shaped noise or multi-talker babble, this total reading time remains unchanged in comparison to a no noise condition. Therefore it appears that singletalker and two-talker backgrounds cause the reading time to increase.

In order to understand further why reading time was increased, the number of fixations and average fixation durations were explored. It was found that both single-talker and two-talker backgrounds increased the number of fixations significantly in comparison to a no noise condition whilst the multi-talker and noise backgrounds did not. The average fixation durations were significantly increased with the single-talker masker but were not increased from the no noise condition with any of the other maskers. The proportion of regressions was also looked at and were found not to increase significantly with any of the backgrounds compared to a no noise condition but were significantly higher with the single-talker and two-talker conditions compared to the speech-shaped noise condition. Taken together these results suggest that the total reading time increased with the single-talker and two-talker backgrounds owing to longer fixation durations and re-reading of the text (increasing the number of fixations). Such results suggest an increase in processing difficulty in these conditions (Rayner 2009).

5.5.3.1 *Relating eye movement data to previous research*

Relating eye movement results to normative data may be rather complex since results will vary according to the group of individuals being tested and the particular set of sentences being read. Number of words within a sentence and the syntactic complexity of sentences for example will also hugely influence eye movement measures (Rayner 2009). One measure which may be consistent however is mean fixation duration. Rayner (2009) reports that on average, fixation durations in reading are 200-250 ms. The results of the current study are thus consistent with this. Comparing the data in the current study to the one previous eye tracking study looking at reading in background speech by Cauchard et al. (2012), the proportion of regressions in the no noise condition and in the single-talker condition are similar. The other reading measure recorded was saccade amplitude (the mean degree of visual angle), thought to reduce when the reading task is more difficult (Rayner 2009). There was however no significant differences found with this measure across any background noise condition. This result may mean that for each condition, participants are progressing through the sentence in a similar manner where the cost is mainly to time spent reading the sentence. Furthermore, saccade amplitude may be highly dependent on the words within the sentence, where short words may be skipped over (Rayner 2009). The reasons for lack of an effect with this measure may be unclear but are in keeping with previous findings by Cauchard et al. (2012) who also found no significant differences in saccade amplitude across background noise conditions.

Unlike the results from previous reading comprehension studies (e.g. Sörqvist et al., 2010; Martin et al., 1988), the effect of reading comprehension was found not to be significant. This finding does however tie in with that found in the eye tracking study by Cauchard et al. (2012).

As previously discussed in Chapter 2, Cauchard et al. (2012) explain that this finding could be attributed to possible ceiling effects. Concerning this, ceiling effects may have also been present in the current study as results were high and simple yes/no questions were used for relatively short sentences. Cauchard et al. (2012) also explain however that since their experiment allowed readers to read with no time constraints, as previous studies employed, the reader may have slowed down in order to achieve accurate comprehension. So it may have been the case in the Cauchard et al. (2012) study and also in the current study that there was no cost to reading comprehension but at the cost of reading efficiency.

To comment on the size of the interference, in comparison to no noise, the presence of a single-talker background had a cost of 2.18 additional fixations per sentence, an additional 8.75 ms per fixation, an additional 575.13 ms total reading time per sentence and about 4% more regressions. Although these cannot be directly compared to any other study, to put them into some sort of context it is interesting to consider the findings from Kirkby et al. (2011). Kirkby et al. (2011) compared total fixation durations when reading in silence between a group of adults, a group of typically developing children (aged 9) and a group of children with dyslexia (matched for age and IQ with the typically developing children). It was found that the dyslexic group took 970 ms longer per sentence than their typically developing peers. The typically developing children also took 1364 ms longer per sentence than the adults. Considering that developmental effects in reading are large and that dyslexia is a significant clinical impairment, the effects found with adults reading in the presence of a single talker may be considered large.

5.5.3.2 *Possible reasons for interference*

The finding that total reading time increased with the single and two-talker backgrounds owing to longer fixation durations and more regressions is one which also agrees with that found by Cauchard et al. (2012). Therefore, since an increase in both these eye movement measures is thought to reflect difficulties in post lexical integration (Reichle 2011), it may be speculated that the interference from the speech background could be occurring at later higher level comprehension processes. It is important to note however that results from this study are not able to definitively determine the exact nature of the interference. Such results may fit in nonetheless with reports from Martin et al. (1988) who suggested that the speech backgrounds could interfere with semantic processing as opposed to phonological processing, since meaningful speech interfered with reading comprehension more than foreign speech and random words. It would be of interest therefore if future research addressed the phonologic and semantic content of background speech utilising eye tracking measures to investigate further the level where interference is occurring.

Concerning the different effects from background sounds, the single-talker and two-talker backgrounds interfered with reading while the speech-shaped noise background did not. Also, with the multi-talker babble condition the interfering effects seen with a single-talker and twotalker backgrounds seems to disappear. Such results suggest that when background sound contains additional linguistic properties, it may interfere with the linguistic processing of the text. Furthermore, background sound may only interfere when it is intelligible (i.e. with the single and two-talker backgrounds as opposed to the multi-talker and noise backgrounds), and it may not matter what level it is presented at. When there are more talkers in the background the phonemes may not be intelligible and semantic content may be distorted so disruption to the reading task does not occur. It cannot be deduced from this study however whether this interference is occurring with phonological processing and/or semantic processing, but disruptions in eye movement data suggest post lexical integration difficulties so it may be likely interference with semantic processing is occurring.

As Marsh et al. (2009) explained, background sounds may interfere if they call upon the same processes involved in reading the text (i.e. linguistic processes like semantic processing). It may be that cognitive capacity for a particular process is stretched if the target and the background call upon the same level of processing. Since the processing of background speech may be automatic and obligatory (Hawley et al. 2004), our eye movements may slow down to keep up with decoding meaning from the text and inhibiting meaning from the background speech.

One particularly interesting finding from the current reading task is that the two-talker background does not interfere any more so than the single-talker background. A single-talker may be understandable enough that we are building a representation of meaning of that sentence which interferes with building a representation of meaning with the target sentence. With more than one talker however, and particularly with multiple talkers, the meaning associated with the background may be less obvious as it is less distinguishable and so we may be better able to ignore it.

Overall, the findings from the reading task alone suggest that greater processing difficulty may be occurring when background speech is intelligible (i.e. with single and two-talker backgrounds), and this was found not to depend on the level of speech (i.e. whether it is quiet 55 dB (A), or loud speech 75 dB (A)). These results have implications for acoustical environments in the workplace or at school where the occurrence of background speech during reading may be an everyday issue.

5.5.4 Comparisons between background effects on reading and speech intelligibility

With regards to the effects each background has on the reading and speech intelligibility tasks, the implications of these findings combined can be considered to further understand the interference effects of speech on communication more generally.

Rosen et al. (2013) said that as the number of talkers in an acoustical mix increases, the effects of energetic masking on speech intelligibility will increase as the spectral and temporal dips become filled and physical masking becomes inevitable. At the same time, the effects of informational masking will decrease as the background becomes less and less similar to the speech target (Rosen et al. 2013).

The first important point to make is that those maskers associated with informational masking in the speech intelligibility literature (i.e. speech maskers) were found to be those most disruptive to a reading task, as opposed to those associated with most energetic masking (i.e. speech-shaped noise maskers). This finding was expected and may be thought to show the effects of linguistic interference at higher cognitive levels, thus solely show the informational masking effects without any confusion from the effects of energetic masking.

One enigma that has emerged however, relates to the findings with the two-talker background. The background sound considered to contribute most informational masking in speech intelligibility tasks, the two-talker masker, was not found to cause most disruption to the reading task. The two-talker masker interfered no more than the single-talker masker. This result is surprising since it was thought the reading task may examine interference at higher cognitive levels, thus the interference from informational masking. Therefore it appears that informational masking perhaps contains at least two components, one which is shared with the reading task (i.e. since the speech backgrounds interfered more than the noise background) and one which is not (i.e. since the two-talker background interfered no more than the single-talker background). This finding may tie in with Shinn-Cunningham's (2008) theories on informational masking and the two components of object formation and object selection. Object formation being the ability to determine the object so failures may stem from uncertainty in segregating the two sounds, and object selection being the ability to attend to the object, so failures may stem from similarities or the compelling nature of the masker making it difficult to attend to the correct sound. It could be that the object selection component of informational masking is impinging on the reading and speech intelligibility task by drawing attention away from the target and introducing semantic competition when the background speech is distinguishable. The object formation component may impinge only on the speech intelligibility task making it difficult to determine the target from the two-talker masker and to parse the auditory scene. In the reading task with the two-talker background there are only 2 talkers present, but in the listening task there are 3 talkers present with the addition of the target talker. Therefore object formation in the reading task may be more easily achieved. Furthermore, with the reading task object formation may be more readily achieved as both target and masker are presented across different sensory modalities, thus object selection may be what is being measured as attention may be diverted from the speech maskers. It would be necessary however to replicate the findings of this study using different stimuli in order to substantiate these inferences.

Concerning measures of listening effort, the results of the current study reflect previous findings of listening effort using pupillometry. The fact that the single-talker background was found to be the most disruptive to the reading task and least disruptive in the speech intelligibility task reflects the findings of Koelewijn et al. (2012b) who showed that the singletalker contributed to increased cognitive load (larger pupil dilation) compared to the noise masker. This suggests that the reading task is tapping into the cognitive demands of speech interference and that top-down processes may be engaged to deal with interfering effects of speech maskers. Furthermore this shows that current measures of speech intelligibility do not examine the cognitive demands of such listening situations. The large effect with the reading task again is likely measuring aspects of informational masking effects as separate from any effects of energetic masking. It would be interesting therefore to investigate cognitive load during speech intelligibility tasks, by measuring pupil dilation to see if listening effort appears also to decrease with an increasing number of talkers. Particularly it would be interesting to examine the pupillary response with the two-talker background to see if the results agree with those found in the reading task or those found with the speech intelligibility task.

5.5.5 Limitations in the study

It is important to consider the limitations that exist in the current study. It is likely that voice quality of the background speech may affect performance as well as the semantic content of the background speech. In the current study owing to the availability of background speech and the random selection process it may have been the case that participants heard parts of the background speech more than once. This could have had an effect on the results in that they may have been more distracted by the background, to further hear the rest of what they may have heard a little of previously, or it may have made them less distracted, as they may have more easily been able to classify it as background to be ignored. It would be of interest to see if the findings from this study with the single and two-talker backgrounds are replicable with different target and background stimuli and perhaps investigate the effects of various manipulations within speech backgrounds.

Furthermore, whilst discrepancies exist between the reading and listening task it is important to stress that differences in the methodologies may complicate comparisons. It was initially the plan to match the two tasks perfectly in terms of the dependant variables. Issues however, were raised with regards to target material not being equally matched and with the reading tasks requiring no talking due to the head restraint for accurate eye movements. Future research should therefore consider developing target material which can be used in both the speech intelligibility and reading tasks to further compare results more directly. The reading experiment unlike the speech intelligibility task also showed no significant effects with the stationary noise and multi-talker backgrounds. The reading task therefore could perhaps be sensitised to determine how such backgrounds affect reading on a local (at the word level) rather than global level (at the sentence level). Also words and interfering speech could be manipulated further, to additionally determine the features of certain sounds and how they may affect the reading process of particular words. It is also important to consider that the speech intelligibility and reading tasks were also measuring different things. The speech intelligibility task measured the discrimination of speech within background sounds while the reading task measured the disruption to the normal reading process within background sounds as well as reading comprehension. Despite the poor internal validity in comparing the two tasks, significant overall effects were seen with both tasks and so each have good external validity. It is important to consider however, that the components of informational masking in speech intelligibility may be more complex and could involve many other aspects relating sound source segregation, working memory, and other cognitive aspects which may be difficult to tease apart. Future studies could benefit from measuring various cognitive factors to see how much they predict the interference effects of background speech on both speech intelligibility and reading.

5.6 Conclusions

- Reading can be disrupted by background sounds that are thought in speech intelligibility tasks to comprise largely informational masking.

- This disruption is independent of either a 55 dB (A) background or a 75 dB (A) background level, but speech intelligibility depends on level as performance decreases at lower SNRs.

Chapter 6: **Conclusions and future research**

6.1 Conclusions

The present study set out to investigate the effect that different background sounds with various acoustic and linguistic properties have on speech intelligibility, particularly to look further into the previously found child/adult differences. It had previously been shown that children may be more affected than adults with speech backgrounds compared to noise backgrounds (Hall et al. 2002; Bonino et al. 2013; Leibold & Buss 2013), and suggested thus that children are more susceptible to informational masking mechanisms. Although definitions of informational masking are unclear, such results have been taken to suggest differences between children and adults which relate to complex cognitive factors such as attention.

Whilst previous studies have documented such differences, they have often used small samples, complex speech intelligibility tasks with high cognitive demands, complex speech backgrounds with talkers of varying number, and often background talkers of the same gender as the target talker. Concerning this, the cognitive skills of children in such studies have also not been considered and there have been no measurements in place to ensure results are not confounded by vocabulary differences. Whilst previous research has also revealed a prolonged developmental trajectory in speech intelligibility tasks in children with speech backgrounds (Hall et al. 2002; Wightman & Kistler 2005; Wightman et al. 2006; Wightman et al. 2010; Bonino et al. 2013; Leibold & Buss 2013), none have considered developmental effects within the same sample of children.

Conclusion 1

Children aged 5-8 years require a more advantageous signal to noise ratio than adults to achieve the same speech intelligibility score in both steady state speech-shaped noise and single-talker backgrounds. This remains true even for both simple single word target stimuli and for more complex sentence target stimuli, for both adaptive and constant stimuli procedures, for different gender target and background talkers, and for children with above average vocabulary scores. Moreover, the difference observed here between children and adults is much larger with the single-talker background than the steady state speech-shaped noise background. There is also some evidence from both the current research (with a longitudinal study) and previous research (with cross sectional studies), that the rate at which speech intelligibility becomes adult-like, as children get older, is slower for the single-talker background compared to the speech-shaped noise background.

Although previous studies have recognised the existence of child/adult differences, no study has considered recent propositions from Bernstein and Grant (2009) and Bernstein and Brungart (2011) that may suggest such findings could be due to an SNR confound when comparing SRTs across child and adult populations, with speech-shaped noise and single – talker backgrounds. Therefore investigating child/adult differences in such backgrounds, taking into account this factor, was necessary to determine the authenticity of such differences and is also the first study of its kind.

Conclusion 2

The larger difference between children and adults for the single-talker background, compared to the steady state speech-shaped noise background, is partly but not entirely due to this experimental confound identified by Bernstein and Grant (2009) and Bernstein and Brungart (2011). This suggests that some child/adult differences exist relating to the difference between the populations with the speech-shaped noise background. The full explanation for such differences however, remains to be determined. Whether it represents a 'deficiency' per se with auditory or language processing or a biologically adaptive leaning strategy is unclear.

One difficulty in resolving the adult-children difference, especially with speech backgrounds, with traditional listening experiments is the difficulty in separating sensory, cognitive and linguistic interference between the target and the masker. Studies of reading in background sounds might contribute to this. Since speech backgrounds are considered to contribute effects of informational masking thought to originate from higher central cognitive levels, it was of interest to investigate the cognitive involvement when processing language within speech backgrounds as separate from the peripheral effects of energetic masking. A reading paradigm was employed in order to further understand the mechanisms of informational masking relating to cognitive involvement, to help further understand the interfering effects of background speech in speech intelligibility tasks. No other study has attempted to address this issue using a reading task tracking eye movements to examine processing difficulties, and only one previous study has investigated the effects of speech backgrounds on the process of reading via eye movement measures (Cauchard et al. 2012). It was thought the reading task could be a tool to separate informational masking effects from energetic masking effects.

Conclusion 3

> The normal process of reading was disrupted and slowed down by the presence of a single-talker and two-talker background (those backgrounds associated with informational masking in speech intelligibility tasks). Such interference was also found not to depend on level. This finding extended previous research by using background speech at two different levels with varying numbers of talkers. A multi-talker and speech-shaped noise background caused no disruption to the normal process of reading. Together these findings suggest perhaps that a background with linguistic properties interferes with the linguistic processing of the text.

The findings from previous speech intelligibility tasks suggested that two-talker backgrounds cause the most disruption to speech intelligibility tasks and therefore thought to contribute large effects of informational masking, as results have been shown to be poorer than expected based on energetic summation alone (Carhart et al. 1969). This finding was not however mirrored in the reading task.

Conclusion 4

> The two-talker background was found to be no more disruptive to the reading task than the single-talker background. The differences between the interference of background across the reading and speech intelligibility tasks may advance our understanding of the mechanisms of informational masking by reinforcing theories that informational masking may contain two elements (Shinn-Cunningham 2008). It could be suggested that the reading task may be evaluating the object selection process and that the greater masking effects of two-talker backgrounds compared to single-talker backgrounds found in speech intelligibility tasks may reflect failures in both the object selection and object formation processes.

6.2 Future research

The implications of the present study suggest that children may be disadvantaged in comparison to adults in their ability to perceive speech whilst others are talking. This is particularly relevant for children since it is likely a common everyday occurrence in an educational setting. As previously discussed, this may have an impact on children when learning. Furthermore, since background speech has also been shown in the current study to affect the normal process of reading with adults, and found to be independent of level with both quiet speech and loud speech interfering to the same extent, this may too have implications for children learning within the classroom. Such effects may be expected to be even stronger with children compared to adults when considering the stronger effects speech backgrounds have with speech intelligibility tasks.

One important area of future research therefore would be to examine the effects of background speech on the normal process of reading amongst children, to see if the findings follow the same pattern as with adults. Preliminary investigations with 5 children (aged 7-10 years) were carried out, although no significant effects were found (the results from which can be found in Appendix G). A full investigation may help to determine possible reasons for child/adult differences in speech intelligibility tasks with the speech backgrounds. If informational masking does consists of two components, and if the reading task is addressing one (i.e. object selection), we may be able to further understand if children are more affected by speech backgrounds owing to failures in object formation or object selection or both.

It would be necessary to use a reading task with reading material of age appropriate levels for children, and to use that same reading material with adults in the presence of both a speechshaped noise and a speech background. Developmental effects are likely to be seen (Kirkby et al. 2011) and it is likely that speech backgrounds would interfere the most, but the presence of a masker*age interaction (with larger effects seen in children with the speech background) could show that children are affected by speech backgrounds when reading in the same way they are affected by speech backgrounds during speech intelligibility. Therefore such a result could indicate that children have difficulty with object selection and difficulties with selective attention. It is important to note however, that we cannot be certain the reading task looks only at object selection. If no interaction emerged however, it may be that children have difficulty with object formation in speech intelligibility tasks relating to aspects of sound source segregation and parsing the auditory scene. A study of this kind could advance our understanding of how communication is affected by background speech in children and provide further insight into how they may cope in noisy classroom environments.

It would also be relevant to understand the developmental trajectory for the purpose of understanding what is “normal” in order to provide a guide for deciphering communication difficulties from general development.

It would however firstly be necessary to replicate the results of the current reading study in order to validate its findings with different stimuli and to see if it is repeatable. It seems that the two-talker background interfered no more than the single-talker and there was a hint in the results that disruption may decrease monotonically with an increasing number of talkers. Therefore it would be interesting to carry out the reading task again with speech backgrounds with increasing number of talkers to see if such a pattern emerges.

Whilst the reading task has shown to mirror findings with studies investigating listening effort using measures of pupil dilation, this suggests both reading and pupillometry measures tap into the cognitive demands of the task. It would be interesting therefore to use pupillometry measures during speech intelligibility tasks to see if findings agree with those of the reading task in the present study and to see if listening effort decreases with increasing number of talkers perhaps as linguistic interference becomes less. In particular it would be of interest to examine the effects between the single-talker and two-talker backgrounds to see if two-talker backgrounds show increased listening effort (which could relate to listeners obtaining poor SRTs) or decreased listening effort (which could relate to listeners allocating fewer resources to the task, perhaps it is too difficult). Relating to this, it would also be of interest to see if children perform more poorly than adults in speech intelligibility tasks with single-talker backgrounds because it is very effortful (which may be shown by increased pupil dilations) or because they are not allocating as many resources to the task (which may be shown by decreased pupil dilations.

As Koelewijn et al. (2012) found, measures of various cognitive factors (i.e working memory capacity, reception of text and inhibition of irrelevant material) were found to be correlated with better speech intelligibility scores and measures of inhibition and text reception showed better results correlating with larger pupil dilations with a single-talker speech background, also seen by (Zekveld et al. 2011). The authors suggest that this results shows better cognitive abilities may enable participants to exert more effort when listening environments are complex, although it does remain unclear whether better cognitive abilities are associated with higher cognitive load or less (e.g. Zekveld et al., 2011). Thus it would be of interest to examine individual differences in cognitive abilities alongside speech intelligibility measures to see if performance can be predicted and if larger individual differences in children in particular can be explained.

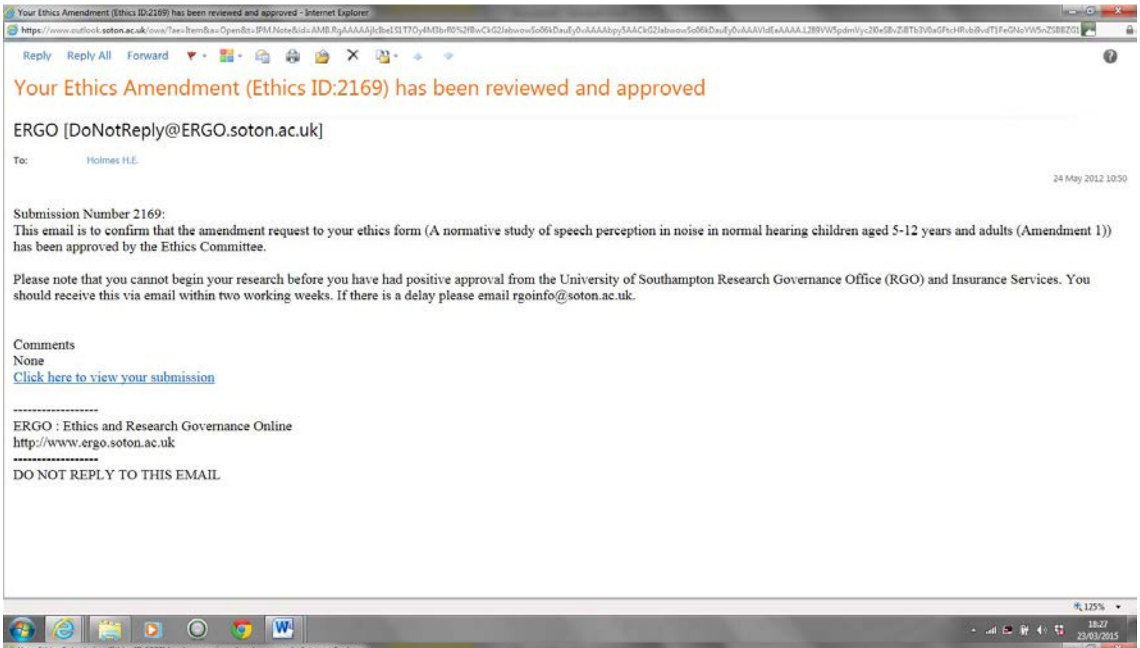
Finally, it seems clear that the speech intelligibility task may not provide any information about the cognitive demands and processing difficulties amongst background speech. Furthermore in current clinical audiological settings evaluation of hearing is often carried out by testing a patient's speech intelligibility in only steady state noise backgrounds. Whilst this may provide insight into hearing acuity at the periphery, it does not incorporate realistic listening situations.

Considering then perhaps cognitive immaturity (i.e. in children) and possible cognitive decline (i.e. in older adults) taking into account the cognitive involvement when communicating amongst background sounds may be particularly important. The development of a clinical test may then be useful to further understand these effects. A speech comprehension task, as opposed to a speech intelligibility task may be a simple measure which could tell us something more about the cognitive involvement. It would be of interest therefore to determine if such a test reflects the findings of the reading task in the present study and previous pupillometry results. It could be thought however that this may be examining mostly long term memory; although it may be more realistic than current speech intelligibility tasks. Investigations incorporating online measures (i.e. eye tracking and pupillometry studies), are likely however to be superior.

Appendices

Appendix A : Safety and Ethics Approval

A.1 Experiment 1



Your Ethics Amendment (Ethics ID:2169) has been reviewed and approved

ERGO [DoNotReply@ERGO.soton.ac.uk]

To: Holmes H.E.

24 May 2012 10:50

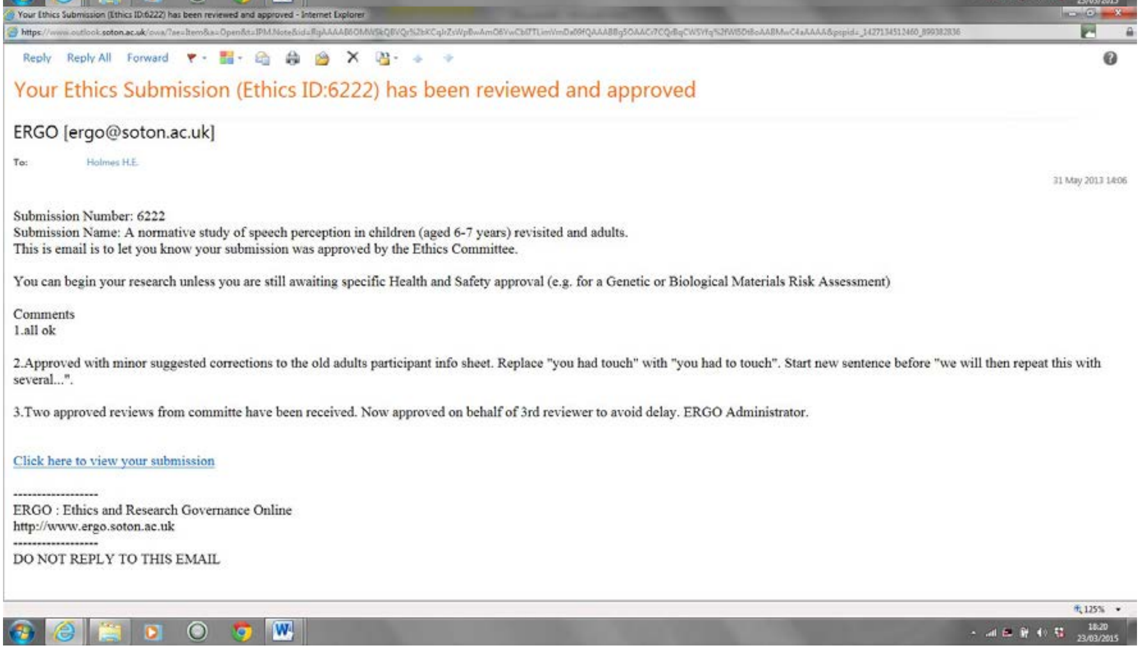
Submission Number 2169:
This email is to confirm that the amendment request to your ethics form (A normative study of speech perception in noise in normal hearing children aged 5-12 years and adults (Amendment 1)) has been approved by the Ethics Committee.

Please note that you cannot begin your research before you have had positive approval from the University of Southampton Research Governance Office (RGO) and Insurance Services. You should receive this via email within two working weeks. If there is a delay please email rgoinfo@soton.ac.uk.

Comments
None
Click here to view your submission

ERGO : Ethics and Research Governance Online
http://www.ergo.soton.ac.uk

DO NOT REPLY TO THIS EMAIL



Your Ethics Submission (Ethics ID:6222) has been reviewed and approved

ERGO [ergo@soton.ac.uk]

To: Holmes H.E.

31 May 2013 14:06

Submission Number: 6222
Submission Name: A normative study of speech perception in children (aged 6-7 years) revisited and adults.
This is email is to let you know your submission was approved by the Ethics Committee.

You can begin your research unless you are still awaiting specific Health and Safety approval (e.g. for a Genetic or Biological Materials Risk Assessment)

Comments
1.all ok

2.Approved with minor suggested corrections to the old adults participant info sheet. Replace "you had touch" with "you had to touch". Start new sentence before "we will then repeat this with several...".

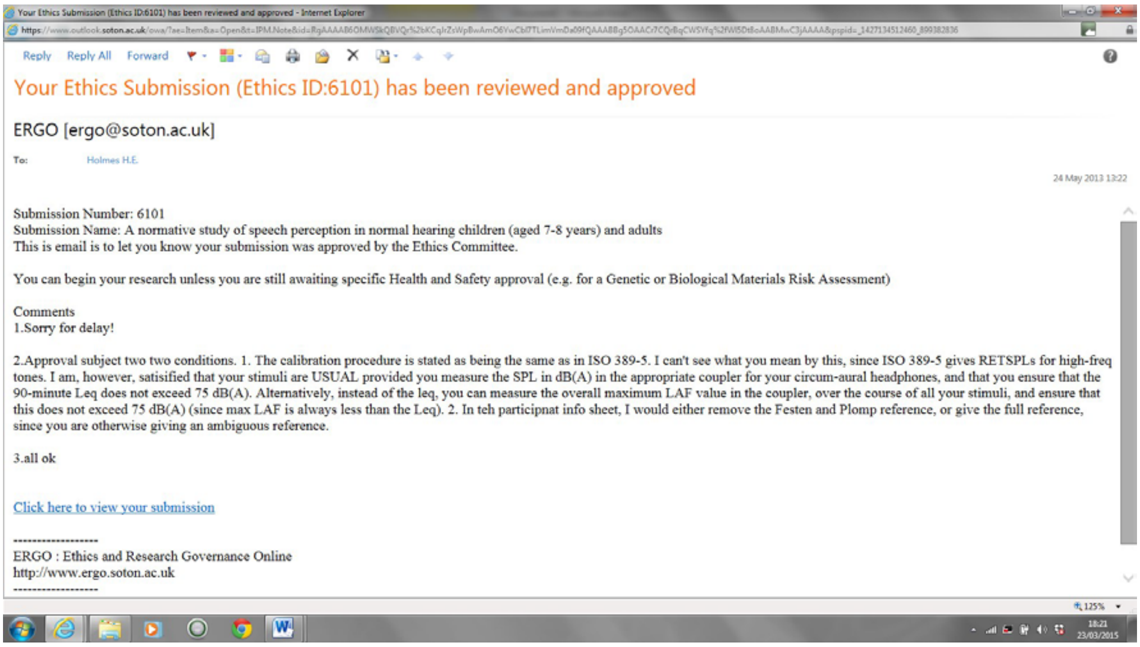
3.Two approved reviews from committe have been received. Now approved on behalf of 3rd reviewer to avoid delay. ERGO Administrator.

Click here to view your submission

ERGO : Ethics and Research Governance Online
http://www.ergo.soton.ac.uk

DO NOT REPLY TO THIS EMAIL

A.2 Experiment 2



Reply Reply All Forward

Your Ethics Submission (Ethics ID:6101) has been reviewed and approved

ERGO [ergo@soton.ac.uk]

To: Holmes H.E.

24 May 2013 13:22

Submission Number: 6101
Submission Name: A normative study of speech perception in normal hearing children (aged 7-8 years) and adults
This is email is to let you know your submission was approved by the Ethics Committee.

You can begin your research unless you are still awaiting specific Health and Safety approval (e.g. for a Genetic or Biological Materials Risk Assessment)

Comments
1.Sorry for delay!

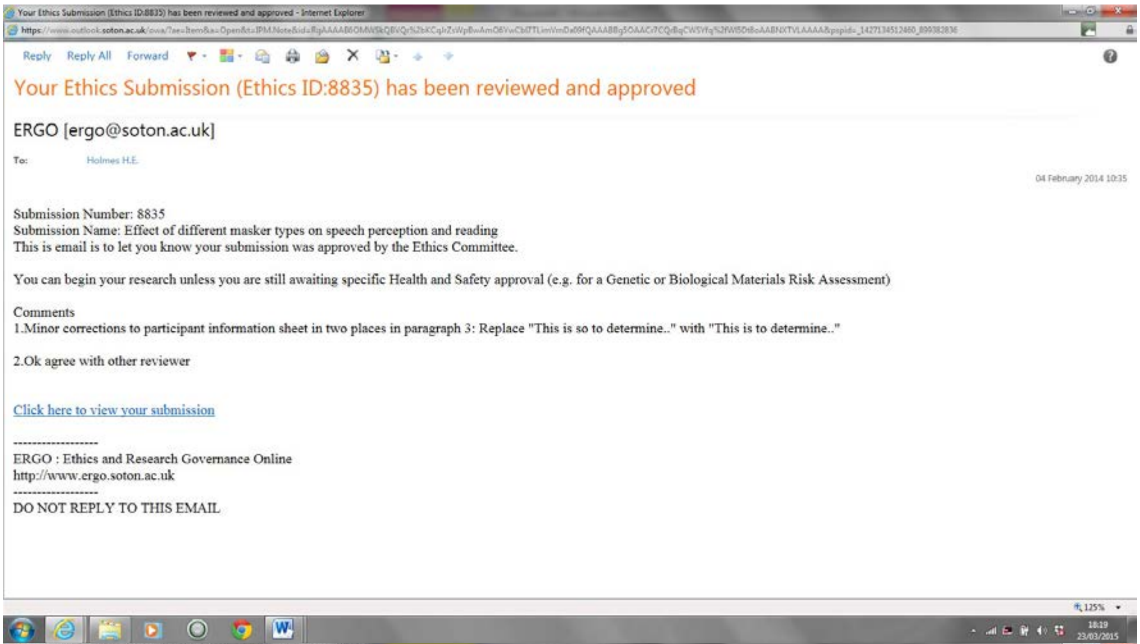
2.Approval subject two two conditions. 1. The calibration procedure is stated as being the same as in ISO 389-5. I can't see what you mean by this, since ISO 389-5 gives RETSPLs for high-freq tones. I am, however, satisified that your stimuli are USUAL provided you measure the SPL in dB(A) in the appropriate coupler for your circum-aural headphones, and that you ensure that the 90-minute Leq does not exceed 75 dB(A). Alternatively, instead of the leq, you can measure the overall maximum LAF value in the coupler, over the course of all your stimuli, and ensure that this does not exceed 75 dB(A) (since max LAF is always less than the Leq). 2. In teh participnat info sheet, I would either remove the Festen and Plomp reference, or give the full reference, since you are otherwise giving an ambiguous reference.

3.all ok

Click here to view your submission

ERGO : Ethics and Research Governance Online
http://www.ergo.soton.ac.uk

A.3 Experiment 3



Your Ethics Submission (Ethics ID:8835) has been reviewed and approved

ERGO [ergo@soton.ac.uk]

To: Holmes H.E.

04 February 2014 10:35

Submission Number: 8835
Submission Name: Effect of different masker types on speech perception and reading
This is email is to let you know your submission was approved by the Ethics Committee.

You can begin your research unless you are still awaiting specific Health and Safety approval (e.g. for a Genetic or Biological Materials Risk Assessment)

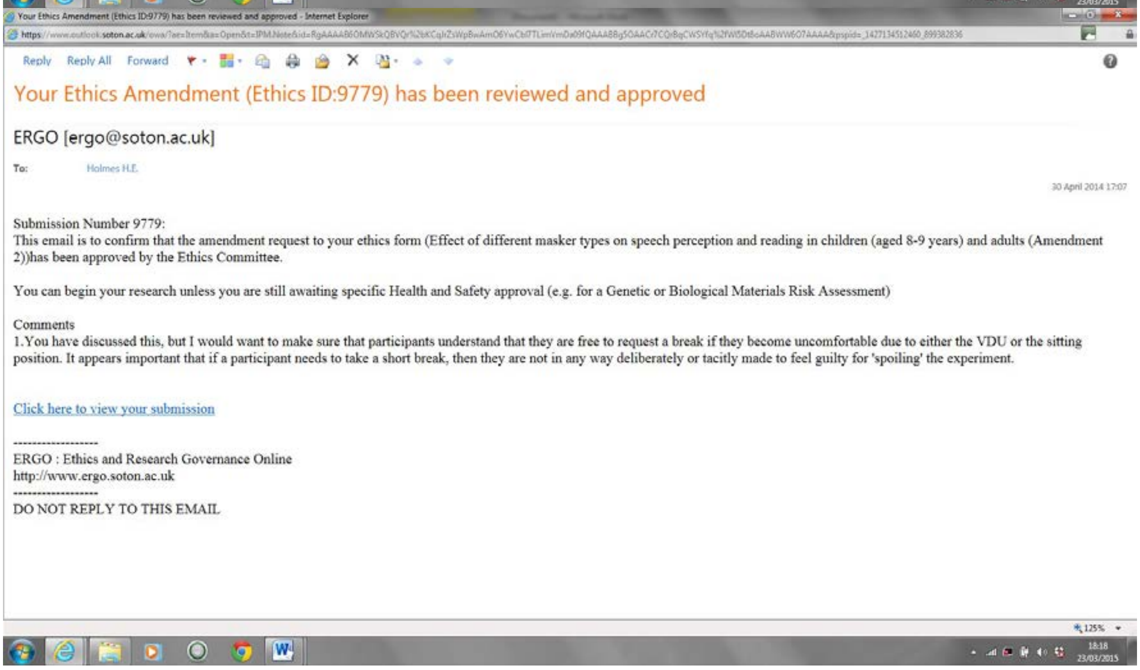
Comments
1.Minor corrections to participant information sheet in two places in paragraph 3: Replace "This is so to determine.." with "This is to determine.."

2.Ok agree with other reviewer

Click here to view your submission

ERGO : Ethics and Research Governance Online
http://www.ergo.soton.ac.uk

DO NOT REPLY TO THIS EMAIL



Your Ethics Amendment (Ethics ID:9779) has been reviewed and approved

ERGO [ergo@soton.ac.uk]

To: Holmes H.E.

30 April 2014 17:07

Submission Number 9779:
This email is to confirm that the amendment request to your ethics form (Effect of different masker types on speech perception and reading in children (aged 8-9 years) and adults (Amendment 2))has been approved by the Ethics Committee.

You can begin your research unless you are still awaiting specific Health and Safety approval (e.g. for a Genetic or Biological Materials Risk Assessment)

Comments
1.You have discussed this, but I would want to make sure that participants understand that they are free to request a break if they become uncomfortable due to either the VDU or the sitting position. It appears important that if a participant needs to take a short break, then they are not in any way deliberately or tacitly made to feel guilty for 'spoiling' the experiment.

Click here to view your submission

ERGO : Ethics and Research Governance Online
http://www.ergo.soton.ac.uk

DO NOT REPLY TO THIS EMAIL

Appendix B : Consent forms and questionnaires

B.1 Experiment 1

B.1.1 Adults

Purpose of study

When adults listen to speech in background noise it is known that they perform better when the background noise is fluctuating (e.g. another talker) as opposed to when the noise is stationary (e.g. a pink noise) (Festen & Plomp 1990). Less is known about the way children hear in the presence of these two types of maskers and there is some controversy within the literature, with some schools of thought suggesting that children perform differently to adults. The purpose of this study is to determine adults' ability to recognise speech using standard clinical tests, in the presence of these two masker types so to identify if there are any age affects between the year 1 and year 6 children (from another study) and the adults from this study. Individual differences between the adults and children will also be examined to see if they perform consistently with both types of maskers.

Prior to becoming a participant

You will need to sign an informed consent form and then undergo a screening process to find out if you are suitable for this experiment. You must be aware that your participation in this study is voluntary and you are able to withdraw from this study at any time, without giving a reason. Any information you give will be kept confidential. At the end of the study, the researcher will be happy to answer any questions that you may have and deal with any concerns you may have expressed.

What you are required to do

You will be required to attend 2 sessions. The sessions will last approximately 20 minutes each. These 2 sessions will occur on separate days.

You will need to fill out a health questionnaire regarding your otological (ear) health. You will also need to sit a hearing screening test (in the first session) which will involve playing sounds into your two ears to find the quietest level at which you can hear the sounds and you will respond by raising your hand. These sounds will be presented via headphones and you will be sat in a quiet room. The results from these screening tests will ensure that only participants who are otologically normal will be included in this study.

This experiment will involve you listening over headphones to words and sentences played amongst different types of background noise maskers and you will be required to either touch, on a touch-screen monitor the corresponding picture of the word you think you have heard or to repeat the sentence you think you have heard after every trial. You will undergo the same test four times in total consisting of speech presented in two difference background noise types. You will be offered a break at the midway point of the experiment where headphones may be removed, and you may have further breaks if necessary.

Risks

- Noise exposure: The loudness of the sounds listened to and the length of listening time will not exceed recommended levels.
- Damage to the ear/spread of infection: This may only occur if correct protocols are not followed, but procedures will be performed by a qualified audiologist.
- Electrocution: This should not occur as all equipment used will be safety checked and approved.
- Trip hazards: These will be minimised as the wires will be tidied before and after each participant and positioned so that participant do not need to walk over them..

Safety and ethics approval

This experiment has been approved by the human experimentation safety and ethics committee and if you wish to make a complaint or have any concerns about this study, you can contact Professor Rosamond Mitchell at: r.f.mitchell@soton.ac.uk or Dr Martina Prude at: m.a.prude@soton.ac.uk and quote the approval number 2169. For more information on this study, please contact the researcher Hannah Holmes at: heh1v07@soton.ac.uk.

Consent form to be completed by adult subjects taking part in an experiment
(Adults are 18 years of age or older.)

Exposure Number: .... **Safety and Ethics approval number:**..2169...

University of Southampton
Institute of Sound and Vibration Research

Before completing this form, please read the list of contra-indications which has been provided by the experimenter on the reverse of this form.

This consent form applies to a subject volunteering to undergo an experiment for research purposes. The form is to be completed before the experiment commences.

I, ...

Date of birth...
of ....
(address or department)

consent to take part in: A normative study of speech perception in noise in normal hearing children aged 5-12 years and adults

to be conducted by: Miss Hannah Holmes (Audiologist and PhD researcher)

Appendix B

during the period of: May 2012 - October 2012

____

The purpose and nature of this experiment have been explained to me. I understand that the investigation is to be carried out solely for the purposes of research. I am willing to act as a volunteer for that purpose on the understanding that I shall be entitled to withdraw this consent at any time, without giving any reasons for withdrawal. My replies to the above and below questions are correct to the best of my belief, and I understand that they will be treated by the experimenter as confidential.

Date: .... Signed: ....
(Volunteer subject)

I confirm that I have explained to the subject the purpose and nature of the investigation which has been approved by the Human Experimentation Safety and Ethics Committee.

Date: .... Signed: ....
(Researcher in charge of experiment)

This form must be submitted to the Secretary of the Human Experimentation Safety and Ethics Committee on completion of the experiment

Otological Health Questionnaire

Thank you for your willingness to participate in this experiment. Before the experiment can begin you are required to fill out this form. Please answer the questions below.

1) Name:____ 2)

Email:____ 3) Gender: ☐

☐

Male Female

4) DOB:____

5) Do you have normal hearing? ____

6) Do you have persistent tinnitus? ____

7) Have you had any recent noise exposure? ____

8) Are you currently suffering from an ear infection? ______________________________________

9) Have you had any recent ear surgery in the last 12 months? ______________________________

10) Are you currently suffering from an upper respiratory tract infection? ______________________

11) Is English you first language? ____

This experiment will involve sitting on a chair in a quiet room with headphones placed on your head. You will be listening to words and sentences for approximately 40-50 minutes at a time, whilst looking at a computer screen and responding to the sounds played to you by either touching an icon on a touchscreen or repeating back the word and sentences you think you have heard.

12) Is there any condition which would prevent you from being able to participate in this experiment?

____

Thank you. Your participation is very much appreciated.

to

B.1.2 Children

Research study offers hearing check for your child

Dear parents/carers,

Our names are Miss Hannah Holmes and Dr Daniel Rowan. We will be visiting Shirley Infant and Junior Schools between April and July 2012 to check the hearing of children in Y1 as part of a research study.
We would like to offer you the opportunity for your child to be involved and to have a hearing check. We will provide you with the result and any advice necessary.

The study involves your child listening to some words played over headphones and either pointing to a picture of the word displayed on a computer screen or repeating the words. It helps to tell us how well a child can hear speech when there is background noise, unlike most hearing checks which involves whistles in quiet. It will take no more than 30 min during the normal school day. This check will also be repeated on another day to look for consistency. Your Head Teachers have agreed for us to visit the school and we are requesting your consent for your child to take part.

We are doing this research to find out the range of results in children of different ages without hearing problems so we can better understand the benefit deaf children receive from hearing devices called cochlear implants, which are available on the NHS. This work will also contribute to PhD research.

More information on the study can be found on our website:
www.southampton.ac.uk/audiology/science/shirleyhearing.html

Please complete the attached form for your child to participate

- Your child will only have the check if you consent to it (see attached form). You may withdraw your permission at any time without giving a reason.
- We will also give your child the quick hearing check that they had when they started school, to check that there aren't obvious hearing problems. We will inform you of these results too.
- Your child's results will be shared with you and no one else. When we report the results (e.g. in a scientific journal), all the children's names will be removed.
- The study has been officially approved by a safety and ethics committee (application 2169) and by the University research office.
- If you have any questions, you can contact Miss Hannah Holmes at heh1v07@soton.ac.uk or Dr Daniel Rowan on 02380 592288 or at audiology@southampton.ac.uk.
- If you have any concerns about the study, please contact ISVR Safety & Ethics Committee, ISVR, University of Southampton, SO17 1BJ, Professor Rosamond Mitchell at r.f.mitchell@soton.ac.uk or Dr Martina Prude at m.a.prude@soton.ac.uk.

Thank you in advance for your interest in and support of this project.

Yours faithfully, Miss Hannah Holmes and Dr Daniel Rowan

Complete and return this section only your child's school
ASAP to consent to your child participating

This consent form applies to the legal guardian of a child volunteering to undergo a study for research purposes. The form is to be completed before the experiment commences.

Appendix B

I, (your name) ____ give consent

for: (your child's name) ____________________________________ (your child's date of

birth)________________________________ of (your child's school)

____

to take part in: A normative study of speech perception in noise in normal hearing children aged 5-12 years.

to be conducted by: Miss Hannah Holmes (Audiologist and PhD researcher) or Dr Daniel Rowan (Audiologist and research supervisor).

during the period 23th April 2012 to 31st July 2012

The purpose and nature of this experiment have been explained to me. I understand that the study is to be carried out solely for the purposes of research. I am willing for my child to participate for that purpose on the understanding that I shall be entitled to withdraw this consent at any time, without giving any reasons for withdrawal. My replies to the above and below questions are correct to the best of my belief, and I understand that they will be treated by the researcher as confidential.

You: Date: _____________ Signed: ______________________________

I confirm that I have explained to the subject the purpose and nature of the investigation which has been approved by the ISVR Human Experimentation Safety and Ethics Committee.

To be completed by the researcher: Date:____________________ Signed:____________________

Please circle 'yes' or 'no' to all five questions:

Do you have any concerns about your child's hearing? Yes No

Do you feel that your child needs the TV volume higher, or needs you to repeat what you say more, than you would expect? Yes No

Does your child currently have ear or hearing problems or has your child had recent ear surgery (within past 12 months)? Yes No

Does your child have any special educational needs? Yes No

Is English your child's first language? Yes No

B.2 Experiment 2

B.2.1 Adults

Research study to check your hearing

to

Dear Participant,

My name is Hannah Holmes (Qualified Audiologist). I am carrying out a study investigating the effect of age on speech perception in different background noises. When adults listen to speech in background noise it is known that they perform better when the background noise is fluctuating (e.g. another talker) as opposed to when the noise is stationary (e.g. a steady-state noise). Less is known about the way children hear in the presence of these two types of maskers and there is some controversy within the literature, with some schools of thought suggesting that children perform differently to adults. The purpose of this study is to determine children's (aged 7-8 years) and adults' ability to recognise speech using a standard clinical test, in the presence of differing masker types and to identify any child adult differences in attempts to determine the reasons for such differences.

The key research questions of this study are to determine whether there are real differences in the way that children and adults perceive speech in speech maskers, or whether the differences are due to an acoustical artefact originating from differences in the baseline (stationary noise masker) conditions.

You will be required to attend 2 sessions. The sessions will last approximately 30-40 minutes each. These 2 sessions will occur on separate days. You will need to fill out a brief health questionnaire regarding your otological (ear) health. You will also need to undergo a vocabulary check to determine the range of your vocabulary, and sit a hearing screening test which will involve playing sounds into your two ears to find the quietest level at which you can hear the sounds and you will respond by raising your hand. These sounds will be presented via headphones and you will be sat in a quiet room. The results from these screening tests will ensure that only participants who are otologically normal will be included in this study.

The main experiment will involve you listening over headphones to sentences played amongst different types of background noise maskers and you will be required to repeat the sentence you think you have heard after every trial. You will be offered a break at the midway point of the experiment where headphones may be removed, and you may have further breaks if necessary.

Please complete the attached form to participate

- You will have the check if you consent to it (see attached form). You may withdraw your permission at any time without giving a reason.
- Your results will be shared with you and no one else. When we report the results (e.g. in a scientific journal), all names will be removed.
- The study has been officially approved by a safety and ethics committee (application 6101) and by the University Research Office.
- If you have any questions, you can contact me at heh1v07@soton.ac.uk or my supervisor, Dr Daniel Rowan, on 02380 592928 or at audiology-enquiries@isvr.soton.ac.uk.
- If you have any concerns about the study, please contact Dr Martina Prude, Research Governance Office, at m.a.prude@soton.ac.uk.

Thank you in advance for your support of this project.

Yours faithfully, Miss Hannah Holmes.

Appendix B Complete and return **this section only** to

consent to participate

This consent form is to be completed before the experiment commences.

I, (your name) ____ (your

date of birth)____ of (department/address)

____ consent to take part in: A

normative study of speech perception in children (aged 7-8 years) and adults.

to be conducted by: Miss Hannah Holmes (Audiologist and PhD researcher)

during the period 1st June 2013 to 31st December 2013.

The purpose and nature of this experiment have been explained to me. I understand that the study is to be carried out solely for the purposes of research. I am willing to participate for that purpose on the understanding that I shall be entitled to withdraw this consent at any time, without giving any reasons for withdrawal. My replies to the above and below questions are correct to the best of my belief, and I understand that they will be treated by the researcher as confidential.

You: Date: ________________ Signed: _______________________________________

I confirm that I have explained to the subject the purpose and nature of the investigation which has been approved by the ISVR Human Experimentation Safety and Ethics Committee.

To be completed by the researcher: Date:___________________ Signed:___________________

Please circle 'yes' or 'no' to all three questions:

Do you have any concerns about your hearing?	Yes	No
Do you currently have ear or hearing problems or have you had recent ear surgery (within past 12 months)?	Yes	No
Is English your first language?	Yes	No

B.2.2 Children

Research study to check the hearing of your child

Dear Parents/Carers,

My name is Miss Hannah Holmes (Qualified Audiologist). I will be visiting Shirley Junior School in June and July 2013 to check the hearing of children in Year 3 as part of a research study. I would like to offer you the opportunity for your child to be involved and to have a hearing check. You will be provided with the result and any advice necessary.

The hearing check involves your child listening to some simple sentences in different background sounds over headphones. Your child will be required to repeat back as much of the sentence that they have heard. This will help to tell us how well children can hear speech when there is background noise, unlike most hearing checks which involve whistles in quiet. This will take 30-40 min during the normal school day, and will be repeated one week later to check for consistency. Your Head Teacher has agreed for us to visit the school and will ensure that it does not interfere with your child's learning. This study is similar to one we conducted at the school in 2012 and is running in parallel to a study in Shirley Infants School.

I am requesting your consent for your child to have this hearing check, to have follow-up checks once per year while your child remains at SJS and for us to collect basic data such as your child's vocabulary range, gender and date of birth. You will be reminded 2 months before hearing checks in future years to give you an opportunity to change your mind.

We are doing this research to find out the range of results in children of different ages without hearing problems so we can better understand the benefit deaf children receive from hearing devices called cochlear implants, which are available on the NHS. This work will also contribute to my PhD research.

Please complete the attached form for your child to participate

- Your child will only have the check and follow-up checks over the next 4 years if you consent to it (see attached form). You may withdraw your permission at any time without giving a reason.
- We will also give your child the quick hearing check that they had when they started school, to check that there aren't obvious hearing problems. We will inform you of these results too.
- Your child's results will be shared with you and no one else. When we report the results (e.g. in a scientific journal), all the children's names will be removed.
- The study has been officially approved by a safety and ethics committee (application 6101) and by the University Research Office.
- If you have any questions, you can contact me at heh1v07@soton.ac.uk or my supervisor, Dr Daniel Rowan, on 02380 592928 or at audiology-enquiries@isvr.soton.ac.uk.
- If you have any concerns about the study, please contact Dr Martina Prude, Research Governance Office, at m.a.prude@soton.ac.uk.

Thank you in advance for your support of this project.

Yours faithfully

Miss Hannah Holmes.

Appendix B

Complete and return **<u>this section only</u>** to

your child's school ASAP to consent to your child participating

This consent form applies to the legal guardian of a child volunteering to undergo a study for research purposes. The form is to be completed before the experiment commences.

I, (your name) ____

give consent for: (your child's name) ____

(your child's date of birth)____ of (your child's

school) ____ to take part in:

A normative study of speech perception in children.

to be conducted by: Miss Hannah Holmes (Audiologist and PhD researcher) during the

period 1st June 2013 to 31st July 2013, then once per year over the next 4 years.

The purpose and nature of this experiment have been explained to me. I understand that the study is to be carried out solely for the purposes of research. I am willing for my child to participate for that purpose on the understanding that I shall be entitled to withdraw this consent at any time, without giving any reasons for withdrawal. My replies to the above and below questions are correct to the best of my belief, and I understand that they will be treated by the researcher as confidential.

You: Date: _________________ Signed: ____

I confirm that I have explained to the subject the purpose and nature of the investigation which has been approved by the ISVR Human Experimentation Safety and Ethics Committee.

To be completed by the researcher: Date:____________________ Signed:___________________

Please circle 'yes' or 'no' to all four questions:

Do you have any concerns about your child's hearing?	Yes	No
Does your child currently have ear or hearing problems or has your child had recent ear surgery (within past 12 months)?	Yes	No
Does your child have any special educational needs?	Yes	No
Is English your child's first language?	Yes	No

B.3 Experiment 3

Research study to look at distractions from background noise

Dear Participant,

My name is Hannah Holmes (Qualified Audiologist and PhD Researcher). I am carrying out a study investigating how participants are able to cope with background sound, both when listening to speech and reading sentences. I am aiming to determine how different types of background sound, e.g. static noise and interfering talkers affect the task of listening and reading to establish how the individual qualities of each background sound affects each task.

You will be required to attend 2-4 sessions. Each session will last approximately 90-120 minutes and will occur on separate days. You will need to fill out a brief health questionnaire regarding your otological (ear) health. You will also need to undergo a vocabulary check to determine the range of your vocabulary, and sit a hearing screening test which will involve playing sounds into your two ears to find the quietest level at which you can hear the sounds; you will respond by raising your hand. These sounds will be presented via headphones and you will be seated in a quiet room. The results from these screening tests will ensure that only participants who are otologically normal are included in this study.

The main experiment will involve you carrying out two tasks. One task entails listening over headphones to sentences played amongst different types of background sounds and you will be required to repeat the sentence you think you have heard after every trial. This is to determine your speech perception accuracy. The other task entails reading silently sentences which will appear on a computer screen. Whilst reading these sentences different types of background sounds will be played to you over headphones and you will be required to try to understand each sentence since you will be asked multiple choice comprehension questions about what you have read after some trials. The reading task requires you also to sit with your head in a chin and forehead rest to keep your head still in order to enable the computer to accurately track your eye movements. This is to determine the fluency of your silent reading. You will be offered a break at the midway point of the experiment where you may sit back and remove the headphones; you may also have further breaks if necessary.

Please complete the attached form to participate

- You will have the check if you consent to it (see attached form). You may withdraw your permission at any time without giving a reason.
- Your results will be shared with you and no one else. When we report the results (e.g. in a scientific journal), all names will be removed.
- The study has been officially approved by a safety and ethics committee (application 8835) and by the University Research Office.
- If you have any questions, you can contact me at heh1v07@soton.ac.uk or my supervisor, Dr Daniel Rowan, on 02380 592928 or at audiology-enquiries@isvr.soton.ac.uk.
- If you have any concerns about the study, please contact Dr Martina Prude, Research Governance Office, at m.a.prude@soton.ac.uk.

Thank you in advance for your support of this project.

Yours faithfully

Miss Hannah Holmes.

Appendix B Complete and return **this section only** to

consent to participate

UNIVERSITY OF Southampton
Institute of Sound and Vibration Research

This consent form is to be completed before the experiment commences.

I, (your name) ____

(your date of birth)____ of

(department/address) ____

consent to take part in: Effect of different masker types on speech perception and reading.

to be conducted by: Miss Hannah Holmes (Audiologist and PhD researcher) during

the period 6th January 2014 to 31st July 2014.

The purpose and nature of this experiment have been explained to me. I understand that the study is to be carried out solely for the purposes of research. I am willing to participate for that purpose on the understanding that I shall be entitled to withdraw this consent at any time, without giving any reasons for withdrawal. My replies to the above and below questions are correct to the best of my belief, and I understand that they will be treated by the researcher as confidential.

You: Date: _________________ Signed: ____

I confirm that I have explained to the subject the purpose and nature of the investigation which has been approved by the ISVR Human Experimentation Safety and Ethics Committee.

To be completed by the researcher: Date:____________________ Signed:____________________

Please circle 'yes' or 'no' to all three questions:

Do you have any concerns about your hearing?	Yes	No
Do you currently have ear or hearing problems or have you had recent ear surgery (within past 12 months)?	Yes	No
Is English your first language?	Yes	No

Appendix C : ANOVA tables

Table C.1: Summary of results from the first four-way mixed measures ANOVA comparing performance between 50 children in year 1 (aged 5-6 years) and adults described in section 3.4.5. Significant effects are highlighted.

Tests of Between-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*P*
Intercept	213766.45	1	213766.45	4940.59	<0.001
Age	12637.32	1	12637.32	292.08	<0.001
Error	4240.21	98	43.27		

Tests of Within-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Masker	37174.47	1	37174.47	1160.92	<0.001
Masker * Age	2069.17	1	2069.17	64.62	<0.001
Error(Masker)	3138.13	98	32.022		
Session	586.88	1	586.88	27.97	<0.001
Session * Age	107.75	1	107.75	5.14	0.03
Error(Session)	2056.33	98	20.98		
Repeat	63.51	1	63.51	3.72	0.06
Repeat * Age	48.31	1	48.31	2.83	0.10
Error(Repeat)	1671.83	98	17.06		
Masker * Session	85.67	1	85.67	5.33	0.02
Masker * Session * Age	26.57	1	26.57	1.66	0.20
Error(Masker*Session)	1573.96	98	16.06		
Masker * Repeat	85.81	1	85.81	5.06	0.03
Masker * Repeat * Age	10.67	1	10.67	0.63	0.43
Error(Masker*Repeat)	1662.87	98	16.97		
Session * Repeat	0.22	1	0.22	0.01	0.91
Session * Repeat * Age	16.30	1	16.30	0.98	0.33
Error(Session*Repeat)	1630.16	98	16.63		
Masker * Session * Repeat	39.78	1	39.78	2.62	0.11
Masker * Session * Repeat * Age	9.77	1	9.77	0.64	0.43

Error(Masker*Session*Repeat)	1489.12	98	15.20		

Table C.2: Summary of results from the second four-way mixed measures ANOVA comparing performance between 34 children in year 2 (aged 6-7 years) and adults described in section 3.4.5. Significant effects are highlighted.

Tests of Between-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Intercept	193517.59	1	193517.59	4606.14	<0.001
Age	5964.44	1	5964.44	141.97	<0.001
Error	3445.06	82	42.01		

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Masker	28377.10	1	28377.10	959.15	<0.001
Masker * Age	2111.16	1	2111.16	71.36	<0.001
Error(Masker)	2426.02	82	29.59		
Session	209.14	1	209.14	8.86	<0.01
Session * Age	4.02	1	4.02	0.17	0.68
Error(Session)	1935.10	82	23.60		
Repeat	0.56	1	0.56	0.03	0.87
Repeat * Age	0.03	1	0.03	0.00	0.97
Error(Repeat)	1574.31	82	19.20		
Masker * Session	65.33	1	65.33	3.74	0.06
Masker * Session * Age	23.85	1	23.85	1.37	0.25
Error(Masker*Session)	1432.58	82	17.47		
Masker * Repeat	13.61	1	13.61	0.83	0.36
Masker * Repeat * Age	2.91	1	2.91	0.18	0.67
Error(Masker*Repeat)	1340.51	82	16.35		
Session * Repeat	34.96	1	34.96	2.23	0.14
Session * Repeat * Age	3.43	1	3.43	0.22	0.64
Error(Session*Repeat)	1287.74	82	15.70		

Masker * Session * Repeat	0.16	1	0.16	0.01	0.91
Masker * Session * Repeat * Age	10.66	1	10.66	0.81	0.37
Error(Masker*Session*Repeat)	1079.07	82	13.16		

Table C.3: Summary of results from third four-way repeated measures ANOVA comparing performance between 50 children when in year 1 (aged 5-6 years) and 34 children when in year 2 (aged 6-7 years) described in section 3.4.5. Significant effects are highlighted.

Tests of Within-Subjects Effects

Source	**Sum of Squares**	**Degrees of freedom**	**Mean Square**	***F***	***P***
Year	506.28	1	506.28	15.48	<0.001
Error(Year)	1079.45	33	32.71		
Masker	13073.20	1	13073.20	457.70	<0.001
Error(Masker)	942.58	33	28.56		
Session	530.88	1	530.88	19.18	<0.001
Error(Session)	913.66	33	27.69		
Repeat	36.96	1	36.96	1.38	0.25
Error(Repeat)	882.97	33	26.76		
Year * Masker	4.24	1	4.24	0.22	0.65
Error(Year*Masker)	649.23	33	19.67		
Year * Session	63.19	1	63.19	2.47	0.13
Error(Year*Session)	845.58	33	25.62		
Masker * Session	18.31	1	18.31	0.86	0.36
Error(Masker*Session)	704.53	33	21.35		
Year * Masker * Session	1.81	1	1.81	0.09	0.76
Error(Year*Masker*Session)	654.86	33	19.84		
Year * Repeat	30.78	1	30.78	1.93	0.17
Error(Year*Repeat)	526.21	33	15.95		
Masker * Repeat	23.81	1	23.81	1.48	0.23
Error(Masker*Repeat)	532.50	33	16.14		
Year * Masker * Repeat	9.37	1	9.37	0.49	0.49
Error(Year*Masker*Repeat)	630.80	33	19.12		
Session * Repeat	0.77	1	0.77	0.04	0.84

Error(Session*Repeat)	590.38	33	17.89		
Year * Session * Repeat	38.97	1	38.97	3.16	0.08
Error(Year*Session*Repeat)	406.42	33	12.32		
Masker * Session * Repeat	5.440	1	5.440	0.31	0.58
Error(Masker*Session*Repeat)	570.83	33	17.30		
Year * Masker * Session * Repeat	32.42	1	32.42	2.41	0.13
Error(Year*Masker*Session*Repeat)	443.62	33	13.44		

Table C.4: Summary of results from the fourth four-way mixed measures ANOVA comparing performance between 34 children in year 2 (aged 6-7 years) (minus those 3 children who had below average vocabulary scores) and adults described in section 3.4.5.

Tests of Between-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*P*
Intercept	183341.91	1	183341.91	4431.58	<0.001
Age	5575.80	1	5575.80	134.77	<0.001
Error	3268.37	79	41.37		

Tests of Within-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Masker	26959.73	1	26959.73	885.69	<0.001
Masker * Age	1961.40	1	1961.40	64.44	<0.001
Error(Masker)	2404.70	79	30.44		
Session	214.41	1	214.41	8.98	<0.01
Session * Age	6.40	1	6.40	0.27	0.61
Error(Session)	1886.06	79	23.87		
Repeat	0.20	1	0.20	0.01	0.92
Repeat * Age	0.19	1	0.19	0.01	0.92
Error(Repeat)	1552.22	79	19.65		
Masker * Session	48.82	1	48.82	2.75	0.10
Masker * Session * Age	31.59	1	31.59	1.78	0.19
Error(Masker*Session)	1404.81	79	17.78		
Masker * Repeat	17.18	1	17.18	1.05	0.31
Masker * Repeat * Age	1.21	1	1.21	0.07	0.79
Error(Masker*Repeat)	1287.75	79	16.30		
Session * Repeat	34.22	1	34.22	2.13	0.15
Session * Repeat * Age	3.62	1	3.62	0.23	0.64

Error(Session*Repeat)	1270.67	79	16.08		
Masker * Session * Repeat	0.69	1	0.69	0.05	0.82
Masker * Session * Repeat * Age	13.05	1	13.05	0.97	0.33
Error(Masker*Session*Repeat)	1059.70	79	13.41		

Table C.5: Summary of results from the three-way mixed measures ANOVA comparing the difference between SRT estimation methods between all 50 children in year 1 (aged 5-6 years) and adults described in section 3.4.6. Significant effects are highlighted.

Tests of Between-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Intercept	95057.71	1	95057.71	4237.88	<0.001
Age	6480.97	1	6480.97	288.94	<0.001
Error	2198.19	98	22.43		

Tests of Within-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Method	356.76	1	356.76	305.37	<0.001
Method * Age	1.66	1	1.66	1.42	0.24
Error(Method)	114.49	98	1.17		
Masker	18651.93	1	18651.93	1118.79	<0.001
Masker * Age	961.02	1	961.02	57.64	<0.001
Error(Masker)	1633.82	98	16.67		
Method * Masker	0.26	1	0.26	0.24	0.63
Method * Masker * Age	2.07	1	2.07	1.90	0.17
Error(Method*Masker)	106.34	98	1.09		

Table C.6: Summary of results from the three-way mixed measures ANOVA comparing the difference between SRT estimation methods between 34 children in year 2 (aged 6-7 years) and adults described in section 3.4.6. Significant effects are highlighted.

Tests of Between-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Intercept	85837.01	1	85837.01	4125.12	<0.001
Age	3232.43	1	3232.43	155.34	<0.001
Error	1706.29	82	20.81		

Tests of Within-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Method	326.93	1	326.93	137.40	<0.001
Method * Age	5.04	1	5.04	2.12	0.15
Error(Method)	195.12	82	2.38		
Masker	13728.82	1	13728.82	1046.63	<0.001
Masker * Age	1129.01	1	1129.01	86.07	<0.001
Error(Masker)	1075.61	82	13.12		
Method * Masker	3.79	1	3.79	1.63	0.21
Method * Masker * Age	1.23	1	1.23	0.53	0.47
Error(Method*Masker)	190.71	82	2.33		

Table C.7: Summary of results from the three-way repeated measures ANOVA comparing the difference between SRT estimation methods between all 50 children when in year 1 (aged 5-6 years) and 34 children when in year 2 (aged 6-7 years) described in section 3.4.6. Significant effects are highlighted.

Tests of Within-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Year	219.45	1	219.45	12.64	<0.001
Error(Year)	572.75	33	17.36		
Method	307.40	1	307.40	115.00	<0.001
Error(Method)	88.21	33	2.67		
Masker	6406.37	1	6406.37	374.81	<0.001
Error(Masker)	564.04	33	17.09		
Year * Method	1.20	1	1.20	0.41	0.53
Error(Year*Method)	96.71	33	2.93		
Year * Masker	11.88	1	11.88	1.25	0.27
Error(Year*Masker)	313.16	33	9.49		
Method * Masker	0.66	1	0.66	0.21	0.65
Error(Method*Masker)	103.03	33	3.12		
Year * Method * Masker	3.97	1	3.97	1.46	0.24
Error(Year*Method*Masker)	89.73	33	2.72		

Table C.8: Summary of results from the two-way mixed measures ANOVA comparing performance between 18 children (7-8 years) and 18 adults described in section 4.4.5 Significant effects are highlighted.

Tests of Between-Subjects Effects

Appendix C

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Intercept	6500.339	1	6500.339	2740.242	<0.001
Age	653.757	1	653.757	275.593	<0.001
Error	80.654	34	2.372		

Tests of Within-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Masker	1624.395	1	1624.395	881.672	<0.001
Masker * Age	230.838	1	230.838	125.292	<0.001
Error(Masker)	62.642	34	1.842		

Table C.9: Summary of results from the one-way repeated measures ANOVA comparing location parameter from fitted logistic functions described in section 5.4.4.2. Significant effects are highlighted.

Tests of Within-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Masker[a]	4331.08	2.15	2015.052	821.06	<0.001
Error(Masker)[a]	152.98	62.33	2.454		

[a]Mauchly's test of sphericity gave *p* <0.001

Pairwise Comparisons

(I) Masker	(J) Masker	Mean Difference (I-J)	Std. Error	*p*[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
Single-talker	**Two-talker**	-15.00*	0.43	<0.001	-16.21	-13.79
	Multi-talker	-14.25*	0.38	<0.001	-15.33	-13.17
	Noise	-11.17*	0.38	<0.001	-12.26	-10.09

	Single-talker	15.00*	0.43	<0.001	13.79	16.21
	Multi-talker	0.76	0.27	0.06	-0.02	1.53
Two-talker	Noise	3.83*	0.35	<0.001	2.83	4.83
	Single-talker	14.25*	0.38	<0.001	13.17	15.33
	Two-talker	-0.76	0.27	0.06	-1.53	0.02
Multi-talker	Noise	3.07*	0.17	<0.001	2.58	3.56
	Single-talker	11.17*	0.38	<0.001	10.09	12.26
Noise	Two-talker	-3.83*	0.35	<0.001	-4.83	-2.83
	Multi-talker	-3.07*	0.17	<0.001	-3.56	-2.58

Based on estimated marginal means

*The mean difference is significant at the 0.05 level.
b Adjustment for multiple comparisons: Bonferroni.

Table C.10: Summary of results from the one-way repeated measures ANOVA comparing slope parameter from fitted logistic functions described in section 5.4.4.2. Significant effects are highlighted.

Tests of Within-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*
Masker	4.42	3	1.48	49.73	<0.001
Error(Masker)	2.58	87	0.03		

Pairwise Comparisons

Measure: MEASURE_1

(I) Masker	(J) Masker	Mean Difference (I-J)	Std. Error	p^b	95% Confidence Interval for Difference[b]

					Lower Bound	Upper Bound
Singletalker	Two-talker	-0.15*	0.04	<0.01	-0.26	-0.04
	Multi-talker	-0.47*	0.05	<0.001	-0.62	-0.33
	Noise	-0.41*	0.03	<0.001	-0.51	-0.32
Two-talker	Single-talker	0.15*	0.04	0.003	0.04	0.27
	Multi-talker	-0.32*	0.06	<0.001	-0.48	-0.16
	Noise	-0.26*	0.04	<0.001	-0.37	-0.14
Multitalker	Single-talker	0.47*	0.05	<0.001	0.33	0.62
	Two-talker	0.32*	0.06	<0.001	0.16	0.48
	Noise	0.06	0.04	1.000	-0.06	0.18
Noise	Single-talker	0.41*	0.03	<0.001	0.32	0.51
	Two-talker	0.26*	0.04	<0.001	0.14	0.37
	Multi-talker	-0.06	0.04	1.000	-0.18	0.06

Based on estimated marginal means

*The mean difference is significant at the 0.05 level.

b Adjustment for multiple comparisons: Bonferroni.

Table C.11: Summary of results from the one-way repeated measures ANOVA comparing background masker on the number of fixations described in section 5.4.5.2. Significant effects are highlighted.

Tests of Within-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*p*

Masker[a]	116.44	1.80	64.85	12.38	<0.001
Error(Masker) [a]	272.81	52.07	5.24		

[a]Mauchly's test of sphericity gave *p* <0.001

Pairwise Comparisons

Measure:MEASURE_1

(I) Masker	(J) Masker	Mean Difference (I-J)	Std. Error	p^b	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
No noise	**Single-talker**	-2.18*	0.57	0.01	-3.92	-0.44
	Two-talker	-1.86*	0.54	0.02	-3.49	-0.24
	Multi-talker	-0.47	0.26	0.79	-1.26	0.32
	Noise	-0.31	0.25	1.000	-1.07	0.44
Single-talker	**No noise**	2.18*	0.57	0.01	0.44	3.91
	Two-talker	0.32	0.32	1.000	-0.66	1.30
	Multi-talker	1.72*	0.43	0.01	0.39	3.03
	Noise	1.87*	0.41	<0.01	0.63	3.11
Two-talker	**No noise**	1.86*	0.54	0.02	0.24	3.49
	Single-talker	-0.32	0.32	1.000	-1.30	0.66
	Multi-talker	1.39*	0.44	0.04	0.05	2.74
	Noise	1.55*	0.38	<0.01	0.40	2.70
Multi-talker	**No noise**	0.47	0.26	0.79	-.315	1.26
	Single-talker	-1.72*	0.43	0.01	-3.029	-0.30
	Two-talker	-1.39*	0.44	0.04	-2.736	-0.05
	Noise	0.16	0.17	1.000	-0.36	0.68
Noise	**No noise**	0.31	0.25	1.000	-0.44	1.07
	Single-talker	-1.87*	0.41	<0.01	-3.11	-0.63
	Two-talker	-1.55*	0.38	<0.01	-2.70	-0.40
	Multi-talker	-0.16	0.17	1.000	-0.68	0.36

Based on estimated marginal means

*The mean difference is significant at the .05 level. b
Adjustment for multiple comparisons: Bonferroni.

Table C.12: Summary of results from the one-way repeated measures ANOVA comparing background masker on the fixation durations described in section 5.4.5.3. Significant effects are highlighted.

Tests of Within-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*P*
Masker[a]	2171.00	2.98	729.29	8.23	<0.001
Error(Masker) [a]	7653.06	86.33	88.65		

[a]Mauchly's test of sphericity gave $p = 0.02$

Pairwise Comparisons

(I) Masker	(J) Masker	Mean Difference (I-J)	Std. Error	p^b	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
No noise	**Single-talker**	-8.77*	2.35	0.01	-15.91	-1.64
	Two-talker	-7.46	2.70	0.10	-15.65	0.73
	Multi-talker	0.09	2.34	1.000	-7.01	7.19
	Noise	-1.57	1.94	1.000	-7.47	4.33
Single-talker	**No noise**	8.77*	2.35	0.01	1.64	15.91
	Two-talker	1.31	1.42	1.000	-2.99	5.61
	Multi-talker	8.86*	1.88	<0.01	3.15	14.58
	Noise	7.20*	1.74	<0.01	1.93	12.47
Two-talker	**No noise**	7.46	2.70	0.10	-0.73	15.65
	Single-talker	-1.31	1.42	1.000	-5.61	2.99
	Multi-talker	7.55*	2.24	0.02	0.75	14.36
	Noise	5.89	2.26	0.14	-0.97	12.76
Multi-talker	**No noise**	-0.09	2.34	1.000	-7.19	7.01

(I) Masker	(J) Masker	Mean Difference (I-J)	Std. Error	p^b	Lower Bound	Upper Bound
	Single-talker	-8.86*	1.88	<0.01	-14.58	-3.15
	Two-talker	-7.55*	2.24	0.02	-14.36	-0.75
	Noise	-1.66	1.82	1.000	-7.18	3.86
Noise	No noise	1.57	1.94	1.000	-4.33	7.47
	Single-talker	-7.20*	1.74	<0.01	-12.47	-1.93
	Two-talker	-5.89	2.26	0.14	-12.76	0.97
	Multi-talker	1.66	1.82	1.000	-3.86	7.18

Based on estimated marginal means

*The mean difference is significant at the 0.05 level. b Adjustment for multiple comparisons: Bonferroni.

Table C.13: Summary of results from the one-way repeated measures ANOVA comparing background masker on the total fixation time described in section 5.4.5.4. Significant effects are highlighted

Tests of Within-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*P*
Masker[a]	8457539.91	1.83	4618126.29	15.37	<0.001
Error(Masker) [a]	15958559.56	53.11	300481.29		

Mauchly's test of sphericity gave $p < 0.001$

Pairwise Comparisons

(I) Masker	(J) Masker	Mean Difference (I-J)	Std. Error	p^b	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
No noise	Single-talker	-575.14*	135.05	<0.01	-985.41	-164.86
	Two-talker	-501.09*	129.98	0.01	-895.97	-106.22
	Multi-talker	-99.87	65.80	1.000	-299.77	100.02
	Noise	-82.27	58.76	1.000	-260.80	96.25
Single-talker	No noise	575.14*	135.05	<0.01	164.86	985.41

	Two-talker	74.05	75.21	1.000	-154.45	302.54
	Multi-talker	475.27*	103.32	<0.01	161.38	789.15
	Noise	492.87*	96.38	<0.001	200.05	785.68
Two-talker	No noise	501.09*	129.98	0.01	106.22	895.97
	Single-talker	-74.05	75.21	1.000	-302.54	154.45
	Multi-talker	401.22*	109.94	0.01	67.22	735.22
	Noise	418.82*	94.86	<0.01	130.62	707.01
Multi-talker	No noise	99.87	65.80	1.000	-100.02	299.77
	Single-talker	-475.27*	103.32	<0.01	-789.15	-161.38
	Two-talker	-401.22*	109.94	0.01	-735.22	-67.22
	Noise	17.60	45.79	1.000	-121.50	156.70
Noise	No noise	82.27	58.76	1.000	-96.25	260.80
	Single-talker	-492.87*	96.38	<0.001	-785.68	-200.05
	Two-talker	-418.82*	94.86	<0.01	-707.01	-130.62
	Multi-talker	-17.60	45.79	1.000	-156.70	121.50

Based on estimated marginal means

*The mean difference is significant at the 0.05 level. [b] Adjustment for multiple comparisons: Bonferroni.

Table C.14: Summary of results from the one-way repeated measures ANOVA comparing background masker on the proportion of regressions described in section 5.4.5.5. Significant effects are highlighted.

Tests of Within-Subjects Effects

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*P*
Masker[a]	0.03	2.29	0.01	5.81	<0.01
Error(Masker) [a]	0.16	66.53	0.00		

[a]Mauchly's test of sphericity gave $p < 0.001$

Pairwise Comparisons

(I) Masker	(J) Masker	Mean Difference (I-J)	Std. Error	p^b	95% Confidence Interval for Difference[b] Lower Bound	95% Confidence Interval for Difference[b] Upper Bound
No noise	Single-talker	-0.04	0.01	0.13	-0.08	0.01
	Two-talker	-0.02	0.01	0.54	-0.06	0.01
	Multi-talker	-0.01	0.01	1.00	-0.04	0.01
	Noise	-0.00	0.01	1.00	-0.03	0.03
Single-talker	No noise	0.04	0.01	0.13	-0.01	0.08
	Two-talker	0.01	0.01	0.68	-0.01	0.04
	Multi-talker	0.02	0.01	0.06	-0.00	0.05
	Noise	0.04*	0.01	<0.01	0.01	0.06
Two-talker	No noise	0.02	0.01	0.54	-0.01	0.06
	Single-talker	-0.01	0.01	0.68	-0.04	0.01
	Multi-talker	0.01	0.01	1.00	-0.02	0.04
	Noise	0.02*	0.01	0.01	0.00	0.04
Multi-talker	No noise	0.01	0.01	1.00	-0.01	0.04
	Single-talker	-0.02	0.01	0.06	-0.05	0.00
	Two-talker	-0.01	0.01	1.000	-0.04	0.02
	Noise	0.01	0.01	0.77	-0.01	0.04
Noise	No noise	0.00	0.01	1.000	-0.03	0.03
	Single-talker	-0.04*	0.01	<0.01	-0.06	-0.01
	Two-talker	-0.02*	0.01	0.01	-0.04	-0.00
	Multi-talker	-0.01	0.01	0.77	-0.04	0.01

Based on estimated marginal means

*The mean difference is significant at the 0.05 level
b Adjustment for multiple comparisons: Bonferroni.

Table C.15: Summary of results from the one-way repeated measures ANOVA comparing background masker on the saccade amplitude described in section 5.4.5.6. Significant effects are highlighted.

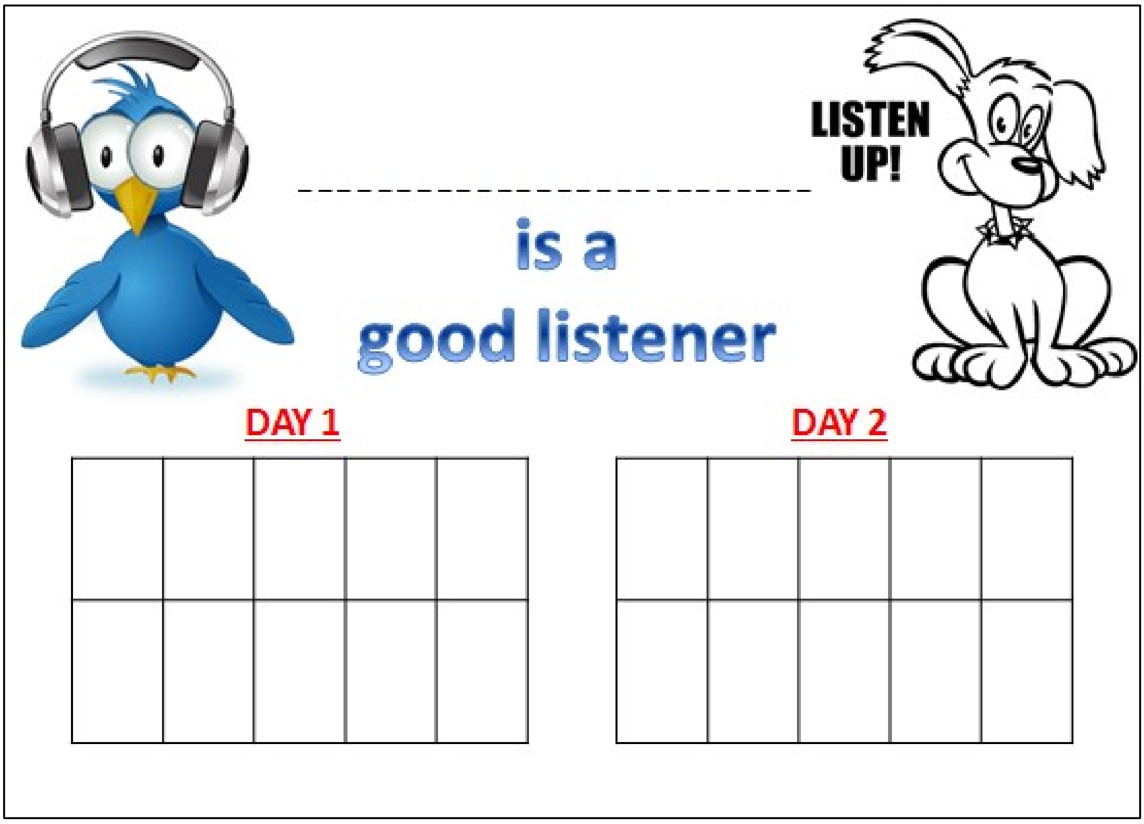
Tests of Within-Subjects Effects

Appendix C

Source	Sum of Squares	Degrees of freedom	Mean Square	*F*	*P*
Masker[a]	0.07	2.54	0.03	0.65	0.56
Error(Masker) [a]	2.93	73.66	0.04		

Mauchly's test of sphericity gave $p < 0.001$

Appendix D : Reward cards for experiment 2



Appendix E : Goodness of fit data

E.1 Experiment 2

Table E.1: Deviance and p-values for the goodness of fit calculated for all individual psychometric functions with adults in speech-shaped noise. The left panel displays the results with session 1, the middle panel session 2 and the right panel both sessions.

Adults: speech-shaped noise		
Participant No.	Session 1	
	Deviance	p-value
1	0.57	0.84
2	1.57	0.51
3	4.25	0.16
4	0.47	0.58

Adults: speech-shaped noise		
Participant No.	Session 2	
	Deviance	p-value
1	1.34	0.57
2	4.15	0.12
3	3.01	0.36
4	0.29	0.81

Adults: speech-shaped noise		
Participant No.	Both sessions	
	Deviance	p-value
1	1.71	0.60
2	5.02	0.10
3	6.82	0.06
4	0.07	0.93

5	0.83	0.67	5	0.68	0.73	5	0.92	0.69
6	4.88	0.15	6	3.08	0.32	6	4.62	0.18
7	6.89	0.05	7	6.87	0.06	7	13.76	0.01
8	0.76	0.8	8	2.66	0.25	8	1.98	0.50
9	7.48	0.01	9	1.81	0.61	9	0.91	0.79
10	2.7	0.08	10	1.87	0.3	10	0.98	0.62
11	4.56	0.02	11	1.16	0.72	11	2.54	0.36
12	0.29	0.84	12	5.33	0.09	12	3.37	0.18
13	0.62	0.84	13	1.34	0.62	13	0.82	0.86
14	0.69	0.61	14	2.7	0.09	14	0.18	0.96
15	3.89	0.21	15	5.33	0.06	15	7.67	0.03
16	2.34	0.23	16	3.29	0.21	16	5.02	0.11
17	1.83	0.62	17	1.46	0.54	17	2.39	0.46
18	4.19	0.17	18	0.46	0.74	18	1.02	0.66
19	2.82	0.28	19	1	0.67	19	2.24	0.40
20	1	0.68	20	0.29	0.78	20	0.97	0.71

Table E.2: Deviance and p-values for the goodness of fit calculated for all individual psychometric functions with adults in the single-talker background. The left panel displays the results with session 1, the middle panel session 2 and the right panel both sessions.

Adults: single-talker			Adults: single-talker			Adults: single-talker		
Participant No.	Session 1		Participant No.	Session 2		Participant No.	Both sessions	
	Deviance	p-value		Deviance	p-value		Deviance	p-value
1	0.68	0.89	1	0.43	0.80	1	0.58	0.89
2	1.00	0.65	2	1.22	0.61	2	1.32	0.63
3	0.37	0.95	3	0.50	0.93	3	0.14	0.98
4	0.59	0.93	4	1.31	0.69	4	1.44	0.75
5	0.56	0.90	5	1.67	0.50	5	0.53	0.93
6	0.48	0.80	6	0.39	0.87	6	0.77	0.78
7	0.60	0.93	7	1.35	0.63	7	0.83	0.82
8	2.09	0.45	8	2.96	0.35	8	3.41	0.36
9	4.07	0.11	9	0.59	0.92	9	2.90	0.32
10	1.55	0.65	10	1.68	0.66	10	2.44	0.54

11	0.36	0.94	11	0.90	0.72	11	0.63	0.86
12	2.22	0.45	12	2.91	0.27	12	0.52	0.92
13	1.19	0.77	13	1.90	0.48	13	0.37	0.91
14	2.72	0.50	14	4.12	0.18	14	3.23	0.33
15	0.90	0.80	15	1.07	0.80	15	0.91	0.80
16	3.54	0.23	16	5.11	0.12	16	4.30	0.17
17	0.05	0.99	17	3.70	0.18	17	1.06	0.77
18	1.32	0.76	18	0.81	0.86	18	1.77	0.65
19	1.54	0.66	19	1.38	0.56	19	0.27	0.93
20	1.26	0.70	20	0.60	0.91	20	1.55	0.65

Table E.3: Deviance and p-values for the goodness of fit calculated for all individual psychometric functions with children in speech-shaped noise. The left panel displays the results with session 1, the middle panel session 2 and the right panel both sessions.

Child: speech-shaped noise			Child: speech-shaped noise			Child: speech-shaped noise		
Participant No.	Session 1		Participant No.	Session 2		Participant No.	Both sessions	
	Deviance	p-value		Deviance	p-value		Deviance	p-value
1	3.61	0.27	1	7.15	0.03	1	9.26	0.03
2	1.29	0.76	2	1.94	0.55	2	0.31	0.92
3	1.48	0.60	3	5.82	0.08	3	5.64	0.11
4	1.14	0.66	4	3.21	0.30	4	3.61	0.21
5	2.62	0.13	5	2.46	0.36	5	2.66	0.42
6	9.29	0.02	6	0.28	0.99	6	3.11	0.31
7	1.92	0.54	7	2.88	0.18	7	4.57	0.13
8	1.09	0.69	8	1.46	0.59	8	2.59	0.45
9	1.50	0.72	9	4.32	0.08	9	2.31	0.54
10	2.87	0.30	10	1.96	0.56	10	3.60	0.29
11	4.28	0.21	11	3.89	0.17	11	7.88	0.04
12	1.66	0.60	12	3.08	0.28	12	2.53	0.50
13	1.99	0.50	13	0.88	0.85	13	2.06	0.56
14	0.84	0.86	14	2.66	0.45	14	0.61	0.91
15	1.66	0.56	15	2.83	0.33	15	3.73	0.18
16	0.62	0.84	16	0.61	0.86	16	1.24	0.74

17	2.42	0.18	17	0.68	0.71	17	2.31	0.46
18	2.31	0.33	18	1.08	0.66	18	2.04	0.45
19	3.21	0.29	19	4.81	0.14	19	5.21	0.15
20	2.36	0.49	20	0.62	0.89	20	1.19	0.71

Table E.4: Deviance and p-values for the goodness of fit calculated for all individual psychometric functions with children in the single-talker background. The left panel displays the results with session 1, the middle panel session 2 and the right panel both sessions.

Child: single-talker			Child: single-talker			Child: single-talker		
Participant No.	Session 1		Participant No.	Session 2		Participant No.	Both sessions	
	Deviance	p-value		Deviance	p-value		Deviance	p-value
1	1.39	0.78	1	1.91	0.66	1	0.22	0.98
2	6.81	0.08	2	0.36	0.95	2	3.66	0.32
3	2.61	0.27	3	0.28	0.98	3	0.42	0.90
4	2.88	0.42	4	1.53	0.66	4	2.59	0.50
5	0.09	0.98	5	2.21	0.43	5	0.35	0.94
6	1.82	0.63	6	5.88	0.14	6	4.59	0.20
7	5.38	0.13	7	3.59	0.34	7	8.39	0.07
8	0.03	1.00	8	5.67	0.06	8	1.81	0.57
9	2.88	0.44	9	5.29	0.18	9	3.76	0.31
10	2.58	0.47	10	2.33	0.49	10	2.79	0.44
11	4.11	0.25	11	1.82	0.63	11	4.49	0.21
12	2.32	0.52	12	3.57	0.30	12	5.76	0.11
13	5.74	0.12	13	4.63	0.23	13	8.15	0.05
14	0.59	0.92	14	3.36	0.40	14	2.34	0.48
15	0.85	0.85	15	0.33	0.97	15	1.09	0.79
16	1.63	0.66	16	6.85	0.04	16	2.77	0.42
17	2.63	0.50	17	4.85	0.21	17	6.20	0.13
18	1.15	0.77	18	1.31	0.76	18	2.35	0.49
19	1.59	0.70	19	1.96	0.59	19	2.65	0.42
20	0.75	0.88	20	0.82	0.84	20	0.35	0.92

E.2 Experiment 3

Table E.5: Deviance and p-values for the goodness of fit calculated for all individual psychometric functions with adults in each background condition. The first panel displays the results with speech-shaped noise, the second panel displays the results with the single-talker background, the thrid panel displays the results with two-talker background and the fourth panel displays the results with the multi-talker background.

No.	**Adults: speechshaped noise**		No.	**Adults: singletalker**		No.	**Adults: two-talker**		No.	**Adults: multi-talker**	
	Deviance	p-value		Deviance	pvalue		Deviance	p-value		Deviance	p-value
1	3.71	0.38	1	0.89	0.85	1	0.93	0.84	1	1.91	0.60
2	5.42	0.07	2	0.41	0.94	2	0.31	0.97	2	1.02	0.80
3	2.83	0.42	3	2.21	0.57	3	1.63	0.46	3	6.79	0.08
4	5.10	0.13	4	2.70	0.34	4	2.27	0.60	4	1.23	0.73
5	2.58	0.51	5	3.58	0.32	5	0.96	0.88	5	2.39	0.52
6	0.26	0.97	6	3.14	0.38	6	0.07	1.00	6	1.71	0.66
7	1.46	0.58	7	0.44	0.97	7	2.00	0.62	7	5.33	0.06
8	3.08	0.43	8	2.98	0.39	8	2.94	0.44	8	0.54	0.92
9	2.19	0.56	9	0.94	0.80	9	0.65	0.88	9	1.21	0.78
10	3.27	0.39	10	2.17	0.56	10	0.68	0.89	10	1.26	0.80
11	2.53	0.50	11	1.18	0.77	11	0.92	0.89	11	6.67	0.09
12	0.44	0.97	12	1.68	0.64	12	0.09	1.00	12	1.52	0.65
13	1.42	0.73	13	0.74	0.87	13	3.13	0.43	13	2.86	0.47
14	0.57	0.83	14	2.47	0.51	14	0.61	0.88	14	0.35	0.96
15	2.61	0.41	15	4.06	0.32	15	3.73	0.36	15	2.24	0.23
16	2.59	0.54	16	1.15	0.75	16	0.87	0.91	16	4.88	0.14
17	1.51	0.71	17	4.96	0.21	17	0.58	0.88	17	1.88	0.63
18	4.39	0.21	18	1.31	0.74	18	3.19	0.41	18	7.20	0.06
19	4.81	0.15	19	4.08	0.25	19	3.55	0.34	19	1.67	0.54
20	1.56	0.69	20	1.52	0.68	20	2.54	0.59	20	2.48	0.44
21	2.89	0.46	21	1.66	0.65	21	1.72	0.43	21	0.82	0.86
22	0.80	0.90	22	0.71	0.89	22	1.82	0.60	22	1.47	0.70
23	3.33	0.35	23	1.17	0.76	23	0.37	0.96	23	3.96	0.16
24	6.48	0.10	24	1.74	0.62	24	3.37	0.36	24	1.23	0.73
25	1.80	0.62	25	3.64	0.30	25	3.00	0.44	25	1.40	0.68
26	3.87	0.19	26	2.21	0.50	26	0.60	0.61	26	1.63	0.67
27	4.26	0.18	27	2.01	0.63	27	1.69	0.65	27	1.14	0.68

28	4.36	0.17
29	7.31	0.09
30	1.35	0.64

28	3.57	0.32
29	3.29	0.24
30	1.96	0.57

28	1.77	0.65
29	1.11	0.76
30	3.10	0.43

28	6.39	0.06
29	2.33	0.47
30	1.37	0.63

Appendix F : Read sentence text stimuli

1 The children say a ghost lives in the spooky house on the hill.
2 The painting was of a yellow chair standing on top of a red table.
3 Betsy always wore a thick dress because the thin ones wore out too fast.
4 Lucy put on her dress before going outside to work in the garden.
5 Leo would like to go on holiday to a quiet house during the summer.
6 You can see the large house sitting on top of the hill from miles away.
7 All he wanted was to eat his apple without getting interrupted again.
8 Fred cut his thumb when he was making breakfast this morning.
9 On stage there was a young child singing with the most amazing voice.
10 Bill was the first one to spot the small child which was on top of the tree.
11 Tim and Adrian had never seen such a young judge before this trial.
12 Joe was the first judge nominated for the prize in nearly fifty years.
13 In the hospital there was a small store where you could buy flowers.
14 Sam was disappointed they did not sell his usual drink anywhere on campus.
15 We stopped at a rural hotel hoping that they accepted credit cards.
16 The inner components are protected by a black metal increasing its lifespan.
17 The victim was killed by a sharp stone thrown at him from quite some distance.
18 When Jenny heard the awful music again she almost started to weep.
19 The wine had a sharp taste which did not make it very popular.
20 Jack quickly cleaned the dirty table before his parents arrived for dinner.
21 A man with a heavy voice announced the end of the football game.
22 Louise saw a kind woman donating one of her kidneys to save her friend.
23 After dinner he told his son a short story which he knew from his childhood.
24 Bob went to another party after leaving his friends on Sunday.
25 The priest could not find anybody in the whole world willing to help him.
26 The officer said that without a basic radio using the boat was useless.
27 Katy bought an expensive phone knowing she would regret buying it later.
28 Sean always tried to drink some juice every morning to improve his health.
29 The child drew a picture of a big tiger after his trip to the zoo.
30 Shaun bought Debbie a new piano before her birthday party.
31 The young owner standing by the door was not very happy about the hotel.
32 Wendy dropped her frame after Fred jumped out and scared her.
33 Lorna wore a small medal given to her especially for the parade.
34 A rusty wheel belonging to Terry has been left in the garden since last summer.
35 Megan took the large plate across the room to put away in the kitchen cupboard.
36 Stephen saw a small sheep during his walk around the field.
37 William spent most of his time in a small cabin while on holiday.
38 The children were excited to see a cute robin during their lunch break.

39 When Josh saw a massive shark while he was at the zoo he was very scared.

40 Lauren has a new knife bought from the large department store.

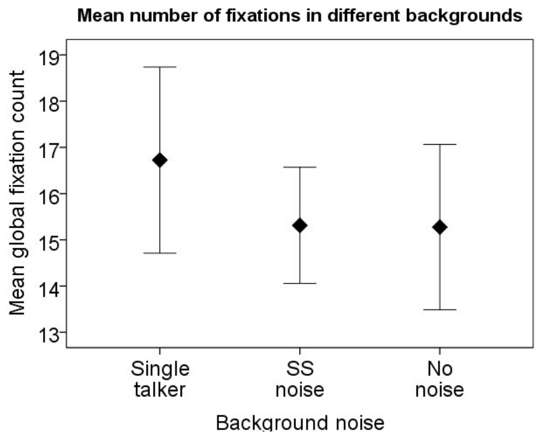
41 David grabbed his glass filled with juice before leaving the house.

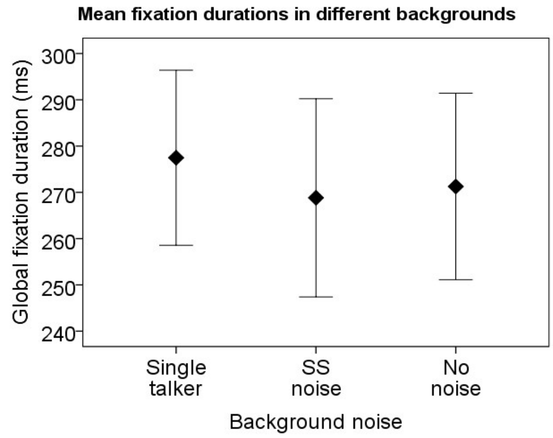
42 Laura played with a toy horse given to her by her grandparents.

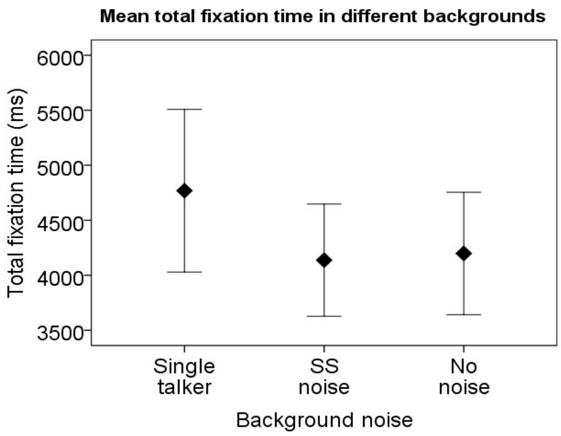
43 Kelly wants to buy a new shirt today to wear to her job interview. 44 Natalie is going to visit the beach later in the day after work.

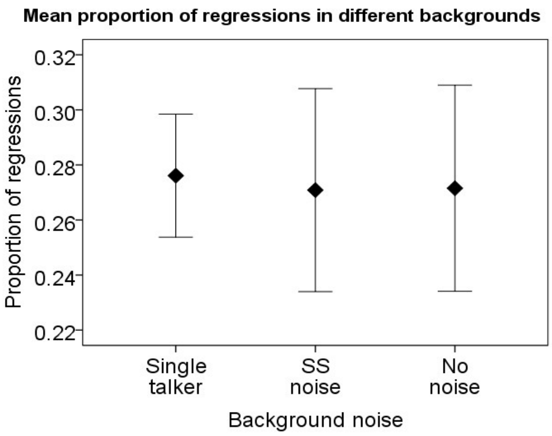
45 Greg wanted to play the part of the slave after reading the script.
46 Paul went to see the movie without his friends at the weekend.
47 The old guard walked the lost little girl back to her parents.
48 Ashleigh bought a large plant which she decided to put in her garden.
49 Peter had never tried bacon before his holiday abroad.
50 The passengers said that the smell towards the back of the train was horrible.
51 The builder went to get the large cable after realising he had left it in his van.
52 Andrew was told to clean the old table before the party guests arrived.
53 The man was woken up by the sound of a radio early this morning.
54 Jenna walked her dog near the large field after finishing dinner.
55 Mark went to buy a new towel while they were on sale at the store.
56 Mike saw they had built a large tower close to the old farm buildings.
57 Becky received a nice shell given to her by Fred while on holiday.
58 Lee always had water alongside his lunch when visiting his mother.
59 Vicky often made a nice treat which she gave to her husband for dessert.
60 Edward had a wooden floor installed in his new house in the country.
61 Emily tripped over the long chain that was lying on the floor in the garage.
62 Jacob always liked to look at the snake when he visited the zoo.
63 Frank didn't feel well, he had a sharp pain in his chest when he coughed.
64 Mike was excited when he found out he was the new coach for the football team.
65 Lily wanted to put the sheet outside on the washing line, but it was raining.
66 Charles decided to visit the restaurant by the shore later on today.
67 Stan took a really good photo of Becky when she was dancing last night.
68 The accountant picked up the important paper from the floor where it had fallen.
69 The book was about a lost crown and how the clever princess was going to find it.
70 The server picked up the cloth and started to clean the tables by the window.
71 Chris laughed when he saw the funny video with the cat dancing on the table.
72 Danny pulled the switch for the lights when he entered the dark basement room.

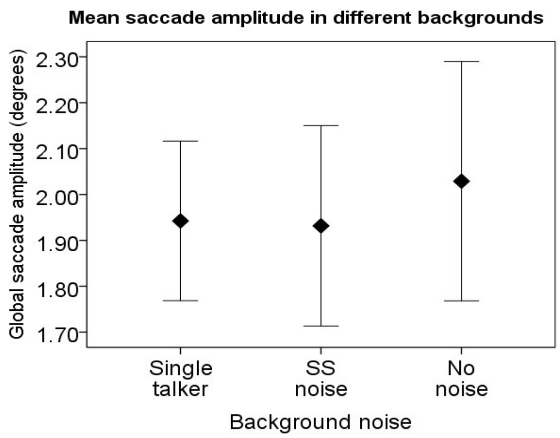
Appendix G : Pilot reading study with children











List of References

Aiken, L.R. (1994). *Psychological testing and assessment*, Needham Heights, MA: Allyn & Bacon.

Akhtar, N. (2005). The robustness of learning through overhearing. *Dev. Sci.*, 8, 199–209.

Allen, P., Wightman, F. (1995). Effects of signal and masker uncertainty on children's detection. *J. Speech Hear. Res.*, 38, 503–11.

Arbogast, T.L., Mason, C.R., Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *J. Acoust. Soc. Am.*, 112, 2086.

Bacon, S.P., Opie, J.M., Montoya, D.Y. (1998). The Effects of Hearing Loss and Noise Masking on the Masking Release for Speech in Temporally Complex Backgrounds. *J. Speech Lang. Hear. Res.*, 41, 549.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annu. Rev. Psychol.*, 63, 1–29.

Baddeley, A., Eldridge, M., Lewis, V. (1981). The role of subvocalisation in reading. *Q. J. Exp. Psychol. Sect. A*, 33, 439–454.

Baddeley, A.D. (1979). Working memory and reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma, eds. *Processing of visible language*. Plenum, New York:

Baddeley, A.D., Lewis, V.J. (1981). Inner active processes in reading: The inner voice, the inner ear, and the inner eye. In & C. A. Perfetti (Eds.), Interactive processes and reading (pp. 107-129). Hillsdale, NJ: Erlbaum. In A. M. Lesgold & C. . Perfetti, eds. *Interactive processes and reading*. (*pp. 107–129*). Hillsdale, NJ: Erlbaum.

Banbury, S.P., Macken, W.J., Tremblay, S., et al. (2001). Auditory distraction and short-term memory: phenomena and practical implications. *Hum. Factors*, 43, 12–29.

Bargones, J.Y., Werner, L.A. (1994). Adults listen selectively; Infants do not. *Psychol. Sci.*, 5, 170–174.

Beaman, C.P., Bridges, A.M., Scott, S.K. (2007). From Dichotic Listening to the Irrelevant Sound Effect: a Behavioural and Neuroimaging Analysis of the Processing of Unattended Speech. *Cortex*, 43, 124–134.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.*, 91, 276–292.

Bench, J., Bamford, J. (1979). *Speech-Hearing tests in the spoken language of hearing-impaired children*, London: Academic Press Inc. Ltd.

Bench, J., Kowal, Å., Bamford, J. (1979). The Bkb (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *Br. J. Audiol.*, 13, 108–112.

Bernstein, J.G.W. (2012). Controlling signal-to-noise ratio effects in the measurement of speech intelligibility in fluctuating maskers. In T. Dau, M. L. Jepsen, T. Poulson, et al., eds. *Speech perception and auditory disorders*. (*pp. 33–44*). Ballerup, Denmark: Danovox Jubilee Foundation.

Bernstein, J.G.W., Brungart, D.S. (2011). Effects of spectral smearing and temporal finestructure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio. *J. Acoust. Soc. Am.*, 130, 473–88.

Bernstein, J.G.W., Grant, K.W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.*, 125, 3358–72.

Bernstein, J.G.W., Summers, V., Iyer, N., et al. (2012). Set-size procedures for controlling variations in speech-reception performance with a fluctuating masker. *J. Acoust. Soc. Am.*, 132, 2676–89.

Best, V., Gallun, F.J., Ihlefeld, A., et al. (2006). The influence of spatial separation on divided listening. *J. Acoust. Soc. Am.*, 120, 1506–1516.

Blandy, S.E., Lutman, M.E. (2005). Hearing threshold levels and speech recognition in noise in 7-year-olds. *Int. J. Audiol.*, 44, 435–443.

Blyth, M. (2013). *Listening effort in different masker types measured by pupil dilation*. University of Southampton.

Bonino, A.Y., Leibold, L.J., Buss, E. (2013). Release from perceptual masking for children and adults: benefit of a carrier phrase. *Ear Hear.*, 34, 3–14.

Boyle, R., Coltheart, V. (1996). Effects of irrelevant sounds on phonological coding in reading comprehension and short-term memory. *Q. J. Exp. Psychol. A.*, 49, 398–416.

Brady, S., Shankweiler, D., Mann, V. (1983). Speech perception and memory coding in relation to reading ability. *J. Exp. Child Psychol.*, 35, 345–67.

Brand, T., Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J. Acoust. Soc. Am.*, 111, 2801–2810.

Bregman, A.S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press.

Bronkhorst, A.W. (2000). The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions. *Acustica*, 86, 117–128.

Brungart, D.S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.*, 109, 1101.

Brungart, D.S., Simpson, B.D. (2004). Within-ear and across-ear interference in a dichotic cocktail party listening task: effects of masker uncertainty. *J. Acoust. Soc. Am.*, 115, 301–310.

Brungart, D.S., Simpson, B.D., Ericson, M. a, et al. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J. Acoust. Soc. Am.*, 110, 2527–2538.

Calandruccio, L., Buss, E., Hall, J.W. (2014). Effects of linguistic experience on the ability to benefit from temporal and spectral masker modulation. *J. Acoust. Soc. Am.*, 135, 1335–43.

Carhart, R., Tillman, T.W., Greetis, E.S. (1969). Perceptual Masking in Multiple Sound Backgrounds. *J. Acoust. Soc. Am.*, 45, 694.

Cassidy, G., MacDonald, R.A.R. (2007). The effect of background music and background noise on the task performance of introverts and extraverts. *Psychol. Music*, 35, 517–537.

Cattermole, C. (2003). *Comparison of speech recognition in noise with sound field and earphone presentation*. University of Southampton, UK.

Cauchard, F., Cane, J.E., Weger, U.W. (2012). Influence of Background Speech and Music in Interrupted Reading: An Eye-Tracking Study. *Appl. Cogn. Psychol.*, 26, 381–390.

Chan, D., Fourcin, A., Gibbon, D., et al. (1995). EUROM - A spoken language resource for the EU. In *European Conference on Speech Communication and Technology*. (*pp. 867–870*).

Chenoweth, N. a., Hayes, J.R. (2003). The Inner Voice in Writing. *Writ. Commun.*, 20, 99–118.

Cherry, E.C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *J. Acoust. Soc. Am.*, 25, 975.

Chomsky, N. (1959). A note on phrase structure grammars. *Inf. Control*, 2, 393–395.

Christiansen, C., Dau, T. (2012). Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise. *J. Acoust. Soc. Am.*, 132, 1655–66.

Colle, H.A. (1980). Auditory encoding in visual short-term recall: effects of noise intensity and spatial location. *J. Verbal Learning Verbal Behav.*, 19, 722–735.

Colle, H.A., Welsh, A. (1976). Acoustic masking in primary memory. *J. Verbal Learning Verbal Behav.*, 15, 17–31.

Conway, A.R. a., Cowan, N., Bunting, M.F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychon. Bull. Rev.*, 8, 331–335.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *J. Acoust. Soc. Am.*, 119, 1562.

Cowen, N. (2005). *Working Memory Capacity*, Hove, UK: Psychology Press.

Cutler, A., Clifton, C. (2000). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort, eds. *The neurocognition of language*. (*pp. 125–166*). Oxford, UK: Oxford University Press.

Dai, H.P., Scharf, B., Buus, S. (1991). Effective attenuation of signals in noise under focused attention. *J. Acoust. Soc. Am.*, 89, 2837–42.

Dale, E., Reichert, D.H. (1957). *Bibliography of vocabulary studies*, Bureau of Educational Research, Ohio State University.

DeCasper, A.J., Fifer, W.P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*, 208, 1174–6.

Desimone, R., Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, 18, 193–222.

Dirks, D.D. (1986). Articulation index predictions of contextually dependent words. *J. Acoust. Soc. Am.*, 80, 82.

Dubno, J.R., Ahlstrom, J.B., Horwitz, A.R. (2000). Use of context by young and aged adults with normal hearing. *J. Acoust. Soc. Am.*, 107, 538.

Dubno, J.R., Horwitz, A.R., Ahlstrom, J.B. (2003). Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing. *J. Acoust. Soc. Am.*, 113, 2084.

Dunn, L.M., Dunn, L.M., Whetton, C., et al. (1997). *The British Picture Vocabulary Scale: Second edition Manual* 2nd ed. L. M. Dunn, L. M. Dunn, C. Whetton, et al., eds., London: GL Assessment Limited.

Duquesnoy, a J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *J. Acoust. Soc. Am.*, 74, 739–743.

Durlach, N. (2006). Auditory masking: Need for improved conceptual structure. *J. Acoust. Soc. Am.*, 120, 1787–1790.

Durlach, N.I., Mason, C.R., Shinn-Cunningham, B.G., et al. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *J. Acoust. Soc. Am.*, 114, 368.

Edwards, B. (2007). The future of hearing aid technology. *Trends Amplif.*, 11, 31–45.

Eggermont, J.J., Brown, D.K., Ponton, C.W., et al. (1996). Comparison of distortion product otoacoustic emission (DPOAE) and auditory brain stem response (ABR) traveling wave delay measurements suggests frequency-specific synapse maturation. *Ear Hear.*, 17, 386– 94.

Eisenberg, L.S., Dirks, D.D., Bell, T.S. (1995). Speech Recognition in Amplitude-Modulated Noise of Listeners With Normal and Listeners With Impaired Hearing. *J. Speech Lang. Hear. Res.*, 38, 222.

Eisenberg, L.S., Shannon, R. V, Martinez, a S., et al. (2000). Speech recognition with reduced spectral cues as a function of age. *J. Acoust. Soc. Am.*, 107, 2704–10.

Elliott, L.L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *J. Acoust. Soc. Am.*, 66, 651.

Eysenck, H.J. (1967). *The Biological Basis of Personality* C. C. Thomas, ed., Springfield, IL:

Fallon, M., Trehub, S.E., Schneider, B.A. (2000). Children's perception of speech in multitalker babble. *J. Acoust. Soc. Am.*, 108, 3023–3029.

Feldman, J. (2003). What is a visual object? *Trends Cogn. Sci.*, 7, 252–256.

Festen, J.M., Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J. Acoust. Soc. Am.*, 88, 1725–36.

Francis, A.L. (2010). Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech. *Atten. Percept. Psychophys.*, 72, 501–516.

Freyman, R.L., Balakrishnan, U., Helfer, K.S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J. Acoust. Soc. Am.*, 115, 2246.

Freyman, R.L., Helfer, K.S., McCall, D.D., et al. (1999). The role of perceived spatial separation in the unmasking of speech. *J. Acoust. Soc. Am.*, 106, 3578–3588.

Furnham, A., Bradley, A. (1997). Music while you work: The differential distraction of background music on the cognitive test performance of introverts and extraverts. *Appl. Cogn. Psychol.*, 11, 445–455.

Garadat, S.N., Litovsky, R.Y. (2007). Speech intelligibility in free field: Spatial unmasking in preschool children. *J. Acoust. Soc. Am.*, 121, 1047.

Gathercole, S.E., Baddeley, A.D. (1993). Language Comprehension. In *Working memory and language*. New York, NY: Psychology Press.

Gathercole, S.E., Pickering, S.J., Ambridge, B., et al. (2004). The structure of working memory from 4 to 15 years of age. *Dev. Psychol.*, 40, 177–90.

George, E.L.J., Festen, J.M., Houtgast, T. (2006). Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.*, 120, 2295.

Gomes, H., Molholm, S., Christodoulou, C., et al. (2000). The Development of Auditory Attention in Children. *Front. Biosci.*, 5, d180–d120.

Gustafsson, H.A., Arlinger, S. (1994). Masking of speech by amplitude-modulated noise. *J. Acoust. Soc. Am.*, 95, 518.

Halin, N., Marsh, J.E., Haga, A., et al. (2014). Effects of speech on proofreading: can taskengagement manipulations shield against distraction? *J. Exp. Psychol. Appl.*, 20, 69–80.

Hall, J.W., Buss, E., Grose, J.H., et al. (2012). Effects of age and hearing impairment on the ability to benefit from temporal and spectral modulation. *Ear Hear.*, 33, 340–8.

Hall, J.W., Buss, E., Grose, J.H. (2005). Informational masking release in children and adults. *J. Acoust. Soc. Am.*, 118, 1605.

Hall, J.W., Grose, J.H. (1994). Development of temporal resolution in children as measured by the temporal modulation transfer function. *J. Acoust. Soc. Am.*, 96, 150–4.

Hall, J.W., Grose, J.H., Buss, E., et al. (2002). Spondee Recognition in a Two-Talker Masker and a Speech-Shaped Noise Masker in Adults and Children. *Ear Hear.*, 159–165.

Hawley, M.L., Litovsky, R.Y., Culling, J.F. (2004). The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *J. Acoust. Soc. Am.*, 115, 833–843.

Hazan, V., Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *J. Phon.*, 28, 377–396.

Heald, S.L.M., Nusbaum, H.C. (2014). Speech perception as an active cognitive process. *Front. Syst. Neurosci.*, 8, 35.

Helfer, K.S., Freyman, R.L. (2005). The role of visual speech cues in reducing energetic and informational masking. *J. Acoust. Soc. Am.*, 117, 842–849.

Henderson, M.T., Crews, A., Barlow, J. (1945). A study of the effect of music distraction on reading efficiency. *J. Appl. Psychol.*, 29, 313.

Hogan, T.P., Catts, H.W., Little, T.D. (2005). The relationship between phonological awareness and reading: implications for the assessment of phonological awareness. *Lang. Speech. Hear. Serv. Sch.*, 36, 285–293.

Howard-Jones, P.A., Rosen, S. (1993). Uncomodulated glimpsing in “checkerboard” noise. *J. Acoust. Soc. Am.*, 93, 2915–2922.

Ihlefeld, A., Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a divided speech identification task. *J. Acoust. Soc. Am.*, 123, 4380–92.

Inhoff, a W., Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: effects of word frequency. *Percept. Psychophys.*, 40, 431–9.

ISO_389-8 (2004). Acoustics -- Reference zero for the calibration of audiometric equipment -- Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones.

Jensen, K.K., Bernstein, J.G. (2013). The fluctuating-masker benefit for normal-hearing and hearing-impaired listeners with equal audibility at a fixed signal-to-noise ratio., 19, 060265–060265.

Jin, S.-H., Nelson, P.B. (2006). Speech perception in gated noise: The effects of temporal resolution. *J. Acoust. Soc. Am.*, 119, 3097.

Johansson, R., Holmqvist, K., Mossberg, F., et al. (2011). Eye movements and reading comprehension while listening to preferred and non-preferred study music. *Psychol. Music*, 40, 339–356.

Johnstone, P.M., Litovsky, R.Y. (2006). Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults. *J. Acoust. Soc. Am.*, 5.

Jones, D.M., Miles, C., Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Appl. Cogn. Psychol.*, 4, 89–108.

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

Kalikow, D.N., Stevens, K.N., Elliot, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J. Acoust. Soc. Am.*, 61, 1337–1351.

Kidd, G., Mason, C.R., Gallun, F.J. (2005). Combining energetic and informational masking for speech identification. *J. Acoust. Soc. Am.*, 118, 982.

Kidd Jr., G., Mason, C.R., Richards, Virginia, M., et al. (2007). Informational Masking. In W. A. Yost, A. N. Popper, & R. R. Fay, eds. *Auditory Perception of Sound Sources*. Springer Handbook of Auditory Research. (*pp. 143–189*). Boston, MA: Springer US.

Kiger, D.M. (1989). Effects of Music Information Load on a Reading Comprehension Task. *Percept. Mot. Skills*, 69, 531–534.

Kingdom, F.A.A., Prins, N. (2009). *Psychophysics: A Practical Introduction*, Academic Press.

Kirkby, J. a, Blythe, H.I., Drieghe, D., et al. (2011). Reading text increases binocular disparity in dyslexic children. *PLoS One*, 6, e27105.

Kleiman, G.M. (1975). Speech recoding in reading. *J. Verbal Learning Verbal Behav.*, 14, 323–339.

Koelewijn, T., Zekveld, A. a, Festen, J.M., et al. (2012a). Processing load induced by informational masking is related to linguistic abilities. *Int. J. Otolaryngol.*, 2012, 1–11.

Koelewijn, T., Zekveld, A.A., Festen, J.M., et al. (2012b). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear Hear.*, 33, 291.

Kuperman, V., Stadthagen-Gonzalez, H., Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behav. Res. Methods*, 44, 978–90.

Leech, R., Aydelott, J., Symons, G., et al. (2007). The development of sentence interpretation: effects of perceptual, attentional and semantic interference. *Dev. Sci.*, 10, 794–813.

Leek, M.R. (2001). Adaptive procedures in psychophysical research. *Atten. Percept. Psychophys.*, 63, 1279–92.

Leibold, L.J. (2012). Development of Auditory Scene Analysis and Auditory Attention. In L. Werner, R. Fay, Richard, & N. Popper, Arthur, eds. *Human Auditory Development*. (*pp. 137–161*). Springer.

Leibold, L.J., Bonino, A.Y. (2009). Release from informational masking in children: effect of multiple signal bursts. *J. Acoust. Soc. Am.*, 125, 2200–8.

Leibold, L.J., Bonino, A.Y., Fleenor, L. (2007). The Importance of Establishing a Time Course for Typical Auditory Development. In *A sound Foundation Through Early Amplification*. (*pp. 35–42*).

Leibold, L.J., Buss, E. (2013). Children's identification of consonants in a speech-shaped noise or a two-talker masker. *J. Speech. Lang. Hear. Res.*, 56, 1144–55.

Leinenger, M. (2014). Phonological Coding During Reading. *Psychol. Bull.*

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.*, 49, Suppl 2:467+.

Levy, B.A. (1978). Speech processing during reading. In A. M. Lesgold, J. W. Pellegrino, J. W. Fokkema, et al., eds. *Cognitive psychology and instruction*. New York: Plenum.

Lewis, D., Hoover, B., Choi, S., et al. (2010). Relationship between speech perception in noise and phonological awareness skills for children with normal hearing. *Ear Hear.*, 31, 761–8.

Litovsky, R.Y. (2005). Speech intelligibility and spatial release from masking in young children. *J. Acoust. Soc. Am.*, 117, 3091.

Lorenzi, C., Husson, M., Ardoint, M., et al. (2006). Speech masking release in listeners with flat hearing loss: effects of masker fluctuation rate on identification scores and phonetic feature reception. *Int. J. Audiol.*, 45, 487–495.

Lovett, R.E.S., Kitterick, P.T., Huang, S., et al. (2012). The developmental trajectory of spatial listening skills in normal-hearing children. *J. Speech. Lang. Hear. Res.*, 55, 865–78.

Lutfi, R.A., Kistler, D.J., Oh, E.L., et al. (2003). One factor underlies individual differences in auditory informational masking within and across age groups. *Percept. Psychophys.*, 65, 396–406.

MacKay, D.G. (1992). Constraints on Theories of Inner Speech. In D. Reisberg, ed. *Auditory Imagery*. Hillsdale, New Jersey: Lawrence Erlbaum Assiciates, Inc.

Macleod, A., Summerfield, Q. (1990). A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *Br. J. Audiol.*

MacPherson, A. (2013). *The factors affectng the psychometric function for speech intelligibility*. University of Strathclyde.

MacPherson, A., Akeroyd, M. a (2014). Variations in the Slope of the Psychometric Functions for Speech Intelligibility: A Systematic Survey. *Trends Amplif.*, 18.

Maddox, R.K., Shinn-Cunningham, B.G. (2012). Influence of task-relevant and task-irrelevant feature continuity on selective auditory attention. *J. Assoc. Res. Otolaryngol.*, 13, 119–29.

Markham, D., Hazan, V. (2004). The effect of talker- and listener-related factors on intelligibility for a real-word, open-set perception test. *J. Speech. Lang. Hear. Res.*, 47, 725–37.

Marsh, J.E., Hughes, R.W., Jones, D.M. (2009). Interference by process, not content, determines semantic auditory distraction. *Cognition*, 110, 23–38.

Marslen-Wilson, W.D., Welsh, A. (1978). Processing Interactions Word Recognition and Lexical Access during in Continuous Speech. *Cogn. Psychol.*, 10, 29–63.

Martin, R.C., Wogalter, M.S., Forlano, J.G. (1988). Reading comprehension in the presence of unattended speech and music. *J. Mem. Lang.*, 27, 382–398.

MATLAB (2012). MATLAB and Statistics Toolbox Release 2012b.

Mattys, S.L., Davis, M.H., Bradlow, A.R., et al. (2012). Speech recognition in adverse conditions: A review. *Lang. Cogn. Process.*, 27, 953–978.

McClelland, J.L., Elman, J.L. (1986). The TRACE model of speech perception. *Cogn. Psychol.*, 18, 1–86.

McCormick, C.E., Signer, S.B., Duncan, S. (1994). KINDERGARTEN PREDICTORS OF FIRST-GRADE READING ACHIEVEMENT: A REGULAR CLASSROOM SAMPLE. *Psychol. Rep.*, 74, 403–407.

McCreery, R., Ito, R., Spratford, M., et al. (2010). Performance-intensity functions for normalhearing adults and children using computer-aided speech perception assessment. *Ear Hear.*, 31, 95–101.

Miller, G.A. (1947). The Masking of Speech. *Psychol. Bull.*, 44, 105–129.

Miller, G.A., Licklider, J.C.R. (1948). The Intelligibility of Interrupted Speech. *J. Acoust. Soc. Am.*, 20, 593.

Mlot, S., Buss, E., Hall, J.W. (2010). Spectral integration and bandwidth effects on speech recognition in school-aged children and adults. *Ear Hear.*, 31, 56–62.

Moore, B.C.J. (2008). *Cochlear hearing loss: Physiological, psychological and technical issues.* 2nd ed., West Sussex: Wiley and Sons Ltd.

Moore, B.C.J., Glasberg, B.R. (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hear. Res.*, 28, 209–225.

Moore, B.C.J., Gockel, H.E. (2012). Properties of auditory stream formation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 367, 919–31.

Moore, J.K., Linthicum, F.H. (2007). The human auditory system: a timeline of development. *Int. J. Audiol.*, 46, 460–78.
Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychol. (Amst).*, 27, 84–92.

Neff, D.L., Green, D.M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Percept. Psychophys.*, 41, 409–15.

Neumann, K., Baumeister, N., Baumann, U., et al. (2012). Speech audiometry in quiet with the Oldenburg Sentence Test for Children. *Int. J. Audiol.*, 51, 157–63.

Nittrouer, S. (2002). From Ear to Cortex: A Perspective on What Clinicians Need to Understand About Speech Perception and Language Processing. *Lang. Speech. Hear. Serv. Sch.*, 33, 237–252.

Nittrouer, S., Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *J. Acoust. Soc. Am.*, 87, 2705.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, 52, 189–234.

Nunnally, J.C., Bernstein, I.H. (1994). *Psychometric theory* 3rd editio., New York, NY: McGraw-Hill.

Oh, E.L., Wightman, F., Lutfi, R.A. (2001). Children's detection of pure-tone signals with random multitone maskers. *J. Acoust. Soc. Am.*, 109, 2888.

Oswald, C.J., Tremblay, S., Jones, D.M. (2000). Disruption of comprehension by the meaning of irrelevant sound. *Memory*, 8, 345–50.

Oxenham, A.J., Simonson, A.M. (2009). Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. *J. Acoust. Soc. Am.*, 125, 457–68.

Papso, C.F., Blood, I.M. (1989). Word Recognition Skills of Children and Adults in Background Noise. *Ear Hear.*, 10, 235–236.

Parfect, V. (2002). *Equivalence Testing of the BKB and IHR sentence lists*. University of Southampton.

Peters, R.W., Moore, B.C., Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *J. Acoust. Soc. Am.*, 103, 577–587.

Pichora-Fuller, M.K., Schneider, B.A., Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.*, 97, 593.

Pichora-Fuller, M.K., Singh, G. (2006). Effects of Age on Auditory and Cognitive Processing: Implications for Hearing Aid Fitting and Audiologic Rehabilitation. *Trends Amplif.*, 10, 29–59.

Plomp, R. (1977). Acoustical aspects of cocktail parties. *Acustica*, 38, 186–191.

Pollack, I. (1975). Auditory Informational Masking. *J. Acoust. Soc. Am.*, 57, S5–S5.

Prins, N., Kingdom, F.A.A. (2009). Palamedes: Matlab routines for analysing psychophysical data. Available at: http://www.palamedestoolbox.org.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol. (Hove).*, 62, 1457–506.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.*, 124, 372–422.

Rayner, K., Liversedge, S.P. (2011). Linguistic and cognitive influences on eye movements during reading. In *The Oxford Handbook of Eye Movements*. (*pp. 753–766*).

Rayner, K., McConkie, G.W., Ehrlich, S. (1978). Eye movements and integrating information across fixations. *J. Exp. Psychol. Hum. Percept. Perform.*, 4, 529–44.

Reichle, E.D. (2011). Serial-attention models of reading. In *The Oxford Handbook of Eye Movements*. (*pp. 767–786*).

Rhebergen, K.S., Versfeld, N.J. (2005). A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *J. Acoust. Soc. Am.*, 117, 2181.

Rönnberg, J., Lunner, T., Zekveld, A., et al. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.*, 7, 31.

Rosen, S., Souza, P., Ekelund, C., et al. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *J. Acoust. Soc. Am.*, 133, 2431–43.

Rudner, M., Lunner, T., Behrens, T., et al. (2012). Working Memory Capacity May Influence Perceived Effort during Aided Speech Recognition in Noise. *J. Am. Acad. Audiol.*, 23, 577–589.

Ryalls, B., Pisoni, D. (1997). The effect of talker variability on word recognition in preschool children. *Dev. Psychol.*, 33, 441–452.

Salamé, P., Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *J. Verbal Learning Verbal Behav.*, 21, 150–164.

Salamé, P., Baddeley, A. (1987). Noise, unattended speech and short-term memory. *Ergonomics*, 30, 1185–94.

Schmuziger, N., Probst, R., Smurzynski, J. (2004). Test-Retest Reliability of Pure-Tone Thresholds from 0.5 to 16 kHz using Sennheiser HDA 200 and Etymotic Research ER-2 Earphones. *Ear Hear.*, 25, 127–132.

Schneider, B. a, Li, L., Daneman, M. (2013). How competing speech interferes with speech comprehension in everyday listening situations. *J. Am. Acad. Audiol.*, 18, 559–72.

Schotter, E.R., Angele, B., Rayner, K. (2012). Parafoveal processing in reading. *Atten. Percept. Psychophys.*, 74, 5–35.

Shield, B., Dockrell, J.E. (2004). External and internal noise surveys of London primary schools. *J. Acoust. Soc. Am.*, 730–738.

Shinn-Cunningham, B. (2013). Understanding informational masking from a neural perspective. In *Proceedings of Meeting on Acoustics.* 19(1), 060143. Acoustical Society of America.

Shinn-Cunningham, B.G. (2008). Object-based auditory and visual attention. *Trends Cogn. Sci.*, 12, 182–6.

Siegel, L.S. (1989). IQ Is Irrelevant to the Definition of Learning Disabilities. *J. Learn. Disabil.*, 22, 469–478.

Simons, D.J., Rensink, R. a (2005). Change blindness: past, present, and future. *Trends Cogn. Sci.*, 9, 16–20.

Slowiaczek, M.L., Clifton, C. (1980). Subvocalization and reading for meaning. *J. Verbal Learning Verbal Behav.*, 19, 573–582.

Snowling, M., Goulandris, N., Bowlby, M., et al. (1986). Segmentation and speech perception in relation to reading skill: a developmental analysis. *J. Exp. Child Psychol.*, 41, 489–507.

Sörqvist, P., Halin, N., Hygge, S. (2010). Individual differences in susceptibility to the effects of speech on reading comprehension. *Appl. Cogn. Psychol.*, 24, 67–76.

Stickney, G.S., Zeng, F.-G., Litovsky, R., et al. (2004). Cochlear implant speech recognition with speech maskers. *J. Acoust. Soc. Am.*, 116, 1081.

Strelcyk, O., Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *J. Acoust. Soc. Am.*, 125, 3328–45.

Stuart, A. (2005). Development of Auditory Temporal Resolution in School-Age Children Revealed by Word Recognition. , 78–88.

Stuart, A. (2008). Reception Thresholds for Sentences in Quiet, Continuous Noise, and Interrupted Noise in School-Age Children. *J. Am. Acad. Audiol.*, 19, 135–146.

Summerfield, Q., Palmer, A.R., Foster, J.R., et al. (1994). Clinical evaluation and test-retest reliability of the IHR-McCormick Automated Toy Discrimination Test. *Br. J. Audiol.*, 28, 165–79.

Sussman, E., Wong, R., Horváth, J., et al. (2007). The development of the perceptual organization of sound by frequency separation in 5-11-year-old children. *Hear. Res.*, 225, 117–27.

Thrift, K. (2014). *Speech Perception in the Presence of a Competing Talker: Can Auditory Training Improve Ability in Children and Adults?* University of Southampton.

Torgesen, J.K., Wagner, R.K., Rashotte, C.A. (1999). *Examiner's Manual: Test of Word Reading Efficiency*, Austin, Texas: Pro-ed Inc.

Trahiotis, C., Bernstein, L.R., Buell, T.N., et al. (1990). On the use of adaptive procedures in binaural experiments. *J. Acoust. Soc. Am.*, 87, 1359–61.

Treiman, R., Clifton, C., J., Meyer, A.S., et al. (2003). Language comprehension and production. In A. F. Healy & R. W. Proctor, eds. *Comprehensive handbook of psychology, Vol. 4: Experimental Psychology .* (*pp. 527–548*). New York, NY: John Wiley & Sons, Inc.

Tremblay, S., Jones, D.M. (1999). Change of intensity fails to produce an irrelevant sound effect: implications for the representation of unattended sound. *J. Exp. Psychol. Hum. Percept. Perform.*, 25, 1005–15.

Vaillancourt, V., Laroche, C., Giguère, C., et al. (2008). Establishment of age-specific normative data for the canadian French version of the hearing in noise test for children. *Ear Hear.*, 29, 453–66.

Venetjoki, N., Kaarlela-Tuomaala, a, Keskinen, E., et al. (2006). The effect of speech and speech intelligibility on task performance. *Ergonomics*, 49, 1068–91.

Warren, R.M. (1970). Perceptual Restoration of Missing Speech Sounds. *Science (80-.).*, 167, 392–393.

Warren, T., McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychon. Bull. Rev.*, 14, 770–5.

Werner, L. a (2007). Issues in human auditory development. *J. Commun. Disord.*, 40, 275–83.

Whetton, C. (1997). *BPVS-II: Technical Supplement* 2nd ed., Windsor: GL Assessment Limited.

Wichmann, F. a, Hill, N.J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept. Psychophys.*, 63, 1293–313.

Wichmann, F. a, Hill, N.J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Percept. Psychophys.*, 63, 1314–29.

Wightman, F., Kistler, D., Brungart, D. (2006). Informational masking of speech in children: Auditory-visual integration. *J. Acoust. Soc. Am.*, 119, 3940.

Wightman, F.L., Callahan, M.R., Lutfi, R. a., et al. (2003). Children's detection of pure-tone signals: Informational masking with contralateral maskers. *J. Acoust. Soc. Am.*, 113, 3297.

Wightman, F.L., Kistler, D.J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *J. Acoust. Soc. Am.*, 118, 3164.

Wightman, F.L., Kistler, D.J., O'Bryan, A. (2010). Individual differences and age effects in a dichotic informational masking paradigm. *J. Acoust. Soc. Am.*, 128, 270–9.

Wilson, R.H., McArdle, R. (2012). Speech-in-noise measures: variable versus fixed speech and noise levels. *Int. J. Audiol.*, 51, 708–12.

Wilson, R.H., McArdle, R. a, Smith, S.L. (2007). An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss. *J. Speech. Lang. Hear. Res.*, 50, 844–56.

Zekveld, A.A., Kramer, S.E., Festen, J.M. (2011). Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response. *Ear Hear.*, 32, 498–510.

Zimmer, J.W., Brachulis-Raymond, J. (1978). Effects of distracting stimuli on complex information processing. *Percept. Mot. Skills*, 46, 791–794.