A developmental perspective on localisation of speech in babble

Rhiannon Brook BSc

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

October 2019

Abstract

The ability to localise speech in background sound is important for listening in real world environments. It is known that children can successfully localise in quiet at an adult-like level by 6 years of age, and children aged 7 to 16 years can localise in broadband noise at a similar level to adults, but are poorer than adults when listening in the presence of speech babble. This may compromise their listening abilities in real world environments, which could in turn affect their language acquisition and educational outcomes. This potential impact means it is important to know more about how children localise in babble.

This thesis includes the first studies to accurately quantify the developmental time course of minimum audible angle (MAA) and localisation of speech in babble abilities of participants aged 6-29 years. The relationships between localisation, detection and identification of speech in babble are also examined.

Firstly, in a series of experiments with 68 participants aged 6-29 years, the time course of typical development of localisation of speech in six talker babble was mapped using a five choice localisation task. It was found that performance could be fitted using an exponential function which had a time constant of 6.6 years. Using a variety of analytical methods, adult-like performance was found to occur in the early to mid-teenage years.

Next, the relationships between detection, identification and localisation of speech in babble were investigated. This was thought to be important because if a listener had difficulty with localisation it would be helpful to know

if they could detect and/or localise the stimulus. This was initially done in 40 participants aged 10-28 years, finding that participants performed best in identification, followed by localisation and detection which were similar. The study was then repeated with 17 participants aged 18-30 years, using more challenging SNRs and an additional localisation task. There were smaller differences between tasks but generally performance was highest on the detection task. Overall, it was found that even if a listener could detect and/or identify a word in babble, they could not necessarily localise it.

Finally, the developmental time course of MAA of speech in babble was mapped in 50 participants aged 6-21 years. Mean performance over all conditions could be fitted with an exponential function with a time constant of 5.8 years and using a variety of analytical methods, adult-like performance occurred in the early to mid-teenage years.

These are the first studies to have found the developmental time courses of localisation and MAA in babble in typically developing and normally hearing children age 6 years and older. These developmental time courses can be used as benchmarks to compare with populations who may have difficulties with spatial hearing in background sound, such as children with hearing impairments or with auditory processing disorder. These studies have also demonstrated the relationships between detection, identification and localisation of speech in babble in adults and children.

Acknowledgements

I would like to thank my supervisors Johanna Barry, Owen Brimijoin, Pádraig Kitterick and Michael Akeroyd for all of their help. I would also like to thank the MRC for funding my studentship. The research team acknowledges the support of the National Institute for Health Research Clinical Research Network (NIHR CRN). I am very grateful to the participants and their parents for taking part in the experiments. Finally, I would like to thank my family and especially Tim, as without his support none of this would have been possible.

Contents

Abstract II
Acknowledgements IV
Contents V
List of Abbreviations XI

Chapter 1
Literature Review 1
1.1 Introduction 1
1.2 The Physics of Spatial Hearing 3
1.2.1 Localisation 3
1.2.1.1 Monaural Cues 4
1.2.1.2 Binaural Cues 5
1.2.2 Methods of Measuring Localisation 6
1.2.2.1 Source discrimination 7
1.2.2.2 Source identification 7
1.3 The Development of Spatial Hearing 9
1.3.1 Neonates and Infants 12
1.3.2 Children 18
1.3.2.1 Localisation in quiet 18
1.3.2.2 Localisation of masked speech 21
1.3.2.3 Masked signals and Spatial Release from Masking 27
1.4 When is adult-like performance reached? 32
1.5 Summary 33

Chapter 2
Experiment 1: Localisation of speech in babble in typically developing children and adults 37
2.1 Introduction 37
2.2 Aims 38
2.3 Methods 40
2.3.1 Participants 40
2.3.1.1 Adults 41
2.3.1.2 Children 41

2.3.2 Equipment 43
2.3.3 Stimuli 46
2.3.3.1 Signals for the localisation tasks 46
2.3.3.2 Babble used for the localisation tasks 47
2.3.4 Procedure 47
2.3.4.1 Localisation tasks 49
2.3.4.2 Pure Tone Audiometry 52
2.3.4.3 Wechsler Abbreviated Scale of Intelligence (WASI) 53
2.3.4.4 Dichotic Digits 53
2.3.4.5 Speech Perception in Noise (SPIN) 54
2.3.4.6 Questionnaires 58
2.4 Results 58
2.4.1 Initial representation of results 59
2.4.2 Comparison of different age groups using a one-way ANOVA 61
2.4.3 Analysis of additional tasks 62
2.4.3.1 Analysis of the Speech Perception in Noise task (SPIN) 62
2.4.4 Analysis of number of tracks used 64
2.4.5 Analysis of gender differences 66
2.5 Discussion 68

Chapter 3
Experiment 2: Detection, Identification and Localisation of speech in babble in typically developing children and adults 71
3.1 Introduction 71
3.1.1 Investigating adult-like performance of localisation of speech in babble 71
3.1.2 Detection, identification and localisation of speech in babble 72
3.2 Aims and Hypotheses 75
3.3 Methods 76
3.3.1 Participants 76
3.3.1.1 Adults 77
3.3.1.2 Children 77
3.3.2 Equipment 78
3.3.3 Stimuli 78

3.3.3.1 Signal 78
3.3.3.2 Masker (babble) 79
3.3.4 Procedure 79
3.3.4.1 Main Localisation tasks 81
3.3.4.2 Detection, Identification and Localisation of speech in babble tasks ($DetB_{2CS}$, $IdentB_{4CS}$, $LocalB_{5CS}$) 83
3.4 Results 86
3.4.1 Localisation of speech in babble adaptive track results ($LocalB_{5AT}$) 87
3.4.1.1 Initial representation of results 87
3.4.1.2 Comparison of different age groups using a one-way ANOVA 88
3.4.2 Detection, identification, localisation of speech in babble results ($DetB_{2CS}$, $IdentB_{4CS}$, $LocalB_{5CS}$) 89
3.4.2.1 Initial signal detection theory analysis (d') 94
3.4.2.2 Bootstrap analysis of d' data 100
3.4.3 Case study: A child with localisation difficulties 105
3.5 Discussion 107
3.5.1 Adaptive track results 107
3.5.2 Detection, identification, localisation of speech in babble ($DetB_{2CS}$, $IdentB_{4CS}$, $LocalB_{5CS}$) 109
3.5.3 Limitations and future work 111

Chapter 4
Analysis of the combined results from Experiments 1 and 2 113
4.1 Introduction 113
4.1.1 How is adult-like performance defined? 113
4.2 Results 117
4.2.1 Initial representation of results from $LocalB_{5AT}$ tasks from Experiments 1 and 2 117
4.2.2 Adult-like performance: Exponential function (direct fit) 118
4.2.3 Adult-like performance: Broken stick analysis 121
4.2.4 Adult-like performance: Comparison of different age groups using an ANOVA 123
4.2.5 Adult-like performance: Exponential function (bootstrap) 125
4.2.6 Summary of results from $LocalB_{5AT}$ 127

4.2.7 Results from additional tasks 128
4.2.7.1 Age-norming of $LocalB_{5AT}$ scores 128
4.2.7.2 Localisation in quiet ($LocalQ_{5CS}$) 131
4.2.7.3 Dichotic Digits 134
4.2.7.4 Wechsler Abbreviated Scale of Intelligence (WASI) 138
4.2.7.5 Digit Span 142
4.2.7.6 Conners Questionnaire 145
4.2.7.7 Evaluation of Children's Listening and Processing Skills (ECLiPS) Questionnaire 148
4.2.7.8 Summary of results from additional tasks 149
4.3 Discussion 151
4.3.1 Developmental time course of localisation of speech in babble . 151
4.3.1.1 Adult like performance levels for localisation of speech in babble 151
4.3.1.2 Why does localisation of speech in babble improve with age? 153
4.3.2 Limitations and future work 157

Chapter 5
Experiment 3: Detection, Identification and Localisation of Speech in Babble in Young Adults 159
5.1 Introduction 159
5.2 Aims and Hypotheses 161
5.3 Methods 162
5.3.1 Participants 162
5.3.2 Equipment 162
5.3.3 Stimuli 165
5.3.3.1 Signal 165
5.3.3.2 Masker 165
5.3.4 Procedure 165
5.3.4.1 Study tasks 165
5.4 Results 167
5.4.1 Overview of results 167
5.4.2 Comparison of task performances 168
5.4.3 d' analysis 170

5.5 Discussion 181
5.5.1 Comparison to findings from Experiment 2 181
5.5.2 Summary and conclusions 183
5.5.3 Limitations and future work 186

Chapter 6
Experiment 4: Minimum Audible Angle in Babble in Children and Young Adults 187
6.1 Introduction 187
6.2 Aims and Hypotheses 189
6.3 Methods 190
6.3.1 Participants 190
6.3.1.1 Adults 190
6.3.1.2 Children 191
6.3.2 Equipment 191
6.3.3 Stimuli 192
6.3.3.1 Accuracy of the Panning 192
6.3.3.2 Signal 200
6.3.3.3 Masker 200
6.3.4 Procedure 200
6.3.4.1 Study tasks 201
6.4 Results 204
6.4.1 Main summary of results 206
6.4.2 Comparing results between conditions 212
6.4.3 Differences between adults and children 214
6.4.4 Adult-like performance: Exponential function (direct fit) 219
6.4.5 Adult-like performance: Broken stick analysis 219
6.4.6 Adult-like performance: Exponential function (bootstrap) 225
6.4.7 Comparison of different analysis methods 231
6.4.8 Comparison to results from LocalB_{5AT} task (Experiments 1 and 2) 233
6.4.9 Additional results of children who had difficulty with the MAA task 234
6.5 Discussion 236

6.5.1 Comparison to other developmental time courses (MAA in quiet and masked speech perception) 237
6.5.2 Comparison to results from Experiments 1 and 2 243
6.5.3 Summary of results and conclusions 245
6.5.4 Limitations and future work 246

Chapter 7
Conclusions 249
7.1 Overview 249
7.2 Key Findings 250
7.2.1 Development of localisation of speech in babble 250
7.2.2 Detection, identification and localisation of speech in babble 253
7.3 Limitations and future work 254
7.3.1 Experiment specific limitations 254
7.3.2 General limitations 256
7.3.3 Future work 257
7.4 Summary 260

References 261

List of Abbreviations

ADHD	Attention Deficit Hyperactivity Disorder
BBN	Broadband Noise
BMLD	Binaural Masking Level Difference
CRM	Coordinate Response Measure
ECLiPS	Evaluation of Children's Listening and Processing Skills
HRTF	Head Related Transfer Function
IEEE	Institute of Electrical and Electronics Engineers
IHR	Institute of Hearing Research
ILD	Interaural Level Difference
ITD	Interaural Time Difference
MAA	Minimum Audible Angle
MRC	Medical Research Council
QMC	Queen’s Medical Centre
RMS	Root Mean Square
SNR	Signal to Noise Ratio
SPIN	Speech Perception in Noise
SRM	Spatial Release from Masking
WASI	Wechsler Abbreviated Scale of Intelligence

Chapter 1

Literature Review

1.1 Introduction

A person's ability to determine the direction that a sound is coming from is a key component of their awareness of their surroundings, which contributes to their safety. For example, when walking through a busy car park it is beneficial to be able to localise the sounds of nearby vehicles in order to safely cross roads. Binaural cues enable spatial hearing and therefore localisation of sounds, which help listeners to hear better in challenging environments. Listening to speech in a noisy environment is a common occurrence for most people, with or without a hearing loss. Typical examples include social environments such as coffee shops and pubs, and in workplaces such as open plan offices and at conferences. Other benefits of binaural hearing include: hearing quieter sounds, less tiredness from listening, reduced negative effects from sudden loud sounds and smoother tone quality (Haggard, 1993).

It is known that localisation abilities mature in early childhood in quiet environments, but localisation in background sound is more challenging for children. Children spend a lot of time in noisy environments, such as classrooms and playgrounds, and this can be a significant problem for them as they are still learning and developing their listening and language abilities.

Background sound in and around the classroom has been shown to have a negative impact on the academic performance of children aged 7 and 11 years, affecting reading ability, memory and cognitive development (Shield and Dockrell, 2008). Klatte, Lachmann and Meis (2010) demonstrated that in environments with unfavourable listening conditions, such as long reverberation times (over 1 s) and background sounds, listening comprehension and speech perception are negatively affected in children aged 6 to 10 years. They also found that children were often not aware of these negative effects on their listening ability. It is clear from these studies that listening to masked speech can cause significant problems for children. However, despite the fact that children spend so much time in noisy environments and the negative impacts of background sound are widely understood, currently little is known about the time course of the development of these localisation in background sound abilities.

This chapter outlines how spatial cues and therefore localisation abilities can help listeners to hear better in challenging environments. Methods of measuring localisation used in previous literature are then reviewed. Following this, the development of spatial hearing in children is discussed, including the impact of background sound. The meaning of adult-like performance and how this can be measured is also considered.

1.2 The Physics of Spatial Hearing

When listening to speech in a noisy environment a listener may need to localise the sound source of interest to improve listening ability. Localisation mainly relies on binaural cues, the use of which can be aided using head movements or orientations. Studies investigating localisation ability and head movements have found that localisation is improved when listeners can move their heads freely (Wallach, 1949; Wightman and Kistler, 1999). These head movements are beneficial because they allow listeners to make use of both monaural and binaural cues (cues which require sound from one and two ears respectively). The following section explains how these cues work and why they are important for localisation and orientation.

Three-dimensional localisation of a sound involves both the distance and direction of the sound source. A review by Kolarik et al. (2016) indicates that level, reverberation, spectral cues, binaural cues, dynamic cues and visual information can all be used to estimate the distance of a sound source. However, distance perception is beyond the scope of this thesis so only the directional aspect of localisation is considered.

1.2.1 Localisation

To describe the location of a sound, two coordinates (other than distance) need to be specified, the azimuth and the elevation of the sound. The azimuth defines the direction of the sound within the horizontal plane relative to straight ahead from the centre of the listener's head, shown in Figure 1.1. The elevation defines the location of the sound within the vertical plane relative to

straight ahead from the centre of the listener's head, shown in Figure 1.2. In order to perceive the azimuth and elevation, and so the full direction of a sound, listeners use both monaural and binaural cues.

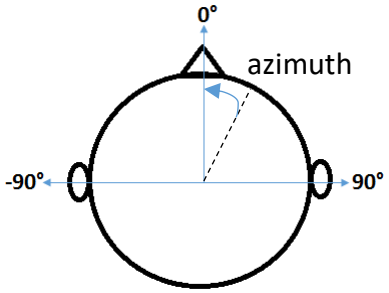


Figure 1.1: Azimuth is the direction of the sound in the horizontal plane, in this example it is approximately 30°

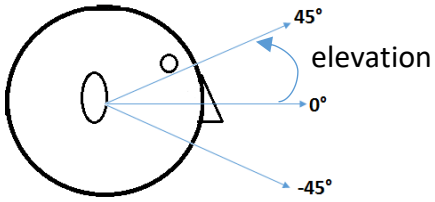


Figure 1.2: Elevation is the direction of the sound in the vertical plane, in this example it is ±45°

1.2.1.1 Monaural Cues

Monaural cues only require sound from one ear. The physical characteristics of the head and pinna and the direction from which the sound arrives alter the way sounds are perceived and the frequency of a sound also affects how it interacts with the head and pinna (Batteau, 1967; Shaw, 1974a, 1974b; Blauert 1969). This direction dependent filtering effect is known as the Head Related Transfer Function (HRTF). Sounds above 6 kHz have a short enough wavelength to interact with the pinna so are filtered and thus can provide useful localisation information. However, sounds with frequencies as low as

500 Hz can still provide monaural localisation information due to their interactions with the head and torso. Pinna cues can be useful for localisation in both the horizontal and vertical planes and can also help resolve front-back confusions for high frequency stimuli (Musicant and Butler, 1984; Middlebrooks and Green, 1991), which occur due to the cone of confusion (described below). Therefore these cues can be valuable for a listener, in conjunction with binaural cues, to aid localisation.

1.2.1.2 Binaural Cues

Binaural cues require the involvement of both ears. They are usually summarised by the "Duplex theory" which was established by Lord Rayleigh (Strutt, 1909). There are two different types of binaural cues: Interaural Time Differences (ITDs) and Interaural Level Differences (ILDs). An ITD is the difference in arrival time across the two ears for a sound and is useful for localising low frequency sounds (below 1.5 kHz). An ILD is the difference in level across the two ears for a sound and is useful for localising high frequency sounds (above 3 kHz). For example, if a sound originates from the left of a listener, it will arrive sooner and be more intense in the left ear. The auditory system can then use the ITD and ILD to determine that the sound is to the left of the listener. The duplex theory is that ITDs are used at low frequencies and that ILDs are used at high frequencies for localisation of sounds. If a listener alters their head orientation the subsequent change in ITDs and ILDs also helps them to better perceive the direction of the sound source (Wallach, 1940). This process underpins the benefit of head movements for sound localisation.

In some cases ITDs and ILDs can be ambiguous as sounds from multiple directions can produce the same ITDs and ILDs. This is referred to as the "cone of confusion" (Moore, 2012) as the ITD and ILD are identical for every direction (Figure 1.3). Identical ITDs will be present for all sound sources which are on the surface of the cone (for this illustration, the cone is drawn to a fixed distance, but in principle it extends to infinity). Common difficulties caused by the cone of confusion are front-back confusions, where the listener is unable to determine whether the sound is behind or in front of them (Brimijoin and Akeroyd, 2016). However, if a listener can move their head then they can overcome the cone of confusion and will be more likely to successfully localise the sound of interest (Brimijoin and Akeroyd, 2012). This shows the importance of using head movements for successful localisation and orientation.

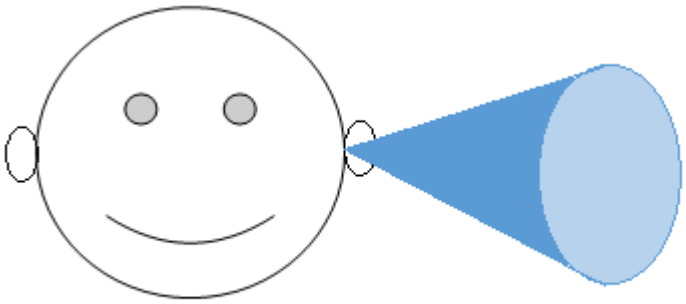


Figure 1.3: Cone of Confusion

1.2.2 Methods of Measuring Localisation

Different studies have used a variety of methods to measure localisation abilities. However, they can all be described as either source discrimination or source identification methods, which are explained in this section.

1.2.2.1 Source discrimination

Akeroyd and Whitmer (2016) provide a summary of the source discrimination method. In this method a listener hears two sounds, each from a different location, and must report the order in which they were presented. The difference between the two locations is varied between presentations. This variation can be made in a number of different ways, including: an 'adaptive track' to generate a psychometric function which gives a threshold, or a 'constant stimuli' task with randomly chosen or pre-chosen lists or locations. A threshold of either the asymptote of the track or an interpolated point on the psychometric function tells us the Minimum Audible Angle (MAA), which is the smallest difference in direction which a listener can detect. This MAA paradigm was first described by Mills (1958).

An example of when this method was used is Perrott and Saberi (1990), who measured MAAs for sound sources on the vertical and horizontal planes. Listeners heard sounds from two different locations and were asked whether the second sound was to the left/right (for presentations in the horizontal plane) or above/below (for presentations in the vertical plane) the first sound. They found that on the vertical plane the mean MAA of four adult participants was 3.65° and in the horizontal plane the mean MAA was 0.97°.

1.2.2.2 Source identification

Akeroyd and Whitmer (2016) summarised the source identification method, where a listener identifies which location a sound came from out of a choice of potential sound source locations. There are multiple outcome measures

which can be taken from this method. A commonly used method is Root Mean Square (RMS) error, where the mean across all trials is taken of the square of the differences between the actual sound location and the reported sound location of a trial and then its square root is calculated. Alternative methods are the percentage of trials answered correctly, mean error (difference between actual sound location and reported sound location), mean error over total number of trials, and the mean of the absolute error over the total number of trials.

An example of this method is described in Hartmann, Rakerd and Gaalaas (1998), where they used source identification to test their model of this method. Participants were asked to identify which loudspeaker, out of an array of either 3, 6, 12 or 24 loudspeakers, was the source of broadband noise. Performance on this task was measured using RMS error, variability and constant error. RMS error increased initially with number of speakers, but was not strongly influenced by number of speakers from 6 speakers onwards. Constant error was found to increase with increasing number of speakers and variability did not seem to vary with number of speakers used (mean results over all participants shown in Table 1.1).

Number of speakers	Mean RMS error	Mean variability	Mean constant error
3	4.6	4.1	2.0
6	5.5	4.2	3.5
12	6.8	4.2	5.4
24	6.9	3.4	6.0

Table 1.1: Table showing results from Hartmann, Rakerd and Gaalaas (1998)

1.3 The Development of Spatial Hearing

In this section, the development of spatial hearing skills will be discussed. The developmental time course of localisation skills is examined for neonates, infants and children. Children's spatial hearing in background sound abilities are also considered. Each study is described in detail, after an opening summary, together with tables of the key design parameters. Figure 1.4 shows a summary of the developmental localisation literature (the features of which are explained in Table 1.2). From this it can be seen that localisation in quiet (and broadband noise) abilities improve from neonates until early childhood, Spatial Release from Masking (SRM) abilities appear to be stable from about two or three years of age (except for in a study by Yuen and Yuan (2014)), and localisation in speech babble does not improve/reach adult-like levels until much later in the teenage years. Localisation in speech babble appears to be the only ability on the figure which shows improvement in the teenage years.

All of the studies discussed here used cross-sectional designs, apart from the neonate studies (Muir and Field 1979; Morrongiello et al., 1994 and Nemeth et al., 2015) which are not cross-sectional as they only look at one age range rather than several age groups. The main drawback to this study design is that you cannot know for sure if the differences between groups are down to the factor of interest (in this case age) or other differences between the

participants in each group. One of the ways to minimise the impact of this bias would be to match participants in each group for factors such as gender, socioeconomic status and IQ. However, this is time consuming and would likely result in the exclusion of many participants and therefore reduction in power of the experiment. Cross-sectional studies can also be prone to cohort effects, where differences between the groups can be due to differences between generations rather than age effects, however this is unlikely in the studies discussed below due to the small range of ages being tested. Longitudinal studies do not have these issues, but require a very long time period over which to test participants and are prone to attrition.

Colour	**Participant group**	**Study type**
Yellow	Neonates	Localisation in quiet
Blue	Infants	Localisation in quiet
Purple	Children	Localisation in quiet
Orange	Children	Spatial release from masking
Green	Children	Localisation of masked signals
Feature	**Meaning**	
Shaded box	Participants of approximately this age tested	
Dashed line	No improvement with age within age range tested	
Line	Upwards slanting line shows improvement with age (and downwards slanting line shows reducing performance with age)	

Table 1.2: Table explaining features of Figure 1.4

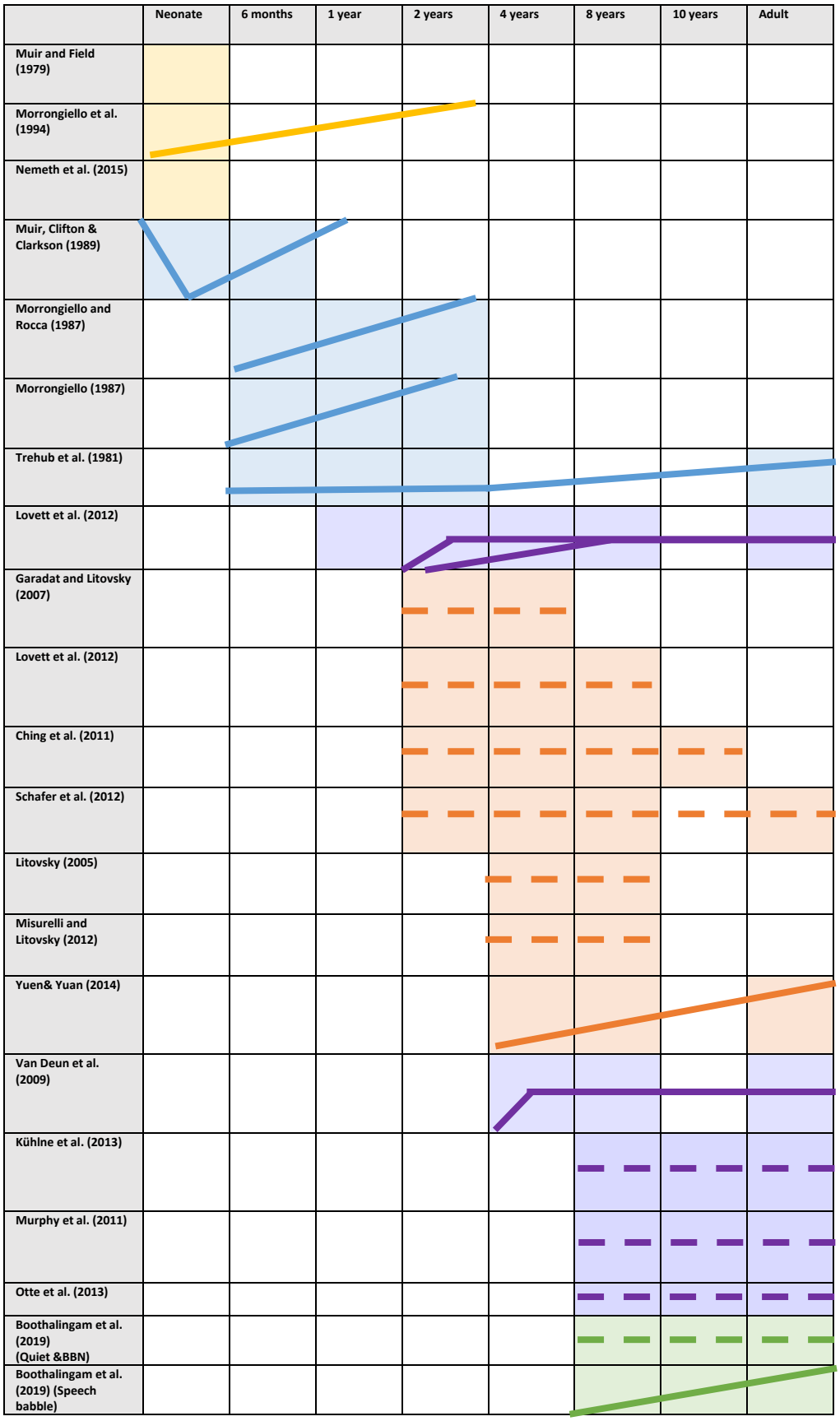


Figure 1.4: Figure summarising development of localisation literature (features explained in Table 1.2)

1.3.1 Neonates and Infants

In summary, the literature suggests that neonates have an innate but inaccurate ability for sound localisation. Although this ability briefly disappears (at around one to three months of age), localisation accuracy then improves throughout infancy (from 6 to 24 months of age), however these improvements may not be evident when speech is masked. The key design parameters of these studies are summarised in Table 1.3 and Table 1.4.

For neonates, Muir and Field (1979) carried out two experiments to investigate head turning responses in neonates to a long duration sound source (5-20 s) of approximately 70-85 dB SPL which was 90° from the left or right of the midline. They first tested 9 three-day-old infants and found that they correctly turned their heads on 87% of presentations. Following this, they changed the experiment to remove visual cues and so that the person holding the neonate did not know the location of the sound. They found that 12 three-day-old infants correctly turned their heads on 74% of presentations. Although this study does not prove the presence of an accurate localisation response in neonates, it does demonstrate that the ability to orient to sounds is present from shortly after birth. A study by Morrongiello et al., (1994) showed that newborns are probably capable of localisation in the horizontal plane, using both source identification and discrimination methods, but their accuracy is much poorer than that of a six-month-old (not measured, but derived from estimates for the localisation accuracy of six-month-old infants). They found that 34 neonates made head movements towards the presented

sounds on 30% of all presentations. Accuracy of head turns was also measured. In this study, the average error of head movements was 26°, compared to the literature data of 16° for six-month-olds and 7° for 18 month-olds (Morrongiello and Rocca, 1987). This decrease in error with age suggests that localisation accuracy improves with age from neonates until 18 months of age. The stimuli were white noise bursts with a duration of 500 ms and level of 66 dB SPL, which were presented from speakers were positioned at 36°, 54°, 72° and 90° on either side of the baby's midline.

Nemeth et al. (2015) used Electroencephalogram (EEG), to suggest that newborns have some localisation ability. They tested 68 newborns (aged one to three days), looking at their encoding of ITDs and ILDs, in both the free field and with monaural sound presentation. Their results suggested that newborns may be able to process cues necessary for localisation, but can't necessarily fully utilise them in the way adults can.

For infants, Muir, Clifton and Clarkson (1989) used head turning responses in the horizontal plane to show that localisation ability is significantly worse at one to three months of age than at birth, but then improves from four to seven months of age. They tested 104 infants aged from 3 days to 7 months. Participants heard a rattle sound with a level of 65 dB SPL and a duration of 20s from one of two loudspeakers, one 90° to the left, and one 90° to the right. The number of presentations in which participants turned towards the stimuli was recorded. This U-shaped developmental function found by Muir, Clifton and Clarkson (1989), showed that participants scored approximately

75% correct at birth, 50% at 2 months and over 90% by 7 months of age. This curious result was explained by Muir and Hains (2004), who later argued that the head turning response to sound seen in newborns is likely to be a reflex which disappears and that the head turning response seen at around four months of age is true auditory localisation.

Morrongiello and Rocca (1987) and Morrongiello (1987) tested infants aged 6 to 18 months and showed that localisation error in the azimuthal and vertical planes decreases with age. Morrongiello (1987) estimated the MAA in the median sagittal plane for 50 infants aged 6 to 18 months using loudspeakers positioned at 0°, 4°, 8°, 12° and 16° (so this was a source discrimination experiment). Three different stimuli were used, namely bandpass noise at <4 kHz, 4-8 kHz and 8-12 kHz. The proportion of correct responses improved on average (over all loudspeaker positions) between the ages of 6 and 18 months by 0.06 in the <4 kHz condition, 0.1 in the 4-8 kHz condition and by 0.25 in the 8-12 kHz condition. Morrongiello and Rocca (1987) tested localisation abilities in the horizontal plane of 80 infants aged 6, 9, 12, 15 and 18 months using a source identification method. They used loudspeakers at 18°, 36°, 54°, 72° and 90° from both left and right of midline to present the stimuli, which were clicks with a level of 65 dB SPL. Head tracking was used to measure performance. It was found that average error decreased from approximately 16° in 6 month olds to 7° in 18 month olds. Trehub et al., (1981) looked into the localisation of masked speech in abilities of adults and infants (118 participants, aged 6 to 24 months old). The signal used was a speech phrase ("Hi there") presented at a variety of signal levels from 23 to 63 dB, while the

masker was white noise presented at 42 and 60 dB. They measured thresholds of intensity levels corresponding to 65% correct. They found that localisation performance was significantly worse for infants than adults, with a difference of 10-12 dB. No significant effect of age was found within the infant group, suggesting that localisation for masked speech shows little improvement between the ages of 6 to 24 months.

Study		Muir and Field, 1979	Morrongiello et al., 1994	Nemeth et al., 2015		Muir, Clifton and Clarkson, 1989
Test		Left/right orienting response in alert newborns	Localisation in newborns	Measurement of Event-Related Brain Potentials (ERPs) using EEG to test sensitivity to binaural cues		Newborns tested as in Muir and Field (1979), others tested as shown below
Listeners	**Number**	**9;12**	**34**	**68**		**104**
	Age range	**2-7 days; 2-4 days**	**7-133 hours**	**1-3 days**		**8 groups: 3 days old and then 1, 2, 3, 4, 5, 6 and 7 months old**
	Mean age	3.2 days; 3.0 days	45.1 hours	-		3; 35; 61; 95; 128; 158; 186; 214 days
	SD age	-	24.9 hours	-		0; 4; 8; 9; 7; 7; 8; 5 days
	Gender (f;m)	6;6 and 6;6	20;14	32;36		5;9, 6;4, 6;4, 5;5, 9;11, 9;11, 6;4, 6;4
Inclusion criteria		Weight >2.5kg, gestational age 38-42 weeks, birth conditions criteria (no complications etc)	Birth weight>2.5 kg, gestational age 38-42 weeks, Apgar scores of ≥7	Healthy and born to term, Apgar ≥9.		-
Speakers	**Number**	2	8	2	Headphones in some presentations	2
	Distance	20cm	65cm	80cm		1.3m
Signal	**Type**	Rattle/recorded rattle	Sequence of white noise bursts	White noise segments		Recorded rattle
	Level	70-85 dB (units not given)	66 dB SPL	80 dB SPL		65 dB SPL
	Duration	5-20s	500ms each	200msec		Max 20s
	Presentation angles	±90°	±36°, ±54°, ±72°, ±90°	±45° and headphones		±90°
Test time (mins)		-	12-15 mins	-		-
Total presentations		5-8	16	-		8
Response		Head turn and maintained position for 5s	Head turn	ERP measurement on EEG		Head turn
Main findings		**Significantly more likely to turn to sound than to wrong side or not to move**	**Newborns shown some ability to localise sounds**	**Newborns may be sensitive to ITDs. Effect sizes: η_p^2=0.26, 0.19, 0.21, 0.16**		**U-shaped developmental function shown, localisation ability much poorer between 1 and 3 months of age than at birth, then improves gradually from 4 to 7 months**

Table 1.3: Table showing Newborn and Infant Localisation studies (Part 1) (Muir and Field, 1979; Mui, Clifton and Clarkson; Morrongiello et al., 1994; Nemeth et al., 2015)

Study		Morrongiello and Rocca, 1987	Morrongiello, 1987	Trehub et al., 1981
Test		Localisation in horizontal plane	Localisation in median sagittal plane (MSP)	Detection/Localisation of masked speech
Listeners	**Number**	**60 from 5 age groups**	**50 from 5 age groups**	**118; 28; 29; 34 infants and 20 adults**
	Age range	6, 9, 12, 15, 18 months	6, 9, 12, 15, 18 months	6, 12, 18, 24 months and adults
	Mean age	0.5; 0.75; 1; 1.3; 1.5 years	0.5; 0.73; 1.02; 1.23; 1.51 years	26.9; 53.1; 79.8; 113.3 weeks; adults age not given
	SD age (years)	16; 13; 13; 11; 14 days	16; 20; 18; 21; 22 days	-
	Gender (f;m)	-	At least 4f and 4m in each group	-
Exclusion criteria		History of ear infections, cold on day of testing	Not full term and healthy	Not full term and healthy, cold on day of testing
Speakers	**Number**	10	9	2
	Distance (m)	1.3	2.3	-
Practice(and feedback, if any)		30	4	Training carried out at 73 dB and audio-visual reward for correct head turn
Signal	**Type**	Clicks x8	White noise bursts	Speech “Hi there”
	Frequency	-	4 kHz and lower, 4-8 kHz and 8-12 kHz.	-
	Level	65 dB C	65 dB (Lin)	With 42 dB masker: 23, 28, 33 and 63 dB (units not specified) With 60 dB masker: 38, 43, 48, 53 and 63 dB (units not specified)
	Duration	-	500ms	1s
	Presentation angles	±18°, ±36°, ±54°, ±72°, ±90°	MSP: 0°, ±4°, ±8°, ±12°, ±16°	±45°
Masker	**Type**	-	-	White noise
	Frequency	-	-	-
	Level	-	-	42 and 60 dB C
	Duration	-	-	-
	Presentation angles	-	-	±45°
Test time (mins)		10	30	-
Total presentations		-	24	25
Response		Head turn	Head turn	Head turn
Main findings		**Localisation improves with age from 6-18 months**	**Localisation in the MSP improves from 6-18 months**	**Localisation significantly worse for infants than adults but no age effect within infant group**

Table 1.4: Table showing Newborn and Infant Localisation studies (Part 2) (Treuhub et al., 1981; Morrongiello, 1987; Morrongiello and Rocca, 1987)

1.3.2 Children

1.3.2.1 Localisation in quiet

It is well established that children have adult-like localisation in quiet abilities from the age of five or six years (Table 1.5, Table 1.6 and Table 1.7). These studies are described in detail below.

Kühlne et al. (2013) tested localisation accuracy in the horizontal plane and frontal hemisphere of 129 participants aged 6 to 18 years old. They found that localisation accuracy of Gaussian noise bursts with a level of 35 dB SL and a duration of 250 ms was stable for children aged from six years old. These results are in agreement with Murphy et al. (2011) who tested 40 children aged 6 to 14 years on a source identification task in the horizontal plane and frontal hemifield. Speech stimuli were presented from loudspeakers 30°, 60° and 120° about the midline. No effect of age on localisation ability was found, as all participants performed at ceiling.

Lovett et al. (2012) found similar results using a source identification task in the horizontal plane. They tested 48 participants aged one to seven years and young adults on a speech localisation task. They used presentation angles of 0°, 15°, 30° and 60° about the midline and the signal was presented at 70 dB A SPL. In the 60° condition age effects were present. It was found that the two year olds scored worse than adults, three year olds scores were not significantly different from adults and the four year olds had a median score of 100% (older children were not tested on this condition). In the 30° condition there were no effects of age. In the 15° condition age effects were

present. Three, four and five year olds were worse than adults. The six and seven year olds performances were not significantly different to that of the adults.

Otte et al. (2013) also found similar localisation abilities in children and adults using a source identification experiment. They tested 18 children aged 7 to 11 years old and 10 young adults using broadband stimuli, which varied in both azimuth and elevation. They found that mean absolute error was higher for children than adults, but the differences were not significant (difference of 2.5° in azimuth condition and 4.8° in elevation condition).

Van Deun et al. (2009) tested the localisation ability of 33 children aged four to six years old and five young adults using a source identification method. They used broadband stimuli presented at 65 dB SPL and presentation angles of 0°, ±15°, ±30°, ±45° and ±60° (in the horizontal plane). They measured RMS error and mean absolute error. They found that the median value across children of the mean absolute errors were 5° for 4 year olds, 2° for 5 year olds, 1° for 6 year olds and 0° for adults. The median RMS errors were 10° for 4 year olds, 6° for 5 year olds, 4° for 6 year olds and 0° for adults. Using Kruskall-Wallis tests, they found that four year olds performed significantly worse than the other children and that children aged five and six years performed at adult like levels. Overall, it seems that localisation accuracy stabilises and reaches an adult like level at approximately six years of age.

Two studies have been carried out to investigate MAAs in quiet in children, with somewhat conflicting findings. Litovsky (1997) found that performance

was adult-like in some five year olds but not in others. Kühnle et al. (2013) showed age effects between 6-7 year olds and 13-18 year olds. These studies will be discussed in greater depth in Chapter 6.

It should be considered whether localisation ability is linked to head size, as infants and children (who have small heads) do not perceive ITDs as well as adults (who have larger heads), meaning they cannot localise as well as adults. Clifton et al. (1998) measured infants head sizes and interaural differences. They wrote that ITDs for a given location will change with increasing age (and head size) for this age group. However, this happens gradually and the plasticity of the auditory system means that localisation ability is recalibrated with increasing head size. It is known that head circumference reflects brain size, and that the brain is 80% of adult size by the time children reach three years of age and is 90% of adult size when children reach seven years of age (Graber, 2019). Growth spurts in head size occur between 7-11 years and then 16-19 years, with the head reaching full adult size by 19 years of age (Huelke, 1998). If head size was strongly linked to localisation ability then we might expect localisation abilities to be much closer to adult-like by three years of age than has been shown in the literature. Therefore, it seems unlikely that increasing head size plays a large part in the development of adult-like localisation in quiet abilities at approximately 6 years of age. Martin, Johnstone and Hendrick (2015) considered the reasons for age-related improvements in localisation in quiet abilities in three to five year olds (comparing them to adult participants). They write that the reasons for children being worse than adults are likely due to central auditory processing,

auditory development and/or higher level or goal directed attention (and that testing conditions and task comprehension are unlikely to be factors). The authors state that further research is required to fully understand the reasons for improvement in localisation in quiet abilities in this age group.

1.3.2.2 Localisation of masked speech

Much less in known about localisation of masked speech abilities for both children and adults. A small number of studies have found that adults are worse at localisation when the signal to noise ratio (SNR) becomes less favourable/more talkers are present (Drullman and Bronkhorst, 2000; Simpson et al., 2006; Kopčo, Best and Carlile, 2010). Boothalingam et al. (2016) carried out the only currently published study looking at speech sound localisation in broadband noise and speech babble in children (although they present what appears to be the same data again in Boothalingam et al. (2019)). This study is summarised in Table 1.6. They tested 21 children aged 7 to 16 years and 21 young adults on localisation in quiet, localisation in broadband noise and localisation in speech babble tasks (using source identification methods). A ring of eight loudspeakers spaced 45° apart was used for this study, the study was testing localisation ability in the horizontal plane and in both the front and rear hemifields. The signal used was the word "baseball", which had a duration of 600 ms and it was presented at SNRs of -12, -6 and 0 dB for the localisation of masked speech tasks. The speech babble consisted of four pairs of concatenated sentences. The maskers were presented diffusely from four of the loudspeakers (45°, 135°, 225° and 315°). Performance was measured as percent correct ± 95% confidence intervals

(but only considering responses from the correct hemifields). Unsurprisingly the study found that there were no significant differences between the adults (96.2% ± 2.4%) and children's (96.5% ± 2.2%) performance in quiet, as previous research has already shown that children on the whole have adult like localisation abilities by the age of six or seven years (Van Deun et al., 2009; Lovett et al., 2012; Kühlne et al, 2013). The high scores of the participants suggest that they may have been performing at ceiling, which is perhaps not surprising given the large angles (45°) between the speakers. Adults and children also performed at similar levels for localisation in the broadband noise condition. Regression analysis showed β=0.05, *p*=0.62 (where β is the beta coefficient of the linear regression equation and a high value represents a stronger effect). However, in the speech babble condition children had reduced localisation accuracy compared to adults. For SNRs -12, -6 and 0 dB the approximate means given were adults: 90, 93% and 95%, children: 80%, 87% and 92%. These adult mean scores are similar to those obtained in quiet (96.2% ± 2.4%), suggesting that adults may still be performing at ceiling level and that the task may have been too easy. The children's mean scores in babble were slightly lower than the scores obtained in quiet (96.5% ± 2.2%). There was no significant effect of age within the group of children for the localisation of speech in babble task (regression analysis showed β=0.06, *p*=0.57). The researchers suggest the reason for children's poorer localisation in speech babble than in broadband noise was due to the informational aspect of the speech babble masker, which children are more susceptible to than adults (Wightman and Kistler, 2005). An

informational masker contains content which competes with the signal, rather than the energy of the masker competing with the signal as is the case in energetic masking (Lentz, 2018). The speech babble masker is more distracting to the listener and requires them to have cognitive and attentional abilities, which enable them to focus on the signal of interest. Speech babble contains linguistic content so children with poor language skills may experience greater masking than those with good language skills. However, if speech babble consists of too many talkers then it essentially becomes an energetic masker so it is important to carefully consider the number of talkers in a speech babble (Simpson and Cooke, 2005). This highlights the importance of choice of masker for any study involving masked speech and the importance of testing cognitive, attentional and language skills. The main drawback of this study was that the task may have been too easy, as indicated by the higher scores achieved by participants (especially in the quiet task). This is most likely because of the large separation angles between the speakers (45°). If the task had been more challenging then the study may have been more successful in demonstrating age effects for localisation of speech in babble.

Study		Van Deun et al., 2009	Murphy et al., 2011	Kühlne et al., 2013
Task		Localisation	Localisation acuity	Localisation accuracy
Measure		RMS error, mean absolute error	Localisation score	Hit accuracy
Children	**Number**	**33**	**40 (26;14)**	**129**
	Age range (years)	**4-6**	**6-10; 11-15**	**6-18**
	Mean age (years)	4.5	8.3; 12.9	9.2
	SD age (years)	0.79	-	2.9
	Gender (f;m)	20;13	-	72;64
Adults	**Number**	5	6	-
	Age range (years)	23-27	-	-
	Mean age (years)	24	-	-
	Gender (f;m)	-	-	-
Exclusion criteria		Hearing loss	Hearing l oss	Hearing loss, glue ear, perceptive disorder
Speakers	**Number**	9	24	47
	Distance (m)	1	1.5	2.35
Practice		Children only	No	3
Feedback		Yes	Yes	Yes
Signal	**Type**	Broadband (bell-ring)	Single nouns (toy names)	Gaussian noise bursts
	Frequency	-	-	Low frequency: 0.3-1.2 kHz, High frequency: 2-8 kHz
	Level	65 dB SPL	57-67 dB A	35 dB SL
	Duration	1s	6s	250 ms
	Presentation angles	±60°, ±45°, ±30°, ±15°, 0°	0°, ±30°, ±60°	±4°, ±26°, ±30°, ±60°, ±64°, ±86°, ±90°
Test time (mins)		Children 14, adults 3	-	45-100
Total presentations		Children 51, adults 27	120	168
Response		Verbally identify signal source speaker	Verbally identify signal source speaker	Point to sound source with laser pointer
Main findings		**Four year olds perform significantly worse than other children. No significant differences for older children and adults.**	**No developmental changes found.**	**Localisation more accurate for low frequency than high frequency signals, no age effects found.**

Table 1.5: Table showing studies investigation localisation ability in children Part 1 (Van Deun et al., 2009; Murphy et al., 2011; Kühlne et al., 2013)

Study		Otte et al., 2013	Boothalingam et al., 2016
Task		Localisation: azimuth and elevation	Localisation of masked speech
Measure		Mean absolute error	Front/back correct %
Children	**Number**	**18**	**21**
	Age range (years)	**7-11**	**7-16**
	Mean age (years)	9.5	11.4
	SD age (years)	1.3	-
	Gender (f;m)	8;10	13;8
Adults	**Number**	10	21
	Age range (years)	20-34	18-30
	Mean age (years)	24.9	22.8
	SD age (years)	4.9	-
	Gender (f;m)	3;7	All female
Exclusion criteria		Hearing loss or hearing concerns	Hearing loss
Speakers	**Number**	58	8
	Distance (m)	-	1.5
Practice		Yes	No
Feedback		No	-
Signal	**Type**	Gaussian white noise	Speech word, SNRs of: -12, -6, 0 dB
	Frequency	High pass filtered at 0.5kHz, low pass filtered at: 5, 7, 11 or 20 kHz	-
	Level	55 dB A	-
	Duration	150 ms	600 ms
	Presentation angles	Azimuth: -75° to +75°, Elevation: -55° to +55°	0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°
Masker	**Type**	-	Broadband noise (BBN), speech babble (four channel)
	Frequency	-	-
	Level	-	66 dB SPL
	Duration	-	Continuous
	Presentation angles	-	45°, 135°, 225°, 315°
Test time (mins)		-	-
Total presentations		Children 72, adults 96	240
Response		Point to sound source with head or laser	Verbally identify signal source, head movement tracking
Main findings		**Children localised sounds in azimuth and elevation as well as adults but require frequencies >11 kHz for elevation localisation.**	**No significant difference between adults and children in quiet, similar performance in BBN. Children performed significantly worse in speech babble. No age effect within group of children.**

Table 1.6: Table showing studies investigation localisation ability in children Part 2 (Otte et al., 2013; Boothalingam et al., 2016)

Study		Lovett et al., 2012				
Task		Left right discrimination	Movement tracking	Localisation	Masked speech	SRM
Measure		% correct	% correct	% correct	SNR for 70.7% correct threshold	SNR for 70.7% correct threshold
Children	Number	56		48		
	Age range (years)	1.5-1.9; 2.0-2.9; 3.0-3.9; 4.0-4.9; 5.0-5.9; 6.0-6.9; 7.0-7.9		2.0-2.9; 3.0-3.9; 4.0-4.9; 5.0-5.9; 6.0-6.9; 7.0-7.9		
	Mean age (years)	1.7; 2.4; 3.5; 4.5; 5.6; 6.3; 7.6		2.4; 3.5; 4.5; 5.6; 6.3; 7.6		
	SD age (years)	0.2; 0.4; 0.3; 0.3; 0.3; 0.3; 0.3		0.4; 0.3; 0.3; 0.3; 0.3; 0.3		
	Gender (f;m)	5;3, 4;4, 3;5, 4;4, 5;3, 6;2, 3;5, 5;3		4;4, 3;5, 4;4, 5;3, 6;2, 3;5, 5;3		
Adults	Number	8, Age 20-24 years (mean 21.6, standard deviation 1.2, 5f;2m)				
Exclusion criteria		Hearing loss				
Speakers	Number	13				
	Distance (m)	1.65				
Practice		4	0	4	Yes	Yes
Feedback		Yes	No	No	No	No
Signal	Type	Speech sentence	Environmental sounds	Speech sentence	Single nouns (toy test)	Single nouns (toy test)
	Frequency	-	L.ow-pass filtered at 5.5 kHz	-	-	-
	Level	70 dBA SPL	71 dBA SPL	70 dBA SPL	50 dBA SPL	50 dBA SPL
	Presentation angles	±30°, ±60°	-90° to +90° moving around edge of ring	0°, ±15°, ±30°, ±60°	0°	0°
Masker	Type	-	-	-	Broadband pink	Broadband pink
	Frequency	-	-	-	-	-
	Level	-	-	-	Varied adaptively	Varied adaptively
	Duration	-	-	-	1.4s	1.4s
	Presentation angles	-	-	-	0°	0°, ±90°
Test time (mins)		6-150				
Total presentations		40	4	90	-	-
Response		Eye / head / body turn	Eye / head / body turn	Verbally identify source	Point to the toy	Point to the toy
Main findings		**±60 condition: significant effect of age, 2 year olds had lower score than adults, ±30 condition: significant effect of age, 1 and 2 year olds had lower scores than adults, 3 year olds not significantly different to adults, all other age groups median of 100% correct in both conditions.**	**No significant effect of age.**	**60° condition: significant effect of age, 2 year olds lower scores than adults, 3 year olds not significantly different to adults, 4 year olds median score 100% correct, 30° condition: no significant effect of age, 15° condition: significant effect of age. Younger children worse, 3, 4 and 5 year olds lower than adults, 6 and 7 year olds not significantly different to adults.**	**Significant effect of age, younger children had higher SRTs, 4 and 5 year olds had significantly higher than adults, older children not significantly different to adults.**	**No significant effect of age, all had SRM of approximately 5.0 dB.**

Table 1.7: Table summarising tasks carried out by Lovett et al. (2012)

1.3.2.3 Masked signals and Spatial Release from Masking

The impact of background sound on speech intelligibility can vary depending on the type of masker. Maskers can contain energetic and/or informational content. Energetic masking occurs when the energy in a masker competes with the energy of a signal, whereas informational masking occurs when the content of a masker competes with the signal (Lentz, 2018). Moore (2012) writes that informational masking is likely to be due to inappropriate direction of attention. As children already have poorer attentional abilities than adults, they are especially vulnerable to the negative effects of informational masking.

Factors such as gender of the speech masker, whether the source is co-located with the target and how many speech maskers are present influence speech intelligibility. Brungart et al. (2001) showed that listeners had greater difficulty listening to speech when the voice of the masker had similar characteristics to the target, such as gender. Simpson and Cooke (2005) carried out a study looking at the relationship between number of talkers in a babble and speech intelligibility, looking at single talker babble, multiple talker babble and speech-shaped noise. They found that speech intelligibility steeply decreased until eight talkers were present, at which point it plateaued. All maskers containing between 8 and 128 talkers babbling had roughly the same effect on speech intelligibility.

When listening to speech in background sound, speech intelligibility can be poor due to masking of the target speech. Binaural hearing helps listeners to

better understand speech in the presence of maskers or background sound. Binaural sensitivity can be measured using the binaural masking level difference (BMLD) (Hirsh, 1948). It measures the difference in detectability between the presentation of signal and maskers with the same phase and level differences for both ears to the same signal and maskers but with differing phase and level differences for each ear (Moore, 2012). More recently this type of task has been demonstrated by Freyman et al. (2001), who showed that when signal and masker were spatially separated by 60°, speech intelligibility was better than when they were co-located. This is called Spatial Release from Masking (SRM) and has since been demonstrated in many other studies, including those investigating SRM in children. The development of SRM in children is relevant to this work as the onset of this ability highlights a point in development where a child can utilise spatial cues to improve how well they can listen to masked speech. SRM studies can also provide information about effects of different signal and masker types.

Several studies have been carried out to investigate the development of SRM in children, outlined in Table 1.7, Table 1.8 and Table 1.9 and summarised below. Lovett et al. (2012) tested children aged two to eight years of age and found that SRM reached an adult-like level at three years of age (SRM was approximately 5 dB for all age groups). Many other studies tested children from the age of three years old and found either no effects or small but non-significant effects of age on SRM. Litovsky (2005) found mean SRMs of 5.3 dB in children aged four to seven years and 3.3 dB in young adults. Garadat and Litovsky (2007) found mean SRMs of 7.7 dB in three year olds and 11 dB in

four to five year olds. Ching et al. (2011) tested children aged 3 to 12 years and found mean SRM of 3.6 dB for words and 2.6 dB for sentences, with no effects of age on performance. Misurelli and Litovsky (2012) tested SRMs of 4 to 6.3 year olds and 6.9 to 8 year olds, finding that the younger children had SRMs of 5.2 dB and 2.8 dB and the older children had SRMs of 6.0 dB and 4.8 dB (for asymmetrical and symmetrical maskers respectively). Schafer et al. (2012) found that mean SRM was 5.8 dB for three year olds, 6.0 dB for four year olds, 5.9 dB for five year olds, 5.5 dB for six year olds and 4.7 dB for adults.

In contrast to the studies summarised above, Yuen and Yuan (2014) found that SRM performance improved with age (from approximately 2 dB at four years of age to approximately 8 dB at nine years of age) and did not reach an adult level until at least nine years of age. The reasons for this study's findings contrasting with the rest of the literature are not clear. One reason may be that the test was carried out in Mandarin, rather than English. Due to the differences in the language such as tonal qualities SRM may be more difficult to achieve in Mandarin than English. It might have been expected that, as the masker used was energetic, it should be less challenging for children and so SRM would be present at a younger age. This comparison to studies using informational maskers, which found SRM to be established at a younger age, shows how important the choice of masker is.

Study		Litovsky, 2005	Garadat and Litovsky, 2007	Ching et al., 2011
Task		SRM	SRM	SRM
Children	**Number**	**36**	**10; 10**	**31**
	Age range (years)	**4.5-7.5**	**3.25-3.8; 4.2-5.5**	**3.5-11.8**
	Mean age (years)	5.5	3.6; 4.9	7.1
	SD age (years)	1	-	2.2
	Gender (f;m)	22;14	4;6 and 3;7	13;18
Adults	**Number**	9	-	-
	Age range (years)	-	-	-
	Mean age (years)	-	-	-
	SD age (years)	-	-	-
	Gender (f;m)	-	-	-
Exclusion criteria		Hearing loss, auditory dysfunction or cognitive disorders, not native English speaker	Hearing loss, ear infections, illnesses or medication taken on day of testing	Hearing loss
Speaker Distance (m)		1.67	1.2	0.8
Practice		Familiarisation only	Familiarisation only	Familiarisation only
Feedback		Yes	Yes	No
Signal	**Type**	Spondaic words	Spondaic words	Single words, BKB sentences
	Frequency	-	-	-
	Level	Adaptive	Adaptive	Adaptive
	Presentation angles	0°	0°	0°
Masker	**Type**	Harvard IEEE list sentences or modulated speech-shaped noise	Harvard IEEE list	8 talker babble
	Frequency	-	-	-
	Level	-	60 dB SPL	Adaptive
	Presentation angles	0°, 90° (right)	0°, 90° (right)	0°, ±90°
Test time (mins)		Children 45, Adults 120	60-120	-
Response		Point to picture of the signal word (from a choice of 4)	Point to picture of the signal word (from a choice of 4)	Point to picture of the signal word (from a choice of 4)
Main findings		**SRM children=5.25 dB, adults=3.34 dB, similar amounts of SRM for adults and children, no age effect.**	**SRM=7.7 dB for younger and 11 dB for older children No significant age effects. SRM established by age 3 years.**	**SRM=3.6 dB (words), 2.6 dB (sentences). Younger children need better SNR for equal performance but no difference in SRM found.**

Table 1.8: Table showing studies investigating SRM in children (part 1) (Litovsky, 2005; Garadat and Litovsky, 2007; Ching et al., 2011)

Study		Misurelli and Litovsky, 2012	Schafer et al., 2012	Yuen and Yuan, 2014
Task		SRM	SRM	SRM
Children	**Number**	**8; 8**	**68**	**26**
	Age range (years)	**4.1-6.3; 6.9-8.8**	**3.0-6.9**	**4-9**
	Mean age (years)	5.0; 7.8	-	7.2
	SD age (years)	0.8; 0.6	-	-
	Gender (f;m)	-	-	13;13
Adults	**Number**	8	17	12
	Age range (years)	-	19.7-42.4	-
	Mean age (years)	-	-	23
	SD age (years)	-	-	1.2
	Gender (f;m)	-	-	6;6
Exclusion criteria		Hearing loss	Hearing loss, non-native English speaker, recurrent otitis media, speech, language, attention or cognitive disorder	Hearing loss
Speakers Distance (m)		1.5	0.9	1
Practice		Familiarisation only	Familiarisation and practice	Training in quiet until score of 100%
Feedback		Negative only	-	-
Signal	**Type**	Spondaic words	Speech sentences	Disyllabic words, lexical tones
	Level	Adaptive	-	Adaptive
	Presentation angles	0°	0°	0°
Masker	**Type**	Two talkers of Harvard IEEE list sentences	Recorded classroom background sound	Speech spectrum weighted noise
	Level	Adaptive	-	Adaptive
	Presentation angles	0°, ±90°	0°, 180°	0°, 90°, 270°
Test time (mins)		60 for adults, 180 for children	-	-
Response		Point to picture of the signal word (from a choice of 4)	Repeat entire speech sentence	Point to picture of the signal word (from a choice of 4/8)
Main findings		**SRM for Younger children= 5.2, 2.8, Older children: 6.0, 4.8 dB. No significant effect of age on SRM found.**	**No significant effect of age on SRM found.**	**SRM for disyllabic words=5.76, lexical tones=4.42. Adults performed significantly better than children. Effect sizes: Cohen's f2= 0.54, 0.69, 0.47, 0.52**

Table 1.9: Table showing studies investigating SRM in children (part 2) (Misurelli and Litovsky, 2012; Schafer et al., 2012; Yuen and Yuan, 2014)

1.4 When is adult-like performance reached?

The studies described above tell us that children's abilities to use spatial cues gradually improve with age, with adult-like localisation in quiet abilities being reached at approximately six years of age. Adult-like performance on localisation in background sound tasks is less well researched, with one study (Boothalingam et al., 2016) showing adult-like performance at seven years of age in broadband noise but poorer performance than adults and no age effects in 7 to 16 year olds in speech babble. Many studies have shown that SRM reaches adult-like levels in early childhood (although a study by Yuen and Yuan (2014) found a longer developmental trajectory).

However, all of these studies raise the question of *how* adult-like performance is defined. The majority of studies use ANOVAs and t-tests to compare groups of children, which can then be described as the same or different to an adult group. This issue is further explored in Chapter 4, where the methods of 17 studies are summarised (Litovsky, 1997; Eisenberg et al., 2000; Johnson, 2000; Stollman et al., 2004; Litovsky, 2005; Stuart, 2005; Garadat and Litovsky, 2007; Vaillancourt et al., 2008; Cameron et al., 2009; Van Deun et al., 2009; Moore et al., 2011; Lovett et al., 2012; Kühnle et al., 2013; Yuen and Yuan, 2014; Martin, Johnstone and Hedrick, 2015; Boothalingam et al., 2016; Jones, 2018). The methods used in these 17 studies were ANOVAs, t-tests, child mean values in previously reported adult range, similar d' values, broken stick functions, Pearson's linear correlations and child values within 95%

population limits of adults. It is clear that there are considerable inconsistencies to method in this area.

It is also necessary to consider that perhaps there is a good reason that children do not have adult-like localisation of speech in babble abilities, rather than that they have a deficit (compared to adults) which they must overcome. It may be the case that in order to maximise language development children listen to any conversations or speech signals around them, rather than focusing in on just one (which would be aided by localisation in babble abilities). A by-product of this may be that localisation of speech in babble abilities are poorer than those of adults.It is also important to note that even when children reach an adult-like performance level on a task they may not have reached true adult-like levels. If adults perform at ceiling on a task then this could indicate that the task is not challenging enough for them to demonstrate the full range of their abilities. Therefore it can be said that children have adult-like abilities on this task, but not necessarily for all tasks of that nature.

1.5 Summary

Listening to speech in a noisy environment is a challenging situation which occurs often, especially for children. The use of spatial cues improve how well listeners localise sounds, meaning that they can focus better on the signal of interest and ignore the background sounds. It is therefore important to know how localisation of speech in challenging environments improves with age. This could be particularly relevant in educational settings, where children are

often expected to listen in less than ideal acoustic surroundings. It could be useful for educators to have an understanding of how localisation in challenging environments improves with age, in order for them to have realistic expectations for children of different ages. Understanding the development of this ability in typically developing children would provide a benchmark which could be used to compare children who may have difficulties in this areas, such as those with a hearing loss or Auditory Processing Disorder, to typically developing children.

The ability to localise improves throughout infancy and early childhood and it is well established that children can localise to an adult-like level in a quiet environment from approximately six years of age. Although it is known that children have greater difficulty than adults with masked speech generally, little is known about how well children can localise in the presence of maskers. Findings from the SRM literature suggest that children have the ability to use spatial cues to improve listening to masked speech from as young as three years of age and it has also been shown that children can localise in broadband noise to an adult-like level by seven years of age. However, the only study to date which has investigated children's localisation of speech in babble did not show any age effects or find the age at which adult-like performance occurs. Therefore, we do not know what the developmental time course of localisation of speech in babble is or at what age adult-like performance is reached. It is also important to note with all of the developmental studies discussed in this chapters that a cross-sectional design was used. This means that it is possible that factors other than age may

have been present and causing what appeared to be improvement with age. However, as these studies are generally in agreement that localisation abilities improve with age in children it seems most likely that this age-related improvement does occur. It may also be the case that there is a factor which develops with age and has a knock-on effect of improving localisation of speech in babble abilities.

This thesis addresses these gaps in the literature, mapping the developmental time course of localisation of speech in babble in children aged six years and older then deriving the age at which adult-like performance is reached. In order to do this both source discrimination and source identification methods are used (summarised in Table 1.10). There is not a clear way to define and therefore measure adult-like performance. To work around this uncertainty of the definition of adult-like performance, it was estimated using age ranges from multiple analysis techniques. Age-ranges were used rather than point estimates, as it seems invalid to use a point estimate for a concept as vague as adult-like performance. Methods of analysis used to estimate the age at which adult-like performance occurs include t-tests, ANOVAs, fitting of exponential functions, bootstrapping and the use of broken stick functions. The relationships between detection, identification and localisation of speech in babble in both adults and children are also investigated (in Experiments 2 and 3).

Experiment	Chapter	SNR	Method
1 and 2	2, 3 and 4	Varied with adaptive track (beginning at approximately 0 dB)	Source identification
4	6	20, 10, 0 and -10 dB	Source discrimination (minimum audible angle)

Table 1.10: Table summarising methods of studies investigating developmental time courses of localisation in babble (using exponential function and RMS method and broken stick functions)

Table 1.11 summarises the methods used by Boothalingam et al. (2016) and in the experiments described in this thesis. From this it can be seen that this thesis uses a method that varies SNR, whilst keeping angles constant and then a method that varies the angles, whilst keeping the SNRs constant (which is similar to the method used by Boothalingam et al. (2016)). It was decided to do the experiments in the chosen order so that the results from Experiments 1 and 2 could inform the SNR choices for Experiment 4.

	Fixed choices	Variable (and constant stimuli or adaptive track)
Boothalingam et al. (2016)	SNR (x3)	Angle by constant stimuli
Experiments 1 and 2 (Chapters 2, 3 and 4)	Angle (broad 5 choices separated by 15°)	SNR by adaptive track
Experiment 4 (Chapter 6)	SNR (x4)	Angle (narrow) by adaptive track

Table 1.11: Table summarising methods used in study by Boothalingam et al. (2016) and in this thesis

Chapter 2

Experiment 1: Localisation of speech in babble in typically developing children and adults

2.1 Introduction

It is well established that children can localise in quiet to an adult like level by approximately six years of age (Van Deun et al., 2009; Kühlne et al, 2013; Lovett et al., 2012). However, little work has been done to examine the localisation in noise abilities of children. Boothalingam et al. (2016) investigated localisation of a single spoken word (''baseball'') in three conditions (quiet, broadband noise and speech babble) in adults and children aged 7 to 16 years. They reported that regression analysis showed no effect of age on performance. In agreement with previous research they found no significant differences between localisation in quiet abilities for adults and children (Van Deun et al., 2009; Kühlne et al, 2013; Lovett et al., 2012). When they measured percent correct on a localisation-in-quiet task children scored 0.27% higher than adults, this was a very small and non-significant difference ($t[40] = 0.03$, $p = 0.97$). They also reported similar performance for adults and

children in the broadband noise condition, but they do not give statistics for this. Further group differences were not provided, however they reported that in both adults and children, performance was better in broadband noise than speech babble. One of the key findings of this study was that children performed worse than adults in the speech babble condition (exact figures not provided).

While the study by Boothalingam et al. (2016) is informative, in telling us that children cannot localise as well as adults in speech babble, it did not find any age effects in the group of children tested (aged 7 to 16 years). This means that the age at which children can localise in babble to an adult-like level is likely to be 16 years or older.

2.2 Aims

This experiment aims to refine our knowledge about the development of localisation in babble abilities in children by mapping the developmental time course of localisation in babble abilities and estimating the age at which adult-like performance is achieved.

It is unlikely that adult-like localisation of speech in babble abilities will occur until children are at least 13 years of age, as age-related improvements in speech perception in babble have been shown to occur between the age groups of 11-13 and 15-17 years (Elliott, 1979). Also, in a review paper Leibold (2017) suggests that speech perception in noise abilities reach adult-like levels at 13 to 14 years of age. Therefore, it was expected that adults would

perform better on this task than children and that performance in children would gradually improve in the age range from 6 to 13 years. Performance is analysed by splitting the children into three age groups (6-7, 8-9 and 10-13 years old) and comparing these groups to each other and to adults. Other factors such as attention, cognition and linguistic skills are also likely to be involved, making it difficult to make a more precise prediction. In order to try to understand the effects of some of these skills, a speech perception in noise task and a series of cognitive tests were also included in the experiment. The speech in noise task was included in order to make comparisons to children's speech in noise abilities. The other tasks used were the Weschler Abbreviated Scale of Intelligence (Wechsler, 1999), digit span (Weschler, 1991) and dichotic digits (Musiek, 1983). Two questionnaires were also included: Evaluation of Children's Listening and Processing Skills (Barry and Moore, 2014) and Conners Parent Rating Scale-Revised (S) (Conners et al., 1998). The reasons for these particular choices are described in the Procedure section below. Analysis of these additional tasks is reported in Chapter 4, as participants in Experiment 2 also completed these tasks (except for the speech perception in noise task, which is reported in this chapter). In Experiment 2 participants completed the same localisation of speech in babble task described in this chapter, along with a slightly different localisation of speech in babble task and detection and identification of speech in babble tasks.

For the localisation of speech in babble tasks, performance was measured using adaptive tracks to estimate the threshold at which 70.7% correct was achieved. Secondary analysis will take place in order to assess the number of adaptive tracks used, to assess whether it would be possible to use fewer tracks in future studies. Analysis will also take place to assess any relationship between participant gender and speech talker gender, in order to inform which condition(s) should be used for future studies.

2.3 Methods

Ethical approval for this experiment (and all others described in this thesis) was obtained from the North West - Greater Manchester Central Research Ethics Committee (IRAS ID: 222656) and was sponsored by the University of Nottingham.

2.3.1 Participants

Participants with normal hearing (20 dB HL or better at 0.25, 0.5, 1, 2, 4 and 8 kHz), who spoke English as their first language and had no learning or language disorders were recruited for this experiment. The latter criterion was set because it was thought likely that learning or language disorders could cause attentional deficits or difficulty understanding the instructions for the task, which would then negatively affect performance on the localisation tasks. As the main purpose of this experiment was to investigate performance in typically developing children, the presence of any learning/language disorders or hearing impairment would have introduced confounds.

Power analysis has been carried out using G power to calculate the required sample sizes for this experiment and all of those described in this thesis. The effect sizes being looked for are approximately F=0.4-0.6. Depending on the design of each stage of the experiment this effect will be observable in a group of 10-40 participants per group. Therefore a minimum of 10 adults and 10 children were required for the experiment.

2.3.1.1 *Adults*

10 adults (four female, six male) aged 18 to 29 years old (mean 24.5, s.d. 3.7), who were all monolingual English speakers with normal hearing and no language or learning difficulties took part in the experiment. They were recruited through posters placed around the University of Nottingham and a post on University of Nottingham Postgraduate Facebook page.

2.3.1.2 *Children*

20 children (9 female, 11 male) aged 6 to 13 years old (mean age 9.3 years, s.d. 1.7) took part in the experiment. The number of children of each age is shown in Table 2.1. Sixteen of the children were monolingual English speakers and four were bilingual in English and another language since birth (other languages were Chinese, Spanish and Greek). All of these children had normal hearing and no confirmed language or learning difficulties. Normal hearing was confirmed through pure tone audiometry and absence of learning/language disorders was confirmed through the use of a locally created family history questionnaire. Participants were recruited through a local primary school and online advertisements using Facebook and Twitter.

Age (years)	Number of participants
6	1
7	4
8	2
9	8
10	3
11	0
12	1
13	1
18-29	10

Table 2.1: Table showing number of participants of each age

An additional nine children were also tested but their results had to be excluded for the reasons shown in Table 2.2. As shown in the table, five children were unable to complete the experiment due to learning English as a second language. During the experiment and in earlier pilot testing of adults it became evident that even if a participant was fluent in English, if they had not been learning English from birth then they found the task more difficult (or impossible in the case of children) compared to participants who had learned English from birth. This is consistent with studies that have found that although early bilingual English speakers can perform as well as monolingual speakers in speech recognition in quiet tasks, their performance in noise (both speech shaped noise and speech babble) is significantly worse than that of the monolingual English speakers (von Hapsburg, Champlin and Shetty, 2004; Tabri, Abou Chacra and Pring, 2011). Two children were unable to complete the localisation tasks; they were both only 6 years old and potentially had not yet fully developed localisation abilities. One child was excluded as it was found that they had a hearing loss, a report was written for the participant's parent to take to their GP. One child was excluded as they had autism spectrum disorder. During recruitment it was made clear that children with

autism spectrum disorder were not able to take part in the experiment, however the parent of this child had not realised this until attending the test session.

Reason	Frequency
Learned English as a second language	5
Unable to complete localisation tasks	2
Hearing loss	1
Autism Spectrum Disorder	1
Total	**9**

Table 2.2: Table showing reasons for exclusion

2.3.2 Equipment

The test session took place in a large sound proofed room, which contained a ring of loudspeakers in the middle of it. Through a door from the test room was an observation room, where the researcher and the parents of the participants sat. Inside the test room there was a one-way window, so the test room could be seen from the observation room. A sketch of the sound-proof test room, containing the loudspeaker ring, is shown in Figure 2.1.

The loudspeaker ring, shown in Figure 2.2, consisted of 24 speakers, spaced 15° apart from each other. The participant sat on a chair in the middle of the ring and had a table with a touchscreen monitor in front of them.

The five speakers in front of the participant (azimuths of -30°, -15°, 0°, 15°, 30°) had pictures of shapes on top of them (circle, crescent, triangle, square, star). These images were chosen because they were thought to be neutral enough that participants, especially children, would not be inclined to choose their favourite image rather than the correct answer (which could be the case

if a more interesting set of images such as animals or fruit had been chosen). It was decided that the pictures should be in black and white as the signal words used for the experiment were the names of colours, so using colourful images might have confused the participants.

The signal chain used for the loudspeaker ring was a Sonible d:24 amplifier and 24 Tannoy VX6 speakers. The Sonible d:24 amplifier contained a digital to analogue converter which allowed the signals from MATLAB to be processed and sent to the Tannoy VX6 speakers, which contained dual concentric drivers. MATLAB R2016a was used to run the code for the localisation in babble task. Figure 2.3 shows the signal chain.

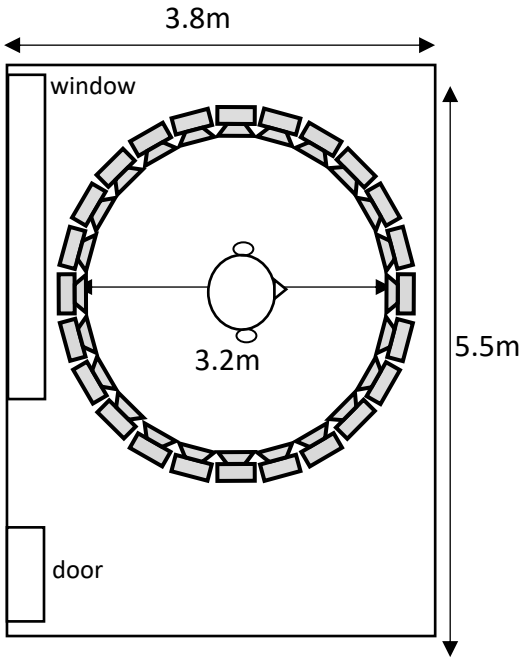


Figure 2.1: Figure showing sketch map of MRC IHR test room in QMC hospital



Figure 2.2: Photograph of loudspeaker ring

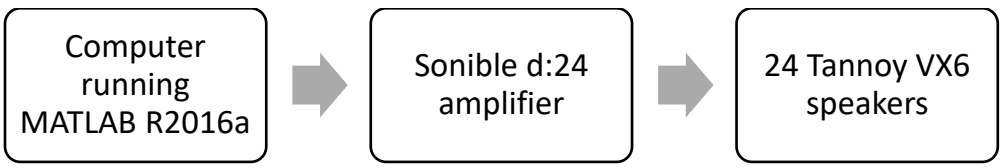


Figure 2.3: Figure showing equipment used for speaker ring

The audiometer used was an Interacoustics AC40 Clinical Audiometer with Sennheiser HD 300 headphones. The dichotic digits and speech perception in noise tasks were run using Psychopy v1.83.04. Sennheiser HD 25-1 headphones were used for these two tasks and for the digit-span task. The MATLAB code used to run the experiment was written by Owen Brimijoin and Michael Akeroyd as part of the overall programming of the laboratory.

2.3.3 Stimuli

2.3.3.1 *Signals for the localisation tasks*

The signals used were single speech words, extracted from the coordinate response measure (CRM) sentence corpus defined by Bolia et al. (2000) and recorded by Kitterick, Bailey and Summerfield (2010), spoken by male and female talkers. This was performed by the researcher, using Audacity to extract the words from the sentences. The words used were: 'red,' 'green,' 'white' and 'blue.' These stimuli were chosen as they were simple words and were available voiced by male and female talkers. Each word had a duration of approximately half a second. There were three different male and three different female talkers for each colour, meaning there was a total of 12 female voiced and 12 male voiced signals. These six talkers were chosen from a selection of eight (four female and four male) as pilot testing showed that the other two were harder to localise than the chosen six.

It was decided to use both male and female stimuli to prevent any bias for a particular gender, as for example a young child raised by a single parent and

educated by a teacher of the same gender may be more used to talkers of that gender.

2.3.3.2 *Babble used for the localisation tasks*

The stimuli used for the babble were male and female voiced Institute of Electrical and Electronics Engineers (IEEE) sentences (when the signal was female voiced then the babble was male voiced and vice versa). Six different IEEE sentences (IEEE, 1969), were played from six different locations in the ring for each presentation (one sentence from each speaker, the locations and sentences changed for each presentation). The ring was split into six equal segments, in each of these six segments an angle was randomly chosen for the sentence to come from. Each sentence had a duration of approximately 2-4 s. The signal was played either 700, 750 or 800 ms after the babble started (this number was chosen randomly for each presentation). The approximate timings of the stimuli are shown in Figure 2.4.

Time (seconds)	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75	3
Babble													
Signal													

Figure 2.4: Table showing approximate timings of babble and signal

2.3.4 Procedure

All participants attended a single session at the Medical Research Council (MRC) Institute of Hearing Research (IHR) laboratory in the Queen's Medical Centre (QMC) Hospital, Nottingham. Adult sessions lasted approximately 45 minutes, child sessions lasted 60-120 minutes. Adult participants were given a

£7.50 inconvenience allowance and any travel expenses were reimbursed. Children were given stickers and £15 in Amazon vouchers to thank them for their time and their parents were reimbursed for any travel expenses. Prior to the experiment beginning consent was obtained.

Table 2.3 shows the tasks which were used in the experiment and whether they were completed by children and/or adults. As shown in the table, children completed all tasks while adults only completed pure tone audiometry and localisation tasks. Participants were given opportunities to take breaks between tasks. The tasks were completed in the order shown in the table.

Any risk of order effects was reduced by counterbalancing the following parts of the experiment: order of conditions in the main localisation task, the order of the right and left condition for the dichotic digits test and the choice of list one or list two for the speech perception in noise task. Results of the additional tasks are reported and analysed in Chapter 4, alongside those of the participants who completed these tasks as part of Experiment 2 (apart from the speech perception in noise task, which is discussed below).

Task	Completed by children	Completed by adults
Pure tone audiometry	✓	✓
Non-verbal parts of the WASI (blocks and matrices)	✓	
Digit span	✓	
Dichotic digits	✓	
Speech perception in noise	✓	
Questionnaires	✓	
Localisation tasks	✓	✓

Table 2.3: Table showing tasks for adults and children

2.3.4.1 *Localisation tasks*

The localisation tasks are summarised in Table 2.4, they are all source identification tasks in the horizontal plane. Prior to these tasks the participant was played the numbers 1 to 24 (voiced by a female talker at 50 dB(A)), each from a different speaker in the ring and moving around the ring in the clockwise direction. This checked that all of the speakers were working whilst familiarising the participant with the environment.

The participant then began the tasks, where they were instructed to remain seated and still and to keep looking straight ahead. They were told they would hear a single word (and babble when appropriate) and that they had to identify which of the five speakers (marked with an image) the word had come from. They responded by pressing the corresponding image on a touchscreen monitor in front of them. The tasks are summarised in Table 2.4 and described in detail below.

Practice task 1: Localisation in quiet (in order)

The signal was presented ten times at 60 dB(A) from the following azimuths in the given order: -30°, -15°, 0°, 15°, 30°, -30°, -15°, 0°, 15°, 30°. No babble was present. This gave the participant an opportunity to hear what the signal sounded like from each location whilst also assessing localisation in quiet abilities. Feedback was given after each presentation. If a participant scored less than 6/10 on this task (or on practice tasks 2 or 3) they repeated it until they scored at least 6/10, to ensure they understood the task. Only one child aged six years old (and no adults) required multiple attempts at this task.

Practice task 2: Localisation in quiet (random)

The signal was presented at 60 dB(A) in a random order ten times choosing from the following azimuths: ±30°, ±15°, 0° (each angle was chosen twice). No babble was present. This basic task allowed the participant to gain confidence in localisation tasks and also acted as a check that they had normal localisation in quiet abilities. Feedback was given after each presentation. As in practice task 1, participants had to score 6/10 in order to progress to the next task. Participants were able to repeat the task if they initially did not score 6/10. Two children (and no adults) required multiple attempts at this task.

Practice task 3: Localisation in low babble

The signal was presented at 60 dB(A) in a random order ten times choosing from the following azimuths: ±30°, ±15°, 0° (each angle was chosen twice). The babble consisted of six individual IEEE sentences. Each sentence was presented at 30 dB(A), meaning the total level of the babble was 38 dB(A) and so the SNR was 22 dB. This task acted as a bridge between the previous familiarisation task and the main task. As it was relatively easy, it ensured that the participant understood the localisation in babble task and that they were prepared for the main task. Feedback was given after each presentation. As in previous tasks, if a participant scored less than 6/10 they had to repeat the task until they did so. One child (and no adults) needed to repeat this task.

*Main task: Localisation of speech in babble (**LocalB**$_{5AT}$)*

The main task was referred to as LocalB$_{5AT}$. The components of this name were chosen for the following reasons: 'Local' from localisation, 'B' from babble and '5$_{AT}$' from 5 alternative forced choice adaptive track. This naming system was introduced to avoid confusion in the following chapter, where another localisation in babble task was introduced along with detection and identification of speech in babble tasks.

In the current experiment each participant completed two conditions:

1. Female voiced signal and male voiced babble
2. Male voiced signal and female voiced babble

The order of the conditions was varied to prevent any order effects. The participant was aware which condition they were doing and also that the babble would be louder than in the familiarisation task.

The signal level of the first presentation was randomly picked from 60-65 dB(A) to reduce predictability. Each of the six IEEE sentences were presented at 55 dB(A), so the total level of the babble was 63 dB(A). The babble stopped between trials and always started again either 700, 750 or 800 ms before the signal. This jittered onset was implemented to make the timing of the signal less predictable.

For the adults each condition consisted of three interleaved two down, one up adaptive tracks. For the children the adaptive tracks were not interleaved, as this allowed the children to take more frequent breaks. This meant each

participant returned three thresholds at which they achieved 70.7% correct for each condition (Levitt, 1971). The adaptive track consisted of two parts. The first part had four reversals of 3 dB. The second part had four reversals of 2 dB. Pilot testing showed that it was important for the step size of the first part to be relatively small, as otherwise the signal level could reach 70 dB(A) after a single error. During pilot testing with children a larger step size caused difficulties as after five presentations at 70 dB(A) (the maximum level allowed) the program would stop and not return a threshold for that track. This was done to ensure that the participant was not presented with too many potentially uncomfortably loud sounds.

This method using variable SNRs was chosen, rather than one using fixed SNRs, as it allowed greater flexibility of signal levels. As there was not sufficient data in the literature to inform choice of SNRs this method was especially beneficial. It meant that a broad range of ages could be tested and complete the task with meaningful results. The results from this experiment can be used to inform the SNR choices for future experiments. However, this can also be a drawback as it makes it more difficult to compare to any future experiments which use fixed SNRs.

2.3.4.2 *Pure Tone Audiometry*

If a participant had a hearing loss then it is likely that they would have struggled with most of the other tasks in the experiment as they would be unable to hear the stimuli at the desired levels. Therefore it was very important to check all participants had hearing within normal limits.

Participants completed pure tone audiometry with the researcher (who is a qualified Audiologist) to confirm they had normal hearing (thresholds of 20 dB HL or better, binaurally at 0.25, 0.5, 1, 2, 4 and 8 kHz).

2.3.4.3 *Wechsler Abbreviated Scale of Intelligence (WASI)*

Non-verbal intelligence may have an impact on localisation of speech in babble abilities, therefore the blocks and matrices part of the Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999) tasks were included to control for non-verbal intelligence. The blocks task involved the participant taking blocks with different colours on their faces and using them to replicate a pattern shown to them. The matrices task involved the child seeing a set of related pictures, shapes or patterns in a sequence and having to pick from a selection of options which one was the correct fit for the sequence.

2.3.4.4 *Dichotic Digits*

When listening and/or localising in background sound a participant must have the ability to separate competing streams of information. The dichotic digits task (Musiek, 1983) measured how well participants were able to do this. The participant heard two digits (from one to nine) at the same time, one in each ear. They were instructed to report the number they heard in one ear at a time. For example, in the first half of the task they reported the number heard in their right ear and in the second half they reported the number heard in their left ear.

2.3.4.5 *Speech Perception in Noise (SPIN)*

A task analogous to the Speech Perception in Noise (SPIN) test (Kalikow, Stevens and Elliott, 1977; Bilger et al., 1984) was included to provide a measure of how well participants could perceive masked speech, as this is likely to have an effect on localisation of speech in babble abilities. The masker was a multitalker babble, made up of text read out continuously by a mix of four male and four female talkers. The participant heard a male voiced sentence in the presence of this masker and had to repeat the final word from it. Six and seven year olds did this for 20 sentences and older children did it for 40 sentences. Half of the sentences were high predictability and half were low predictability; they are all shown in Table 2.5.

Test		Localisation in quiet (in order)	Localisation in quiet (random)	Localisation in low babble	Localisation in babble
Abbreviation		-	-	-	**LocalB$_{5AT}$**
Purpose		Practice	Practice	Practice	**Main task**
Description		Stimuli presented from azimuths in the following order: -30°, -15°, 0°, 15°, 30°, -30°, -15°, 0°, 15°, 30°	Stimuli presented from each of the 5 azimuths twice in a random order	Stimuli presented from each of the 5 azimuths twice in a random order, but with a low level of babble present	**3 thresholds collected for each condition (female voiced signal with male voiced babble, male voiced signal with female voiced babble)**
Method		Constant stimuli	Constant stimuli	Constant stimuli	**Adaptive track (2-down 1-up)**
Number of tracks		1 track with 10 trials (all female voiced)	1 track with 10 trials (all female voiced)	1 track with 10 trials (all female voiced)	**6 tracks (3 female voiced, 3 male voiced)**
Response		Pick 1 location from a choice of 5	Pick 1 location from a choice of 5	Pick 1 location from a choice of 5	**Pick 1 location from a choice of 5**
Signal	**Type**	Word	Word	Word	**Word**
	Location	±30°, ±15°,0°	±30°, ±15°, 0°	±30°, ±15°, 0°	**±30°, ±15°, 0°**
	Level	60 dB(A) every trial	60 dB(A) every trial	60 dB(A) every trial	**60-65 dB(A) starting level**
Masker	**Type**	None	None	6 IEEE sentences (all male voiced)	**6 IEEE sentences (voiced by opposite gender to signal)**
	Location	None	None	6 locations around the ring	**6 locations around the ring**
	Level	None	None	Each sentence 30 dB(A), total 38 dB(A)	**Each sentence 55 dB(A), total 63 dB(A)**
Outcomes		% correct	% correct	% correct	**Mean SNR at threshold**

Table 2.4: Table showing protocol for localisation tasks

The participants heard the sentences at a variety of SNRs, though for the purposes of this experiment the results from the ten low predictability sentences presented at a SNR of 2 dB were of interest. The other sentences, which were not of interest, were high predictability sentences (they were not useful as they could be influenced by language ability) and sentences which had a different SNR depending on the age of the child (they were not useful for comparing performance across all ages of children). As it was not possible to only present the ten trials of interest (low predictability sentences at 2 dB SNR) the participants completed the task for all 20 or 40 sentences and the results from the sentences which weren't of interest were excluded from analysis.

Target word	High predictability	Low predictability
Horn	The angry driver beeped her horn	The angry writer gripped her horn
Crash	The racecars collided in a crash	The programme concluded with a crash
Seeds	The farmer sowed the seeds	The doctor chose the seeds
Roar	The lion gave an angry roar	He is thinking about the roar
Ducks	Helen went to their pond to feed the ducks	Helen went to their land to see the ducks
Smoke	They noticed the fires when the smelt the smoke	They covered the tyres when they smelt the smoke
Speech	The best man was preparing jokes for his speech	The old man was expecting cash for his speech
Test	Simon received his marks for the test	Simon agreed his time for the test
Post	I'll send you a letter in the post	I'll leave you a message in the post
Prize	The winner was rewarded with the prize	The singer was offended by the prize
Head	Daisy wore a helmet on her head	Daisy saw a feather on her head
View	I climbed the mountain to see the view	I promised my aunt she'd see the view
Fire	Simon put out the flames of the fire	Simon looked at the things in the fire
Door	The visitors knocked on the door	The shopkeepers looked at the door
Voice	Ellen tried to speak but she had lost her voice	Ellen saw her face but didn't know her voice
Land	The sailors were happy at the sight of land	The builders were angry at the price of land

Road	The car forced the cyclist off the road	The man forced the children off the road
Bed	The patient spent ages lying in bed	The teacher spent ages thinking in bed
Size	Paul returned the shirt to get a bigger size	Jess revealed the plan to get the bigger size
Style	Her trendy brother had a great sense of style	The working mother had a strong sense of style
Shop	Tim bought some grapes at the shop	Tim bought his friends at the shop
Plants	Chris went to the garden to water his plants	Chris went to the centre to order his plants
Film	Bill went to the movies to see a film	Bill wanted a reason to see a film
Mile	The car broke down so they walked the last mile	The man felt low so he skipped the last mile
Horse	Seth went to the stables to feed her horse	Beth thought of the lady to keep her horse
Note	The singer hit a very high note	The mother held a very long note
Sea	The captain sailed the ship across the sea	The chaplain saw the hill across the sea
Fur	A bear has a thick coat of fur	They knew about the fur
Fish	Salmon is one of the biggest types of fish	Simon is one of the keenest cooks of fish
Card	He knew it was her birthday so he sent a card	He saw it was a problem so he left the card
Sun	The earth and the planets go around the sun	The boys and Janet know about the sun
Stars	The night was lit by the moon and the stars	The wife was sick of the gloom of the stars
Band	Sam was the lead singer of the band	Sam was the last runner in the band
Tea	Carl liked to have biscuits with a cup of tea	Carl had to read textbooks with a cup of tea
Jar	The cookies were kept in a jar	He's glad you called about the jar
Play	Jane watched the actors as they rehearsed for a play	Jane judged the brothers as they prepared for a play
Songs	Sam writes the lyrics for the songs	Sam checks the library for the songs
Sale	The shop offered mass reductions in the sale	The man witness mass consumption in the sale
Boat	The captain threw the anchor off the boat	The chaplain knew the answer was the boat
Cow	The farmer sent Joe to milk the cow	The father went home to see the cow

Table 2.5: Table showing sentences used in SPIN task

2.3.4.6 *Questionnaires*

During the test session the child's parent or guardian was asked to complete the following questionnaires:

- Evaluation of Children's Listening and Processing Skills (ECLiPS) (Barry and Moore, 2014): This identifies areas of language or learning difficulties which a child may be experiencing, which could impact on their localisation of speech in babble abilities.
- Conners Parent Rating Scale-Revised (S) (Conners et al., 1998): This includes a measure of attention, which is important to have as if a participant has poor attentional abilities they are likely to do poorly on the task. This is the attention deficit hyperactivity disorder (ADHD) index. It also measures oppositional behaviour, cognitive problems/inattention and hyperactivity.

2.4 Results

All adult participants were able to fully complete the task, returning six thresholds. Nine child participants returned six thresholds, the other 11 returning between one and five thresholds, as shown in Table 2.6.

Number of thresholds returned	Frequency
1	2
2	1
3	1
4	3
5	4
6	9

Table 2.6: Table showing number of thresholds returned in children

2.4.1 Initial representation of results

The boxplot in Figure 2.5 shows the localisation in babble data for adults and different age groups of children. From visual inspection it can be seen that mean SNR at threshold improves for each age group. A statistical analysis is below.

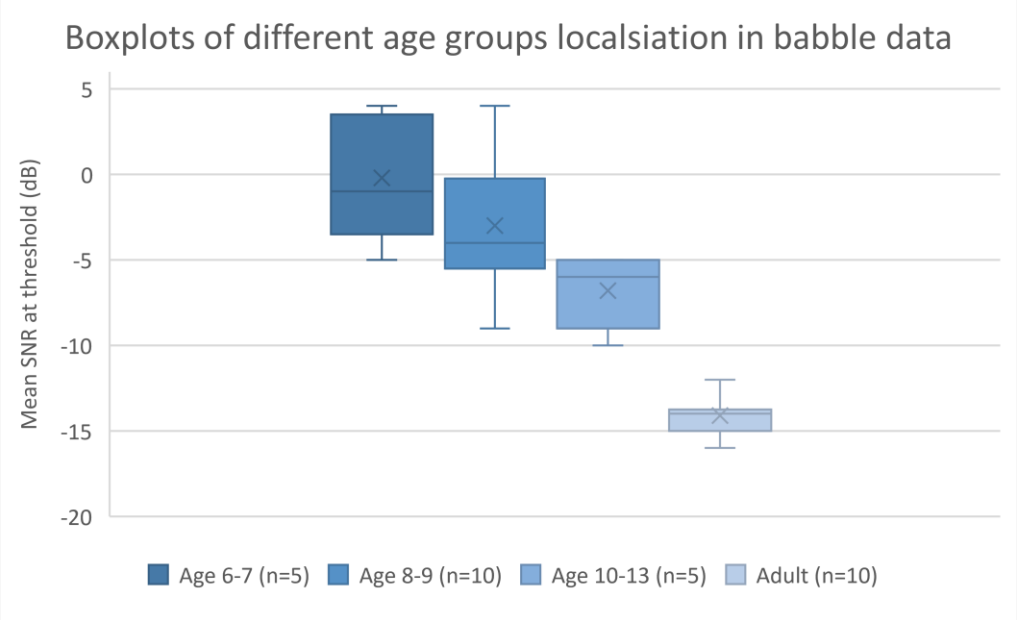


Figure 2.5: Boxplots of different age groups localisation in babble data

The mean SNR at threshold was calculated for each participant by taking the mean of all their finished thresholds and subtracting 63 dB, being the level of the babble. From this an exponential curve was fitted using the Solver function in Excel, shown in Equation 2.1. This allows interpolation to predict performance at different ages. Its parameters are: -15.3 dB is the asymptote, 44.9 dB/year is the slope and 6.6 years is the time constant. R^2 was calculated using the equation shown in Equation 2.2 (Lewis-Beck, Bryman and Futing

Liao, 2004), this value is then multiplied by 100 to give a percentage. For this equation R^2=74%.

$$Predicted\ mean\ SNR\ at\ threshold\ (dB) = -15.3 + (44.9 * e^{\left(-\frac{age}{6.6}\right)})$$

Equation 2.1: Equation for predicted mean SNR at threshold

$$R^2 = 1 - \frac{Sum\ of\ squared\ errors}{Total\ sum\ of\ squares}$$

Equation 2.2: Equation for R^2 (Lewis-Beck, Bryman and Futing Liao, 2004)

Figure 2.6 shows the mean SNR at threshold for each participant, plotted against their age. The dashed line separates the data points of the children and the adults (an adult being a person aged 18 years or older). From this graph it can be seen that the children have much worse SNRs than the adults, and that the SNRs of the children improve with age whereas the adults SNRs are stable with age.

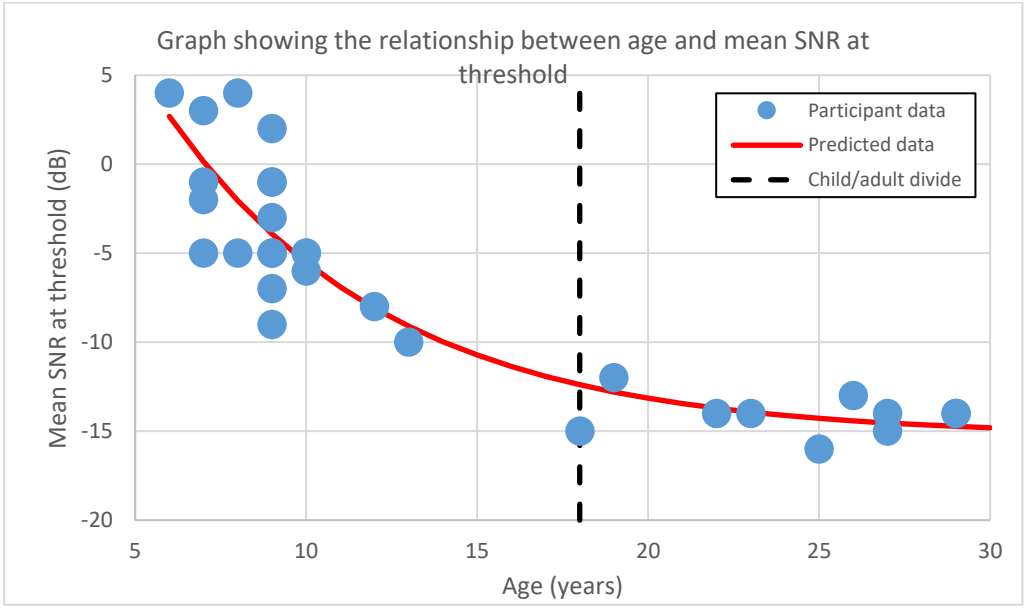


Figure 2.6: Graph showing relationship between performance and age

2.4.2 Comparison of different age groups using a one-way ANOVA

A one-way Welch ANOVA was conducted to determine if average SNR at threshold was different for groups of different age ranges. This type of ANVOA was chosen over the regular one due to unequal variances, as assessed by Levene's Test of Homogeneity of Variance (p = 0.005). For this analysis participants were classified into four groups: age 6-7 (n = 5), age 8-9 (n = 10), age 10-13 (n = 5) and adults (n = 10). There were no outliers and the data was normally distributed for each group, as assessed by the boxplot (Figure 2.5) and Shapiro-Wilk test ($p > 0.05$ for all groups), respectively.

It was found that the average SNR at threshold was significantly different for different age groups: Welch's $F(3, 9.325) = 50.2$, $p < 0.0005$ (this is the smallest value SPSS statistics can give).

The mean SNR at threshold decreased from the age 6-7 group (-0.4 dB ± 3.7), to the age 8-9 group (-2.97 dB ± 3.9) to the age 10-13 group (-6.8 dB ± 2.2) and to the adult group (-14.2 dB ± 1.1), in that order. A Games-Howell post hoc analysis (Games and Howell, 1976) was used due to the unequal variances, which revealed that the decrease from age 6-7 to age 10-13 (6.44, 95% CI (-12.9 to 0.0052)) was statistically significant (p = 0.05), as well as the decrease from age 6-7 to adults (13.81, 95% CI (-20.36 to -7.27), p = 0.003), as well as the decrease from age 8-9 to adults (11.27, 95% CI (7.34 to 15.18), p = 0.005) and the decrease from 10-13 to adults (7.37, 95% CI (-20.3

to -7.27), p = 0.003). These results are summarised in Table 2.7. Overall, they suggest that there is a gradual improvement over these age groups and that all age groups tested, scored significantly worse than the adults.

Age group	8-9 years old	10-13 years old	Adult
6-7 years old	Δ = 2.5 dB 95% CI = -9 to 4 p = 0.619	Δ = 6.4 dB 95% CI = -13 to -0.005 ***p* = 0.05**	Δ = 13.8 dB 95% CI = -20.3 to -7.3 ***p* = 0.003**
8-9 years old	x	Δ = 3.9 dB 95% CI = -0.75 to 8.5 p = 0.113	Δ = 11.3 dB 95% CI = 7.3 to 15.2 ***p* < 0.00052**
10-13 years old	x	x	Δ = 7.4 dB 95% CI = -20.3 to -7.3 ***p* = 0.003**

Table 2.7: Table summarising findings from the ANOVA

2.4.3 Analysis of additional tasks

The analysis of results from all of the additional tasks outlined in the Procedure section are reported in Chapter 4, when combined with the results from Experiment 2 (apart from the SPIN task).

2.4.3.1 *Analysis of the Speech Perception in Noise task (SPIN)*

The relationships between performance on the low predictability (SNR of 2 dB) section of the SPIN and both age and mean SNR at threshold on the LocalB$_{5AT}$ task have been analysed. All 20 participants who completed this task scored between three and eight out of ten. Figure 2.7 shows the relationship between age and SPIN score. It was found that SPIN score improved gradually with age. However, this relationship was examined further using a Pearson's correlation as the relationship is linear. This showed a non-significant correlation of $r(18)$= -0.43, p = 0.06.

Figure 2.8 shows the relationship between mean SNR at threshold on the LocalB$_{5AT}$ task and SPIN score, where lower (better) scores on the LocalB$_{5AT}$ task correlate with higher (better) scores on the SPIN task. This relationship was examined further using a Pearson's correlation as the relationship is linear and the data is normally distributed. This showed a non-significant correlation of $r(18)= -0.30$, $p = 0.20$.

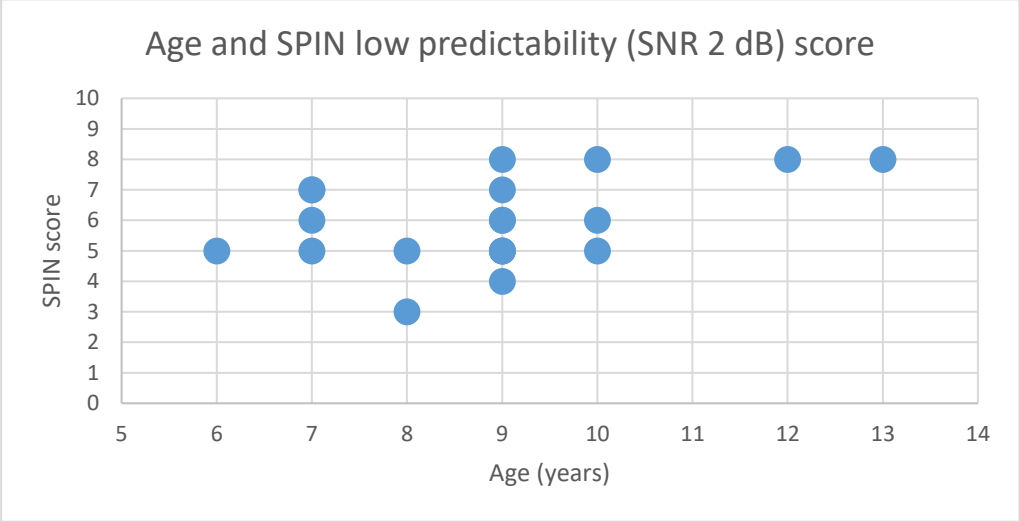


Figure 2.7: Graph showing relationships between SPIN score and age

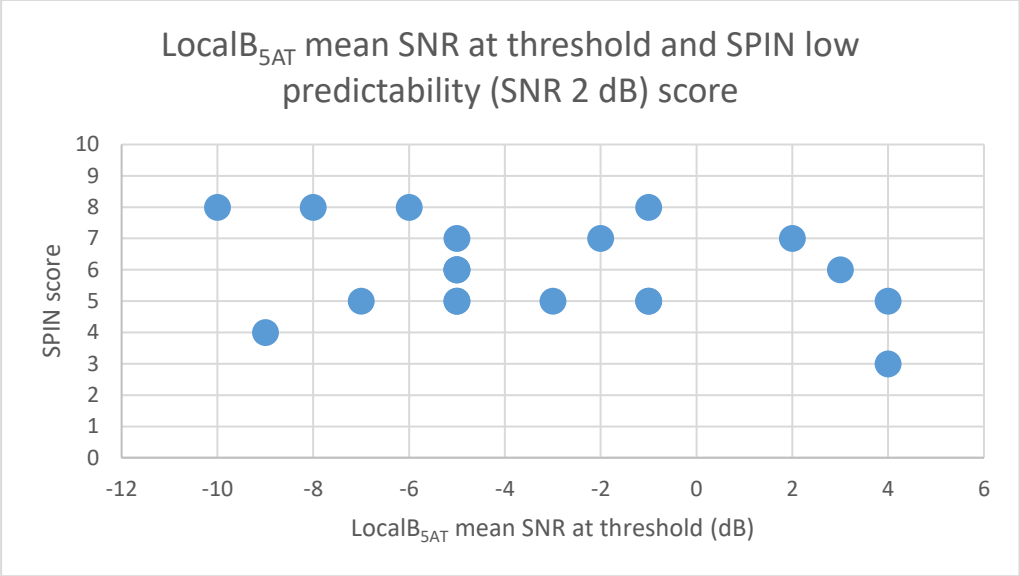


Figure 2.8: Graph showing relationships between SPIN score and LocalB$_{5AT}$ mean SNR at threshold

2.4.4 Analysis of number of tracks used

As stated in the Methods section, the mean SNR at threshold was calculated for each participant from six thresholds (or as many as were returned). It was decided to analyse this data to investigate the benefit of returning six thresholds, rather than just three or four which would enable quicker data collection.

For this analysis the children's thresholds were arranged in the order in which they were collected. The data is shown in Figure 2.9. This graph demonstrates the running mean threshold after each additional track (i.e. the threshold from the first track, then the mean of the thresholds from the first and second tracks, then the mean of the thresholds from the first, second and third tracks and so on). Here it can be seen that within the first few tracks the threshold fluctuates by a relatively large amount for many participants, and then for the last few trials the thresholds are more stable, as would be expected. From the data shown in Table 2.8 it can be seen that the mean difference in threshold generally decreases as the number of thresholds increases.

As the adult's thresholds were collected using two blocks of three interleaved trials, it was not possible to arrange their thresholds in the same way. All three thresholds (of each block) were collected at the same time, so the two blocks must be compared to each other. No significant correlation was found between the two blocks, as shown in Figure 2.10 ($r(8) = -0.04$, $p = 0.916$). This lack of correlation suggests that it is necessary for at least two blocks to be completed by each participant as the data is quite variable.

The data from the child participants suggested that collecting four or five thresholds may be sufficient. However, the lack of correlation between the blocks of adult data suggests variability of the data, such that it would be beneficial to collect a greater number of thresholds. Therefore it was decided that future studies should continue with the collection of six thresholds per participant; as the amount of time saved by collecting fewer thresholds doesn't justify the reduction in quality of the results.

	Mean	Standard Deviation
Difference between threshold 1 and mean of thresholds 1 to 6 (dB)	1.4	0.8
Difference between mean of thresholds 1 and 2 and mean of thresholds 1 to 6 (dB)	0.8	0.6
Difference between mean of thresholds 1 to 3 and mean of thresholds 1 to 6 (dB)	0.5	0.4
Difference between mean of thresholds 1 to 4 and mean of thresholds 1 to 6 (dB)	0.4	0.3
Difference between mean of thresholds 1 to 5 and mean of thresholds 1 to 6 (dB)	0.5	0.3

Table 2.8: Table showing difference in mean thresholds for children

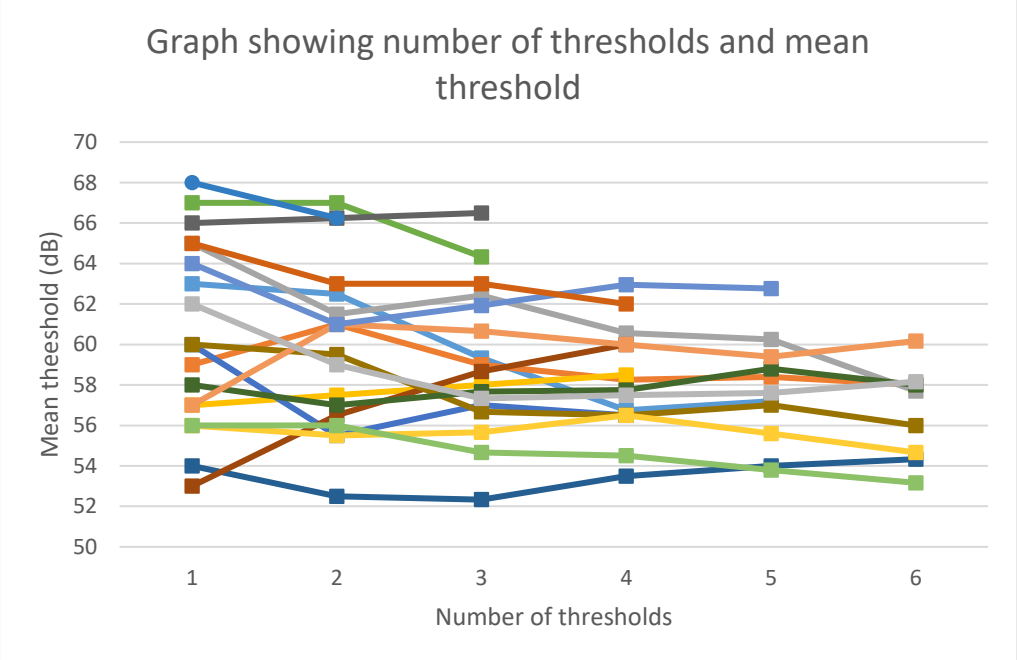


Figure 2.9: Graph showing change in threshold after additional tracks in children

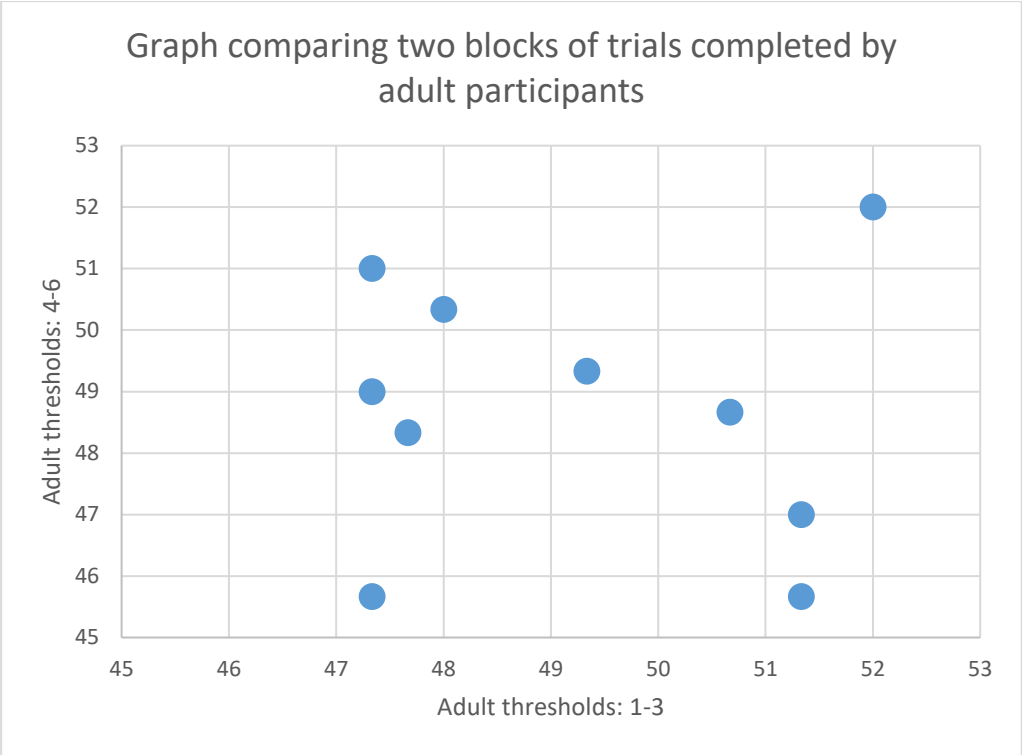


Figure 2.10: Graph showing correlation between the two blocks of tracks in adult participants

2.4.5 Analysis of gender differences

Next, an analysis was carried out to investigate the relationship between participant and speech talker gender. This was done as it was thought that for younger children especially, if they were more used to a particular gender talking (for example if they were raised by a female single parent and had a female teacher at school) then they may be better able to listen to and localise that gender. To do this two independent samples t-tests were carried out, one for the female signal talker condition and the other for the male signal talker condition.

There were no outliers in the data, as assessed by inspection of the boxplot shown in Figure 2.11. The mean SNRs at threshold for each level of gender were normally distributed, as assessed by Shapiro-Wilk's test ($p > 0.05$). There

were 17 male and 13 female participants. The data are mean ± standard deviation (summarised in Table 2.9).

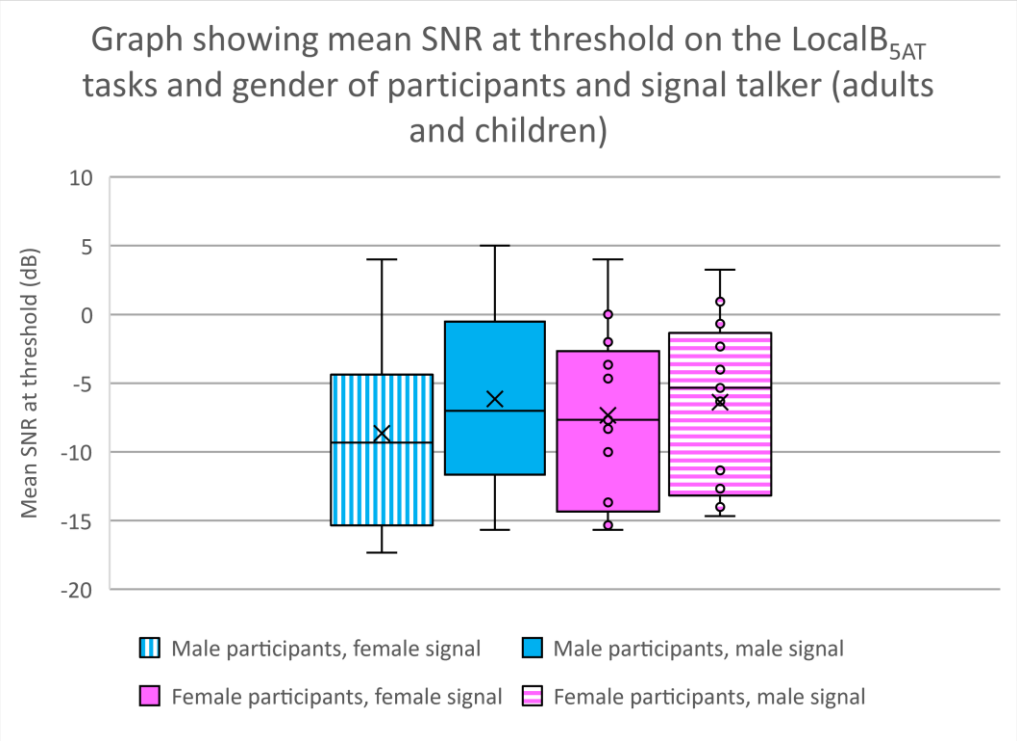


Figure 2.11: Boxplots showing mean signal level at threshold and gender of participants and signal talker

		Participant	
		Female	**Male**
Signal Talker	**Female**	-7.3 dB	-8.7 dB
	Male	-6.4 dB	-6.2 dB

Table 2.9: Table showing mean SNRs for signal talker gender and participant gender

For the female talker condition the male participants (-8.7 dB ± 6.6) had better SNRs at threshold than the female participants (-7.3 dB ± 6.4), which is a difference of 1.4 dB (95% CI, -6.3 to 3.6). There was homogeneity of variances for mean threshold for female talker condition for males and females, as assessed by Levene's test for equality of variances (p = 0.987). However, there was not a statistically significant difference in mean thresholds for male and female participants, $t(27)$ = -0.545, p = 0.590.

For the male talker condition the male participants (-6.2 dB ± 6.2) had better SNRs at threshold than the female participants (-6.4 dB ± 6.2). This is a difference of only 0.2 dB so no further analysis was necessary.

From the results of these t-tests (summarised in Table 2.9) it can be seen that there are no significant effects of gender on mean signal level at threshold for the conditions with different signal talker genders. Therefore it was decided that in order to simplify the task for similar future experiments only one condition (female signal with male babble) would be used for all participants.

2.5 Discussion

This experiment only looked at children aged from 6 to 13 years. Following this, Experiment 2 tested a wider age range to ensure that the developmental time course from six years of age to early adulthood was fully mapped. Therefore, as this experiment will continue as part of Experiment 2 with more children from a wider age range, further discussion takes place in Chapters 3 and 4.

In the localisation of speech in babble task, adults performed significantly better than the children. For children there was a gradual improvement of mean SNR at threshold with age, which was fitted with an exponential function with a time constant of 6.6 years (this is not the age at which adult-like performance is reached). This means that if the decay rate of the exponential function was to become constant (at any point in the function), it would reach the asymptote in 6.6 years and every 6.6 years performance

improves by 63.2%. An ANOVA showed significant differences between 6-7 year olds and 10-13 year olds, 6-7 year olds and adults, 8-9 year olds and adults and 10-13 year olds and adults. This suggests that adult-like performance does not occur until at least 13 years of age. This is in line with the findings of Boothalingam et al. (2016) who found that localisation of speech in babble probably does not reach adult-like levels until at least 16 years of age. Also, this is much later than the age at which localisation in quiet abilities reach adult-like levels (6 years of age), showing how children's localisation abilities are negatively affected by informational masking.

The results from this experiment confirm the hypothesis that localisation in babble abilities continue developing throughout childhood, so are better in adults than in children. ANOVA tests showed that adult performance was significantly better than all of the subgroups of children (6-7 years old, 8-9 years old and 10-13 years old). The results suggest a gradual improvement over the age range tested, demonstrated by the exponential function fitted to the data. They also show that even the oldest children who participated (13 years old) could not consistently perform at an adult-like level on the task. Boothalingam et al. (2016) found that children aged 7 to 16 years had poorer localisation in babble abilities than adults but did not show any age effects. The results from this experiment build upon those obtained by Boothalingam et al. (2016) by showing age effects in children aged 6 to 13 years.

A limitation of this experiment was that the SPIN task used may not have been adequate for assessing the masked speech perception abilities of children. It

was surprising that no significant correlations were found between performance on this task and either age or performance on the $LocalB_{5AT}$ task. A previous study by Elliott (1979) showed age-related performance in children aged 9 to 17 years on a task very similar to this one, whereas in this experiment no significant effect of age was found. This may be because the age group tested here was generally younger than those in the Elliott (1979) study. However, it would be expected to still see some age effects as a review paper by Leibold (2017) found that masked speech perception abilities reach adult-like levels by 13 to 14 years of age. It is possible that the number of trials contained in this task (10) was not sufficient to see any effect that may have been present. Due to the differences between the SPIN and $LocalB_{5AT}$ tasks, such as difference in number of talkers in the babble and variety of words used, it cannot be known whether no correlation was found due to these abilities not being related or due to the differences in the methods of these tasks. Therefore, if a child performed poorly at the localisation in babble task it cannot be confidently known whether this was caused by poor masked speech perception, meaning they were unable to perceive the stimulus to localise it, or if they could adequately perceive the speech in babble but could not localise. This limitation is addressed in Experiment 2.

Chapter 3

Experiment 2: Detection, Identification and Localisation of speech in babble in typically developing children and adults

3.1 Introduction

3.1.1 Investigating adult-like performance of localisation of speech in babble

This experiment builds on the results of Experiment 1, to complete the developmental time course of localisation of speech in babble in children aged six years and older. Experiment 1 only tested children aged 6 to 13 years, so this new experiment adds to Experiment 1 by testing children aged 10 to 17 years. As in Experiment 1, cognitive abilities will be tested in children using the non-verbal parts of the WASI, dichotic digits, digit span and questionnaires (ECLiPS and Conners). The results from these two experiments will be combined for joint analysis in Chapter 4.

3.1.2 Detection, identification and localisation of speech in babble

In this experiment a new speech identification in babble task was introduced, to replace the one used in Experiment 1 which was found to have limitations. The purpose of this task was to compare the ability of participants to localise and identify speech in babble. In order to better compare these abilities, a new localisation of speech in babble task was also introduced. To complement these tasks a detection of speech in babble task was also added to the experiment. Therefore the relationships between detection, identification and localisation of speech in babble will be investigated.

Little work has been done to investigate the relationships between these tasks. Hawkins and Stevens (1950) measured thresholds of speech intelligibility and detection of four adult participants, using white noise as the masker. Surprisingly, they found that performance was better on identification, by an average of 9 dB. These findings are supported by those of Liu and Jin (2019), who tested vowel identification in long-term speech shaped noise in ten adult participants. They found that thresholds were dependent on which vowel was presented and that performance was better for identification than detection on average (over all vowels by approximately 8 dB, but they did not state if this difference was statistically significant).

It would appear that no studies have yet investigated detection and localisation, and just one has looked at identification and localisation in

adults. Hawley, Litovsky and Colburn (1999) compared the performance of 12 participants aged 18 to 21 years on speech intelligibility and localisation in background sound tasks, but with inconclusive findings. They used male voiced IEEE sentences for the target and the masker sentences, which were each approximately 62 dB(A). They used one, two or three masker sentences and when more than one was present they came from different locations. Both tasks were carried out in a sound field and over headphones. In the speech intelligibility task, participants were asked to identify the target sentence, from a known location, and ignore the masker sentences. The target came from either -90°, 0° or 90°. The masker(s) came from -90°, -60°, -30°, 0°, 30°, 60° or 90°. Performance was measured using the percentage of key words incorrect for each condition (error rate). In the localisation task participants were asked to identify the location of a known sentence and ignore the masker sentences. The target and masker sentences could have been located at -90°, -60°, -30°, 0°, 30°, 60° or 90°. Performance was measured using RMS error, percent correct and the correlation coefficient *r* for the least-squares linear regression (between actual target location and the participant's judged target location). Due to the large separation angles and high levels of the target sentences, the localisation results for all participants were close to maximum performance. This means it was not possible to make useful comparisons to the speech intelligibility performance data. Therefore, the relationships between identification and localisation of speech in babble are still unknown in both adults and children. In adults it has been found that

identification of masked speech is better than detection of masked speech, but in children this has not yet been established.

The experiment described in this chapter investigated for the first time the relationships between detection, identification and localisation of speech in babble in children and adults using the tasks described below and in Table 3.1 and Table 3.2:

- Detection in babble ($DetB_{2CS}$): This task was referred to as $DetB_{2CS}$. The components of this name were chosen for the following reasons: 'Det' from detection, 'B' from babble and '2CS' from 2 choice constant stimuli task. The participant was required to respond yes/no to whether a signal (colour word: red, green, white or blue) was present in the background of babble. The signal was present 50% of the time and always came from the same location (0°).
- Identification in babble ($IdentB_{4CS}$): This task was referred to as $IdentB_{4CS}$. The components of this name were chosen for the following reasons: 'Ident' from identification, 'B' from babble and '4CS' from 4 choice constant stimuli task. The participant was required to identify which word the signal was (colour word: red, green, white or blue) in the background of babble. This signal was present 100% of the time and always came from the same location (0°).
- Localisation in babble ($LocalB_{5CS}$): This task was referred to as $LocalB_{5CS}$. The components of this name were chosen for the following reasons: 'Local' from localisation, 'B' from babble and '5CS' from 5

choice constant stimuli task. The participant was required to identify the location of the signal (colour word: red, green, white or blue) in the background of babble. The signal came from either -30°, -15°, 0°, 15° or 30°, and it was present 100% of the time.

	Direction	**Word**	**Response**
Detection ($DetB_{2CS}$)	0° (when signal present)	Red, green, white or blue	Yes/no
Identification ($IdentB_{4CS}$)	0°	Red, green, white or blue	Red, green, white or blue
Localisation ($LocalB_{5CS}$)	-30°, -15°, 0°, 15° or 30°	Red, green, white or blue	-30°, -15°, 0°, 15° or 30°

Table 3.1: Table showing design between tasks

	Direction	**Word (red, green, white or blue)**	**Presence**
Detection ($DetB_{2CS}$)	Ignore	Ignore	**Report**
Identification ($IdentB_{4CS}$)	Ignore	**Report**	Ignore (always present)
Localisation ($LocalB_{5CS}$)	**Report**	Ignore	Ignore (always present)

Table 3.2: Table showing which parts of the tasks participants were required to report/ignore

3.2 Aims and Hypotheses

The aims of this experiment were to address the limitations of Experiment 1, firstly by testing a wider age range of children, and secondly, by adding in new $DetB_{2CS}$, $IdentB_{4CS}$ and $LocalB_{5CS}$ tasks.

The introduction of these tasks also allow us to have greater understanding about the relationship between these abilities, in both typically developing children and those who experience difficulties with listening in babble. This is the first study to compare performance on these tasks in adults and children. It is hypothesised that children who have difficulties with speech detection

and/or identification in babble will perform poorly in the LocalB$_{5CS}$ task. The reason is that if they cannot tell that a word is present then it will probably be difficult for them to localise it, and if they perform poorly on either of these tasks it could suggest that they have difficulties with processing in babble so will struggle with the LocalB$_{5CS}$ task also. It is expected that participants will perform better in identification than detection of speech in babble tasks in line with previous findings (Hawkins and Stevens, 1950; Liu and Jin, 2019).

It is also hypothesised that children in the age range already tested (6 to 13 years of age) will perform to the same level as in Experiment 1, and it is predicted that older children aged 14 to 17 years of age will perform at the levels predicted by the exponential function fitted to the data from Experiment 1, shown in Equation 3.1.

$$Predicted\ mean\ SNR\ at\ threshold\ (dB) = -15.3 + (44.9 * e^{\left(-\frac{age}{6.6}\right)})$$

Equation 3.1: Equation for predicted mean SNR at threshold in Experiment 1

3.3 Methods

3.3.1 Participants

As in Experiment 1, participants with normal hearing (20 dB HL or better at 0.25, 0.5, 1, 2, 4 and 8 kHz), who spoke English as their first language and had no learning or language disorders (except for mild dyslexia and/or dyspraxia which caused them few difficulties) were recruited for this experiment. The latter criterion was set because it was thought likely that learning or language disorders could cause attentional deficits or difficulty understanding the

instructions for the task, which would then negatively affect performance on the localisation tasks. As the main purpose of this experiment was to investigate performance in typically developing children, the presence of any learning/language disorders or hearing impairment would have introduced confounds.

3.3.1.1 Adults

Five adults (three female, two male) aged 23 to 29 years (mean 25.1, s.d. 1.8) took part in the experiment. They were all monolingual English speakers with normal hearing and no language or learning difficulties. These five participants were PhD students in the same university department as the researcher, but were unaware of the purpose of the experiment.

3.3.1.2 Children

35 children (14 female, 21 male) aged 10 to 17 years (mean age 13.7 years, s.d. 1.7) took part in the experiment. The number of children of each age is shown in Table 3.3. All of the children were monolingual English speakers, with normal hearing and no confirmed language or learning difficulties (except two who had mild dyslexia and/or dyspraxia). They were recruited through online advertisements using Facebook and Twitter.

An additional four children were also tested but their results had to be excluded for the reasons shown in Table 3.4. For one participant full testing was not possible due to technical problems with the equipment. One child was unable to complete the localisation tasks so was excluded. One child was excluded for having ADHD, dyslexia and possible autism spectrum disorder

and another was excluded for having ADHD, dyslexia and dyspraxia. Although the experiment advertisement made it clear that participants should not have any learning/language disorders, some parents did not realise this until attending the test session.

Age (years)	Number of participants
10	2
11	6
12	3
13	10
14	6
15	4
16	3
17	1
Adults	5

Table 3.3: Table showing ages of participants

Reason	Frequency
Insufficient data collected due to technical problems	1
Insufficient data collected due to child being unable to complete task	1
ADHD, dyslexia and possible autism spectrum disorder	1
ADHD, dyslexia and dyspraxia	1
Total	**4**

Table 3.4: Table showing reasons for exclusion

3.3.2 Equipment

The same equipment was used as in Experiment 1. The MATLAB code used to run the experiment was written by Owen Brimijoin and Michael Akeroyd as part of the overall programming of the laboratory.

3.3.3 Stimuli

3.3.3.1 Signal

The signals used were the same single speech words extracted from the CRM sentence corpus as used in Experiment 1, defined by Bolia et al. (2000) and

recorded by Kitterick, Bailey and Summerfield (2010). As before, the words used were 'red,' 'green,' 'white' and 'blue.' However, for this experiment only female voiced words were used for the signal. It was decided to only use the female voice/male babble condition in order to simplify the experiment, as in Experiment 1 no difference in performance was found between the different participant and talker genders.

3.3.3.2 Masker (babble)

As in Experiment 1 the stimuli used were six IEEE sentences (IEEE, 1969) and they were played from six different locations in the ring (one sentence from each speaker). Unlike in Experiment 1, only male voiced sentences were used for the babble.

3.3.4 Procedure

All participants attended a single session at the MRC Institute of Hearing Research laboratory in the QMC Hospital, Nottingham. Adult sessions lasted approximately one hour, child sessions lasted 90 to 120 minutes. Adult participants were given £7.50 inconvenience allowance and any travel expenses were reimbursed. Children were given £15 in Amazon vouchers to thank them for their time and their parents were reimbursed for any travel expenses. Parental consent was obtained prior to the experiment beginning.

Table 3.5 shows the tasks which were used in the experiment, whether they were completed by children and/or adults and if they were carried out in the same way as in Experiment 1. As shown in the table, children completed all

tasks while adults only completed pure tone audiometry and speaker ring tasks. Participants were given opportunities to take breaks between tasks.

As in Experiment 1 the order of the right and left condition for the dichotic digits test were counterbalanced to reduce the risk of order effects.

Task	Completed by children	Completed by adults	Same as in Experiment 1
Pure tone audiometry	✓	✓	✓
Non-verbal parts of the WASI (blocks and matrices)	✓		✓
Digit span	✓		✓
Dichotic digits	✓		✓
Questionnaires	✓		✓
Speaker ring tasks (detection, identification and localisation)	✓	✓	

Table 3.5: Table showing study tasks for adults and children

Pure tone audiometry, non-verbal parts of the WASI, digit span, dichotic digits and questionnaires were all carried out in the same ways and for the same reasons as stated in Experiment 1.

The localisation task carried out in Experiment 1 was repeated using nearly the same method. The exceptions were:

1. In all presentations the signal was female voiced and the babble was male voiced.
2. During the familiarisation tasks, localisation in quiet was only tested in randomly presented order, unlike in Experiment 1 where it was also tested in order (azimuths of -30°, -15°, 0°, 15°, 30°).
3. Following the localisation in quiet practice task, localisation in quiet was tested in a constant stimuli task with 30 trials.

3.3.4.1 Main Localisation tasks

As in Experiment 1 prior to the tasks shown in Table 3.6 the participant was played the numbers 1 to 24 (voiced by a female talker at 50 dB(A)), each from a different speaker in the ring and moving around the ring in the clockwise direction. This checked that all of the speakers were working whilst familiarising the participant with their environment. The participants were given the same instructions as in Experiment 1.

Practice task 1: Localisation in quiet (random)

This task was the same as practise task 2 used in Experiment 1.

Mini task/Practice task 2: Localisation in quiet (LocalQ_{5CS})

This task was referred to as LocalQ_{5CS}. The components of this name were chosen for the following reasons: 'Local' from localisation, 'Q' from quiet and '5CS' from 5 choice constant stimuli task. It was decided that it would be useful to have a measure of localisation in quiet (as well as the familiarisation task) in order to compare performance in quiet and in babble. The signal was presented at 60 dB(A) six times in a random order from each of the following azimuths: ±30°, ±15°, 0°. No babble was present. This task had not been carried out in Experiment 1.

Practice task 3: Localisation in low babble

This task was the same as practice task 3 used in Experiment 1.

Main task: Localisation of speech in babble adaptive track (LocalB$_{5AT}$)

This task was the same as the main task used in Experiment 1, except only the female voiced signal and male voiced babble condition was carried out. Six thresholds were still collected.

Test		**Localisation in quiet (random)**	**Localisation in quiet**	**Localisation in low babble**	**Localisation in babble**
Abbreviation		-	LocalQ$_{5CS}$	-	**LocalB$_{5AT}$**
Purpose		Practice	Task	Practice	**Main task**
Description		Stimuli presented from each of the 5 azimuths twice in a random order	Stimuli presented in a random order from 1 of the 5 signal locations 30 times (each location is used 6 times)	Stimuli presented from each of the 5 azimuths twice in a random order, but with a low level of babble present	**6 thresholds collected**
Method		Constant stimuli	Constant stimuli	Constant stimuli	**Adaptive track (2-down 1-up)**
Number of tracks		1 track with 10 trials (all female voiced)	1 track with 30 trials (all female voiced)	1 track with 10 trials (all female voiced)	**6 tracks (all female voiced)**
Response		Choose from one of five options	Choose from one of five options	Choose from one of five options	**Choose from one of five options**
Signal	**Type**	Word	Word	Word	**Word**
	Location	±30°, ±15°, 0°	±30°, ±15°, 0°	±30°, ±15°, 0°	**±30°, ±15°, 0°**
	Level	60 dB(A) every trial	60 dB(A) every trial	60 dB(A) every trial	**60-65 dB(A) starting level**
Masker	**Type**	None	None	6 IEEE sentences (all male voiced)	**6 IEEE sentences (all male voiced)**
	Location	None	None	6 locations around the ring	**6 locations around the ring**
	Level	None	None	Each sentence 30 dB(A), total 38 dB(A)	**Each sentence 55 dB(A), total 63 dB(A)**
Outcomes		% correct	% correct	% correct	**Mean SNR at threshold**

Table 3.6: Table showing protocol for localisation tasks

3.3.4.2 Detection, Identification and Localisation of speech in babble tasks ($DetB_{2CS}$, $IdentB_{4CS}$, $LocalB_{5CS}$)

In addition to the localisation in babble tasks shown above, each participant completed 3 constant stimuli tasks: $DetB_{2CS}$, $IdentB_{4CS}$ and $LocalB_{5CS}$ (described in Table 3.7). These tasks were carried out using the same equipment and stimuli as the previous localisation tasks. These tasks were always carried out after the previous localisation tasks.

Test		**Detection in babble**	**Identification in babble**	**Localisation in babble**
Abbreviation		$DetB_{2CS}$	$IdentB_{4CS}$	$LocalB_{5CS}$
Number of Presentations		160 (80 with signal present, 80 with signal absent; 16 at each of 5 levels)	80 (16 at each of 5 levels)	75 (15 of each of 5 levels)
Response		Choose from one of two options	Choose from one of four options	Choose from one of five options
Response Options		Was sound present: yes or no	Word: red, green, blue or white	±30°, ±15°, 0°, represented by shapes
Signal	**Type**	Word	Word	Word
	Location	0°	0°	±30°, ±15°, 0°
	Level	-9, -6, -3, 0 and 3 dB in relation to threshold from initial localisation in babble task	-9, -6, -3, 0 and 3 dB in relation to threshold from initial localisation in babble task	-9, -6, -3, 0 and 3 dB in relation to threshold from initial localisation in babble task
Noise	**Type**	6 IEEE sentences	6 IEEE sentences	6 IEEE sentences
	Location	6 locations around the ring	6 locations around the ring	6 locations around the ring
	Level	Each individual sentence: 55 dB(A), Total: 63 dB(A)	Each individual sentence: 55 dB(A), Total: 63 dB(A)	Each individual sentence: 55 dB(A), Total: 63 dB(A)
Practice		5 trials (no feedback)	5 trials (no feedback)	6 trials (no feedback)
Outcomes		% correct at each level	% correct at each level	% correct at each level

Table 3.7: Table showing Detection, Identification and Localisation of speech in babble constant stimuli tasks

Stimuli level choice for $DetB_{2CS}$, $IdentB_{4CS}$, $LocalB_{5CS}$ tasks

For these tasks the level of the babble was constant, as in the previous localisation tasks. Each of the six sentences was presented at 55 dB(A), making the total level 63 dB(A). Signal level was based on the approximated mean signal level at threshold in the previous localisation task (adaptive track). The signal levels used were -9, -6, -3, +0 and +3 dB relative to the approximated mean signal level at threshold. For example, if a participant's mean signal level at threshold was 50 dB, the levels used in the constant stimuli tasks would be 41, 44, 47, 50 and 53 dB(A). These levels were chosen following piloting, as they were found to be challenging but not impossible.

Identification in babble ($IdentB_{4CS}$)

The participant sat in the centre of the speaker ring and was instructed that they had to listen for a word and report which word they heard. They were told the babble would be present, as in the previous localisation task ($LocalB_{5AT}$). The participants were told that the level of the word would vary, so sometimes it would be easy as the word would be loud and sometimes the word would be very quiet so it would be difficult. After each presentation the participant used the touchscreen to press the words 'red,' 'green,' 'white' or 'blue' to indicate which colour they heard, making this a one interval-4 choice task. The signal word was always presented from 0°. The babble was presented from six locations around the ring, which changed for each presentation as in previous tasks ($LocalB_{5AT}$). This task consisted of 80 presentations, 16 at each of the five levels (explained above). Prior to the task

beginning the participant completed five practice trials with no feedback, to ensure they understood the task. Performance was measured as percentage of responses correct for each level.

Detection in babble ($DetB_{2CS}$)

The participant sat in the centre of the speaker ring and was instructed that they would hear a word in half of the presentations and they had to report whether they heard it or not. They were told the babble would be present, as in previous tasks ($LocalB_{5AT}$ and $IdentB_{4CS}$). The participants were told that the level of the word would vary, so sometimes it would be easy as the word would be loud and sometimes the word would be very quiet so it would be difficult. After each presentation the participant used the touchscreen to press the word 'Yes' if they thought the word was present and 'No' if they thought only babble was presented, making this a one interval-yes/no task. When the word was present it was always presented from 0°. The babble was presented from six locations around the ring, which changed for each presentation, as in the previous localisation tasks. This task consisted of 160 presentations, 80 with the signal present and 80 with the signal absent. The signal was presented 16 times at each of the five different levels (explained above). Prior to the task beginning the participant completed five practice trials with no feedback, to ensure they understood the task. Performance was measured as percentage of correct responses for each level.

Localisation in babble (LocalB$_{5CS}$)

The participants were instructed that this task was essentially the same as the previous localisation in babble task (LocalB$_{5AT}$), but that they may find it a little more challenging due to the signal levels (explained above). The instructions to use the touchscreen to identify the location of the signal were as before. The same locations were used as in the previous localisation tasks (±30°, ±15°, 0°), so it was still a one interval-five choice task. Prior to the task beginning the participant completed six practice trials with no feedback, to ensure they understood the task. Performance was measured as percentage of correct responses for each level.

3.4 Results

All five adult participants completed the tasks, giving six thresholds. 32 child participants gave six thresholds, the other three returned four or five thresholds, as shown in Table 3.8.

Number of thresholds returned	Frequency
1	0
2	0
3	0
4	1
5	2
6	32

Table 3.8: Table showing number of thresholds returned in children

3.4.1 Localisation of speech in babble adaptive track results (LocalB$_{5AT}$)

3.4.1.1 Initial representation of results

The boxplot in Figure 3.1 shows the LocalB$_{5AT}$ data for adults and different age groups of children. From visual inspection it can be seen that mean SNR at threshold improves for each age group.

The mean SNR at threshold was calculated for each participant by taking the mean of all their finished thresholds and subtracting 63 dB (the level of the babble). From this an exponential curve was fitted using the Solver function in Excel, shown in Equation 3.2. This allows interpolation to estimate performance at different ages. Its parameters were: -15.3 dB is the asymptote, 48 dB/year is the slope and 6.0 years is the time constant. For this equation R^2=24%.

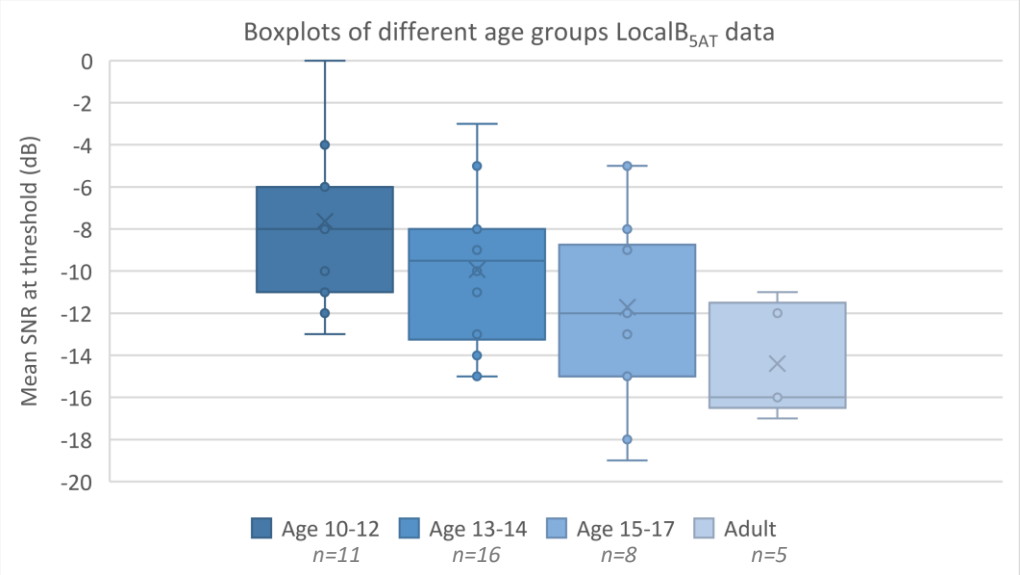


Figure 3.1: Boxplots of different age groups LocalB$_{5AT}$ data

$$Predicted\ mean\ SNR\ at\ threshold\ (dB) = -15.3 + (48.0 * e^{\left(-\frac{age}{6.0}\right)})$$

Equation 3.2: Equation for predicted mean SNR at threshold

Figure 3.2 shows the mean SNR at threshold for each participant, plotted against their age. The dashed line separates the children's and the adult's data points. From this graph it can be seen that the general trend is that performance improves with age, although some of the older children (aged 15 and 16 years) performed better than any of the adults.

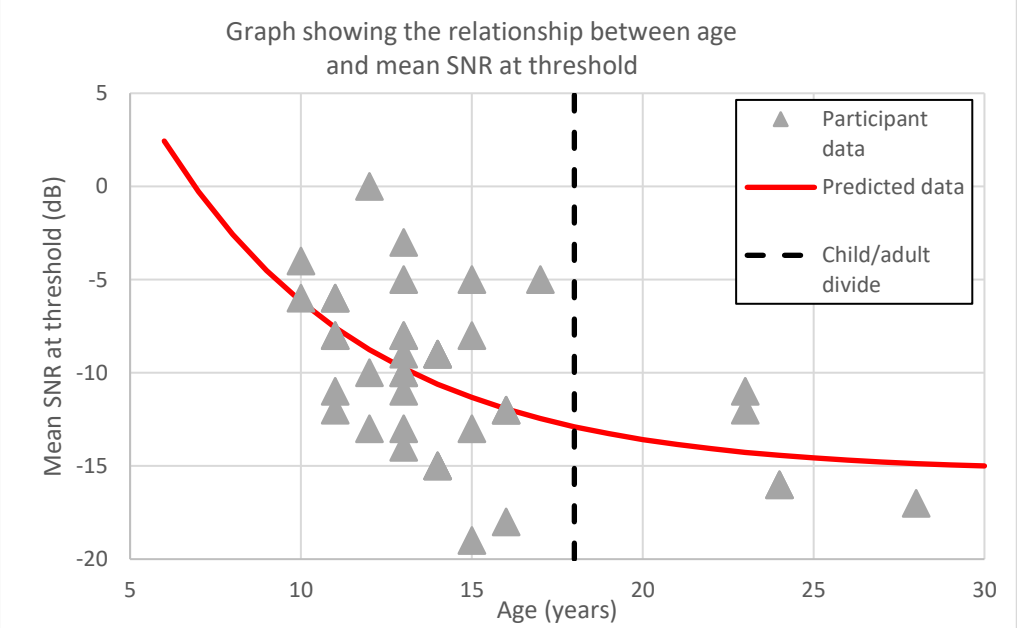


Figure 3.2: Graph showing relationship between performance and age for Experiment 2

3.4.1.2 Comparison of different age groups using a one-way ANOVA

A one-way ANOVA was conducted to determine if average SNR at threshold was different for groups of different age ranges. For this participants were classified into four groups: aged 10-12 (*n* = 11), aged 13-14 (*n* = 16), aged 15-17 (*n* = 8) and adults (*n* = 5). There were no outliers and the data was normally distributed for each group, as assessed by boxplot (shown in Figure 3.1) and

Shapiro-Wilk test ($p > 0.05$ for all groups). Homogeneity of variances was not violated, as assessed by Levene's Test of Homogeneity of Variance ($p = 0.4$).

The average SNR at threshold was significantly different for each of the age groups, $F(3, 36) = 3.7$, $p = 0.02$.

Average SNR at threshold decreased from the age 10-12 group (-7.6 dB ± 3.8), to the age 13-14 group (-10.2 dB ± 3.5) to the age 15-17 group (-11.5 dB ± 5.3) and to the adult group (-14.4 dB ± 2.7), in that order. Tukey post hoc analysis revealed that the decrease from age 10-12 to adult (6.8, 95% CI (1.0 to 12.5 dB)) was statistically significant ($p = 0.015$) with a small effect size ($\omega^2 = 0.17$). No other differences were statistically significant.

3.4.2 Detection, identification, localisation of speech in babble results ($DetB_{2CS}$, $IdentB_{4CS}$, $LocalB_{5CS}$)

The mean results from each of the $DetB_{2CS}$, $IdentB_{4CS}$ and $LocalB_{5CS}$ tasks with results split by age group (adults and children), are shown in Figure 3.3, Figure 3.4 and Figure 3.5. Note that the numbers on the x axis (-9, -6, -3, 0 and 3 dB) are not the SNRs, but are the signal level in relation to the mean SNR at threshold from the $LocalB_{5AT}$ task for the participant. It is not possible to look at the developmental time courses of these results as each participant had different signal levels presented to them, so their results are not easily comparable to each other.

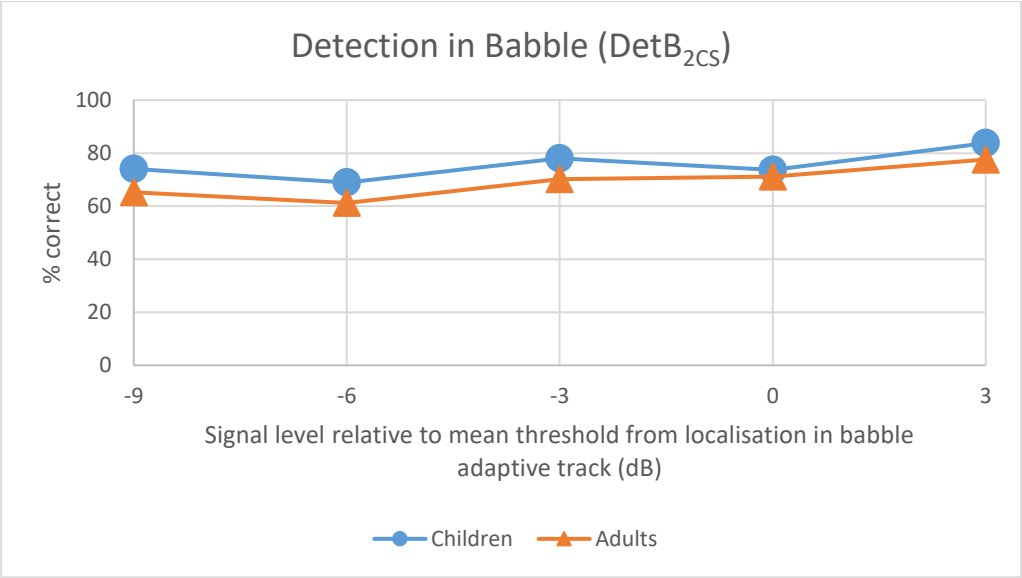


Figure 3.3: Detection in babble (adults and children)

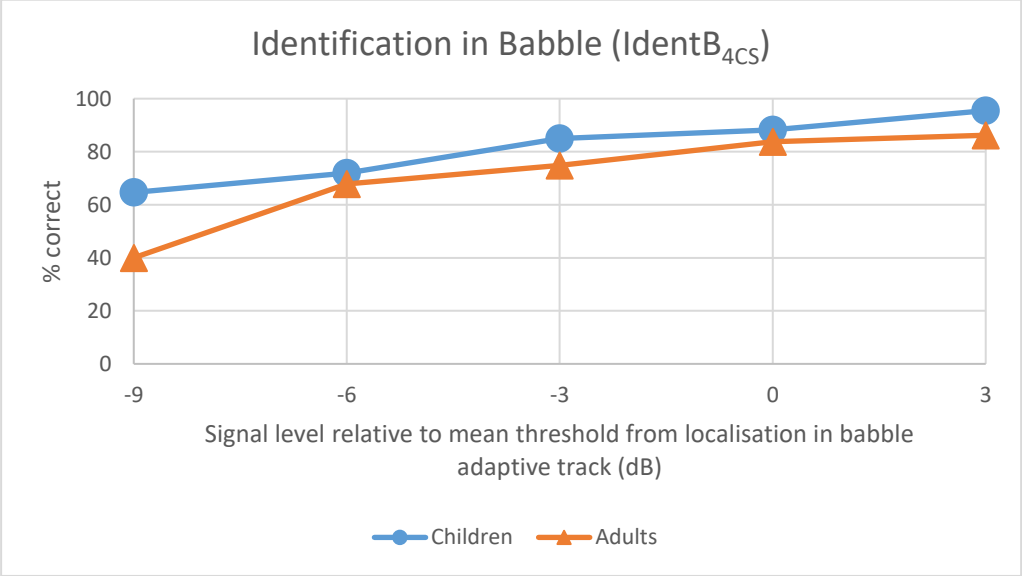


Figure 3.4: Identification in babble (adults and children)

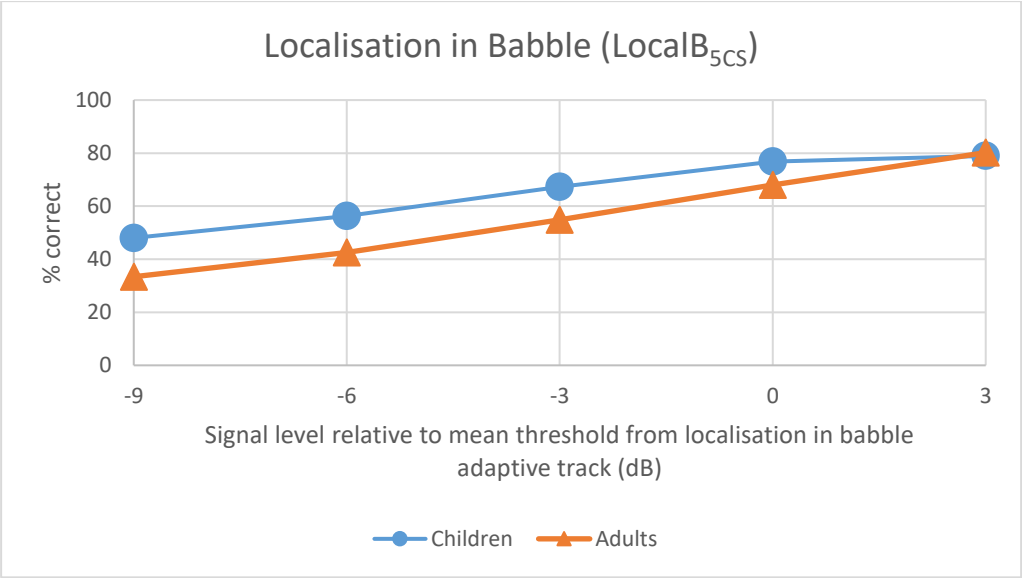


Figure 3.5: Localisation in babble (adults and children)

In Figure 3.3 it can be seen that the performance for $DetB_{2CS}$ task is fairly constant across the range tested, with little difference between adults and children. Figure 3.4 shows a gradual improvement in performance with level on the $IdentB_{4CS}$ task with, with only a small difference between adults and children. Figure 3.5 shows the $LocalB_{5CS}$ task results. This graph shows a gradual improvement in performance with level, with little difference between the adults and the children.

It is not surprising that the adult and child mean scores are similar. As the signal levels used for these tasks were chosen for each participant based on their threshold from the $LocalB_{5AT}$ task, in order to compensate for age-related differences in abilities, the tasks should have been of equal difficulty for each participant. The mean performance of all participants (adults and children together) are shown for all three tasks in Figure 3.6. Overall,

performance was generally highest in the $IdentB_{4CS}$ task, followed by the $DetB_{2CS}$ task and then the $LocalB_{5CS}$ task.

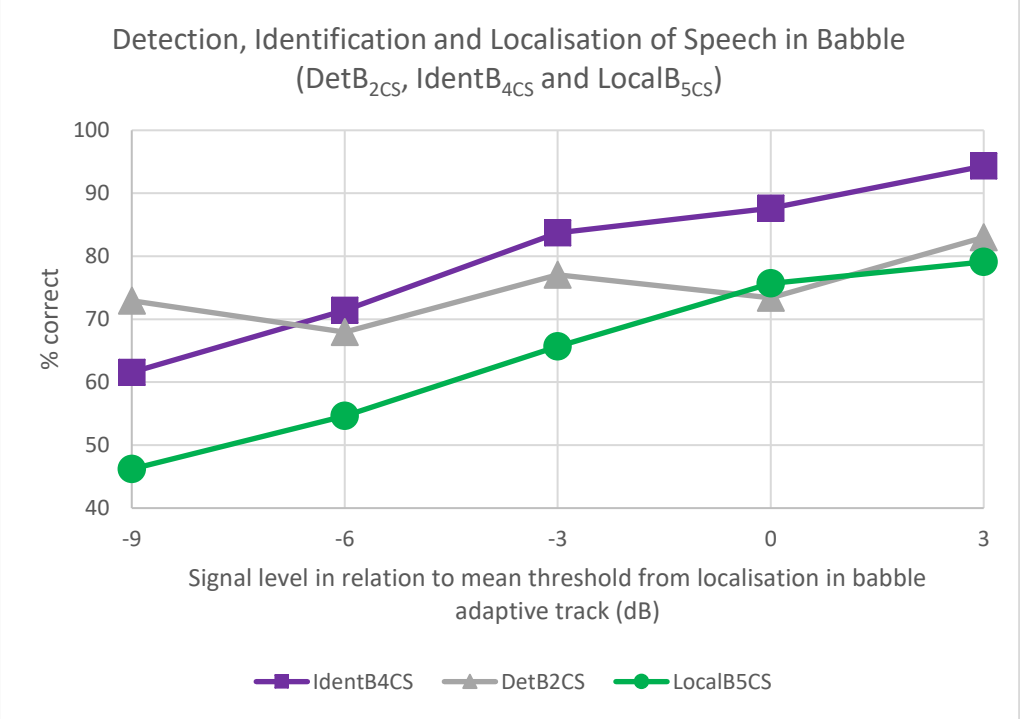


Figure 3.6: Graph showing mean performance for all participants on all constant stimuli tasks

A one-way repeated measures ANOVA was carried out to investigate any differences between mean scores (% correct) across the SNRs for each task. All task averages were normally distributed as assessed by the Shapiro-Wilk test of normality, except for $IdentB_{4CS}$ (p = 0.05). Three outliers were present, according to visual assessment of boxplots shown in Figure 3.7, but despite this they were not removed as they are plausibly genuine values (some children did very well on the tasks and appeared to find them easy, whereas others struggled). The assumption of sphericity was violated (according to Mauchly's test of sphercitiy, p=0.007). Epsilon (ε) was 0.812, as calculated according to Greenhouse and Geisser (1959), and was used to correct the

one-way repeated measures ANOVA. Task performance was significantly different for the different tasks,

$F(1.62, 63.35) = 48.0$, $p < 0.0005$, $\omega^2 = 0.44$.

The post hoc testing results of comparisons between the tasks are shown in Table 3.9. From this, it can be seen that there were significant differences between all tasks. The ANOVA showed that performance on the $IdentB_{4CS}$ task was significantly better than both the $DetB_{2CS}$ and the $LocalB_{5CS}$ tasks, and performance on the $DetB_{2CS}$ task was significantly better than the $LocalB_{5CS}$ task.

	$IdentB_{4CS}$	**$DetB_{2CS}$**
$LocalB_{5CS}$	$IdentB_{4CS}$ significantly better Mean difference=15.4 95% CI, 11.0 to 19.9 p<0.0005	$DetB_{2CS}$ significantly better Mean difference=10.6 95% CI, 6.1 to 15.2 p<0.0005
$IdentB_{4CS}$	x	$IdentB_{4CS}$ significantly better Mean difference=4.9 95% CI, 1.9 to 7.8 p<0.0005

Table 3.9: Table showing results from one-way repeated measures ANOVA

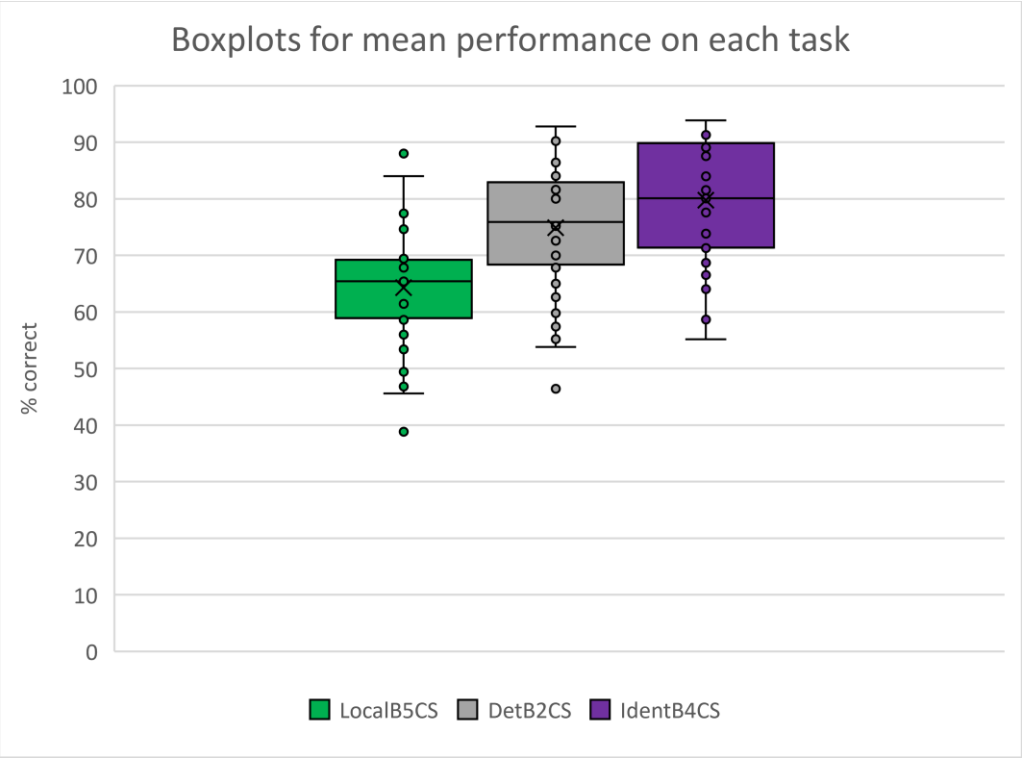


Figure 3.7: Boxplots for mean performance on each task (for all participants)

3.4.2.1 Initial signal detection theory analysis (d')

The data from the $DetB_{2CS}$ task can also be analysed using d', described by MacMillan and Creelman (2008). d' analysis uses the false alarm and hit rates to calculate a measure of sensitivity, which is less biased than the raw data. It also allows data from tasks with a different number of response options to be compared. It gives a better measure of how detectable the signal is, removing the bias caused by participants who each approach the task with their own strategy. A worked example for one participant is shown below, using Table 3.10 which shows the data from one participant (RB201) on the $DetB_{2CS}$ task using the d' results. In Table 3.10 the columns shows the five SNRs and the rows show the steps in the d' calculation.

Signal level	-9	-6	-3	0	3	No signal
Number correct	8/16	9/16	8/16	12/16	9/16	59/80
Signal present and correctly detected ("hit")	8	9	8	12	9	-
Signal present and not detected ("miss")	8	7	8	4	7	-
Signal absent and correctly stated as absent ("correct rejection")	59	59	59	59	59	-
Signal absent and incorrectly stated as detected ("false alarm")	21	21	21	21	21	-
Hit rate	0.50	0.56	0.50	0.75	0.56	-
False alarm	0.26	0.26	0.26	0.26	0.26	-
d'	0.64	0.79	0.64	1.31	0.79	-

Table 3.10: Table showing d' data for RB201 detection in babble task

- **Signal level:** This is the signal level in relation to threshold from the LocalB_{5AT} task, -9 is the most challenging and 3 is the easiest.
- **Number correct:** This is how many trials the participant got right.
- **Signal present and correctly detected ("hit"):** These numbers are taken from the above row.
- **Signal present and not detected ("miss"):** These numbers are calculated by subtracting the corresponding Signal present and correctly detected cell from 16 (which was the total number of presentation of each signal level).
- **Signal absent and correctly stated as absent ("correct rejection"):** These numbers are taken from The Number correct/No signal cell.
- **Signal absent and incorrectly stated as detected ("false alarm")**: These numbers are calculated by subtracting the number in the Signal absent and correctly stated as absent column from 80 (the total number of word absent trials).

- **Hit rate:** Signal present and correctly detected divided by 16 (number of trials per signal level).
- **False alarm:** Signal absent and incorrectly stated as detected divided by 80 (number of word absent trials).
- **d':** The inverse of the cumulative normal distribution function of Hit rate minus the inverse of the cumulative normal distribution function of false alarm. In this example it was between 0.64 and 1.31.

This process was then carried out on the $DetB_{2CS}$ data from all participants. The mean values for all participants are shown in Table 3.11. From this it can be seen that d' generally increases with increasing signal level.

In order to meaningfully compare these results to those from the $IdentB_{4CS}$ and $LocalB_{5CS}$ results, they must also be converted to d'. This was done by converting the percent correct for each signal level to a d' score using a conversion table (Table A5.7, MacMillan and Creelman, 2008). A typical contingency table normally used for d' analysis (showing hit rate, false alarm etc) cannot be used for these tasks as they do not have conditions where the signal is absent. This is shown for both $IdentB_{4CS}$ and $LocalB_{5CS}$, using the average data from all participants, in Table 3.12 and Table 3.13 respectively.

Signal level	-9	-6	-3	0	3
% correct	76	75	77	76	78
d’	1.42	1.26	1.58	1.43	1.79

Table 3.11: Table showing % correct to d' conversion for detection of speech in babble ($DetB_{2CS}$)

Signal level	-9	-6	-3	0	3
% correct	62	71	84	88	94
d’	1.22	1.53	2.09	2.32	2.8

Table 3.12: Table showing % correct to d' conversion for identification of speech in babble ($IdentB_{4CS}$)

Signal level	-9	-6	-3	0	3
% correct	46	55	66	76	79
d’	0.9	1.17	1.52	1.89	2.01

Table 3.13: Table showing % correct to d' conversion for localisation of speech in babble ($LocalB_{5CS}$)

Figure 3.8 shows the mean d’ values of $DetB_{2CS}$, $IdentB_{4CS}$, $LocalB_{5CS}$ from all participants. Like in the raw data (% correct) performance was generally highest in $IdentB_{4CS}$, however unlike in the raw data (where $DetB_{2CS}$ was generally better than $LocalB_{5CS}$) after d’ analysis $DetB_{2CS}$ and $LocalB_{5CS}$ are fairly equal.

As a d’ value of 1 means that performance is at threshold, it can be seen that for all tasks, at each difficulty level, participants are performing at threshold or above. At the easier signal levels a d’ of 2 is reached for $LocalB_{5CS}$ and $IdentB_{4CS}$ (and very nearly for $DetB_{2CS}$), meaning that participants found the task easy. However, none of the tasks reach a d’ of 4, which would indicate that the participants found the task very easy.

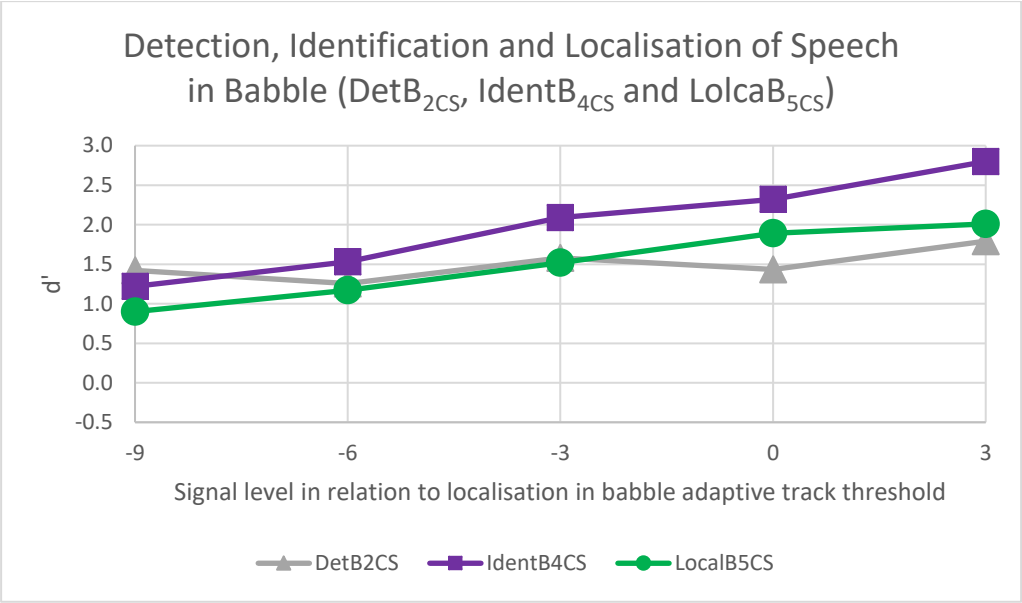


Figure 3.8: Graph showing d' of detection, identification and localisation of speech in babble (for all participants)

Comparison of results from the LocalB$_{5CS}$ and LocalB$_{5AT}$ tasks

The LocalB$_{5AT}$ task estimated the mean SNR at threshold at which a participant ideally achieved 70.7% correct. Therefore it would be expected that at the 0 level in LocalB$_{5CS}$ task participants should score approximately 70.7% correct. The mean score for all participants at this level was actually 75.7%.

A one-sample t-test was carried out to investigate whether the mean localisation scores from the constant stimuli task were different to the expected score (70.7%). There were no outliers in the data, as assessed by inspection of a boxplot, shown in Figure 3.9. Performance scores were normally distributed, as assessed by Shapiro-Wilk's test ($p > 0.05$). The mean localisation score (75.7 ± 13.1 s.d.) was significantly higher by 5.0 (95% CI, 0.82 to 9.18) than the expected population localisation score of 70.7, $t(39) = 2.418$, $p = 0.020$, $d = 0.38$.

Although this is a fairly small effect size, it shows that participants performed better on the LocalB$_{5CS}$ task than on the LocalB$_{5AT}$ task. This is a surprising result due to the similarity of the tasks. It may be caused by practice effects, as participants always completed the LocalB$_{5CS}$ task after the LocalB$_{5AT}$ task. It may be caused by implementation differences between the two tasks, as the LocalB$_{5AT}$ task used a staircase method, whereas the LocalB$_{5CS}$ task presented the stimuli (at the five different levels) in a random order. However, it may be due to the fact that sometimes the adaptive track in the LocalB$_{5CS}$ task may not have reached the asymptote (so performance appears worse than participants are actually capable of). With this knowledge and the possibility of learning effects being present it is most likely that these differences are not of importance and it can be concluded that performance is generally consistent between the two tasks.

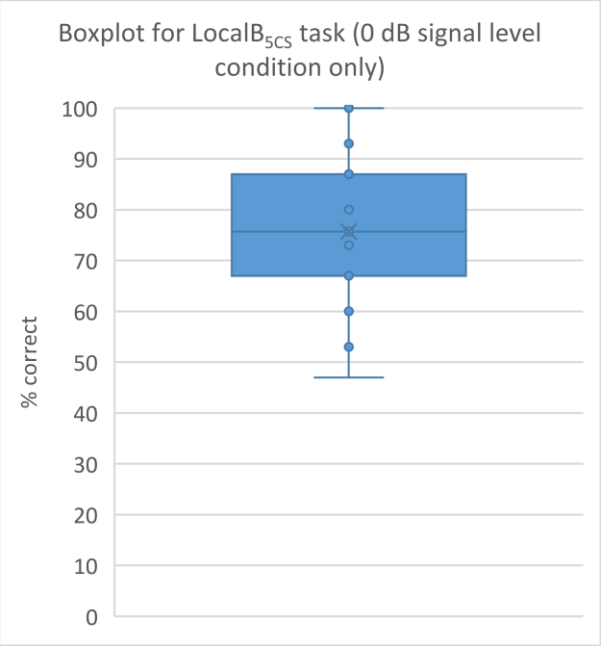


Figure 3.9: Boxplot for LocalB$_{5CS}$ data (0 dB signal level condition only)

3.4.2.2 Bootstrap analysis of d' data

A bootstrap analysis was carried out on the d' scores of all participants in order to fit curves to the data, using the equation shown in Equation 3.3. Bootstrap analysis estimates properties of the sampling distribution from the sample data (Efron and Tibshirani, 1993; Field, 2013). This is done by taking many repeated samples from the data and replacing each one before taking the next, so a particular sample could appear multiple times. The property of interest is then calculated for each bootstrap sample. This is then repeated many times (in this case, 1000 times), giving a distribution of m and k from which confidence intervals can be calculated. This was carried out in MATLAB, with code written by Michael Akeroyd as part of the laboratory statistical software.

$$d' = m * power^{k}$$

Equation 3.3: Equation for bootstrap curve fit (McFadden, 1966)

The m values indicates the offset from the point (0,0) along the x axis and the k value indicates how steep the curve is. The median m and k values and their 95% confidence intervals are shown in Table 3.14. The curve fits for $DetB_{2CS}$, $Ident_{4CS}$ and $LocalB_{5CS}$ are shown in Figure 3.10, Figure 3.11 and Figure 3.12 respectively. On each of the graphs the mean participant data is plotted, along with a function directly fitted to the mean participant data (here called the 'direct fit') and the bootstrap curve fit. Looking at these graphs it can be seen that the direct fit is generally very similar to the bootstrap curve fit, as would be expected as the bootstrap data is taken from the actual data. This is also

shown in the R^2 values for the direct fit and bootstrap fits, where for each task the R^2 value is the same or very similar for the direct and bootstrap fits. These values are: for $DetB_{2CS}$ R^2=55%, for $Ident_{4CS}$ R^2=98/97% and for $LocalB_{5CS}$ R^2=96/95% (given as direct fit/bootstrap where these results differ). This shows that the curve fits are better for the $Ident_{4CS}$ $LocalB_{5CS}$ tasks than for the $DetB_{2CS}$ task. The curve fits for all three tasks are shown on the same graph in Figure 3.13. From these graphs it can been seen that the curve fits are fairly shallow and are very similar to the raw d' data. Boxplots with error bars showing 95% confidence intervals for the m and k values are shown in Figure 3.14 and Figure 3.15 respectively. From these it can be seen that all three tasks gave different m values for all tasks and none of the 95% confidence intervals overlap. The k values are very similar and the 95% confidence intervals overlap for $IdentB_{4CS}$ and $LocalB_{5CS}$, whereas the k value for $DetB_{2CS}$ is lower and its confidence intervals do not overlap with those of the other tasks. This is not surprising, given that visually the $IdentB_{4CS}$ and $LocalB_{5CS}$ slopes appear to have similar steepness and the $DetB_{2CS}$ slope is much flatter.

Task	m value	Upper 95% confidence interval	Lower 95% confidence interval	k value	Upper 95% confidence interval	Lower 95% confidence interval
DetB$_{2CS}$	**1.59**	1.48	1.69	**0.09**	0.03	0.16
IdentB$_{4CS}$	**2.52**	2.39	2.65	**0.29**	0.24	0.35
LocalB$_{5CS}$	**1.81**	1.70	1.92	**0.28**	0.22	0.34

Table 3.14: Table showing m and k values and their 95% confidence intervals for all tasks

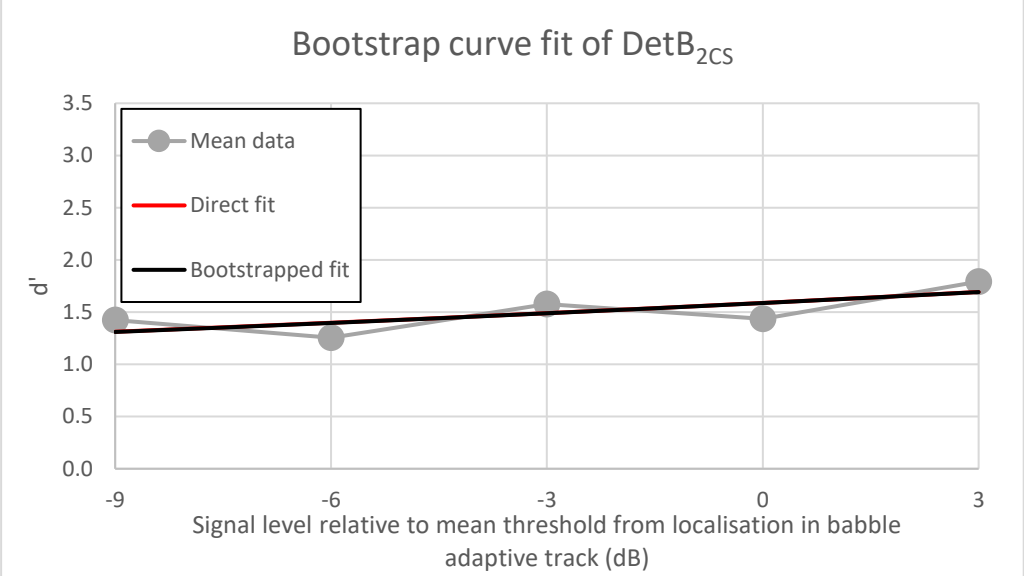


Figure 3.10: Figure showing curve fit from bootstrap analysis for the DetB$_{2CS}$ data (all participants)

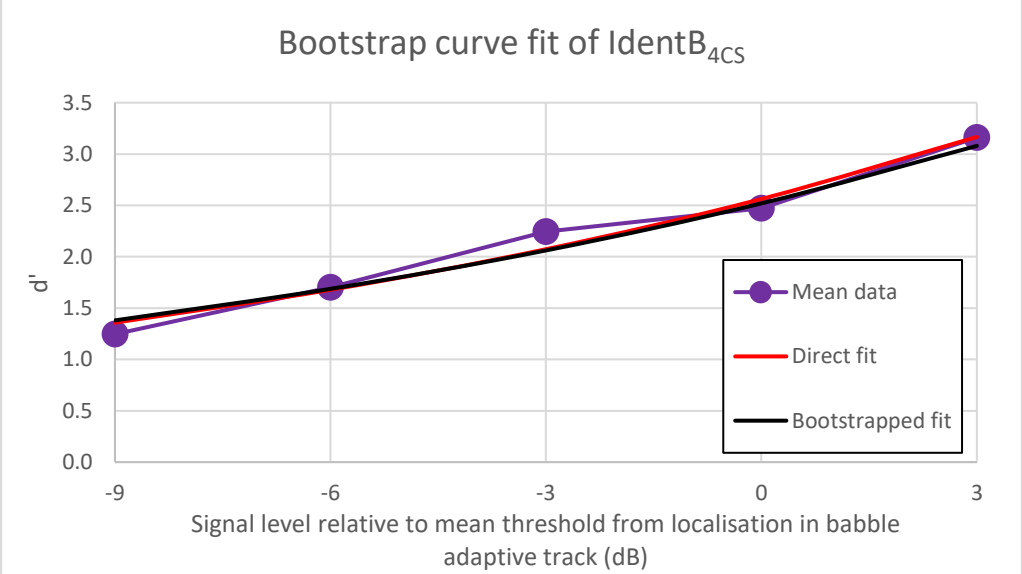


Figure 3.11: Figure showing curve fit from bootstrap analysis for the IdentB$_{4CS}$ data (all participants)

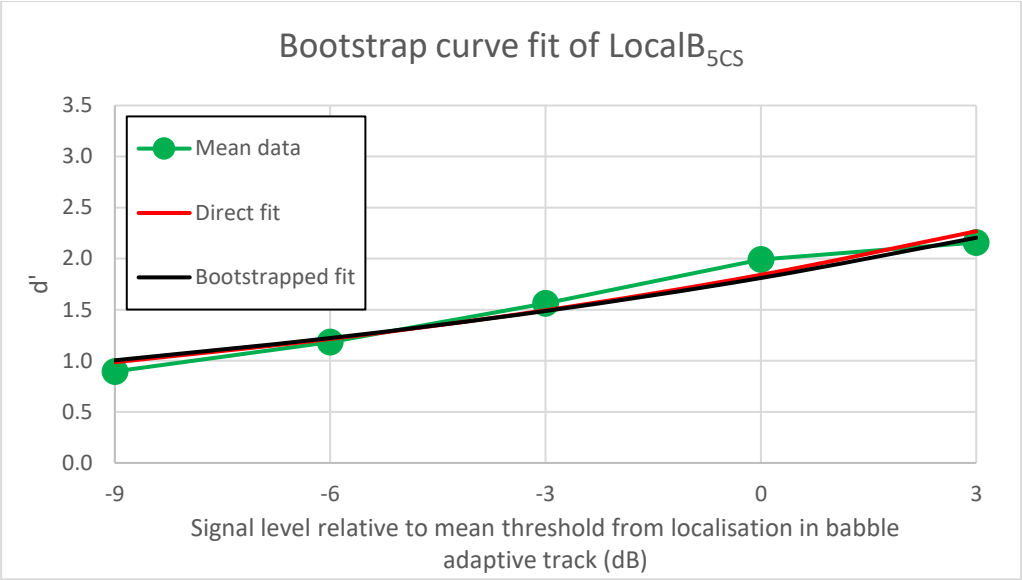


Figure 3.12: Figure showing curve fit from bootstrap analysis for the LocalB$_{5CS}$ data (all participants)

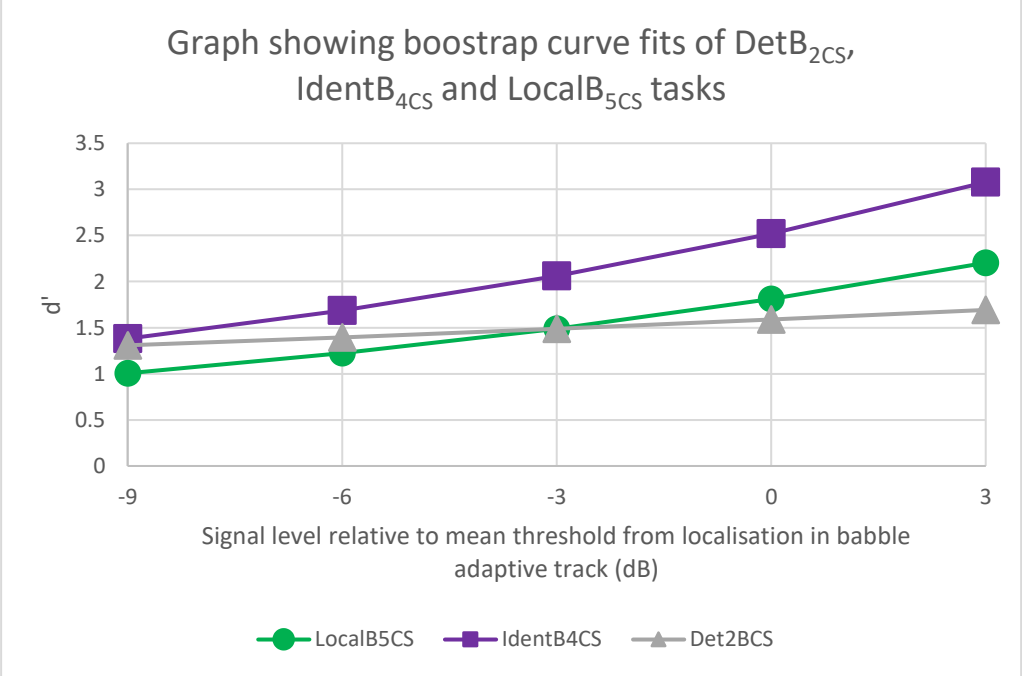


Figure 3.13: Graph showing bootstrap curve fits for all tasks (all participants)

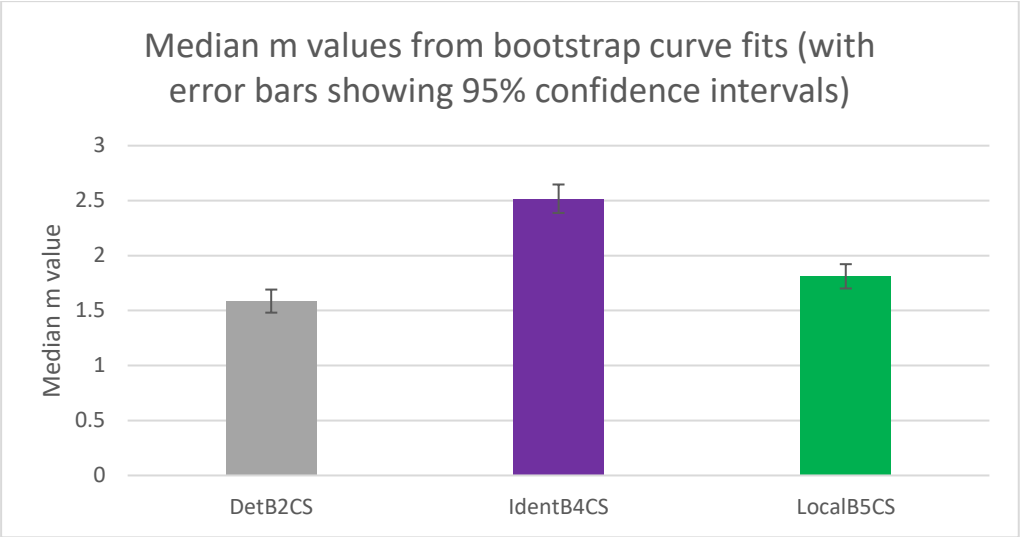


Figure 3.14: Figure showing median m values from the bootstrap curve fits for all tasks

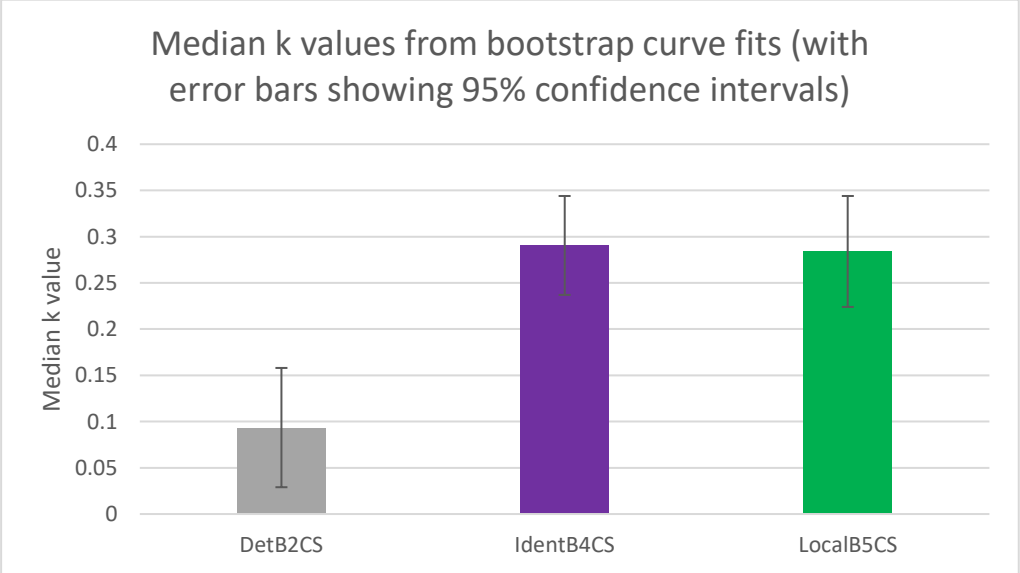


Figure 3.15: Figure showing median k values from the bootstrap curve fits for all tasks

3.4.3 Case study: A child with localisation difficulties

In this experiment one 11 year old female child (RB241) failed to complete the $LocalB_{5AT}$ task. This child spoke English as their first language, had normal hearing and no diagnosed learning or language disorders. Her parent reported that she was doing well at school and had not noticed any difficulties with her hearing. This participant had an identical twin who also took part in the experiment and performed as would be expected for her age.

The data from this participant is presented in Table 3.15. Normally the signal levels used for the constant stimuli tasks were chosen based on the threshold from the $LocalB_{5AT}$ task. As this participant did not complete the $LocalB_{5AT}$ task she did not get a threshold, so the signal levels for the $DetB_{2CS}$, $IdentB_{4CS}$ and $LocalB_{5CS}$ tasks were chosen based on mean SNR at threshold from other 11 year olds who had taken part in the experiment so far, which was -9.0 dB. This turned out to be fairly close to the expected mean SNR at threshold for 11 year olds at the end of the experiment when all participants had been tested, which was -7.4 dB.

Task	Scale	Result	Average for age	Interpretation
Non-verbal IQ (scaled)	53-157	97 (42% percentile)	100	Slightly below average
Digit span total (scaled score)	1-19	5	10	Fairly poor
Dichotic digits right (% correct)	0-100%	50%	98%	Very poor
Dichotic digits left (% correct)	0-100%	100%	98%	Good
Localisation in quiet (% correct)	0-100%	50%	87%	Very poor
Localisation in babble adaptive track: Mean SNR at threshold (dB)	Maximum/worst possible score of 7 dB (best score on this task was -19 dB)	No data	-7.4 dB	Poor
Constant stimuli: localisation in babble (% correct)	0-100%	26%	66.4%	Poor
Constant stimuli: identification in babble (% correct)	0-100%	80%	85%	Fairly good
Constant stimuli: detection in babble (d')	0 (0 or lower indicates guessing)-4.5	0.03	1.35	Participant was guessing

Table 3.15: Table for child with localisation difficulties

From Table 3.15, it can be seen that this participant scored reasonably well in the non-verbal IQ, dichotic digits left and $IdentB_{4CS}$ tasks. However she did poorly in the other tasks: namely digit span, dichotic digits right, $LocalQ_{5CS}$, $LocalB_{5CS}$, $LocalB_{5AT}$ and $DetB_{2CS}$. Although this participant presented with no listening difficulties it is clear from these results that she has difficulties with spatial hearing. It is interesting that she had difficulty with detection in babble (d' scores suggest she guessed most of the time) but not with identification in babble. These results fit in with the general trend of participants scoring higher in identification than localisation and detection in babble. As the participant scored well on identification it is clear that she is capable of detecting the word. This suggests that a deficit in either of these abilities was

not responsible for the poor localisation ability of this participant. The participant scored poorly on the localisation in quiet task which suggests that her difficulties with localisation in babble could be based on localisation deficits rather than speech in background sound problems. She scored poorly on the dichotic digits right task, despite scoring well on left (where in the right task she reported the number heard in her right ear and ignores the number heard in her left ear), demonstrating an unreliable ability to spatially separate stimuli.

The ECLiPS questionnaire indicated that the child had listening difficulties in multiple areas (especially in pragmatic & social skills), but not at a clinically significant level. The Conners questionnaire indicated some oppositional behaviour, but no attention issues. Nothing of note was found in the family history questionnaire.

3.5 Discussion

3.5.1 Adaptive track results

In the LocalB$_{5AT}$ task a gradual improvement with age was found, although some teenagers achieved better scores than the young adults (this may be due to experiment error or chance). This age-related improvement in performance is very similar to the results from Experiment 1. Table 3.16 shows the age groups and equations for predicted mean SNR at threshold (dB) for both experiments. From this table it can been seen that the equations are very similar, with the same asymptotes and differences of only 0.6 years for

the time constant and 3.1 dB/year for the multiplier of the exponential. These differences suggest that for the participants tested in Experiment 2 the rate of change was slightly faster, which would be expected due to older children taking part, and that the expected adult-like performance level was slightly worse than in Experiment 1. It must be noted that the R^2 from the exponential function for Experiment 1 (74%) was much higher than that for Experiment 2 (24%), indicating a better fit to the data in Experiment 1.

Experiment	Age group (years)	Equation for predicted mean SNR at threshold (dB)
Experiment 1	6-13, 18-29	$Predicted\ mean\ SNR\ at\ threshold\ (dB) = -15.3 + (44.9 * e^{\left(-\frac{age}{6.6}\right)})$
Experiment 2	10-17, 23-29	$Predicted\ mean\ SNR\ at\ threshold\ (dB) = -15.3 + (48.0 * e^{\left(-\frac{age}{6.0}\right)})$

Table 3.16: Table showing age groups and equations for predicted mean SNR at threshold (dB) for experiments 1 and 2

Earlier, it was predicted that in Experiment 2 older children would perform at the levels predicted by the exponential function fitted to the data from Experiment 1. Figure 3.16 shows the exponential functions fitted to the data from Experiment 1 and Experiment 2. From this it can be seen that the results are very similar and eventually asymptote together. The mean values for each age (from Experiment 2 data) plotted on this graph shows that although some of the predictions based on the data from Experiment 1 were fairly accurate, the predictions for the older teenagers (16 and 17 years old) were less so. However, there was only one 17 year old participant in Experiment 2 so if they did not have average localisation of speech in babble abilities for their age (which from looking at their results seems possible) then it may be that the

predictions from Experiment 1 are fairly accurate. This will be returned to in Chapter 4, where data from Experiments 1 and 2 will be combined for further analysis.

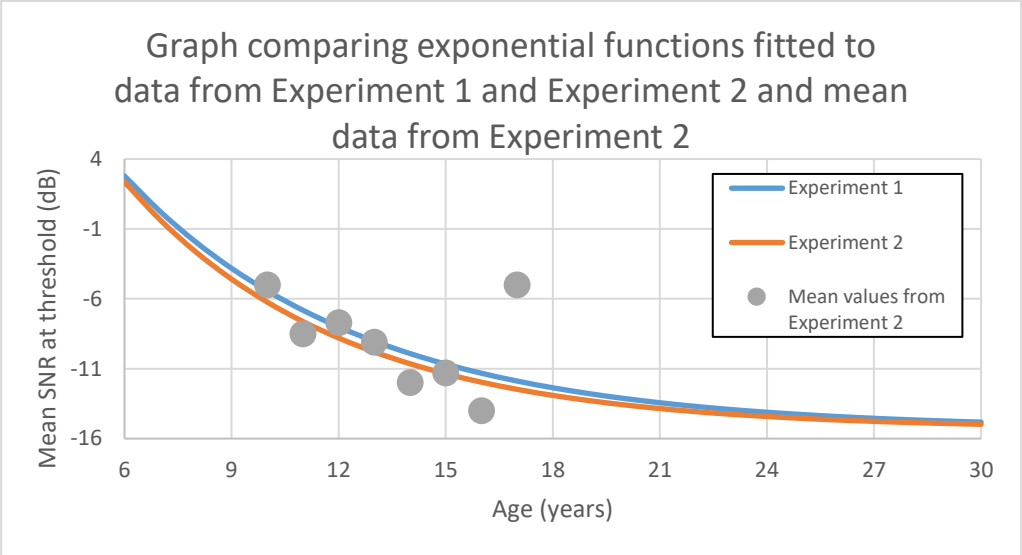


Figure 3.16: Graph showing exponential function fitted to data from Experiment 1 and Experiment 2 and mean data from Experiment 2

3.5.2 Detection, identification, localisation of speech in babble ($DetB_{2CS}$, $IdentB_{4CS}$, $LocalB_{5CS}$)

In the $DetB_{2CS}$, $IdentB_{4CS}$, $LocalB_{5CS}$ tasks, performance was similar for adults and children due to the method of signal level choice, which compensated for age-related differences. Generally, performance was highest for $IdentB_{4CS}$ and this improved as the signal level increased. Performance was overall better for $DetB_{2CS}$ than $LocalB_{5CS}$. $LocalB_{5CS}$ improved slightly more with signal level than $DetB_{2CS}$, and $DetB_{2CS}$ was fairly stable with change in signal level. They were similar at the two easiest signal levels (0 and 3 dB). Following the d' analysis the results still showed that participants scored on average better in the identification task than the detection and localisation tasks. The better

performance on identification than localisation is a novel finding and may suggest that identification is necessary for localisation. As performance was very similar on the $DetB_{2CS}$ and $LocalB_{5CS}$ tasks following d' analysis (and both were worse than performance on the $IdentB_{4CS}$ task) it is not possible to conclude whether localisation is dependent on detection or not.

Previous research has shown that participants perform better at identification than detection of masked speech (Hawkins and Stevens, 1950; Liu and Jin, 2019). The results from the experiment described in this chapter are in agreement with these previous findings. These results seem surprising as it might be thought that if a participant can identify a sound then they should also be able to detect it. This result may be because identification is the task most similar to the listening requirements of daily life. In this experiment it may also be due to the $DetB_{2CS}$ task taking twice as long as the others, so participants may have lost interest and therefore performed poorly. Earlier it was hypothesised that children who have difficulties with speech detection and/or identification in babble will perform poorly in the $LocalB_{5CS}$ task. One participant was very poor at the $LocalB_{5CS}$ task; her results on the other tasks are discussed above. The most important findings were that this participant performed fairly well on the $IdentB_{4CS}$ task and poorly on the $DetB_{2CS}$ task and the $LocalQ_{5CS}$ task. For this participant it seems that poor localisation in babble performance was due to a deficit in localisation abilities, rather than speech in babble problems. However, as this is just one participant this result cannot be generalised to all children with localisation in babble difficulties.

3.5.3 Limitations and future work

One of the limitations when comparing the results of the $DetB_{2CS}$, $IdentB_{4CS}$ and $LocalB_{5CS}$ tasks was the difference in task design. The $IdentB_{4CS}$ task was a one interval-four choice task, whereas the $LocalB_{5CS}$ task was a one interval-five choice task. Although d' analysis helps allow for this, it is difficult to meaningfully compare the results. Therefore in Experiment 3 (Chapter 5), a set of tasks looking at identification and localisation of speech in babble will be carried out where both identification and localisation of speech in babble will be one interval-four choice tasks. The $DetB_{2CS}$ task was a one interval-yes/no task, and although it would be possible to turn it into a four choice task (but with more than one interval) it would be quite different to the other tasks and may be difficult for younger participants. It would not be possible to turn the $LocalB_{5CS}$ or $IdentB_{4CS}$ tasks into one interval-two choice tasks and still get meaningful results as the localisation task would become a left/right discrimination task and the identification task would probably be too easy.

A second limitation was that the $DetB_{2CS}$ task was twice as long as both the $IdentB_{4CS}$ task and the $LocalB_{5CS}$ task. This may have caused the children to get bored and therefore perform worse on this task. The reason the $DetB_{2CS}$ task was so long was so that it had the same number of trials for each signal level as in the other tasks, and to have the same number of trials with signal absent as present. For future experiments with children it would be recommended to split this long task into two shorter ones so children are able to take a break. For adults this length of task should not be so problematic.

Future work in Experiment 3 (Chapter 5) will also expand on the constant stimuli tasks from this experiment by investigating performance at more challenging SNRs. One of the limitations of the results from this task was caused by the signal levels varying for each participant. Although this was a good way to ensure that the tasks were not impossible for any participant to complete, it made it difficult to compare performance across participants. Therefore, in Experiment 3 all participants will do the tasks at the same signal levels. The results from this experiment will be used to guide which signal levels will be used for these future tasks.

In summary, this experiment added to the results of Experiment 1 by testing the localisation of speech in babble abilities of 10-29 year olds using an adaptive track task. The results from these tasks showed improvement of localisation of speech in babble abilities with age. A one-way ANOVA showed that 10 to 12 year olds were significantly worse than adults at this task, but 13 to 14 year olds and 15 to 17 year olds were not significantly different to adults. An exponential function was fitted to the data from this experiment and had very similar parameters to the exponential function fitted to the data from Experiment 1. The relationships between detection, identification and localisation of speech in babble were investigated. It was found that participants scored highest in identification, followed by detection and localisation (which were similar according to d' analysis).

Chapter 4

Analysis of the combined results from Experiments 1 and 2

4.1 Introduction

Both Experiment 1 and Experiment 2 included the LocalB$_{5AT}$ task and additional cognitive tasks. In this chapter the data from all of these tasks are combined and analysed. This allows us to map the developmental time course of localisation of speech in babble from 6 to 29 years of age. Combining the data also enables the achievement of the aims and investigation of the hypotheses originally stated in Chapter 2, related to non-auditory factors which may be linked to the development of localisation of speech in babble in children.

4.1.1 How is adult-like performance defined?

When discussing adult-like performance, the age at which a person is an adult must be considered. Although the legal definition of an adult is 18 years of age, some aspects of brain development are not complete for some individuals until the mid to late twenties (Spear, 2000). Research suggests that the transition from adolescence to adulthood is a gradual process and occurs at a range of ages. The majority of auditory studies use 18 years of age (or

slightly older if no 18 year olds participate in the experiment) as the point of adulthood. Therefore, in order to keep with convention the experiments described in this thesis do the same.

To consider how adult-like performance is defined the methods of 17 studies looking at adult-like performance levels on auditory tasks were examined (Litovsky, 1997; Eisenberg et al., 2000; Johnson, 2000; Stollman et al., 2004; Litovsky, 2005; Stuart, 2005; Garadat and Litovsky, 2007; Vaillancourt et al., 2008; Cameron et al., 2009; Van Deun et al., 2009; Moore et al., 2011; Lovett et al., 2012; Kuhnle et al., 2013; Yuen and Yuan, 2014; Martin, Johnstone and Hedrick, 2015; Boothalingam et al., 2016; Jones, 2018). Fifteen of the studies used t-tests (eight studies) and/or ANOVAs (ten studies) of some variety to compare groups. The following methods were also used:

- Pearson's linear correlation against age (two studies)
- Child mean values in previously reported adult range (one study)
- Similar d' values (one study)
- Broken stick function (one study)
- 95% population limits of adults (one study)

Therefore, from the literature it can be seen that most studies use t-tests or ANOVAs to define adult-like performance. Although this can be a good way of comparing groups of younger children to adults, it is not necessarily suitable when comparing older teenagers to adults, as if there is no significant difference between groups then this does not mean that they are actually the same (Altman and Bland, 1995). This means that it can suggest age groups

where adult-like performance may be occurring but it cannot confirm this. In this chapter a variety of methods are used in order to inform conclusions about when adult-like performance is reached. They are t-tests, ANOVAs, broken stick functions and exponential curve fits (directly to the data and following a bootstrap analysis). These methods are briefly summarised in Figure 4.1.

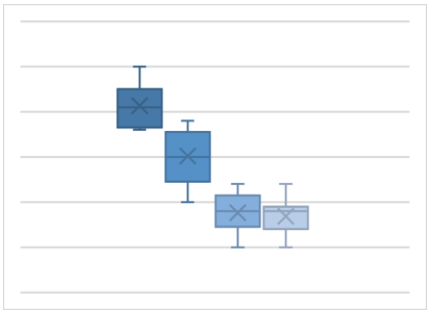
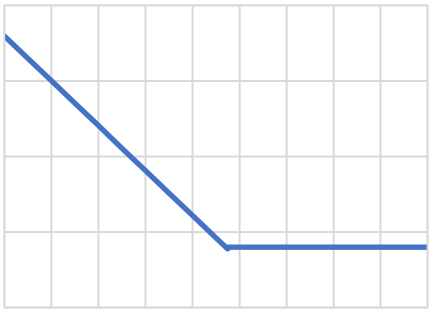
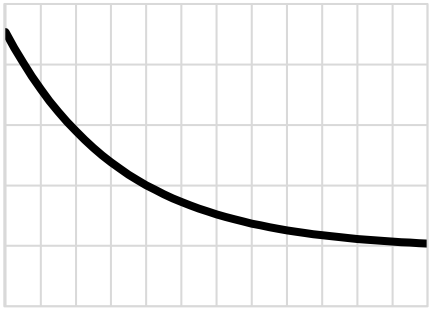
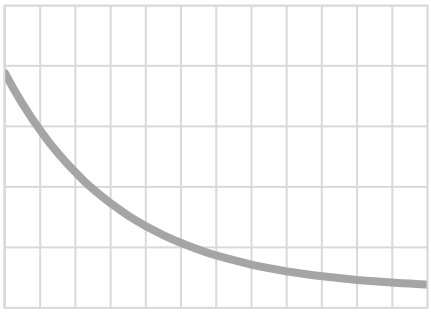
	Diagram	Test for adult-like performance	Confidence intervals
ANOVA/ t-test		Non-significant difference between children and adults	✗
Broken stick		Change point	✓
Exponential function (direct fit)		Difference between children and adults (range of 1-2 RMS errors)	✗
Exponential function (bootstrap)		Difference between children and adults (2 RMS errors)	✓

Figure 4.1 Figure showing different analysis methods

4.2 Results

4.2.1 Initial representation of results from LocalB$_{5AT}$ tasks from Experiments 1 and 2

From this point onwards the data from Experiments 1 and 2 will be combined for analysis.

Two participants took part in both Experiments 1 and 2. One of them was an adult; the other was 12 years and 10 months of age when they participated in Experiment 1 and 13 years and 4 months of age when they participated in Experiment 2. It was decided to use only the mean threshold data from Experiment 1 from both participants in the main analysis, in case any learning effects were present in the data from Experiment 2.

This means that the data from 14 adults (six female, eight male) aged 18-29 years old (mean 24.5, s.d. 3.2) and 54 children (23 female, 31 male) aged 6 to 17 years (mean 11.6, s.d. 2.7) can be used to create a developmental time course for localisation of speech in babble. For the LocalB$_{5AT}$ task all adults were able to give six thresholds. The number of thresholds returned by the children is shown in Table 4.1.

Number of thresholds returned	Frequency
1	2
2	1
3	1
4	4
5	6
6	40

Table 4.1: Table showing number of thresholds returned in children (Experiments 1 and 2)

Figure 4.2 shows boxplots for each year of age and young adults performance on the LocalB$_{5AT}$ task. This graph demonstrates gradual improvement with age up to and including the teenage years. Full boxes are not plotted for age 6 and 17 years as there was only one participant at each of these ages.

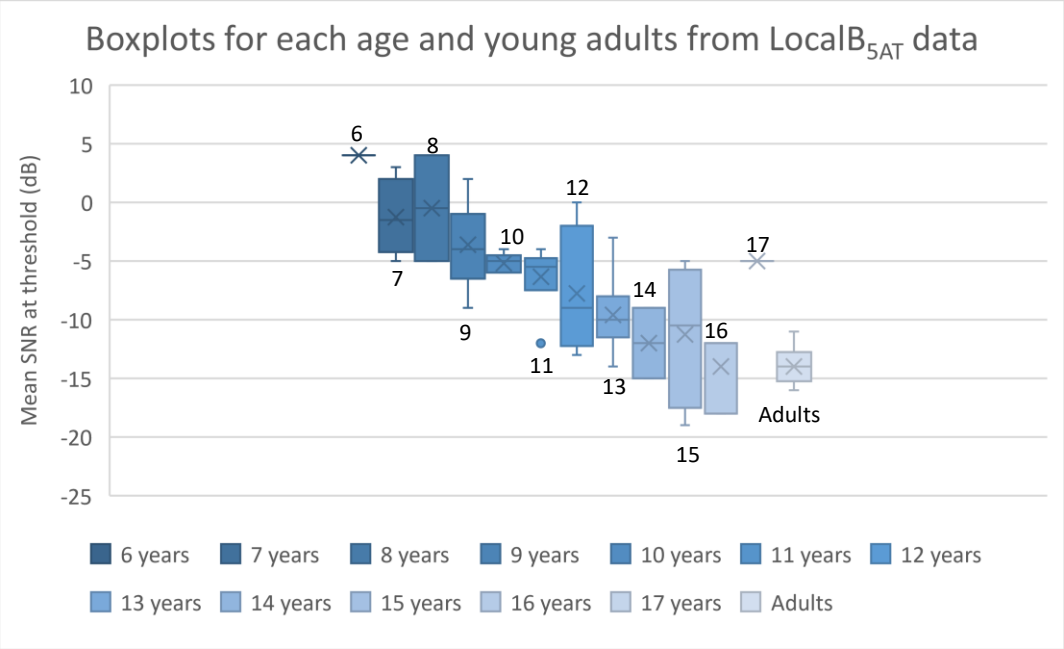


Figure 4.2: Boxplot for each age and young adults LocalB$_{5AT}$ data

4.2.2 Adult-like performance: Exponential function (direct fit)

Figure 4.3 shows the average SNR at threshold for each participant, plotted against their age. It shows a gradual improvement in performance with age and less variability in adults than in children. From this data, an exponential curve was calculated using the Solver function in Excel, shown in Equation 4.1, which allows interpolation to predict performance at different ages. Its parameters were: -15.3 dB is the asymptote, 44.9 dB/year is the slope and 6.6 years is the time constant. This is the same as the time constant found in Experiment 2, and slightly higher than the time constant found in Experiment

1 (6.0 years). The R^2 value for this equation is 64% (which is lower than that for Experiment 1, but higher than for Experiment 2).

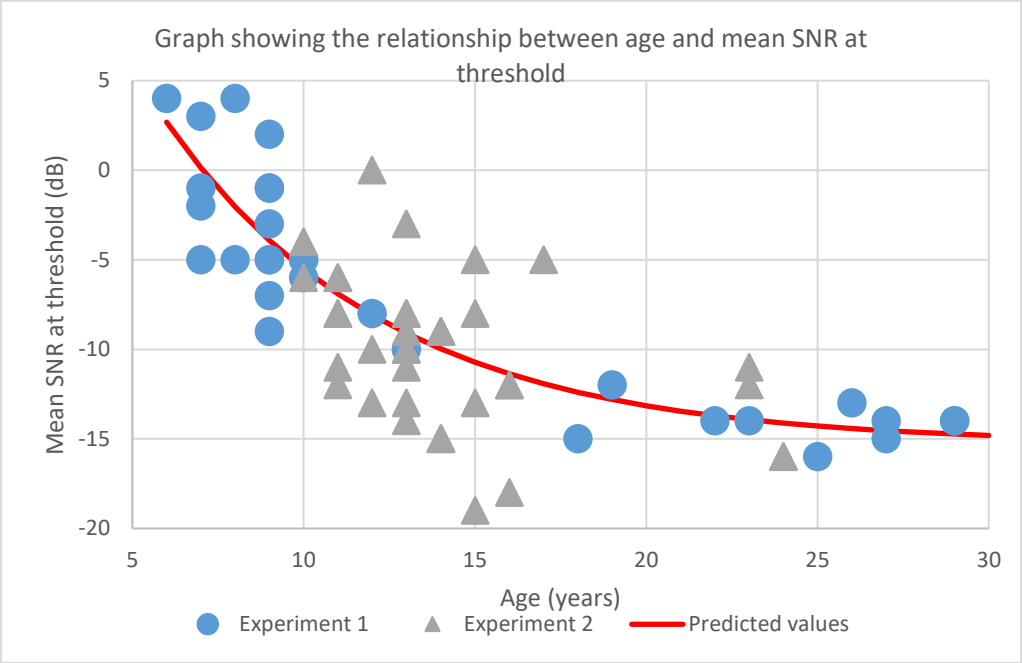


Figure 4.3: Graph showing relationship between performance and age (Experiments 1 and 2)

$$Predicted\ mean\ SNR\ at\ threshold\ (dB) = -15.3 + (44.9 * e^{\left(-\frac{age}{6.6}\right)})$$

Equation 4.1: Equation for predicted mean SNR at threshold for Experiments 1 and 2

One of the ways which the age range during which adult-like performance is reached is defined here using the exponential function fitted to the data and the RMS error of the data (3.3 dB). Figure 4.4 sketches how adult-like performance is defined, showing:

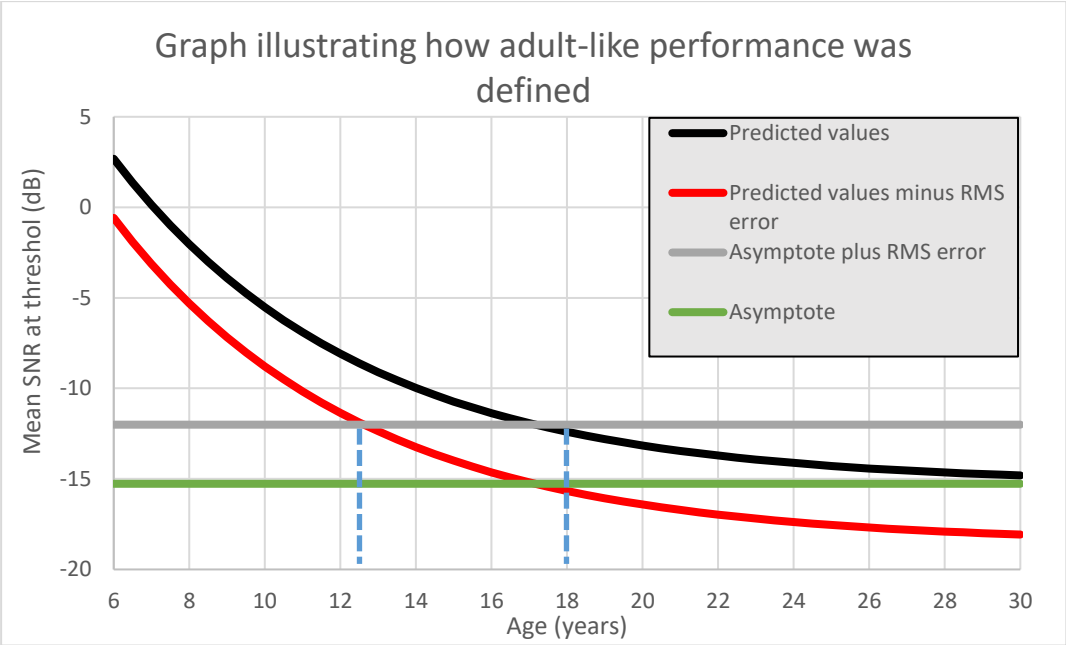


Figure 4.4: Graph illustrating how adult-like performance was defined for the exponential function (direct fit)

- Predicted data from the exponential function fitted to the data (shown in black).
- Predicted data from the exponential function minus RMS error, representing the performance expected from the highest performing participants (shown in red).
- The asymptote of the exponential function representing general expected adult performance (shown in green).
- The above asymptote plus RMS error representing the worst performance expected from adults (shown in grey).

The points at which these lines cross each other indicate the lower and upper age limits when adult-like performance is achieved. The lower limit (including 2 RMS errors) is where the red and grey lines cross (predicted values minus RMS error and asymptote plus RMS error), which indicate the point at which

the best-performing children perform as well as the worst-performing adults (assuming that the data is symmetrical). This occurs at 13 years of age (to the closest 0.5 years). The upper limit (including 1 RMS error) is represented by the intersection of the black and grey lines (predicted values and asymptote plus RMS error) which is the same as the intersection of the red and green lines (predicted values minus RMS error and asymptote). These intersections represent the age at which the average performing children perform as well as the worst performing adults, or the best performing children perform as well as the average adults. This occurs at 17.5 years of age (to the closest 0.5 years). Therefore, using the exponential function (direct fit) the age-range within which adult-like performance is achieved is 13.0-17.5 years of age.

4.2.3 Adult-like performance: Broken stick analysis

A broken stick analysis was also carried out. This is a two stage linear regression, where the slope after the change point is fixed to zero (using R code from Schwarz (2015)). The R analysis fitted a piecewise linear function to the data, found the change point and used bootstrapping to find the slope of the functions before the change point and the confidence intervals for the change point. This type of broken stick function was chosen over the type where there are different (non zero) slopes either side of the change point because adult-like performance should be stable in young adults.

The graph is plotted in Figure 4.5 and the data from the analysis is shown in Table 4.2. This analysis found a change point at 14.8 years of age, with 95% confidence intervals from 13.3 to 18.9 years (or 13.5 to 19.0 years to the

nearest 0.5 years). These values may be taken as the range for age at which adult-like performance is reached, and is very similar to the age range found earlier using the exponential function (direct fit) (13.0-17.5 years). The R^2 value for the broken stick function is 63%, which is very close to the values from the exponential functions (64%).

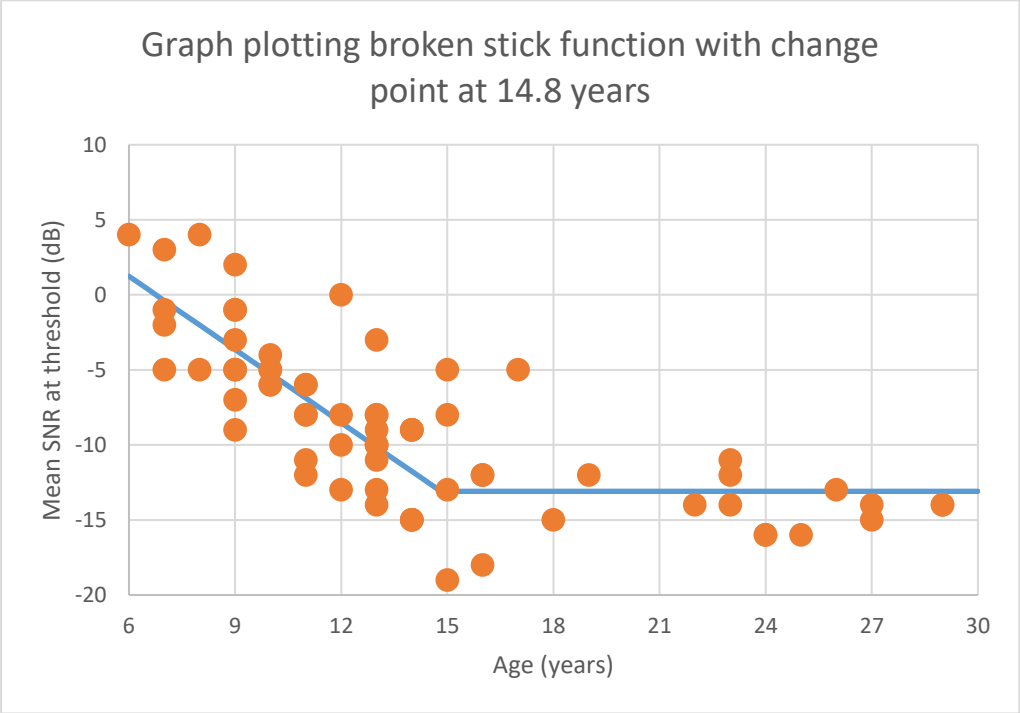


Figure 4.5: Graph showing broken stick function for localisation of speech in babble task

	Change point	Slope before change point
Value	14.8 years	-1.6 dB/year
Lower 95% confidence interval	13.3 years	-2.2 dB/year
Upper 95% confidence interval	18.9 years	-0.95 dB/year

Table 4.2: Table showing data from R broken stick analysis

4.2.4 Adult-like performance: Comparison of different age groups using an ANOVA

A one-way Welch ANOVA was conducted to determine if average SNR at threshold was different for groups of different age ranges, due to unequal variances. Participants were classified into six groups: age 6-8 (n =7), age 9-10 (n=13), age 11-12 (n=10), age 13-14 (n =17), age 15-17 (n=8) and adults (n=15).

No outliers were present, as assessed by the boxplot shown in Figure 4.6. The data was normally distributed for each group, shown by the Shapiro-Wilk test ($p > 0.05$). Homogeneity of variances was violated, as assessed by Levene's Test of Homogeneity of Variance (p =0.043), therefore Games-Howell post hoc tests were used (Games and Howell, 1976). It was found that the average SNR at threshold was significantly different for different age groups, Welch's $F(5, 23.309) = 32.556$, $p < 0.0005$. Average SNR at threshold decreased from the age 6-8 group (-0.29 ± 3.99), to the age 9-10 group (-4.23 ± 2.89), to the age 11-12 group (-8.20 ± 3.74), to the 13-14 group (-10.18 ± 3.41), to the 15-17 group (-11.50 ± 5.32) to the adult group (-14.20 ± 1.70). A Games-Howell post hoc analysis revealed that the difference in mean SNR at threshold was significant for several group pairs, shown in Table 4.3. Adults were significantly different to all groups except the 15-17 year old group.

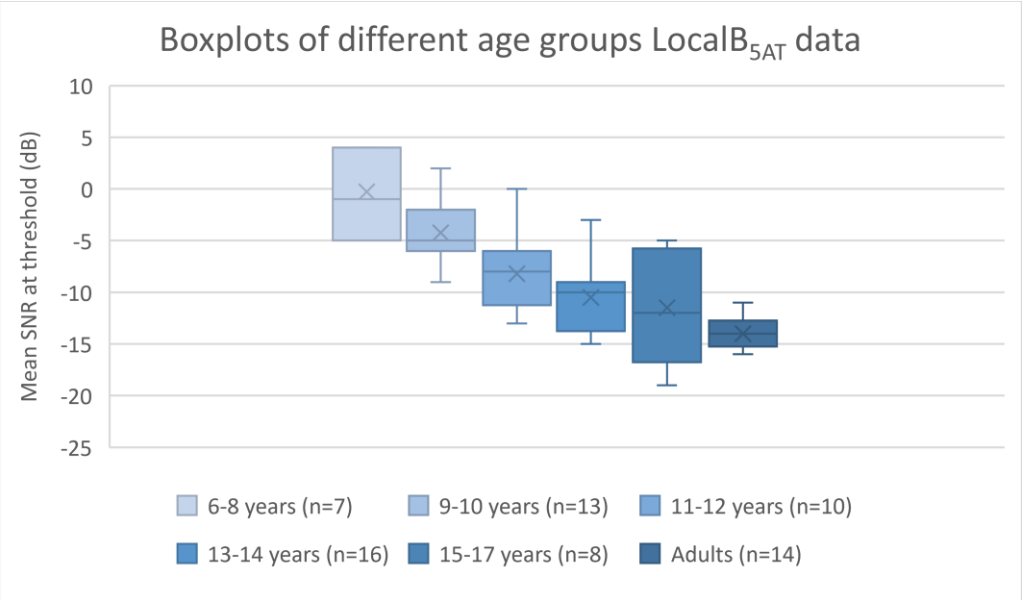


Figure 4.6: Boxplots of different age groups $LocalB_{5AT}$ data from Experiments 1 and 2

Age group	9-10 years old	11-12 years old	13-14 years old	15-17 years old	Adult
6-8 years old	Δ = 4.0 95% CI = -.02 to 9.9 *p* = 0.277	Δ = 7.9 95% CI = 1.5 to 14.3 ***p* = 0.013**	Δ = 9.9 95% CI = 3.9 to 5.9 ***p* = 0.002**	Δ = 11.2 95% CI = 3.2 to 19.2 ***p* = 0.005**	Δ = 13.9 95% CI = 8.0 to 19.9 ***p* < 0.0005**
9-10 years old	x	Δ = 4.0 95% CI = 0.6 to 8.6 *p* = 0.111	Δ = 6.0 95% CI = 2.4 to 9.5 ***p* <0.0005**	Δ = 7.3 95% CI = 0.2 to 14.4 ***p* = 0.046**	Δ = 10.0 95% CI = 7.1 to 12.9 ***p* < 0.0005**
11-12 years old	x	x	Δ = 2.0 95% CI = -2.6 to 6.6 *p* = 0.743	Δ = 3.3 95% CI = -4.1 to 10.7 *p* = 0.679	Δ = 6.0 95% CI = 1.7 to 10.3 ***p* = 0.005**
13-14 years old	x	x	x	Δ = 1.3 95% CI = -5.8 to 8.5 *p* = 0.984	Δ = 4.0 95% CI = 1.1 to 6.9 ***p* = 0.003**
15-17 years old	x	x	x	x	Δ = 2.7 95% CI = -4.4 to 9.8 *p* = 0.728

Table 4.3: Table summarising findings from the ANOVA (statistically significant findings in bold)

4.2.5 Adult-like performance: Exponential function (bootstrap)

Bootstrapping was carried out using MATLAB code written by Michael Akeroyd as part of the laboratory statistical software. Bootstrap analysis estimates properties of the sampling distribution from the sample data (Efron and Tibshirani, 1993; Field, 2013). A bootstrap analysis was carried out on the data in order to fit a curve to the data and also make an estimate of when adult-like performance occurred (when adults and children are separated by 2 RMS errors). Bootstrap analysis takes many small samples from the data (with replacement) and then calculates the parameter(s) of interest for each sample. This process is repeated many times, it was done 1000 times for this analysis.

The full dataset was input and the values used to plot an exponential function similar to that previously fitted to the data (slope, time constant and asymptote). This graph is plotted in Figure 4.7 along with the exponential function described above (fitted to the data using Solver in Excel). From this it can be seen that the functions are very similar to each other, but are most similar for the adults. This may be because the data was more variable for the youngest children tested.

Additional data from the bootstrap analysis is shown in Table 4.4. The median slope is 54.7 dB/year, so is higher than the value from the original exponential function (44.9 dB/year). The mean time constant is 5.6 years, so is lower than the value from the original exponential function (6.6 years). The mean asymptote is -14.9 years, so is very similar to the value from the original

exponential function (-15.3 years). The crossover point (where children and adults are separated by 2 RMS errors) is 11.9 years, with confidence intervals (95%) from 10.3 to 14.5 years. This suggests that the age range during which adult-like performance likely occurs is between approximately 10 and 14.5 years of age. The R^2 value for this curve fit was 64%, which is same as that of the direct fit exponential function.

	Median	Standard deviation	Lower confidence interval	Upper confidence interval
Slope (dB/year)	54.7	15.4	33.5	90.1
Time constant (years)	5.6	1.1	4.1	8.5
Asymptote (dB)	-14.9	0.8	-16.6	-13.6
Crossover (years)	11.9	1.0	10.3	14.5

Table 4.4: Table showing data from exponential function (bootstrap)

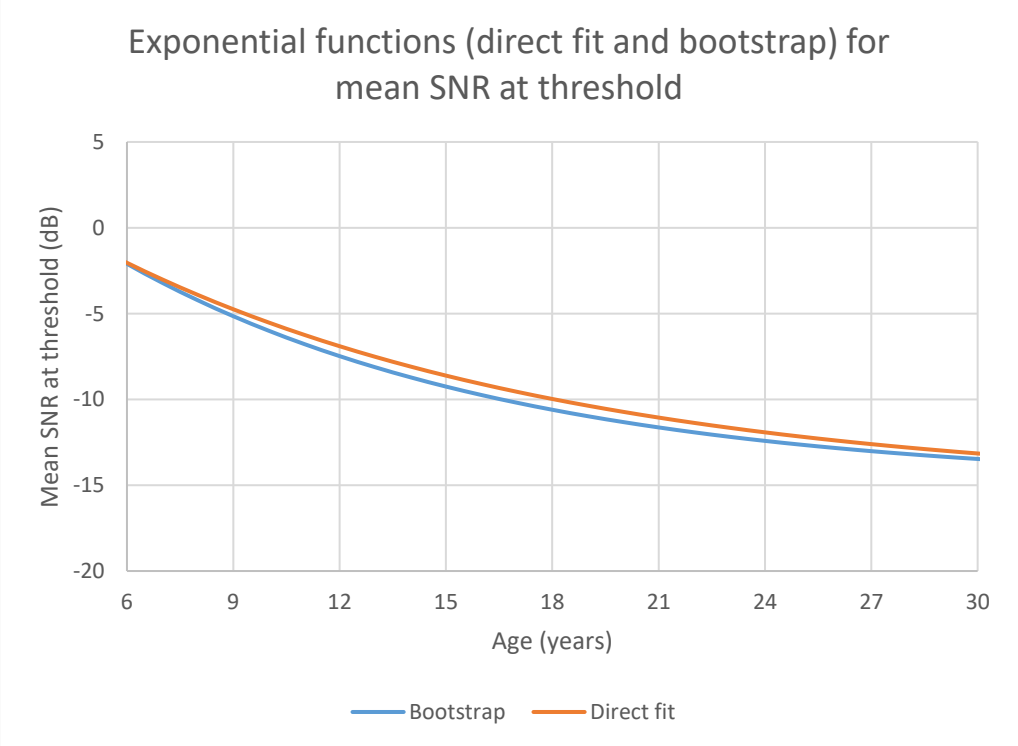


Figure 4.7: Exponential functions for mean SNR at threshold from direct fit and bootstrap

4.2.6 Summary of results from LocalB$_{5AT}$

It is clear that performance on the LocalB$_{5AT}$ task improves with age, but the point at which adult-like performance is reached does not appear to be clearly definable. Table 4.5 shows that different analysis methods gave varying estimates of when adult-like performance was achieved. Therefore it seems best to give a range at which it occurs rather than a point estimate.

Figure 4.8 shows the exponential function (direct fit), exponential function (bootstrap) and broken stick methods on the same graph (along with the participant data). From this graph, it can be seen that the exponential curve fits (direct fit and bootstrap) are very similar. The second (flat) part of the broken stick function is higher than the asymptotes for both curve fits. This means that the adult-like performance level estimated using this method (-13.1 dB) is higher (worse) than that estimated by the direct fit (-15.3 dB) and bootstrap fit (-14.9 dB). The change point of the broken stick function (14.8 years, with 95% confidence intervals of 13.3 to 18.9) is older than adult-like performance estimated by the exponential function (bootstrap) (11.9 years, 95% confidence intervals of 10-14.5 years). The exponential function (direct fit) gave an estimate of 13 to 17.5 years, which is very similar to that given by the broken stick function. Overall, it seems likely that adult-like performance is reached in the early to mid-teenage years.

Method	Findings
Exponential function (direct fit)	The age-range within which adult-like performance is achieved is **13.0-17.5 years** of age.
Broken stick function	The change point was at **14.8 years** of age, with 95% confidence intervals from **13.5 to 19.0 years**.
ANOVA	Adults significantly different to all groups except **15-17 year olds**.
Exponential function (bootstrap)	The crossover point was **12 years**, with 95% confidence intervals giving a range of adult-like performance from **10-14.5 years**.

Table 4.5: Table showing results from different analysis methods

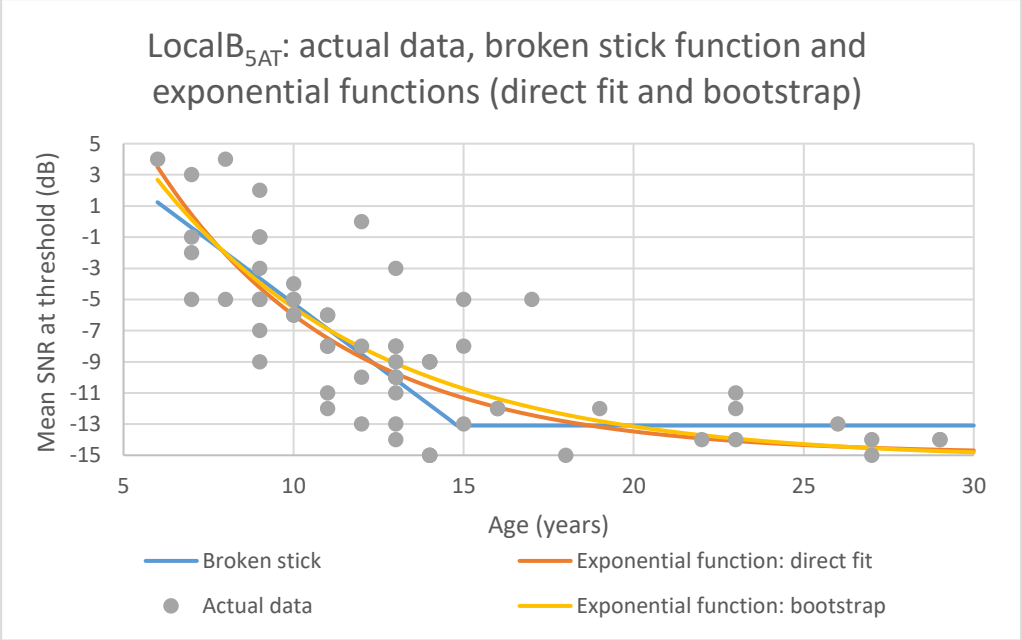


Figure 4.8: Graph comparing different methods for estimating adult-like performance

4.2.7 Results from additional tasks

4.2.7.1 Age-norming of LocalB$_{5AT}$ scores

In this section, results from the additional tasks are summarised and analysed alongside the localisation of speech in babble data from the LocalB$_{5AT}$ task. When tasks have age-normed results, the results which they are compared to also need to be age-normed. Therefore, the LocalB$_{5AT}$ data has been age-normed for this purpose. This was done by subtracting the actual mean SNR at threshold from the predicted mean SNR at threshold for each participant

(based on the exponential function shown in section 4.2.2). For example, if a participant was 14 years old then their predicted mean SNR at threshold would be -10 dB. If their actual mean SNR at threshold was -12 dB, then their age-normed LocalB_{5AT} score would be -10 dB minus -12 dB equalling 2 dB.

The amounts by which participant's performance differed from the predicted performance values based on the exponential function fitted to the data was plotted. Figure 4.9 contains the combined data from the adults and children, which shows a roughly normally shaped distribution around approximately 0 (no difference between actual and predicted values). Figure 4.10 shows just the data from the adults and Figure 4.11 shows just the data from the children, from these graphs it is clear that the adult differences from the predicted values are all much closer to 0 (no difference between actual and predicted values) than those of the children. This tells us that the estimates for adults were more accurate than those made for children. This is due to children performing more variably than adults.

Figure 4.12 shows a histogram with a normal distribution for the age-normed scores. From this it can be seen that the data fits the normal curve fairly well, but not perfectly. Two standard deviations above the mean is 6 dB, meaning that the two participants who each had age-normed scores of 7 and 8 were good at the task. Two standard deviations below the mean is -7 dB, so the one participant who scored -8 was particularly poor at the task.

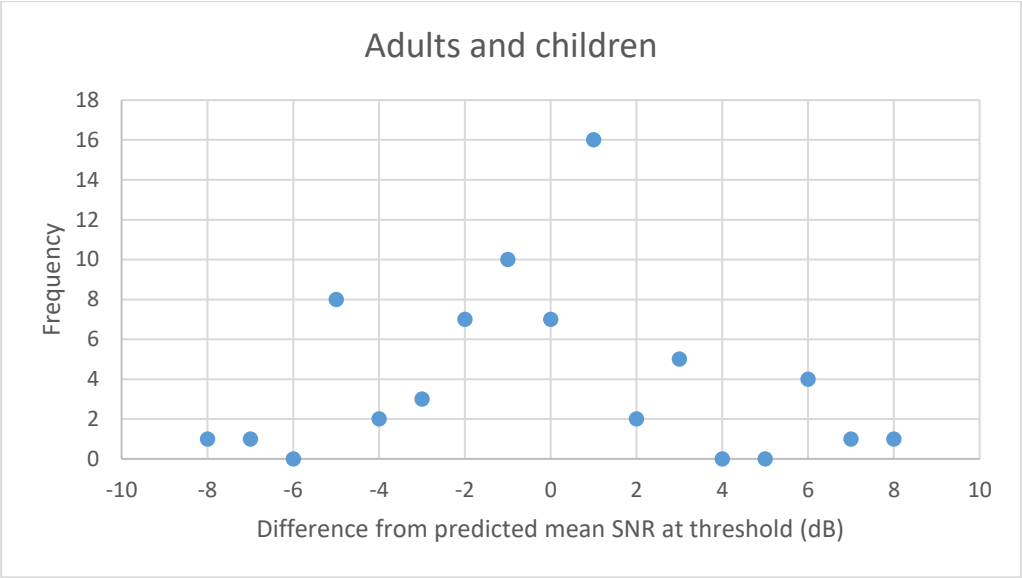


Figure 4.9: Graph showing how all participants performance varied from predicted values

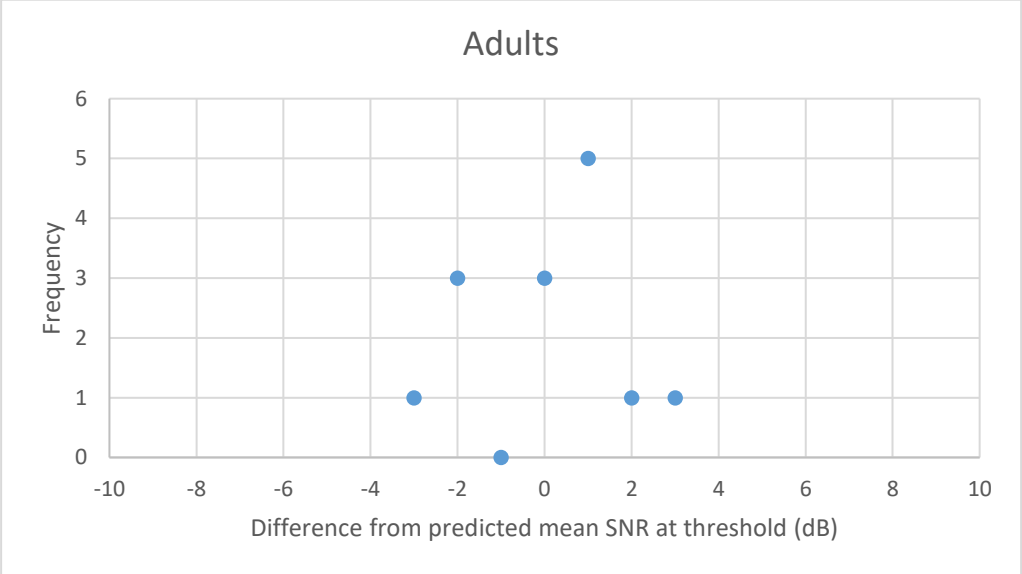


Figure 4.10: Graph showing how adult participants performance varied from predicted values

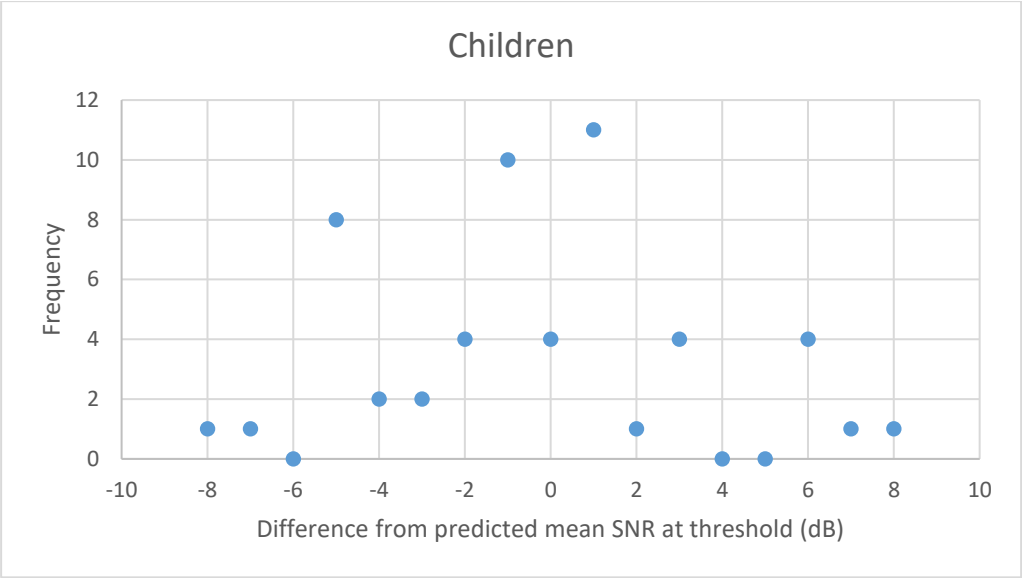


Figure 4.11: Graph showing how child participants performance varied from predicted values

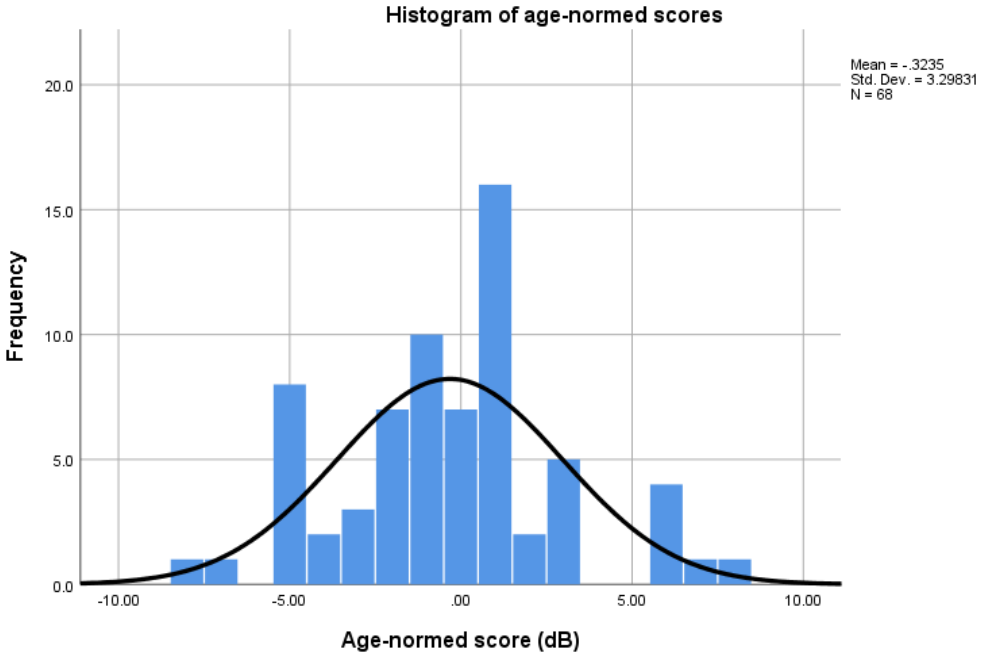


Figure 4.12: Histogram for age-normed scores with a normal curve fitted

4.2.7.2 Localisation in quiet (LocalQ$_{5CS}$)

The LocalQ$_{5CS}$ task was used to assess localisation in quiet abilities of participants. 40 participants aged 10 to 28 years (only those who took part in Experiment 2, as the LocalQ$_{5CS}$ task was not used in Experiment 1) completed the LocalQ$_{5CS}$ task. As can be seen in Figure 4.13 many participants were at

ceiling on the $LocalQ_{5CS}$ task, and those with lower scores generally also had worse scores on the $LocalB_{5AT}$ task.

A Spearman's rank correlation was chosen because the data is monotonic, but not necessarily linear. It showed a medium to large sized statistically significant correlation of $r_s(38) = -0.593$, $p < 0.0005$ between performance on the $LocalB_{5AT}$ and $LocalQ_{5CS}$ tasks. This shows that participants with poor localisation in quiet abilities also have poor localisation of speech in babble skills. This correlation may be influenced by age, as it has been shown that localisation of speech in babble abilities improve with age. However, all participants who completed the $LocalQ_{5CS}$ should have adult-like localisation in quiet abilities as they were all aged 10 years or older and localisation in quiet abilities reach adult-like levels at approximately six years of age. Figure 4.14 shows age and performance in the $LocalQ_{5CS}$ task. From this it can be seen that although there is not a trend of improvement between the ages of 10 and 17 years (the age group of children who completed this task), the children's results appear to be more variable than the adult's results.

Next, a Mann-Whitney U test was run to determine if there were differences in $LocalQ_{5CS}$ score between adults and children. This test was chosen as it does not require data to be normally distributed (which it was not, as assessed by the Shapiro-Wilk test of normality, $p<0.05$ for both adults and children). Distributions of the $LocalQ_{5CS}$ scores for adults and children were similar, as assessed by visual inspection of Figure 4.15. Median $LocalQ_{5CS}$ score was not

significantly different between adults (100%) and children (93%), $U = 43$, $z = -1.854$, $p = 0.064$.

Overall it appears that performance is correlated on the $LocalQ_{5CS}$ and $LocalB_{5AT}$ tasks, and although there is a difference of 7% between mean adult and child scores this was not found to be statistically significant.

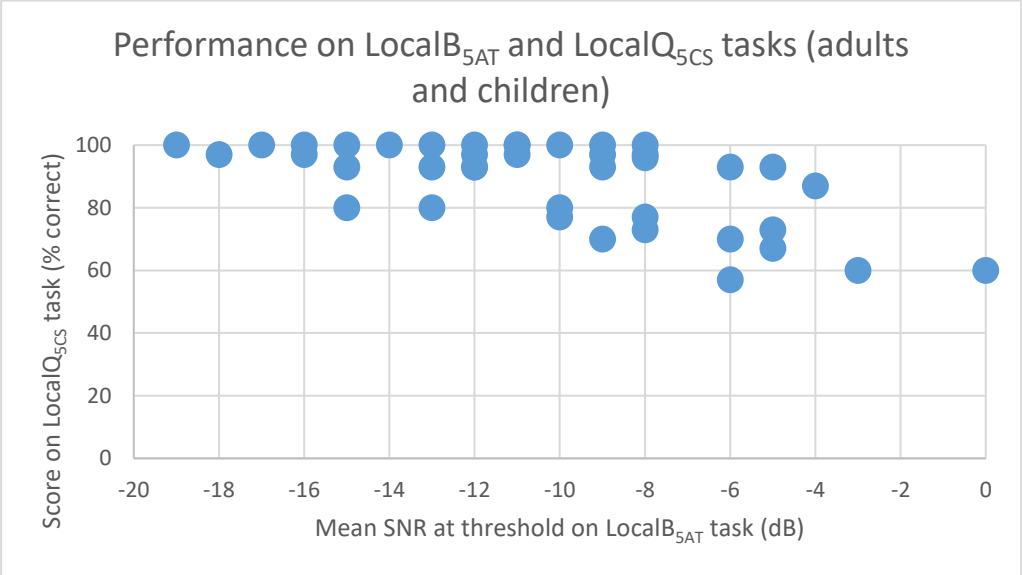


Figure 4.13: Graph showing performance on $LocalB_{5AT}$ and $LocalQ_{5CS}$ tasks

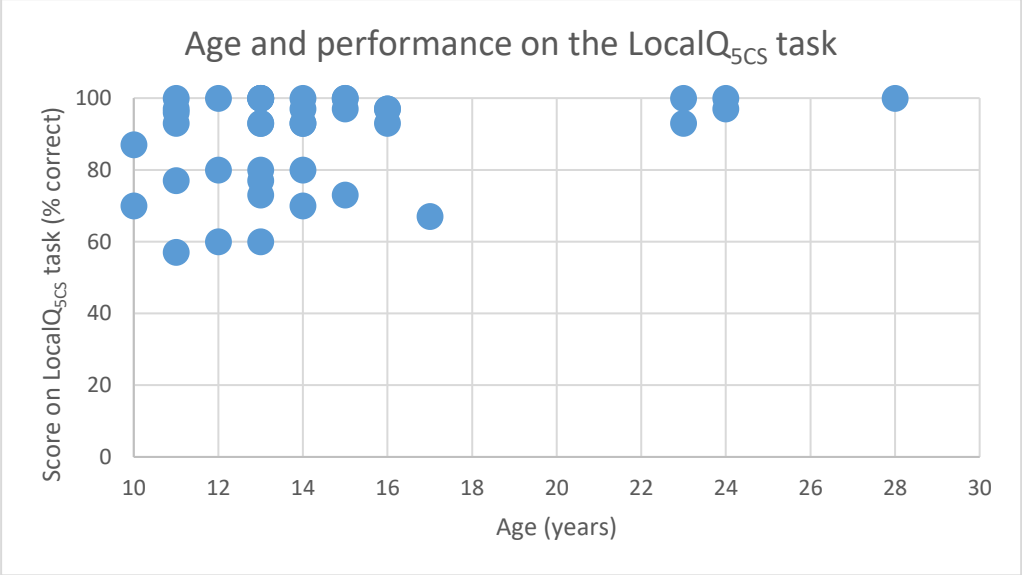


Figure 4.14: Graph showing age and performance on the $LocalQ_{5CS}$ task

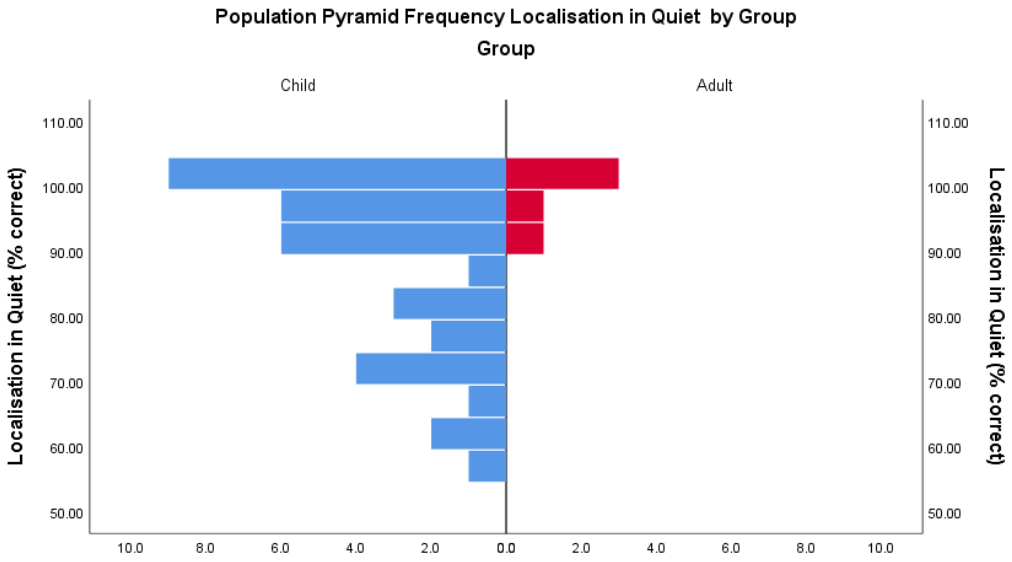


Figure 4.15: Population pyramid graph for localisation in quiet for adults and children

4.2.7.3 Dichotic Digits

The dichotic digits task was carried out to measure how well participants could separate competing stimuli presented simultaneously at each ear. In the left task participants were asked to report the number they heard in their left ear and ignore the number they heard in their right ear. In the right task participants were asked to report the number they heard in their right ear and ignore the number they heard in their left ear.

54 participants aged 6 to 17 years completed this task. The mean data is shown in Table 4.6. From this it can be seen that many participants reached ceiling on this task, with 21 scoring 100% on average and 20 scoring 97.5% on average (so only making one mistake over 40 presentations). In Table 4.7 it can be seen that the poorest performers are young (6-8 years old), however three 7-8 year olds performed very well (97.5% or 100%). It can be seen that performance was on average worse and also more variable on the left task, as mean performance was slightly lower for left than right (a difference of 2.8)

and the left task had a higher standard deviation than the right (a difference of 7.6). A Wilcoxon signed-rank test was conducted as it is the most appropriate test because the data was not linear or normally distributed, but was symmetrically shaped. It found a non-significant median difference of zero between the right and left task scores, z = -1.7, p = 0.088.

	Dichotic digits right (% correct)	Dichotic digits left (% correct)	Dichotic digits average (% correct)
Mean	97.5	94.7	96.1
Standard deviation	5.0	12.6	8.2

Table 4.6: Table showing means and standard deviations for dichotic digits tasks

Mean Score	Frequency	Age of participants (years)	Average age (years)
100%	21	7, 8, 10, 10, 11, 11, 11, 12, 12, 13, 13, 13, 13, 14, 14, 14, 14, 15, 15, 16, 16	12.5
97.5%	21	7, 9, 9, 9, 9, 10, 10, 11, 11, 12, 13, 13, 13, 13, 13, 14, 14, 15, 15, 16, 17	12.0
95%	6	9, 9, 10, 12, 13, 13	11
92.5%	3	9, 9, 11	9.7
85%	2	7, 8	7.5
77.5%	1	6	6
45%	1	7	7

Table 4.7: Table showing age and scores on the dichotic digits task

As the dichotic digits task is not normed for age, scores were compared with the non-age-normed LocalB$_{5AT}$ score. These comparisons are shown in Figure 4.16, Figure 4.17 and Figure 4.18. All of these graphs show similar trends of most participants scoring at or close to 100%, but those who scored less than this tended to have higher (worse) performance on the LocalB$_{5AT}$ task. Table 4.8 shows the four poorest performers on the dichotic digits task (those who scored 80% or lower on either the left or right dichotic digits task). From this it can be seen that these four participants are all very young (aged six, seven, seven and eight years). Two of them had worse performance in the LocalB$_{5AT}$

task than would be expected for their age and the other two had better performance in the $LocalB_{5AT}$ task than would be expected for their age (but none of them were more than two standard deviations from mean performance). It is not possible to draw firm conclusions about why these participants scored poorly on the dichotic digits task due to the small number of participants, especially as it is not clear from their results if they have poor localisation abilities overall. It is possible that they actually had good binaural abilities and had the potential to achieve 100% on the dichotic digits task, but due to lapses in attention during the task (which younger children are more prone to than older children or adults) they achieved low scores. It is not possible to run a correlation test on these tasks ($LocalB_{5AT}$ and dichotic digits) as the data is not linear or monotonic.

Overall there was no difference between performance on the right and the left tasks. Most participants scored at ceiling, but those who didn't tended to be younger participants who also scored fairly poorly on the $LocalB_{5AT}$ task.

ID	Age (years)	Dichotic digits right (% correct)	Dichotic digits left (% correct)	Dichotic digits average (% correct)	SNR (dB)	Predicted SNR for age (dB)	Difference between actual and predicted SNR
RB506146022	6	90	65	77.5	3.5	2.7	Actual SNR worse by 0.8 dB
RB507144020	7	70	20	45	-1.5	0.1	Actual SNR better by 1.6 dB
RB507148008	7	100	70	85	-5	0.1	Actual SNR better by 5.1 dB
RB508151017	8	95	75	85	3.6	-2.0	Actual SNR worse by 5.6 dB

Table 4.8: Table showing dichotic digit and LocalB$_{5AT}$ data for participants who scored lowest of the dichotic digits tasks

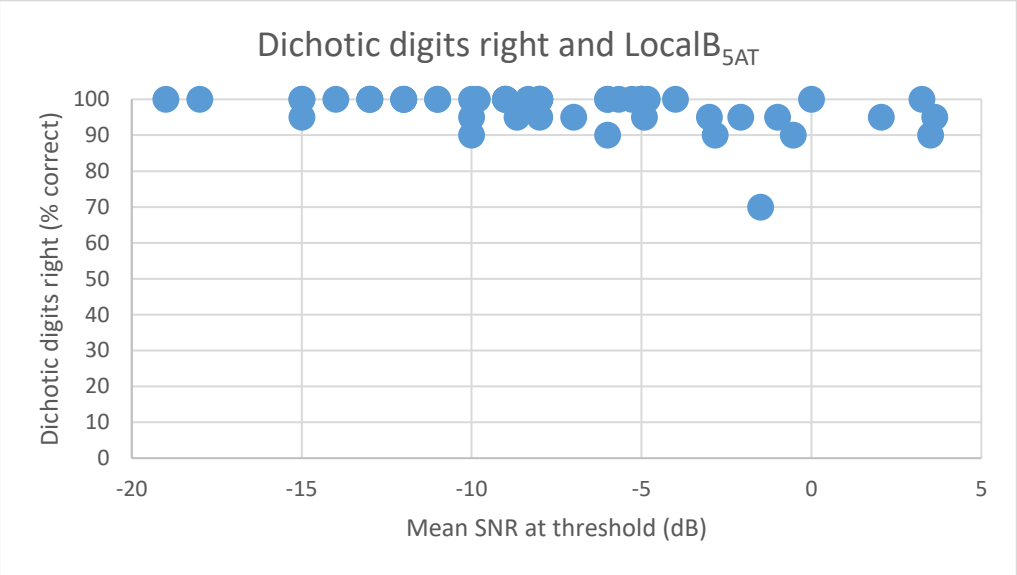


Figure 4.16: Graph showing dichotic digits right and LocalB$_{5AT}$

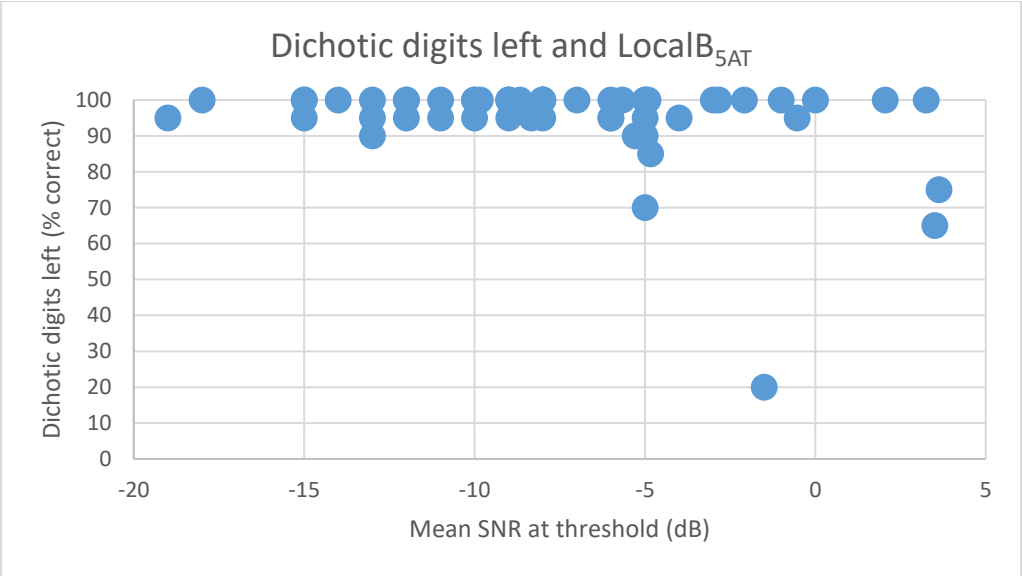


Figure 4.17: Graph showing dichotic digits left and $LocalB_{5AT}$

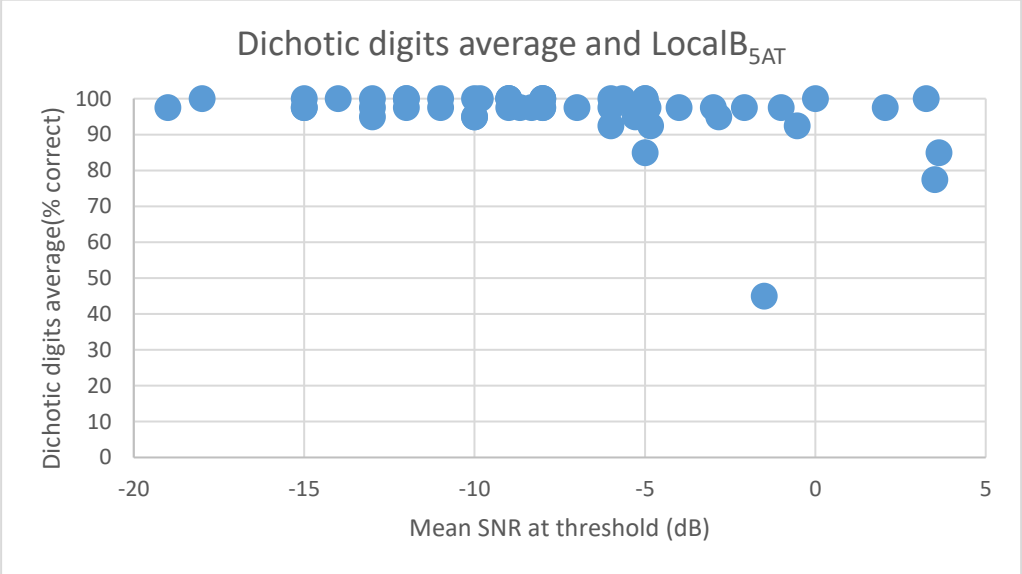


Figure 4.18: Graph showing dichotic digits average and $LocalB_{5AT}$

4.2.7.4 Wechsler Abbreviated Scale of Intelligence (WASI)

The matrices and blocks tasks, which form the non-verbal part of the WASI, were carried out in order to assess non-verbal intelligence of the participants. The t-scores from these tasks were analysed for data from 54 participants. The raw score was converted to a t-score using the WASI manual (Weschler, 1999). T-scores were normed for age and range from 20 to 80, where 50 is the

mean score and the standard deviation is 10. For the matrices task it was found that the t-score ranged from 24 to 76, with an average of 49.5 and a standard deviation of 8.7. For the blocks task it was found that the t-score ranged from 31 to 80, with an average of 57.6 and a standard deviation of 9.8. These values are summarised in Table 4.9. From this it can be seen that on average participants in these experiments scored close to the expected mean on the matrices task and above the expected mean on the blocks task. The performance IQ was also generated for each participant using the sum of the t-scores. The mean of this was 106 (with a standard deviation of 14.2, a minimum of 75 and a maximum of 151), which is close to the expected mean of 100.

	Minimum t-score	**Maximum t-score**	**Average t-score**	**Standard deviation of t-scores**
Standard values	*20*	*80*	*50*	*10*
Matrices	24	76	49.5	8.7
Blocks	31	80	57.6	9.8

Table 4.9: Table showing standard t-score values and t-scores for the matrices and blocks tasks from the WASI

Figure 4.19, Figure 4.20 and Figure 4.21 show the scatterplots for WASI blocks, WASI matrices and WASI performance IQ plotted against age-normed $LocalB_{5AT}$ score. From these graphs it can be seen that there does not appear to be correlation between any of the WASI scores and age-normed $LocalB_{5AT}$ score. Pearson correlations were carried out and confirmed that no significant correlations were present. For the WASI matrices and age-normed $LocalB_{5AT}$ r = -0.1, n = 54, p = 0.4. For the WASI blocks and age-normed $LocalB_{5AT}$ r = 0.1,

n = 54, p = 0.6. For the WASI performance IQ and age-normed LocalB$_{5AT}$ r = -0.03, n = 54, p = 0.8.

Overall, no correlation was found between any of the WASI scores and the age-normed LocalB$_{5AT}$ score.

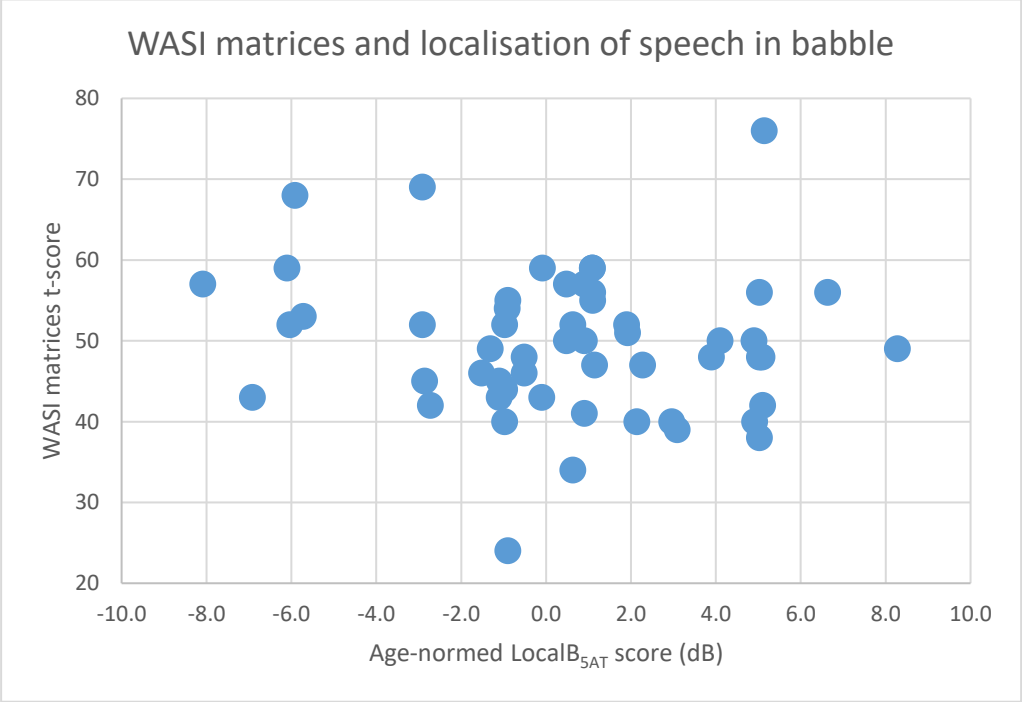


Figure 4.19: Graph showing WASI matrices t-score and age-normed LocalB$_{5AT}$ score

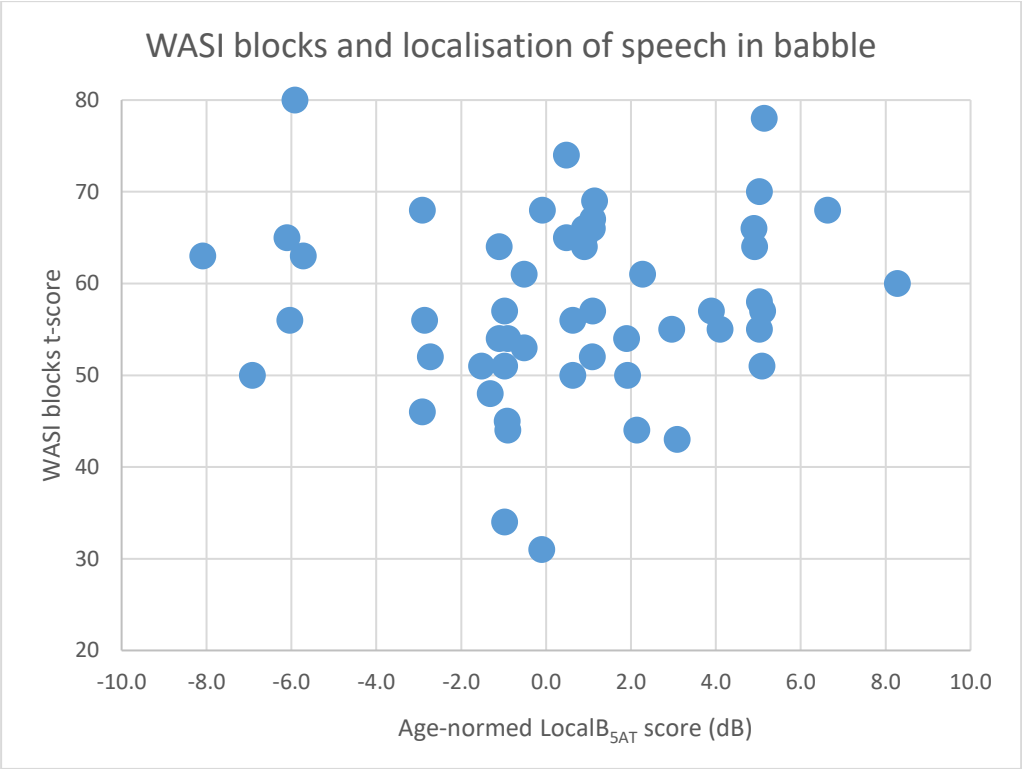


Figure 4.20: Graph showing WASI blocks t-score and age-normed $LocalB_{5AT}$ score

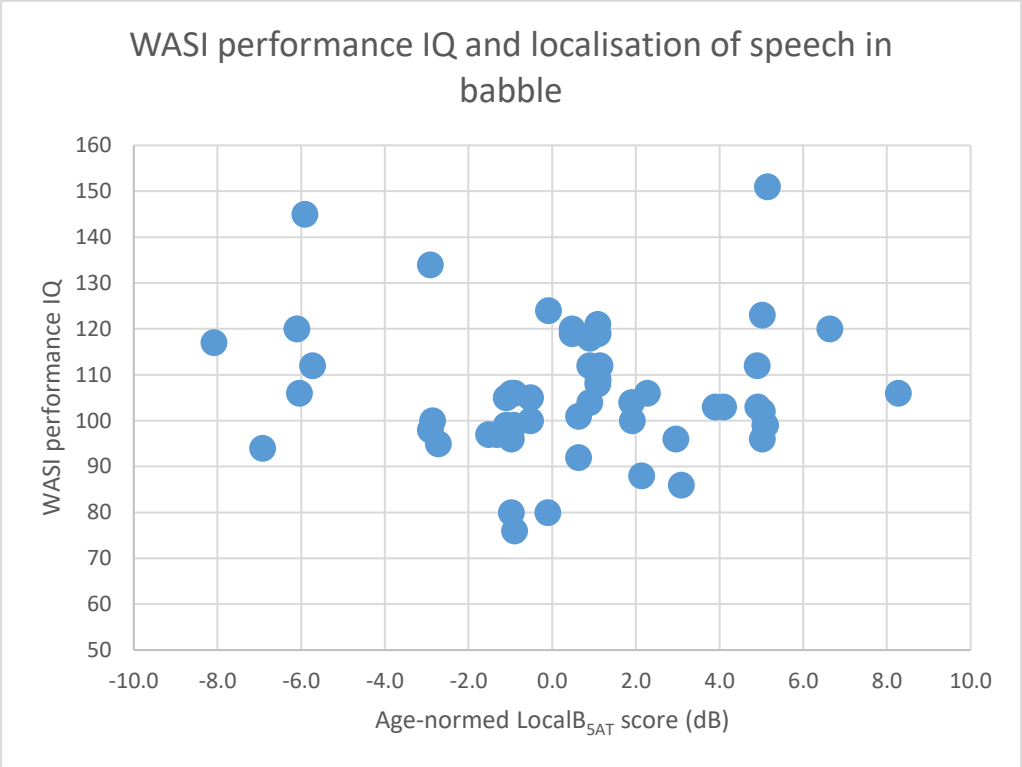


Figure 4.21: Graph showing performance IQ and age-normed $LocalB_{5AT}$ score

4.2.7.5 Digit Span

The digit span task was used to assess working memory. The scaled scores from the digit span (forwards, backwards and total) were analysed. Data from 54 participants was included. The raw score was converted to scaled scores using the WISC manual (Weschler, 1991). The scaled scores were normed for age and range from 1 to 19, where a scaled score of 10 is equal to the median score for age. For the forwards condition the median score was 9 and the mean score was 10, with a standard deviation of 3.3 and a range of 3 to 17. For the backwards condition the median score was 9 and the mean score was 10, with a standard deviation of 3.3 and a range of 4 to 18. For the total score the median score was 10 and the mean score was 10, with a standard deviation of 3.4 and a range of 4 to 19. These values are summarised in Table 4.10. From this it can be seen that on average participants in these experiments scored exactly the expected score of 10.

	Minimum scaled score	**Maximum scaled score**	**Mean scaled score**	**Standard deviation of scaled scores**
Digit span forwards	3	17	10	3.3
Digit span backwards	4	18	10	3.3
Digit span total	4	18	10	3.4

Table 4.10: Table showing scaled scores from the digit span tasks

Figure 4.22, Figure 4.23 and Figure 4.24 show the scatterplots for digit span forwards, backwards and total plotted against age-normed LocalB_{5AT} score. From these graphs it can be seen that there does not appear to be correlation between any of the digit span scores and age-normed LocalB_{5AT} score. Pearson correlations were carried out and confirmed that no significant

correlations were present. For the digit span forwards and age-normed LocalB$_{5AT}$ r = -0.1, n = 54, p = 0.6. For the digit span backwards and age-normed LocalB$_{5AT}$ r = 0.1, n = 54, p = 0.6. For the digit span total and age-normed LocalB$_{5AT}$ r = 0.01, n = 54, p = 0.9.

Overall, no correlation was found between any of the digit span scores scores and the age-normed LocalB$_{5AT}$ score.

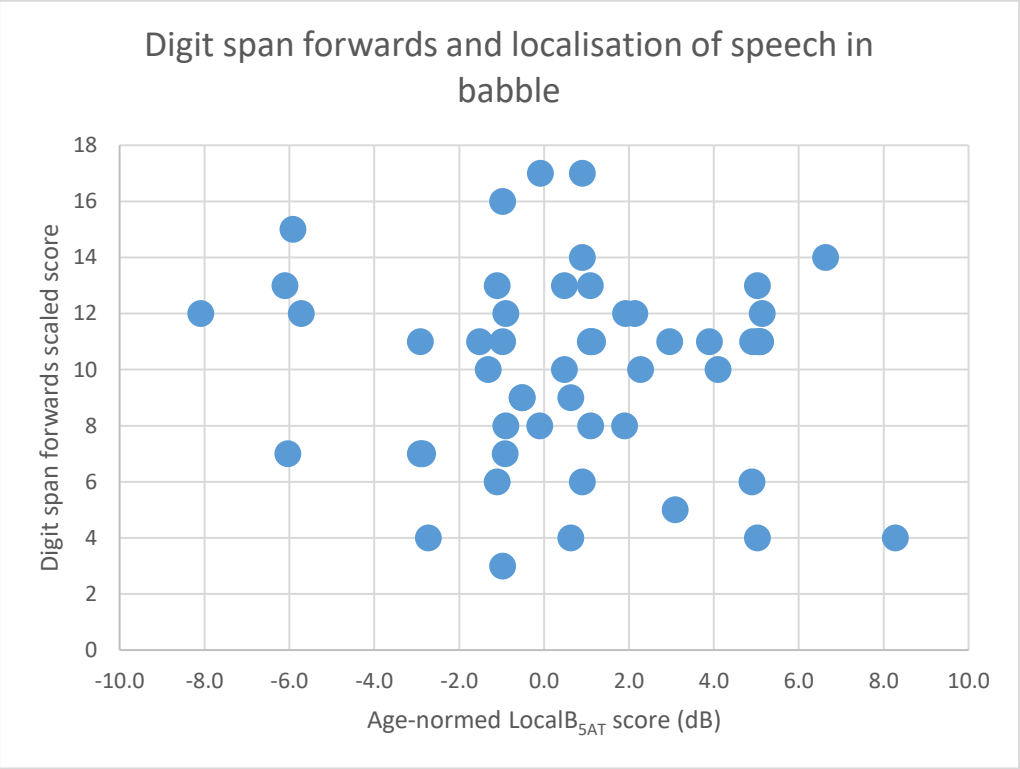


Figure 4.22: Graph showing digit span forwards and age-normed LocalB$_{5AT}$ score

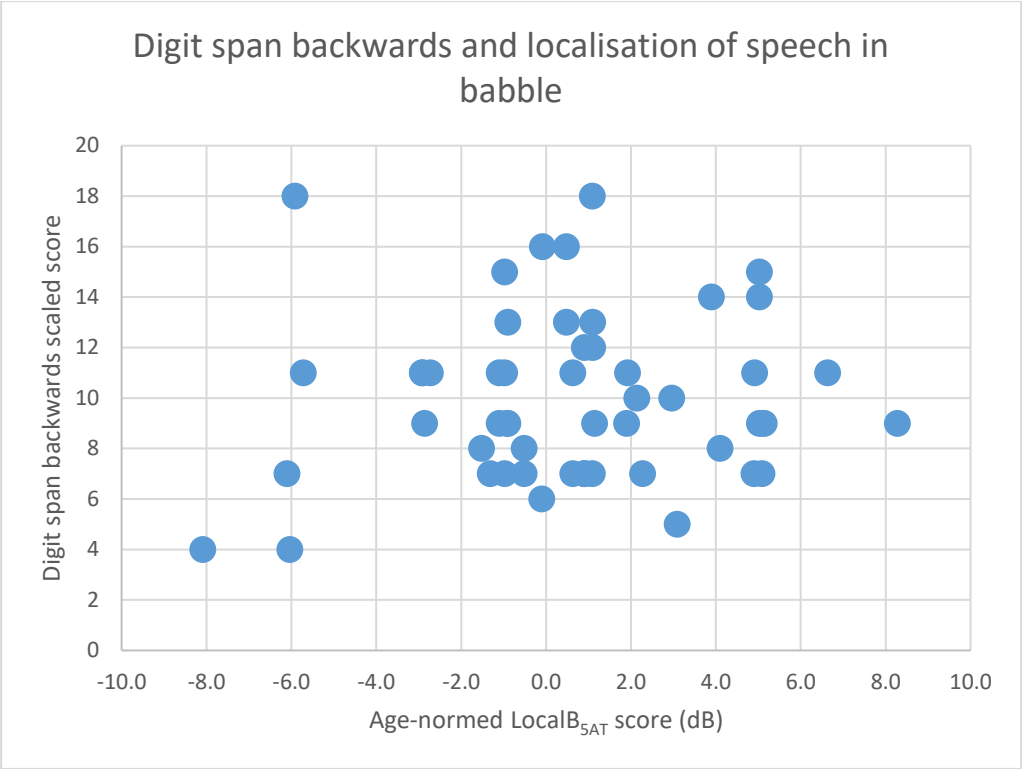


Figure 4.23: Graph showing digit span backwards and age-normed $LocalB_{5AT}$ score

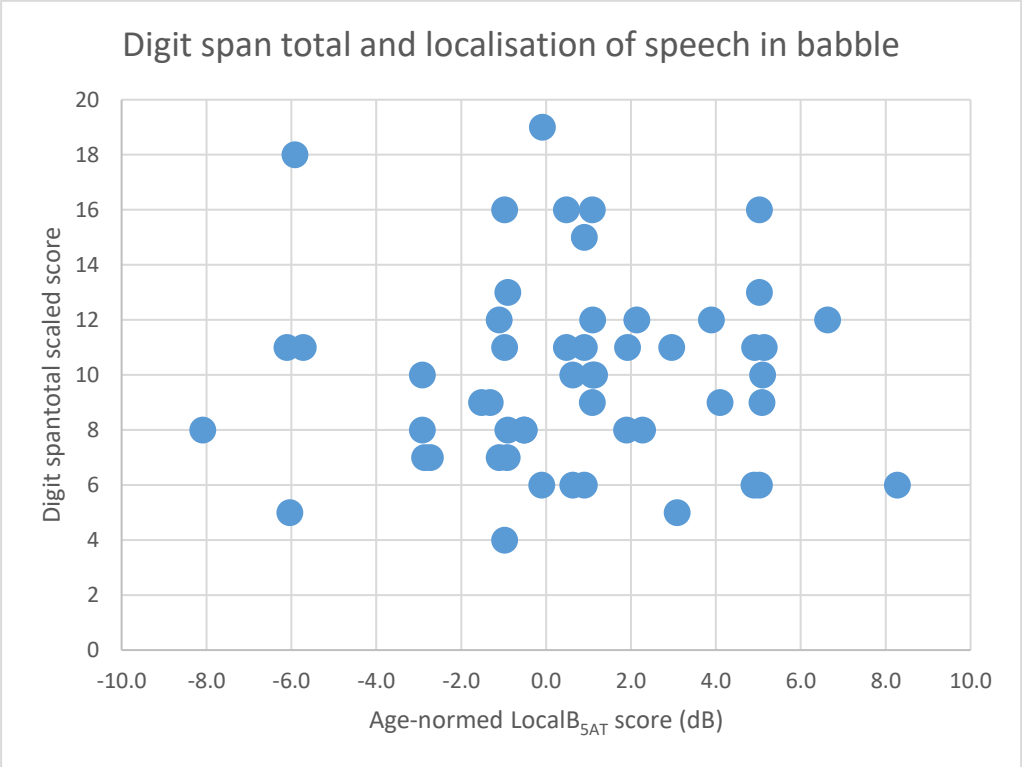


Figure 4.24: Graph showing digit span total and age-normed $LocalB_{5AT}$ score

4.2.7.6 Conners Questionnaire

The Conners questionnaire was completed by the parents of 36 of the participants. Some parents declined to complete this questionnaire. It assessed oppositional behaviour, cognitive problems/inattention, hyperactivity and ADHD index.

Scatter graphs were created for all subsections of the Conners questionnaire to investigate their relationships with age-normed performance on the $LocalB_{5AT}$ task, as the Conners questionnaire is age-normed. These are shown in Figure 4.25 (A: Oppositional), Figure 4.26 (B: Cognitive Problems/Inattention), Figure 4.27 (C: Hyperactivity) and Figure 4.28 (D: ADHD Index). Scores on the Conners subsections could be between 40 and 90 and were age-normed using the conversion provided as part of the questionnaire paperwork.

Spearman's rank correlations were carried out between the age-normed $LocalB_{5AT}$ scores and each of the subsections (A, B, C and D) of the Conners questionnaire. This test was chosen because the data was monotonic but not normally distributed. Table 4.11 shows the results from these correlation tests. From this it can be seen that B (Cognitive problems and inattention) and D (ADHD Index) had small to medium statistically significant correlations with age-normed $LocalB_{5AT}$ scores. Following this, multiple comparisons were tested using the Bonferroni adjustment, giving a new level of significance of (0.05/4 = 0.0125). Both the significance levels for B (Cognitive problems and

inattention) and D (ADHD Index) are greater than 0.0125, therefore they are not actually statistically significant.

	Correlation Coefficient	Significance
A: Oppositional	-0.10	0.56
B: Cognitive Problems/Inattention	-0.35	0.04
C: Hyperactivity	-0.07	0.67
D: ADHD Index	-0.37	0.03

Table 4.11: Table showing results from Spearman's rank correlations for Conners questionnaire and age-normed $LocalB_{5AT}$ score

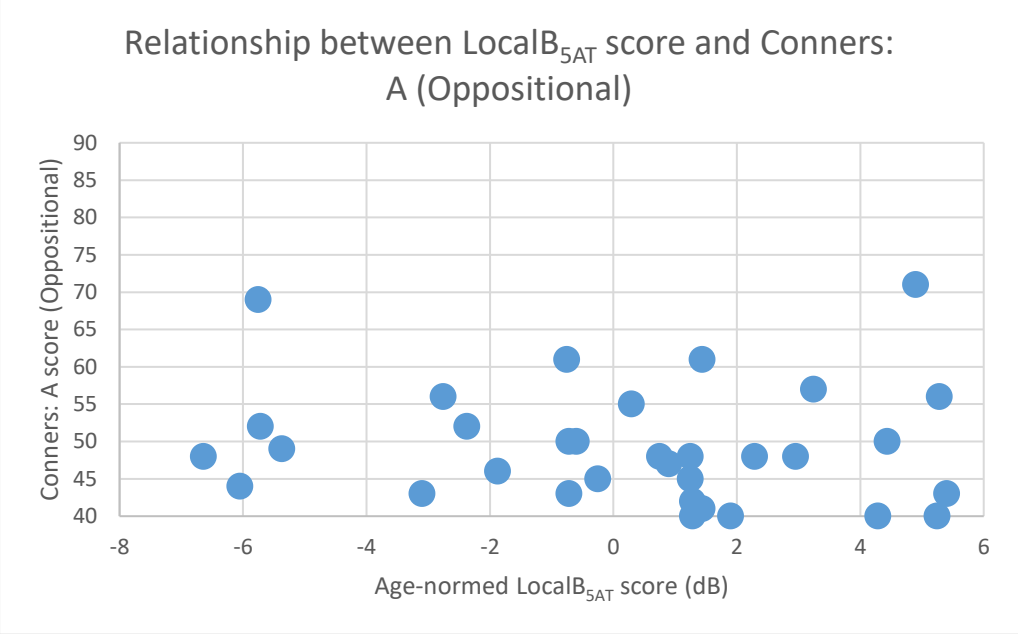


Figure 4.25: Graph showing relationship between $LocalB_{5AT}$ score and Conners: A (Oppositional)

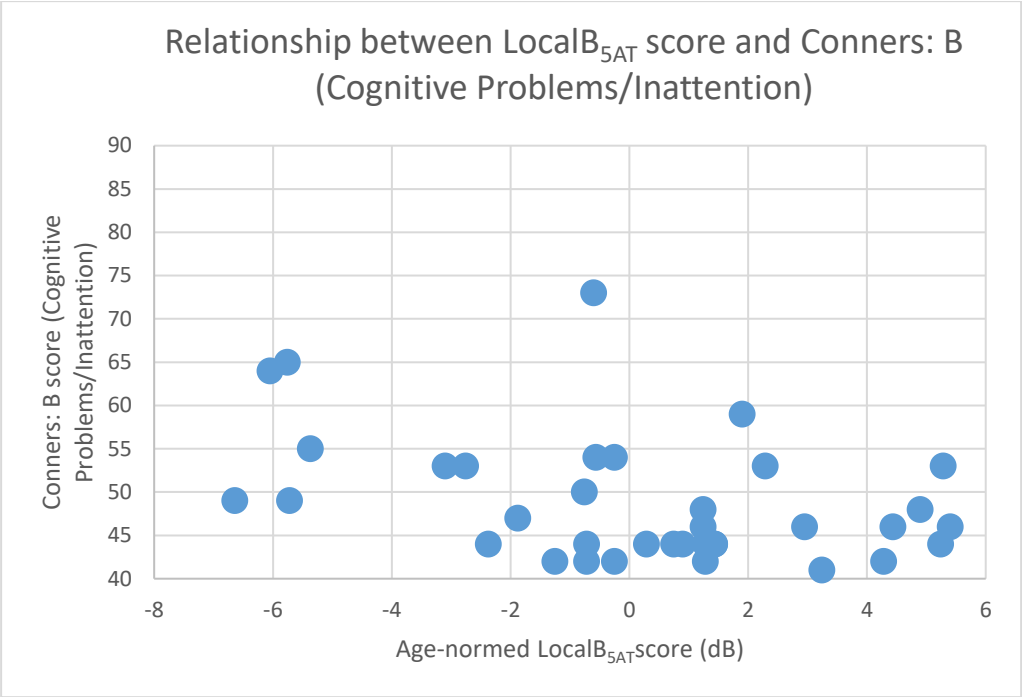


Figure 4.26: Graph showing relationship between LocalB$_{5AT}$ score and Conners: B (Cognitive Problems/Inattention)

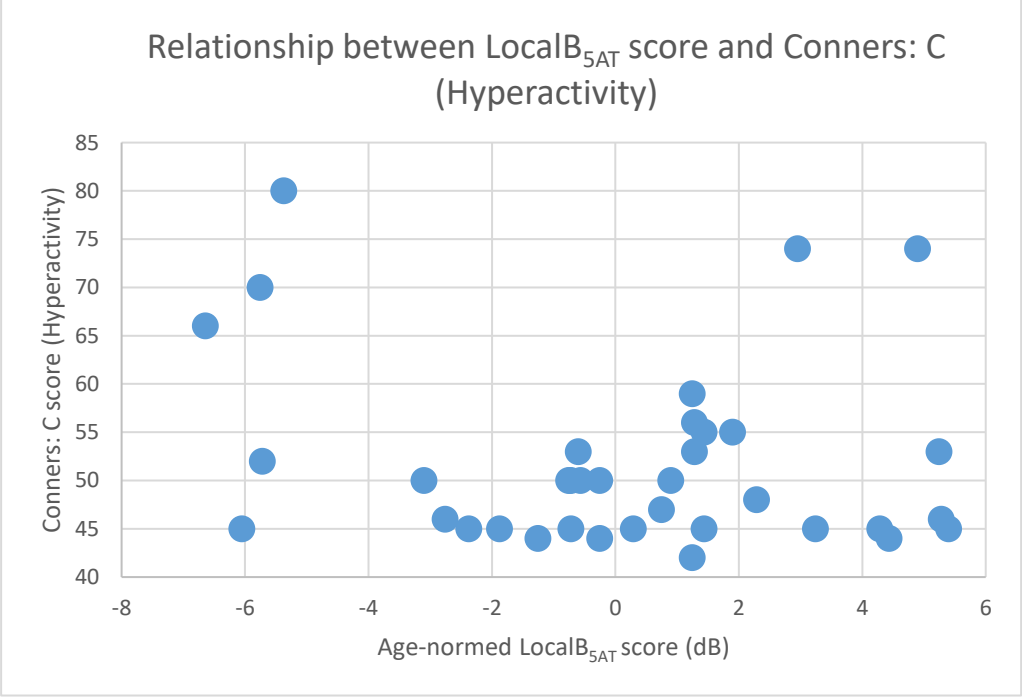


Figure 4.27: Graph showing relationship between LocalB$_{5AT}$ score and Conners: C (Hyperactivity)

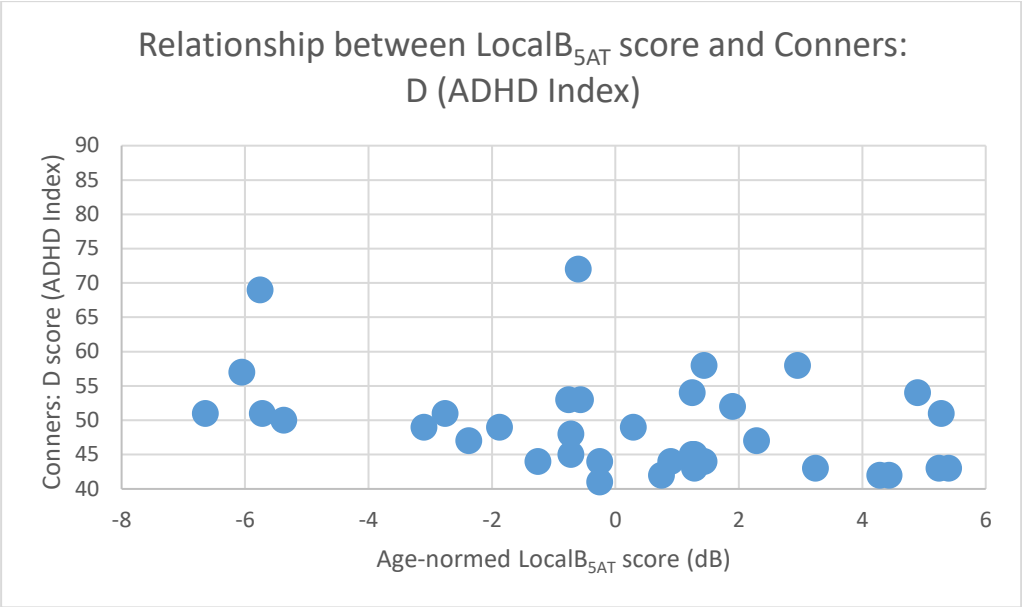


Figure 4.28: Graph showing relationship between LocalB$_{5AT}$ score and Conners: D (ADHD Index)

4.2.7.7 Evaluation of Children's Listening and Processing Skills (ECLiPS) Questionnaire

The ECLiPS questionnaire was used to assess listening and processing skills of pariticpants. Data from the ECLiPS questionnaires of 25 participants aged 6 to 12 years was analysed. This is less than the number of children who participated in the experiment as the ECLiPS is not designed for use with children over the age of 12 years and some parents of participants declined to fill in the questionnaires about their children. Data from the ECLiPS was compared to age-normed LocalB$_{5AT}$ performance as the ECLiPS is an age-normed questionnaire.

Scatter graphs were created for all subsections of the ECLiPS questionnaire to investigate their relationships with age-normed performance on the LocalB$_{5AT}$ task, shown in Figure 4.29. From this there do not appear to be any significant relationships between the ECLiPS subsections and age-normed LocalB$_{5AT}$

performance. Spearman's rank correlations confirmed this observation and no significant correlations were found, results are shown in Table 4.12.

Subsection	Correlation coefficient (r_s)	Significance values
Speech and auditory processing	-0.2	0.3
Language, literacy and laterality	0.02	0.9
Memory and attention	-0.1	0.6
Pragmatic and Social skills	-0.2	0.5
Environment and auditory sensitivity	-0.01	1.0
Listening	-0.1	0.6
Language	-0.3	0.9
Social	-0.2	0.3

Table 4.12: Table showing results from Spearman's rank correlations for the ECLiPS questionnaire and age-normed performance on the LocalB$_{5AT}$ task

4.2.7.8 Summary of results from additional tasks

Table 4.13 shows a summary of the findings of the results from the additional tasks. From this it can be seen that the only task with significant relationships to performance on the LocalB$_{5AT}$ task was the LocalQ$_{5CS}$ task.

Task	Findings
LocalQ$_{5CS}$	Correlation in performance between on the LocalQ$_{5CS}$ and LocalB$_{5AT}$ tasks. No significant difference between performance for adults and children.
Dichotic digits	No difference between performance on the left and right tasks. Younger children tended to perform worst and also performed poorly on the LocalB$_{5AT}$ task.
WASI (blocks and matrices)	No correlation between any of the WASI scores and the age-normed LocalB$_{5AT}$ score.
Digit span	No correlation between any of the digit span scores and the age-normed LocalB$_{5AT}$ score.
Conners questionnaire	No correlation between any of the Conners questionnaire scores and age-normed LocalB$_{5AT}$ scores.
ECLiPS questionnaire	No significant relationships between the ECLiPS subsections and age-normed LocalB$_{5AT}$ performance.

Table 4.13: Table summarising findings from additional tasks

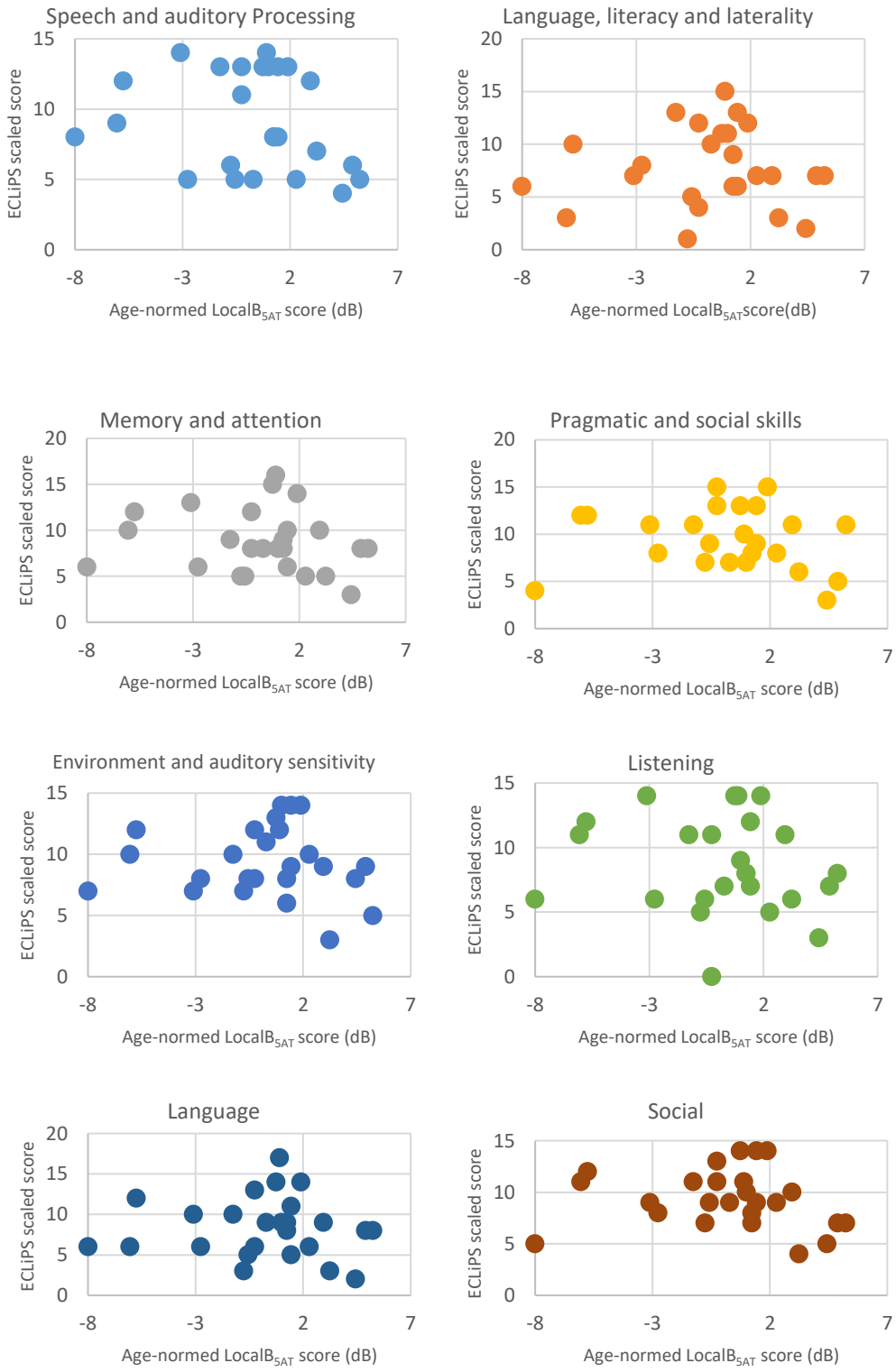


Figure 4.29: Figure showing scatterplots for all subsections of the ECLiPS questionnaire with age-normed $LocalB_{5AT}$ scores

4.3 Discussion

4.3.1 Developmental time course of localisation of speech in babble

This experiment was successful in mapping the developmental time course of localisation in speech in babble, for localisation in the horizontal plane and frontal hemifield, using a source identification method. The results showed a gradual improvement with age in the age range of 6 to 29 years.

The only other study which has investigated localisation of speech in babble in children is Boothalingam et al. (2016), although they present what appears to be the same data again in Boothalingam et al. (2019). They tested 21 children aged 7 to 16 years and 21 adults aged 18-30 years. They found that children performed worse than adults but did not show any age affects or identify the age at which adult-like performance in reached. It may be that because Boothalingam et al. (2016) had a relatively small sample of 21 children, this was not sufficient to demonstrate any age effects.

4.3.1.1 Adult like performance levels for localisation of speech in babble

One of the aims of these experiments was to define the age at which adult-like performance was reached for localisation of speech in babble task. Most studies which have investigated adult-like performance levels have done so using t-tests or ANOVAs. Several different methods have been described in this chapter. They are summarised in Table 4.14 and discussed below.

Method	Age at which adult-like performance is reached (years)
Exponential function (direct fit)	13.0-17.5 (to the nearest 0.5 years)
Broken stick function	13.5-19.0 (to the nearest 0.5 years)
One-way Welch ANOVA	15+
Exponential function (bootstrap)	10.0-14.5 (to the nearest 0.5 years)

Table 4.14: Table summarising different methods of analysis for finding age at which adult-like performance was reached

Looking at the results from the ANOVA it can be seen that the data of 15-17 year olds were not significantly different to the data of the adult group, whereas all of the other groups were. This suggests that adult-like performance should occur at some point beyond 15 years of age. The ANOVA showed a significant difference of 4 dB between the 13-14 year olds and adults. It also showed a non-significant difference of 1.3 dB between the 13-14 year olds and the 15-17 year olds. Next, with the exponential function (direct fit), the age at which adult-like performance was achieved was estimated to be between the ages of 13 and 17.5 years of age. These results seem to be generally in agreement with each other in suggesting that adult-like performance is likely reached between 13 and 17.5 years of age for most children, with the majority of children developing adult-like performance levels for localisation of speech in babble in the lower end of this age range. This is also very similar to the 95% confidence intervals of the change point in the broken stick analysis which were 13.5 and 19 years of age. However, 95% confidence intervals from the exponential function (bootstrap) suggested adult-like performance was reached between 10 and 14.5 years of age. This estimate is similar but slightly lower than the ones discussed above, this

difference is due to the different method used by the bootstrap analysis code (finding the age at which children were 2 RMS errors different to adults).

Overall, these analyses demonstrate that the age at which adult-like performance is reached is unlikely to be measureable at a single age. Instead, it is a range of about four years. Given that, the age range at which adult-like performance is reached is around the early to mid-teenage years for most children. This is much older than the age at which adult-like performance is reached in quiet, which is six years of age (Van Deun et al., 2009; Kühlne et al, 2013; Lovett et al., 2012).

4.3.1.2 Why does localisation of speech in babble improve with age?

Now that the developmental time course of localisation in babble has been mapped and an age effect found, the reasons for this improvement with age can be considered. It is unlikely to be due to age-related maturation in localisation ability per se, as previous research has shown localisation ability in quiet is adult-like by six years of age and localisation in broadband noise (not speech babble) to be mature by seven years of age (Boothalingam et al., 2016). Results from the experiments described in this chapter also support previous findings that children can localise to an adult-like level in quiet by six years of age, as there was no significant difference in performance between adults and children on the localisation in quiet task. However, higher scores on the localisation in quiet task were moderately correlated with better performance on the localisation in babble task. Although ceiling effects are

likely to be present in these results, they do show that localisation in quiet ability is associated with localisation in babble ability.

In a review paper, Leibold (2017) suggests that speech perception in background sound abilities (with one or two streams of competing speech) became adult-like by 13-14 years of age. This is at the lower end of the age-range at which adult-like performance of localisation of speech in babble is expected to be achieved. It is possible that age-related improvements in both perception (or identification) and localisation of masked speech may be linked to improvements in cognitive abilities and in glimpsing abilities (Leibold, 2017).

Glimpsing abilities

Glimpsing refers to the ability to use sparse cues in order to recognise a target stimulus, essentially listening to a signal of interest between the gaps in the competing signals. Buss et al. (2017) investigated the effect of glimpsing abilities on speech reception thresholds (SRTs) in background sound in children aged 5 to 16 years and in young adults. They presented disyllabic words in one and two talker maskers, with and without a speech shaped noise masker and used a four alternative choice task to estimate SRT. They showed age effects in all conditions, with children reaching an adult-like level in the one talker maskers (10-12.9 years) earlier than in the two talker masker (16.1-16.8 years). They found that younger children's SRTs were less negatively affected by the presence of speech shaped noise than the SRTs of the adults in both the one and two talker masker conditions. When the speech shaped

noise was present it was harder for participants to use glimpsing abilities, so this result tells us that the younger children were using glimpsing less than the adults were. Glimpsing abilities appear to continue developing throughout the teenage years, a developmental time course similar to that of localisation of speech in babble. Leibold (2017) writes that glimpsing abilities become less useful as the number of talkers in a masker increase (as there are less gaps in the masker overall). Therefore, the impact of glimpsing on the tasks described in this chapter, which used a six talker babble, are smaller than if the tasks had used a one or two talker masker. However, it is possible that glimpsing abilities are helpful in these localisation of speech in babble tasks and it could be possible that improvement in both of these abilities are driven by a common factor.

Cognitive abilities

Although none of the cognitive abilities tested using the WASI or digit span tasks showed a significant correlation to performance on the LocalB_{5AT} task it is possible that other cognitive abilities could be influential in the development of localisation of speech in babble. Much like localisation in babble, working memory, sustained and selective attention all have long developmental periods (Klenberg, Korkman and Lahti-Nuuttila, 2001; Luna et al., 2004; Crone et al., 2006; Rebok et al., 2007; Karns et al., 2015; Egami et al., 2015; Fortenbaugh et al., 2017; Lufi and Haimov, 2018). Selective attention develops in the early teenage years, whilst sustained attention

continues until the later teenage years and working memory potentially does not reaching adult-like levels until the early twenties.

Although these abilities follow similar developmental time courses to localisation of speech in babble it cannot be concluded what the relationships between them are. It may be that they all develop at similar times due to a common factor which is linked to the maturation of localisation of speech in babble.

No significant correlations were found with localisation of speech in babble abilities and the dichotic digits task, which measured the ability to separate competing stimuli at each ear. It may be that this was due to ceiling effects in the dichotic digits task. Alternatively these abilities may not actually be relevant to each other as the dichotic digits task is much simpler than the localisation of speech in babble task. The Conners questionnaire, which measured oppositional behaviour, cognitive abilitites/inattention, hyperactivity and ADHD index, showed no significant correlations with localisation of speech in babble abilities for any of the subsections. It is worth noting that these measures are based on a parental questionnaire and so are not direct measures of the characteristics. So although they suggest that the characteristics tested are not correlated with performance on the $LocalB_{5AT}$ task, it cannot be firmly concluded that this is the case.

It is surprising that none of the aspects of the ECLiPS questionnaire are correlated with localisation of speech in babble, given that it is a measure of listening skills (Barry et al., 2015). It may be that these questionnaires would

be good predictors of localisation of speech in babble performance in children who are not typically developing, for example if they have Auditory Processing Disorder or clinically significant attention deficits, but that they are not appropriate tools in this context for typically developing children.

4.3.2 Limitations and future work

These studies successfully mapped the developmental time course of localisation of speech in babble from the age of six years to early adulthood. A variety of methods were used to estimate the age at which adult-like performance was reached, with a probable range within the early to mid-teens.

The experiment described in this chapter had a small gap in the ages of participants tested, with only four participants in the age range of 17 to 22 years and with no participants aged 20 or 21 years. In the next experiment greater efforts will be made to ensure a better age range of participants, which should lead to more accurate results. As the results from the questionnaires and cognitive tasks were not found to be very informative they will not be used in any further experiments.

As discussed in Chapter 1, there are drawbacks to using cross-section designs (as was used in this experiment). The main drawback with comparing participants with different ages is that there may be other factors which differ between the groups which could be causing participants to vary in their ability to perform on the task (in addition to age). The children who participated in the experiments are likely to be representative of the general population of

the age groups tested as they were recruited widely from social media and a school in a diverse area of Nottingham. The range and mean values of the children's IQ scores showed that they formed a diverse group, in this respect, representative of the general population. This means that it is reasonable to compare these different age groups of children and assume that any differences in their localisation of speech in babble performance is likely due to age. However, the young adults were all university students or graduates so may not be wholly representative of the general population of this age group. Therefore, the young adults may be less comparable to the children than the children are to each other as the children are likely from a more varied population. In this sample of participants tested in these experiments adult-like performance was found to occur in the early to mid-teenage years. It must be considered whether this can be assumed to be true for the general population. It could be speculated that this participant group have higher cognitive abilities than the general population and therefore would perform better on the task than an average person of the same age, meaning that the adult-like performance level identified in these experiments is higher/more difficult than would be measured in a more varied population. This means that the age at which adult-like performance is reached is potentially over-estimated by these experiments.

Chapter 5

Experiment 3: Detection, Identification and Localisation of Speech in Babble in Young Adults

5.1 Introduction

This experiment was carried out in order to expand on the findings of the detection ($DetB_{2CS}$), identification ($IdentB_{4CS}$) and localisation ($LocalB_{5CS}$) tasks from Experiment 2. There it was found that performance was best in identification, followed by detection and localisation. On one hand, it would perhaps be expected that performance would be highest in detection, as if a listener can identify and/or localise a sound then they should be able to detect it. However, previous studies looking into differences between identification and detection of masked speech/vowels have found that participants perform better on identification than detection tasks (Hawkins and Stevens, 1950; Liu and Jin, 2019). Although, as neither of these studies used informational masking (babble) it is important to further investigate these relationships. Only one study has investigated the relationship between identification and localisation, but made no meaningful comparisons (Hawley, Litovsky and Colburn, 1999). They tested 12 young adults on speech

intelligibility and localisation in background sound tasks, using IEEE sentences at 62 dB(A) for the target and distractor sentences. They used one, two and three talker babbles from a range of locations. They tested participants in a sound field and over headphones. The angles used were ±90°, ±60°, ±30° and 0°. Due to these large separation angles and fairly easy SNRs results for all participants were at ceiling. This meant that the localisation data could not be compared to the identification scores. Therefore, it is important to further investigate the relationships between detection, identification and localisation of speech in babble.

Firstly in Experiment 2 it was thought that any findings based on comparisons between the $IdentB_{4CS}$ and $LocalB_{5CS}$ tasks may have been affected by the fact that they had a different number of response choices (four and five). Therefore in this experiment an additional localisation task, which was a one interval-four choice task, was added to the test battery. This allowed more meaningful comparison between performances on identification and localisation tasks.

Secondly, it was decided to investigate more challenging SNRs than those used in Experiment 2. This allowed for greater understanding of the relationship between performance on different tasks, as in Experiment 2 there seemed to be more complex trends for the more challenging SNRs used. These trends showed that performance on identification appeared to be dropping to the level of (or even worse than) detection and localisation for the hardest two signal levels. Conversely, for all other signal levels

performance on identification was much higher than on the other tasks. In Experiment 2 the SNRs used for each participant were based on their individual performance in the $LocalB_{5AT}$ task. Whilst this meant that the tasks should not have been too difficult for any of the participants, it meant that comparing results between participants was not ideal. For the current experiment it was decided to only test young adults as they are less likely to be affected by difficulty understanding the task or lapses in attention. Only testing young adults meant that the same SNRs could be used for all participants, allowing for more meaningful comparison of data.

5.2 Aims and Hypotheses

The first aim of this experiment was to investigate the relationship between performance on identification and localisation tasks when altering the number of response options. It was hypothesised that participants would perform better on the one interval-four choice $LocalB_{4CS}$ task than the one interval-five choice $LocalB_{5CS}$ task, due to increased probability of guessing correctly by chance. It was also expected that there would be a smaller (if any) difference between performance on the one interval-four choice $LocalB_{4CS}$ and $IdentB_{4CS}$ tasks than the one interval-four choice $IdentB_{4CS}$ and one interval-five choice $LocalB_{5CS}$ tasks.

The second aim of this experiment was to explore the relationships between task performances at more challenging SNRs than those used in Experiment 2. It was expected that the trends would be different to those seen in Experiment 2, in which performance on all SNRs was around threshold or

higher, due to the more challenging SNRs used in the new experiment. The experiment also aimed to find the SNR at which task performance was at floor level and observe the relationships between task performances from this SNR upwards.

5.3 Methods

5.3.1 Participants

17 adults (12 female, 5 male) aged 18 to 30 years (mean 23.6, s.d. 2.8) participated in the experiment. All participants had normal hearing and spoke English as their first language. Participants were friends/colleagues of the researcher or were recruited via posters placed in the University of Nottingham Student Union building or posts on University of Nottingham Student Facebook pages. Participants were unaware of the purpose of the experiment. One additional participant attended a test session but was unable to take part in the experiment due to a previously undiagnosed hearing loss. They were given a report to give to their GP.

5.3.2 Equipment

The same speaker ring was used as in Experiments 1 and 2, although the ring (and all other equipment) was moved to a newly built soundproof test room in the Hearing Sciences building at the University of Nottingham campus. The background level of the test room was 24 dB(A) measured using a Brüel & Kjær 2260 Investigator sound level meter on the slow setting. A photo of the test room is shown in Figure 5.1, and a sketch is shown in Figure 5.2. This

room was similar to the previous one, but instead of having an observation window, CCTV cameras were used and the feed from these could be viewed on a screen in the adjoining observation room. A photo of the observation room is shown in Figure 5.3.

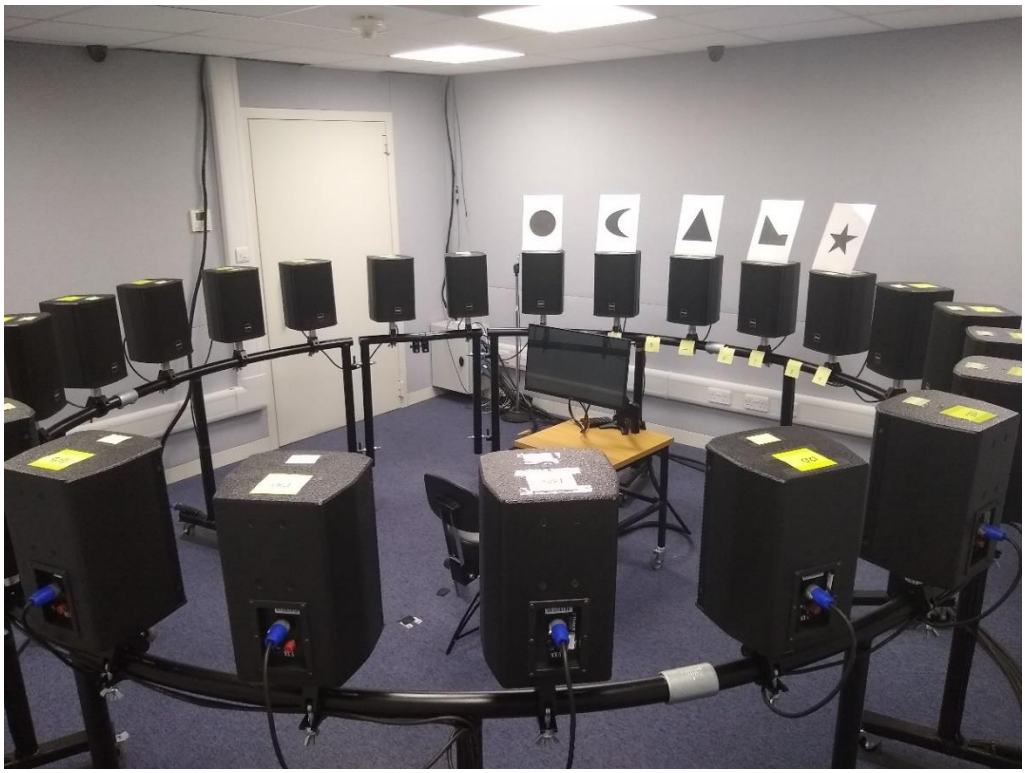


Figure 5.1: Photo of the new test room

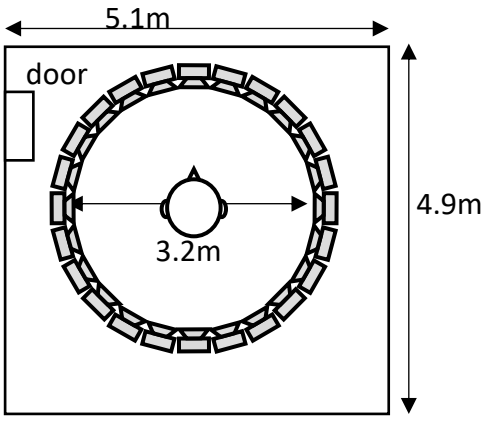


Figure 5.2: Figure showing sketch map of test room in Hearing Sciences, University of Nottingham

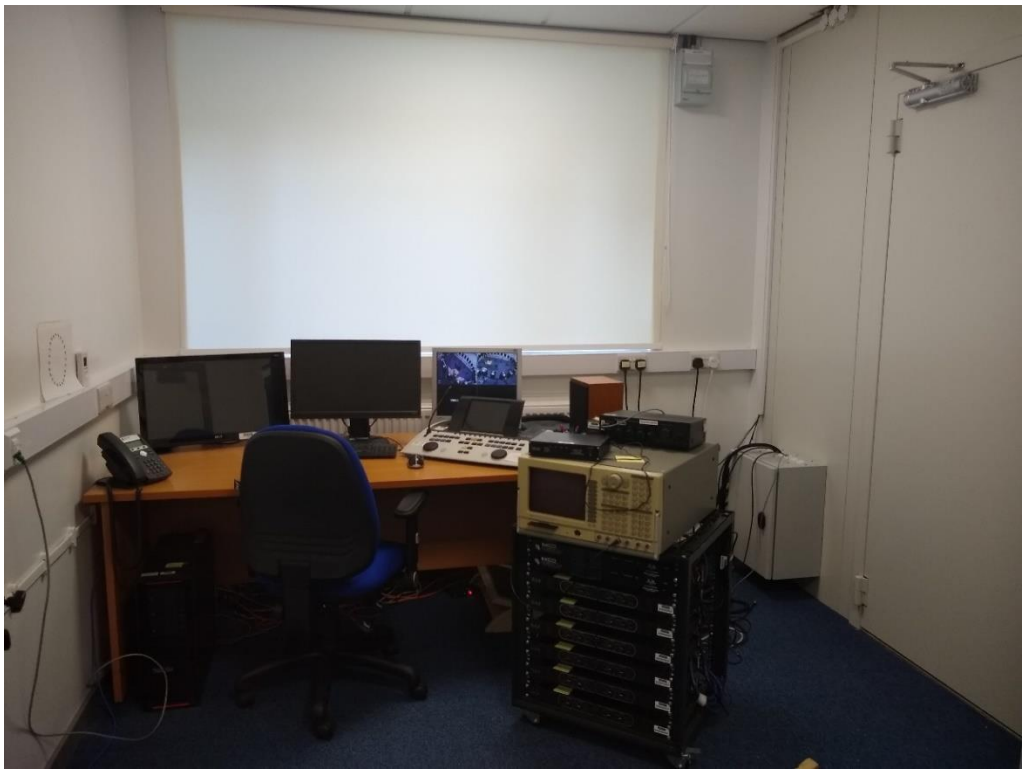


Figure 5.3: Photo of the observation room

All of the equipment was the same as was used in Experiments 1 and 2, except that the Sonible d:24 amplifier had to be replaced with six ART SLA 4 Studio Linear 4 channel amplifiers due to an unfixable fault with it. Two Ferrofish A16 MK-II AD/DA Converters were required in order for this new set up to work. The signal chain is shown in Figure 5.4. The MATLAB code used to run the experiment was written by Michael Akeroyd as part of the overall programming of the laboratory.

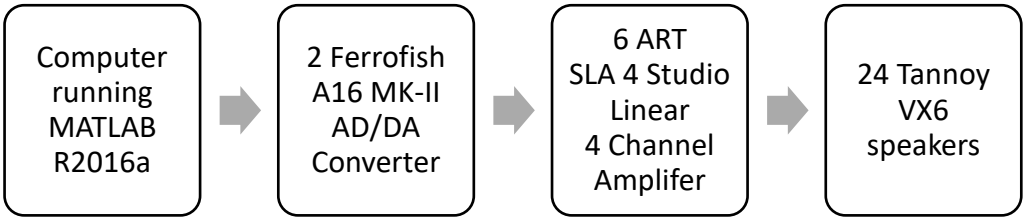


Figure 5.4: Figure showing equipment used for speaker ring

5.3.3 Stimuli

5.3.3.1 Signal

The signals used were the same single speech words from the CRM sentence corpus as used in Experiments 1 and 2. As in Experiment 2, only female voiced words were used for the signal.

5.3.3.2 Masker

As in Experiments 1 and 2 the stimuli used were IEEE sentences and they were played from six different locations in the ring (a different sentence from each of the six speakers). As in Experiment 2, only male voiced sentences were used for the masker.

5.3.4 Procedure

All participants attended a single one hour session in the Hearing Sciences department at the University of Nottingham. Participants were given £7.50 inconvenience allowance and any travel expenses were reimbursed. Prior to the experiment beginning consent was obtained.

5.3.4.1 Study tasks

Prior to the study tasks all participants underwent pure tone audiometry, to check that they had normal hearing. The tasks are explained below.

Practice tasks

The first practice task was a localisation in quiet task ($LocalQ_{5CS}$). This was the same one as used in Experiment 2. The second task was a localisation in babble task with an easy SNR of 15 dB. This was the same one as the one

carried out in Experiments 1 and 2. The exact signal levels used varied slightly compared to those used in the previous experiments, as the lab was calibrated differently when it was moved to the new location. These values are shown in Table 5.1. As these tasks were only for familiarisation and were clearly audible for all participants, these slight differences were not important. Sound level measurements in the new test room were made using a hand held sound level meter, measuring in dB(A) on the slow setting.

	Signal level (dB(A))	**Low background sound level (dB(A))**	**SNR for Localisation in quiet task (dB)**	**SNR for Localisation in low background sound task (dB)**
Experiments 1 and 2	60	38	60	22
Experiment 3	53	38	53	15

Table 5.1: Table showing levels and SNRs for practice tasks

Main tasks

The $DetB_{2CS}$, $IdentB_{4CS}$ and $LocalB_{5CS}$ tasks were the same as those used in Experiment 2, except that more challenging SNRs were used and an additional localisation task (one interval-four choice, referred to as $LocalB_{4CS}$) was added. The details of these tasks are shown in Table 5.2.

Test		Localisation in babble (5 choice)	Localisation in babble (4 choice)	Identification in babble	Detection in babble
Abbreviation		LocalB$_{5CS}$	LocalB$_{4CS}$	IdentB$_{4CS}$	DetB$_{2CS}$
Number of Presentations		75 (15 of each of 5 levels)	80 (16 at each of 5 levels)	80 (16 at each of 5 levels)	160 (80 with signal present, 80 with signal absent; 16 at each of 5 levels)
Response		Choose from one of five options	Choose from one of four options	Choose from one of four options	Choose from one of two options
Response Options		±30°, ±15°, 0°, represented by shapes	±30° and ±15° represented by shapes	Word: red, green, blue or white	Was sound present: yes or no
Signal	Type	Word	Word	Word	Word
	Location	±30°, ±15°, 0°	±30°, ±15°	0°	0°
	Level	30, 33, 36, 39, 42 dB(A)	30, 33, 36, 39, 42 dB(A)	30, 33, 36, 39, 42 dB(A)	30, 33, 36, 39, 42 dB(A)
Masker	Type	6 IEEE sentences	6 IEEE sentences	6 IEEE sentences	6 IEEE sentences
	Location	6 locations around the ring	6 locations around the ring	6 locations around the ring	6 locations around the ring
	Level	Each sentence 55 dB(A), total 63 dB(A)	Each sentence 55 dB(A), total 63 dB(A)	Each sentence 55 dB(A), total 63 dB(A)	Each sentence 55 dB(A), total 63 dB(A)
SNRs		-33, -30, -27, -24, -21 dB	-33, -30, -27, -24, -21 dB	-33, -30, -27, -24, -21 dB	-33, -30, -27, -24, -21 dB
Practice		5 trials (no feedback)	5 trials (no feedback)	5 trials (no feedback)	5 trials (no feedback)
Outcomes		% correct at each level	% correct at each level	% correct at each level	% correct at each level

Table 5.2: Table showing Detection, Identification and Localisation of speech in babble constant stimuli tasks

5.4 Results

5.4.1 Overview of results

All 17 participants completed all of the tasks. The average results for all participants are shown in Table 5.3 and Figure 5.5. From this graph it appears that performance is similar at the two most challenging SNRs (-33 and -30 dB), then increases as the SNRs increase. The task which participants overall

performed best at was $DetB_{2CS}$, followed by $LocalB_{4CS}$ and $IdentB_{4CS}$ and then performance was worst in $LocalB_{5CS}$.

SNR (dB)	$LocalB_{5CS}$ (% correct)	$LocalB_{4CS}$ (% correct)	$IdentB_{4CS}$ (% correct)	$DetB_{2CS}$ (% correct)
-33	21	24	25	35
-30	21	26	25	38
-27	24	33	36	53
-24	31	43	41	63
-21	38	47	55	71

Table 5.3: Table showing mean scores for all participants for each task at each SNR

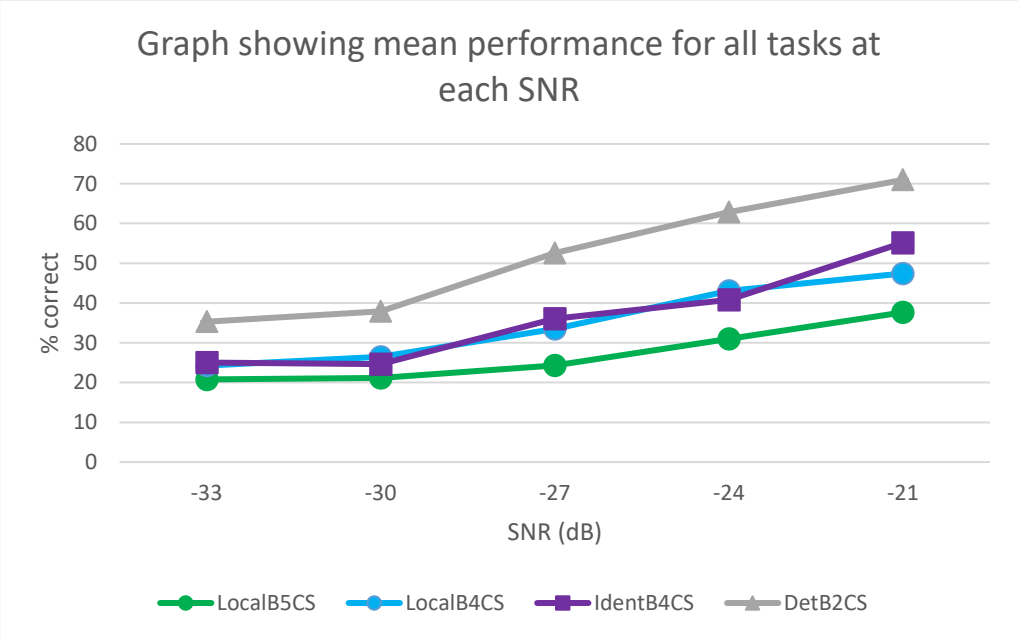


Figure 5.5: Graph showing mean performance for all tasks at each SNR

5.4.2 Comparison of task performances

A one-way repeated measures ANOVA was carried out to investigate any differences between mean scores (% correct) across the SNRs for each task. All task averages were normally distributed, as assessed by the Shapiro-Wilk test of normality. No outliers were present, according to visual assessment of boxplots shown in Figure 5.6. The assumption of sphericity was not violated (according to Mauchly's test of sphercitiy, p=0.164). The ANOVA showed that

performance was significantly different across the tasks $F(1,16) = 49.071$, $p < 0.0005$, $n^2 = 0.754$, $w^2 = 0.68$. Post hoc testing results of comparisons between the tasks are shown in Table 5.4. From this, it can be seen that there were significant differences between all tasks, except between $IdentB_{4CS}$ and $LocalB_{4CS}$.

The ANOVA showed that performance on the $LocalB_{4CS}$ task was significantly better than the $LocalB_{5CS}$ task, as was previously hypothesised. $IdentB_{4CS}$ was significantly better than $LocalB_{5CS}$, but not significantly better than $LocalB_{4CS}$. $DetB_{2CS}$ was significantly better than $IdentB_{4CS}$, $LocalB_{5CS}$ and $LocalB_{4CS}$.

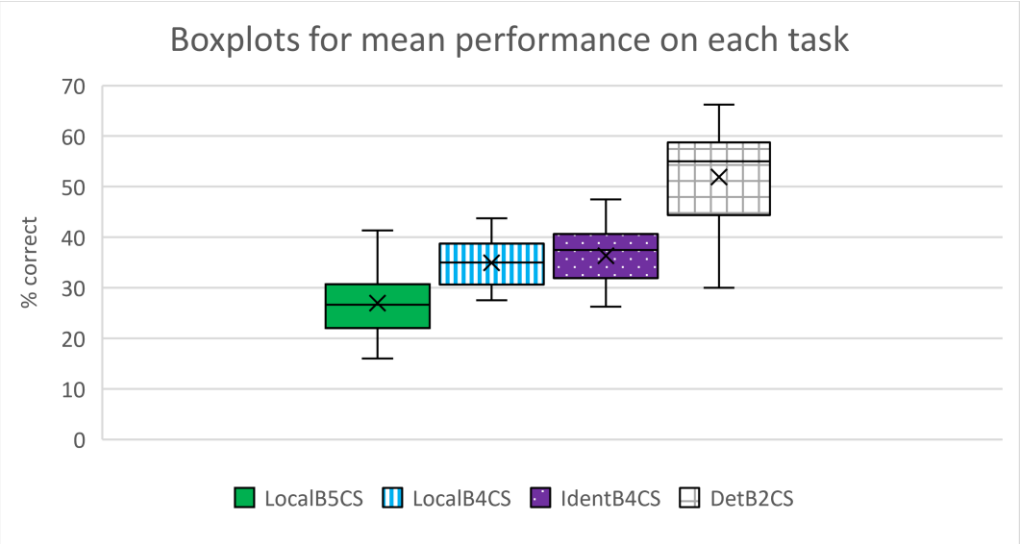


Figure 5.6: Boxplots for mean performance on each task

	LocalB$_{4CS}$	**IdentB$_{4CS}$**	**DetB$_{2CS}$**
LocalB$_{5CS}$	LocalB$_{4CS}$ significantly better, mean difference=8.1 (95% CI, -12.5 to -3.7), p<0.0005	IdentB$_{4CS}$ significantly better, mean difference=9.6 (95% CI, -15.8 to -3.3), p=0.002	DetB$_{2CS}$ significantly better, mean difference=25.0 (95% CI, -33.1 to -16.8), p<0.0005
LocalB$_{4CS}$	x	IdentB$_{4CS}$ non-significantly better, mean difference=1.5 (95% CI, -7.0 to 4.0), p=1	DetB$_{2CS}$ significantly better, mean difference=16.9 (95% CI, -23.4 to -10.2), p<0.0005
IdentB$_{4CS}$	x	x	DetB$_{2CS}$ significantly better, mean difference=15.4 (95% CI, 9.0 to 21.8), p<0.0005

Table 5.4: Table showing results from one-way repeated measures ANOVA

5.4.3 d' analysis

The data underwent d' analysis, as described in Chapter 3. The d' scores tell us about how easy the participants found the tasks, taking into account the number of response options. The calculations were carried out for the detection data, and conversion tables were used for all of the other data. The results from this are shown in Figure 5.7.

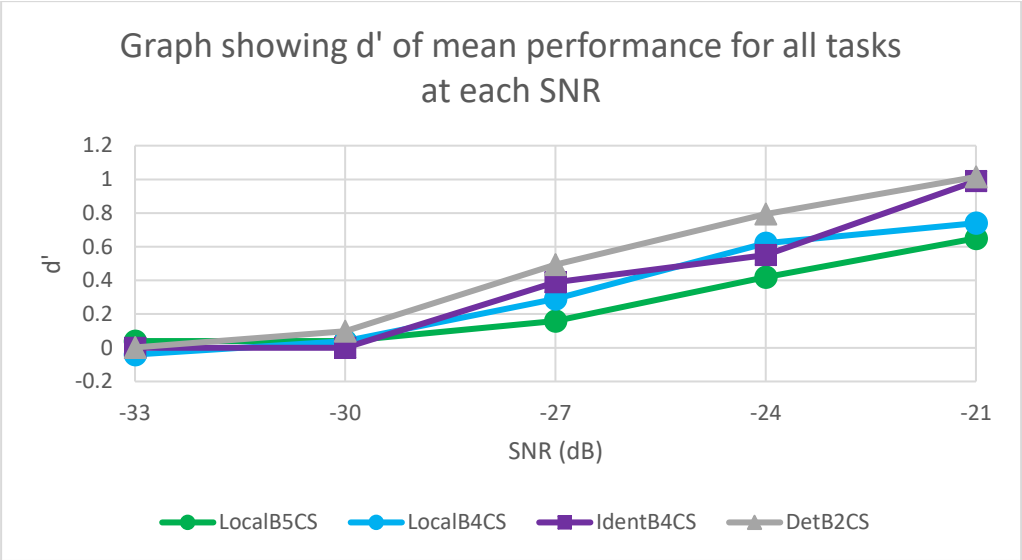


Figure 5.7: Graph showing d' of mean performance for all tasks at each SNR

Results overall are quite similar to the raw percent correct data. For all tasks it appears that performance is similar at the two most challenging SNRs, but then increases sharply for the easier SNRs. The relationships between the task performances also seem similar, as $DetB_{2CS}$ was the task participants performed best in, followed by $IdentB_{4CS}$ and $LocalB_{4CS}$ (which were similar) and then performance was worst in $LocalB_{5CS}$. The main difference appears to be that following the d' analysis the $DetB_{2CS}$ and $IdentB_{4CS}$ scores seem to converge at the easiest SNR (-21 dB), whereas in the raw data performance is higher on the $DetB_{2CS}$ task than the $IdentB_{4CS}$ task at this SNR.

The results are summarised in Table 5.5. For the two hardest SNRs (-33 and -30 dB) the d' values are all around 0, which tell us that the participants were performing at chance. At the -27 dB SNR the d' score for $DetB_{2CS}$ is 0.49, which means the participants were doing better than chance but had not yet reached threshold. For $IdentB_{4CS}$, $LocalB_{4CS}$ and $LocalB_{5CS}$ the d' scores at -27 dB SNR indicate that the participants are still performing at chance levels. At

the -24 dB SNR the d' scores tell us that performance on LocalB$_{4CS}$, DetB$_{2CS}$ and IdentB$_{4CS}$ are all above chance but below threshold. At this SNR LocalB$_{5CS}$ is still at chance level. At -21 dB SNR (the easiest SNR) the d' for detection and identification tell us that the participants are performing at threshold, as the d' scores are around 1. At this SNR the d' scores for LocalB$_{4CS}$ and LocalB$_{5CS}$ both suggested that participants are performing better than at chance but have not yet reached threshold. The criterion from the DetB$_{2CS}$ increase as the SNR increases, meaning that for the more challenging SNRs participants are less likely to have false alarms.

SNR (dB)	LocalB$_{5CS}$ (% correct)	LocalB$_{4CS}$ (% correct)	IdentB$_{4CS}$ (% correct)	DetB$_{2CS}$ (% correct)
-33	Chance performance	Chance performance	Chance performance	Chance performance (criterion -0.2)
-30	Chance performance	Chance performance	Chance performance	Chance performance (criterion -0.1)
-27	Chance performance	Chance performance	Chance performance	Above chance but below threshold (criterion 0.3)
-24	Chance performance	Above chance but below threshold	Above chance but below threshold	Above chance but below threshold (criterion 0.6)
-21	Above chance but below threshold	Above chance but below threshold	Threshold	Threshold (criterion 0.8)

Table 5.5: Table showing meaning of d' scores for all tasks at each level

Following this, curve fits using the mean participant data were carried out for each of the tasks, these are shown in Figure 5.8, Figure 5.9, Figure 5.10 and Figure 5.11 (for DetB$_{2CS}$, IdentB$_{4CS}$, LocalB$_{4CS}$ and LocalB$_{5CS}$ respectively). Figure 5.12 shows the curve fits for all tasks on the same graph. Curves were fitted to

Equation 5.1 using code written by Michael Akeroyd as part of the laboratory statistical software. Table 5.8 shows the R^2 values for these curves.

$$d' = m * power^k$$

Equation 5.1: Equation for curve fits (McFadden, 1966)

From the graphs it can be seen that the curve fits follow the same patterns shown above for the raw d' analysis, which would be expected given that the same data is being used. Performance is best in $DetB_{2CS}$, followed by $IdentB_{4CS}$ and $LocalB_{4CS}$ (which are very similar, although $IdentB_{4CS}$ is better at -21 dB SNR) and then is worst in $LocalB_{5CS}$. The m and k values for these curve fits are shown in Table 5.6. The k values indicate how steep the curve is (McFadden, 1966) and the m values indicate how hard the participants found the task (when compared to curves with the same k values).

As these results were different to those found in Experiment 2, where participants scored highest in $IdentB_{4CS}$, followed by $DetB_{2CS}$ and $LocalB_{5CS}$ the curves were extrapolated to easier SNRs (with d' limited to a maximum of 3 as this indicates when the participants found the task very easy). Figure 5.13 shows the mean data and the curve fits extrapolated to -9 dB SNR. From this it can be seen that before reaching a d' of at least 3, the order of task performance is $IdentB_{4CS}$, $DetB_{2CS}$, $LocalB_{5CS}$ and then $LocalB_{4CS}$. This is similar to the results from Experiment 2, where participants scored highest on the identification task.

Next, the curve fits were carried out with the k value fixed at 1.0 (as done by McFadden (1966)). The m values for these fits are shown in Table 5.7. Figure

5.14 shows these curve fits extrapolated to -9 dB SNR. From this it can be seen that before reaching a d' of at least 3, the order of task performance is $DetB_{2CS}$, $IdentB_{4CS}$, $LocalB_{4CS}$, $LocalB_{5CS}$. This order of performances is more consistent with the -33 to -21 dB SNR data patterns than the one seen in the extrapolated data when k was not fixed (shown in Figure 5.13).

Task	$DetB_{2CS}$	$IdentB_{4CS}$	$LocalB_{4CS}$	$LocalB_{5CS}$
m	39	208	56	198
K	0.7	1.1	0.9	1.2

Table 5.6: Table showing m and k values from curve fit

Task	$DetB_{2CS}$	$IdentB_{4CS}$	$LocalB_{4CS}$	$LocalB_{5CS}$
m	146	130	107	76

Table 5.7: Table showing m values when k is fixed at one

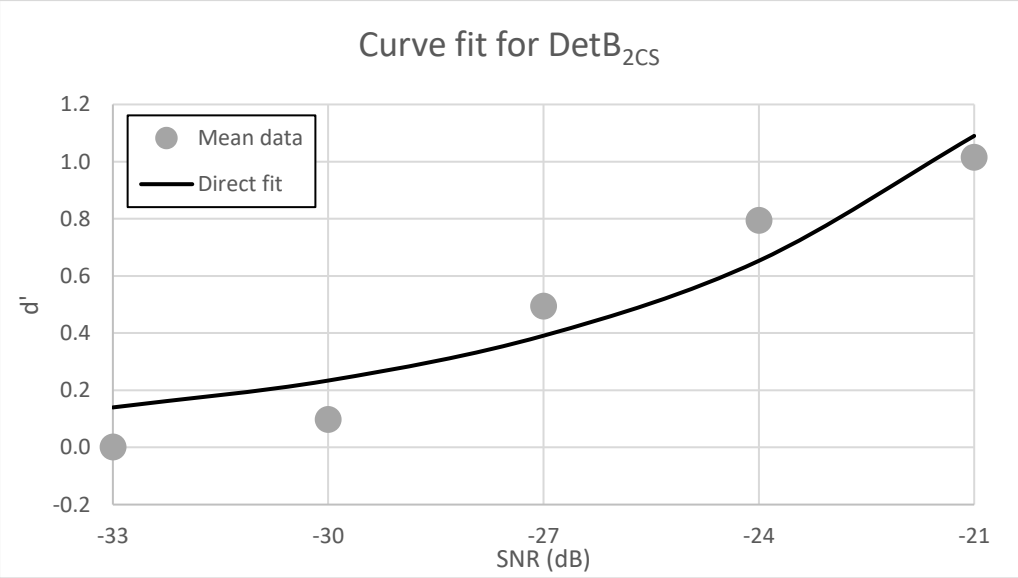


Figure 5.8: Figure showing curve fit for the mean data from the $DetB_{2CS}$ task

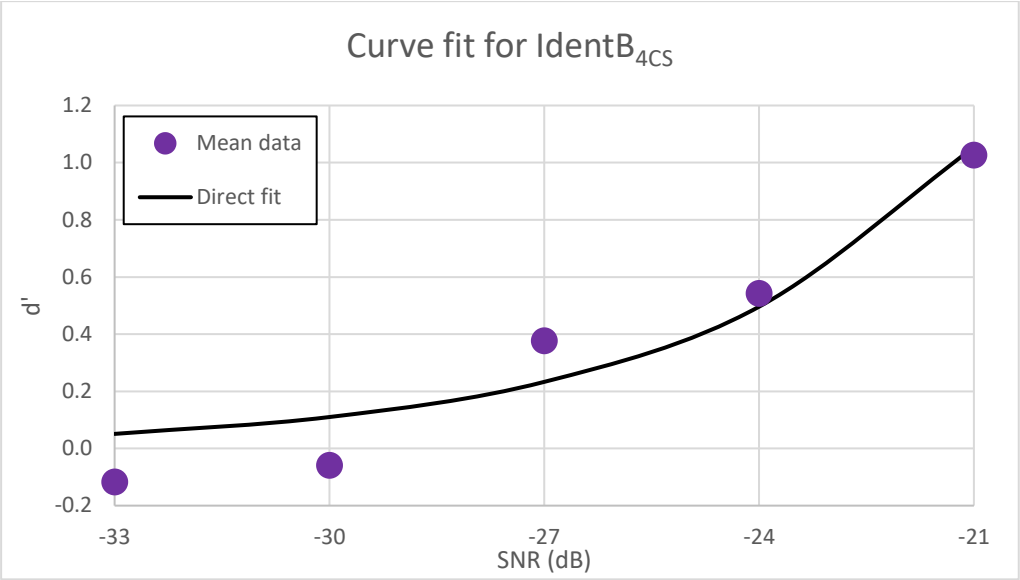


Figure 5.9: Figure showing curve fit for the mean data from the IdentB$_{4CS}$ task

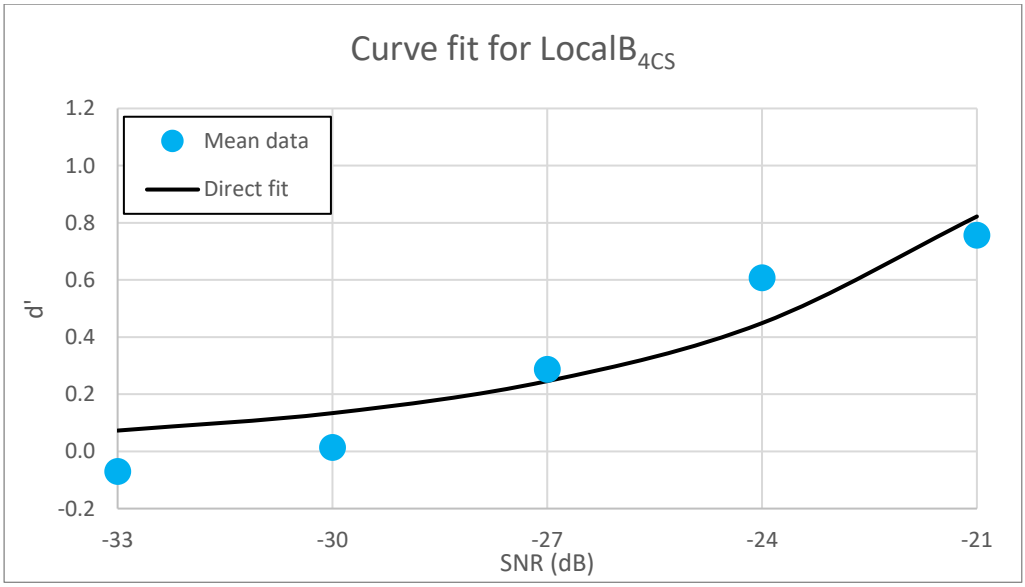


Figure 5.10: Figure showing curve fit for the mean data from the LocalB$_{4CS}$ task

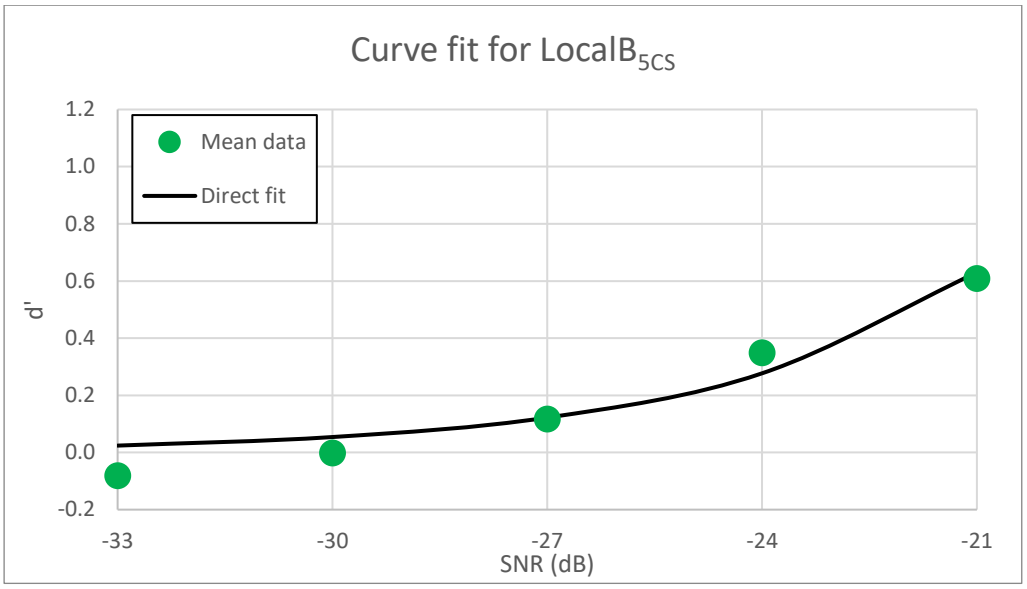


Figure 5.11: Figure showing curve fit for the mean data from the LocalB$_{5CS}$ task

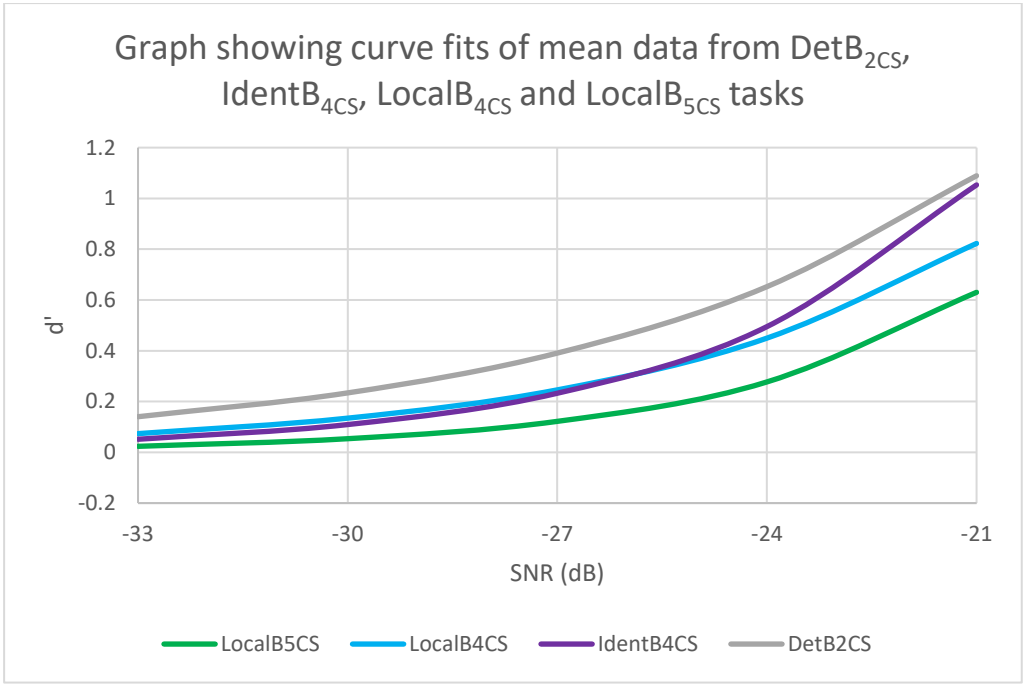


Figure 5.12: Graph showing curve fits to mean data for all tasks

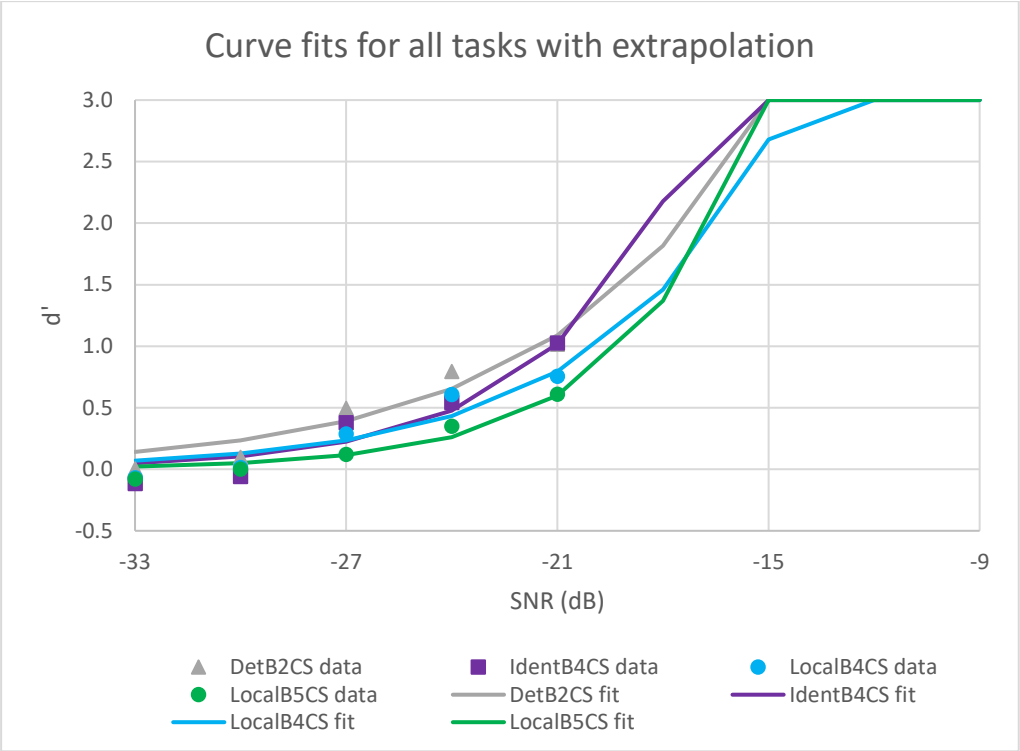


Figure 5.13: Graph showing curve fits for all tasks with extrapolation

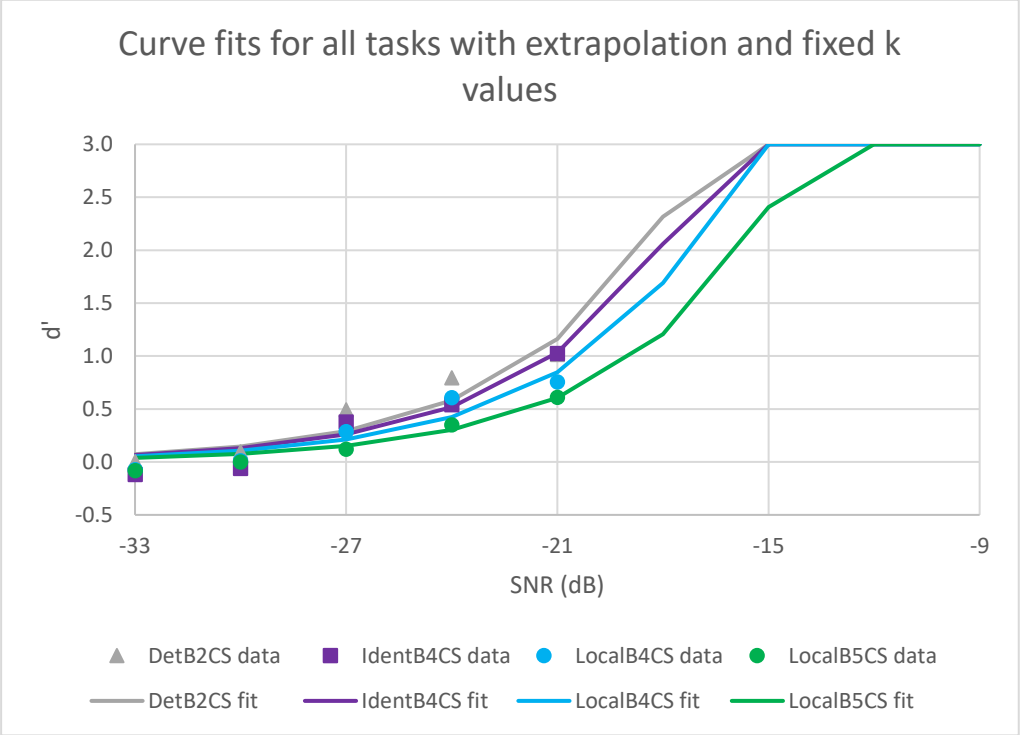


Figure 5.14: Graph showing curve fits for all tasks with extrapolation (k=1 for all tasks)

	R^2 for direct fit (%)	R^2 for bootstrap fit (%)
$DetB_{2CS}$	90	90
$IdentB_{4CS}$	91	90
$LocalB_{4CS}$	87	86
$LocalB_{5CS}$	94	93

Table 5.8: Table showing R^2 values for direct fit and bootstrap fit curves

Next, a bootstrap analysis was carried out on the d' scores of all participants in order to fit curves (and give estimates of confidence intervals for the parameters of the curve) to the data using Equation 5.1 with code written by Michael Akeroyd as part of the laboratory statistical software. This was done in the same way as described in Chapter 3.

The m and k values for each task are shown in Table 5.9. The curves are plotted in Figure 5.15. It can be seen that performance is highest in the $DetB_{2CS}$ task, followed by the $IdentB_{4CS}$ task and then worst in the $LocalB_{4CS}$ and $LocalB_{5CS}$ tasks (which appear very similar). Next, these curves were extrapolated up to a SNR of -15 dB, shown in Figure 5.16. In this graph it can been seen that at the easier SNRs performance is highest in the $IdentB_{4CS}$ task, which is the same trend as was seen in Experiment 2. Table 5.8 shows the R^2 values of these curves, which are very similar/identical to the R^2 values from the direct fit curves.

Figure 5.17 shows the k values from the bootstrap analysis with 95% confidence intervals. From this it can be seen that the confidence intervals for the k values are fairly wide, suggesting that the curves may all have very similar k values. In Table 5.10 it can be seen that the confidence intervals for $IdentB_{4CS}$ and $LocalB_{5CS}$ are very wide, suggesting that these m estimates are

not reliable. This is likely due to a limitation of the bootstrap analysis code in dealing with negative d' values.

Task	$DetB_{2CS}$	$IdentB_{4CS}$	$LocalB_{4CS}$	$LocalB_{5CS}$
m	40	241	48	228
K	0.7	1.1	0.9	1.2

Table 5.9: Table showing m and k values from the mean data and the bootstrap analysis

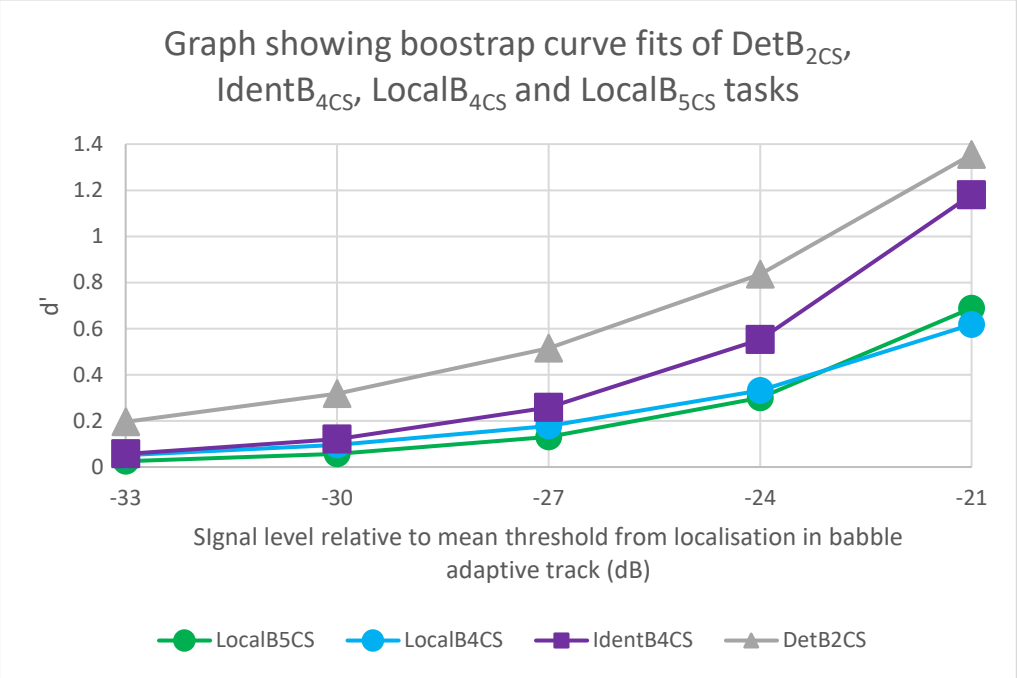


Figure 5.15: Graph showing bootstrap curve fits for all tasks

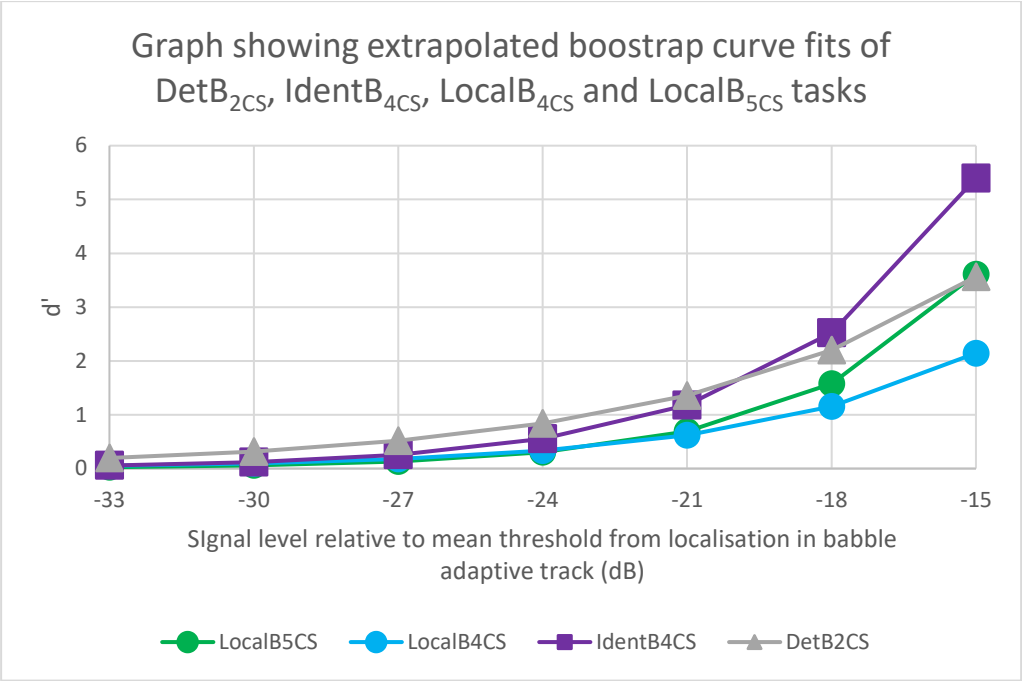


Figure 5.16: Graph showing extrapolated bootstrap curve fits for all tasks

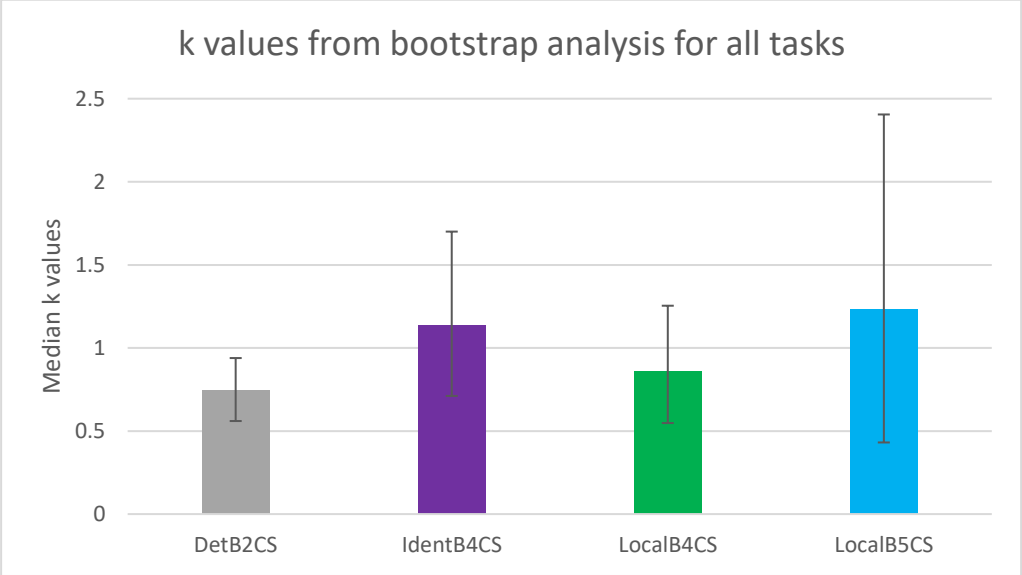


Figure 5.17: Graph showing bootstrap analysis k values with 95% confidence intervals for all tasks

Task	Median	Lower confidence interval	Upper confidence interval
$DetB_{2CS}$	40	14	117
$IdentB_{4CS}$	241	26	4323
$LocalB_{4CS}$	48	10	368
$LocalB_{5CS}$	228	5	99999 (maximum value)

Table 5.10: Table showing bootstrap analysis m values with 95% confidence intervals for all tasks

5.5 Discussion

5.5.1 Comparison to findings from Experiment 2

Although Experiment 2 included child participants and Experiment 3 only involved adults, it is still possible to compare their results. It is also appropriate to compare the two experiments, despite Experiment 2 using relative levels and Experiment 3 using fixed SNRs. This is because the signal levels used in Experiment 2 were personalised for each participant, based on their localisation of speech in babble ability (from the $LocalB_{5AT}$ task). The fixed SNRs used in Experiment 3 were appropriate as all participants were normal hearing adults and their basic localisation ability had been checked and confirmed as typical. Therefore, participants should all have been tested at an appropriate difficulty level.

In the current experiment (Experiment 3) d' analysis showed similar performance for all tasks at the two most challenging SNRs, followed by gradually increasing performance over the other three SNRs. Overall, from the d' analysis, performance appeared to be best on $DetB_{2CS}$ and worse on $LocalB_{5CS}$. In Experiment 2 the corresponding analysis (shown in Figure 5.18) showed that performance was generally highest in $IdentB_{4CS}$, followed by $DetB_{2CS}$ and $LocalB_{5CS}$ which were fairly equal to each other. So for both of these experiments performance was worst on the one interval-five choice $LocalB_{5CS}$ tasks. The biggest difference between the two experiments was that in Experiment 2 the $DetB_{2CS}$ task was one which participants performed very poorly on, whereas in Experiment 3 it was the task on which performance was

highest. This may be due to the different SNRs used (as supported by the bootstrap analysis), or possibly due to the younger children in Experiment 2 finding the $DetB_{2CS}$ task more boring than the other tasks, as it had a longer duration.

Figure 5.18 shows the d' results on graphs for both Experiments 2 and 3. From this it is clear that performance was much worse in Experiment 3 than in Experiment 2. Performance appeared similar on the hardest signal level of Experiment 2 (-9 dB) and the hardest SNR of Experiment 3 (-33 dB), with similar d' values and trends in task performance.

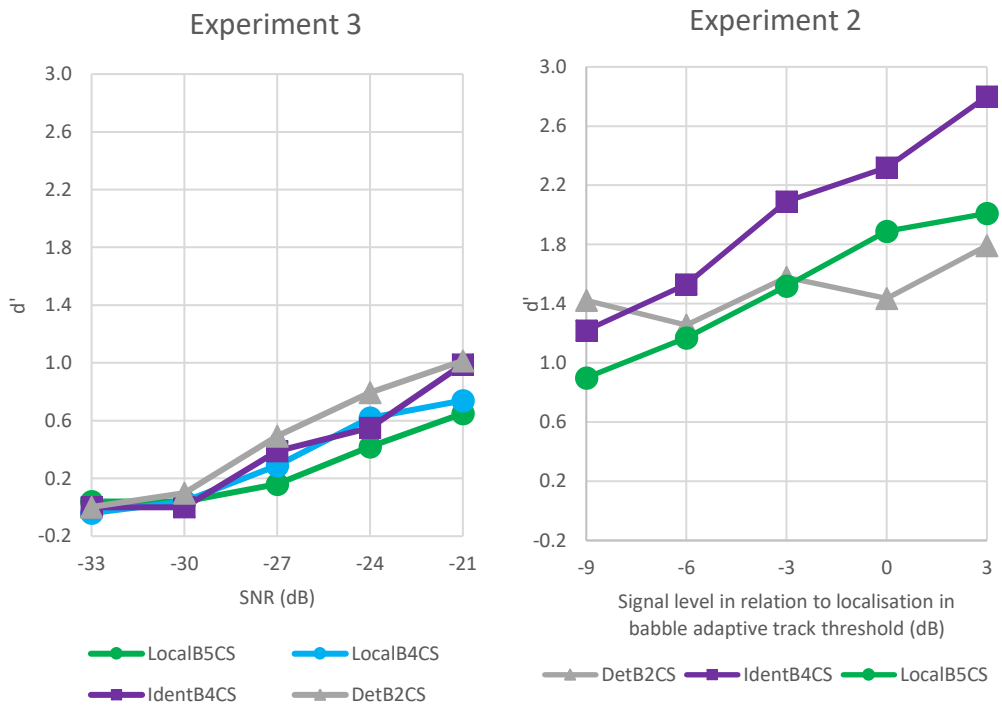


Figure 5.18: Graphs showing d' analysis results for Experiments 2 and 3

5.5.2 Summary and conclusions

For all tasks performance appeared to be at floor at the two hardest SNRs (-33 and -30 dB), with performance then increasing for the following SNRs. The general pattern of the data was that performance was highest on $DetB_{2CS}$, followed by $IdentB_{4CS}$ and $LocalB_{4CS}$ (which were similar to each other) and then $LocalB_{5CS}$ was the task in which performance was worst. A one-way repeated measures ANOVA showed that performance was significantly better on the $LocalB_{4CS}$ task than the $LocalB_{5CS}$ task. Performance on the $IdentB_{4CS}$ task was significantly better than the $LocalB_{5CS}$ task, but not the $LocalB_{4CS}$ task. Performance on the $DetB_{2CS}$ task was significantly better than $IdentB_{4CS}$, $LocalB_{4CS}$ and $LocalB_{5CS}$ tasks. Extrapolation of the curve fits (with and without fixed k values) gave different results, suggesting that the extrapolated curve fits may not be reliable.

It was hypothesised that participants would perform better on the $LocalB_{4CS}$ than the $LocalB_{5CS}$ task. This was confirmed with a one-way repeated measures ANOVA (difference=8.1, $p<0.005$). It was expected that any difference between the $IdentB_{4CS}$ and $LocalB_{4CS}$ tasks would be smaller than any difference in performance between the $IdentB_{4CS}$ and $LocalB_{5CS}$ tasks. This was also confirmed by the one-way repeated measures ANOVA, which found a significant difference of 9.6 ($p=0.002$) between the $IdentB_{5CS}$ and $LocalB_{4CS}$ tasks, and a non-significant difference of 1.5 ($p=1$) between the $IdentB_{4CS}$ and $LocalB_{5CS}$ tasks.

This experiment also aimed to explore the difference between task performances at more challenging SNRs than those used in Experiment 2. Raw data from these tasks showed that performance was highest on $DetB_{2CS}$, followed by $IdentB_{4CS}$ and $LocalB_{4CS}$ (which were similar to each other) and then $LocalB_{5CS}$ was the task in which performance was worst. A one-way repeated measures ANOVA showed that performance on the $LocalB_{4CS}$ was significantly better than on the $LocalB_{5CS}$. $LocalB_{4CS}$ and $IdentB_{4CS}$ were not significantly different and $DetB_{2CS}$ was significantly better than $IdentB_{4CS}$, $LocalB_{4CS}$ and $LocalB_{5CS}$.

An analysis of d' showed similar results, as $DetB_{2CS}$ was the task participants perform best in, followed by $IdentB_{4CS}$ and $LocalB_{4CS}$ (which were similar) and then performance was worst in $LocalB_{5CS}$. The d' analysis also showed that performance appeared to be at floor for the hardest two SNRs, performance gradually improved for all tasks for the following three SNRs. Comparing the d' results for Experiments 2 and 3 highlights a number of similarities, such as performance being worst on the $LocalB_{5CS}$ task, but also some differences. In Experiment 2 the $DetB_{2CS}$ task was one which participants performed very poorly on, whereas in Experiment 3 it was the task on which performance was highest. Although in both experiments performance was worst on the $LocalB_{5CS}$ tasks.

The key findings of this experiment are that performance in all of the tasks were at floor until at least -27 dB SNR, and that threshold was only reached for $IdentB_{4CS}$ and $DetB_{2CS}$ at -21 dB SNR (and was not reached even at this SNR

for $LocalB_{4CS}$ and $LocalB_{5CS}$). Therefore, the results from this experiment suggest that listeners are better at detection than identification and localisation, and that even if a listener can detect and possibly identify a sound they cannot necessarily localise it. This is in conflict with Experiment 2, which found that performance was better on identification than detection, however this may have been due to the inclusion of children and their more limited attention spans in Experiment 2. The results from the current experiment are also in conflict with the findings of Hawkins and Stevens (1950) and Liu and Jin (2019), who found that participants performed better on identification than detection of masked speech tasks. However, neither of these studies used informational maskers so this may affect their results and make it difficult to compare to the results from this experiment. Hawley, Litosvky and Colburn (1999) previously investigated the relationship between localisation and identification of speech in babble. However, due to the large separation angles used (minimum of 30°) participants performed at ceiling on the localisation task. The experiment described in this chapter builds on this previous work as smaller separation angles were used (15°) so comparisons were successfully made between performance on localisation and identification of speech in babble tasks. Overall, the findings of this experiment are novel as it is the only experiment to have investigated the differences between detection, identification and localisation of speech in babble.

5.5.3 Limitations and future work

This experiment aimed to use challenging SNRs, however it may have been beneficial to have carried out more pilot testing to further refine the choice of SNRs. Looking at the data, it may have been better to have tested -18 dB (3 dB easier than the easiest SNR actually used) rather than -33 dB (the hardest SNR tested). Using -18 dB SNR may have meant that participants reached threshold for all tasks, which was not the case with the SNRs actually used. However, it was useful to see that participants appeared to be performing at floor at -33 and -30 dB, which may not have been clear if -33 dB had not been tested. As is generally the case, testing a greater number of SNRs with more trials would have improved the experiment. However, due to time restrictions and the difficulty of recruiting participants for long studies or studies which require multiple sessions, a compromise had to be made by limiting the number of SNRs and trials used.

Chapter 6

Experiment 4: Minimum Audible Angle in Babble in Children and Young Adults

6.1 Introduction

This experiment was carried out in order to expand on the findings of Experiments 1 and 2, to further understanding of the developmental time course of the development of spatial hearing in babble in children. This experiment used a Minimum Audible Angle (MAA) task, with an adaptive track, which altered the target separation angles rather than the signal level (as was done in the LocalB$_{5AT}$ task in Experiments 1 and 2).

The MAA paradigm was first introduced by Mills (1958) and is now a commonly used source discrimination method for measuring localisation ability (Akeroyd and Whitmer, 2016). It aims to find the smallest angle between two sound sources which a listener can detect (separation angle) in the horizontal plane. Two signals are played consecutively from different locations and the listener is asked to identify the order in which they occurred. By varying the separation angle between the two sound sources the MAA can be found. The value of MAAs varies with signal frequency and with the locations of the sound sources relative to the listener (Mills, 1958). The

smallest expected MAA for an adult listener in a quiet environment is approximately 1°, with a signal of 500-750 Hz that is located directly infront of the listener.

Studies investigating MAAs in both adults and in children have so far only been carried out in quiet conditions. In the developmental literature some conflict is present regarding the age at which children have adult-like MAAs. Litovsky (1997) used wideband noise bursts (0.5-8.5 kHz) to investigate MAAs in both adults and children and found that although some five year olds had MAAs within the same range as adults, others could not yet perform at this level. The mean MAA for adults was 0.78° and for 5 year olds was 1.55°. The authors suggest that the variance within the group of five year olds could be due to attention, or this could be the age at which adult-like performance can occur for some children but is just prior to the age at which development to an adult-like level occurs for others. Kühnle et al. (2013) carried out a MAA in quiet study using Gaussian noise bursts (0.3-1.2 kHz or 2-8 kHz), which demonstrated age effects between children aged 6-7 and 13-18 years of age.

The results from these two studies suggest that MAA in quiet in children does not reach an adult-like level until sometime between 5 and 13 years of age depending on the child, although most likely towards the younger end of this range. No study has yet investigated MAA in background sound in children.

The age at which adult-like performance is reached on this task will be estimated using three of the same methods described in Chapter 4 (broken stick functions and exponential functions (direct fit) and (bootstrap)), the

choice of which was made based on a brief review of similar literature (described in Chapter 4).

Mapping the developmental time course of MAA in babble in children will contribute to the knowledge of spatial hearing in background sound in children, an area where there is currently little published research. Understanding more about the development of spatial hearing in background sound in children is important and this knowledge could be used for investigating the spatial hearing in background sound abilities of children with suspected spatial hearing deficits, such as those with a hearing loss or Auditory Processing Disorder. It is possible that in the future an MAA in babble task over headphones could be used as an assessment tool for these patient groups.

6.2 Aims and Hypotheses

The aim of this experiment was to further understand the developmental time course of spatial hearing in babble, by measuring the MAA in babble in participants aged 6 to 21 years of age. The MAA task was carried out at four different SNRs in order to vary the difficulty of the task. It was hypothesised that performance on the tasks would improve with age and follow a similar developmental time course to the one found in Chapter 4.

The additional tasks used in Experiments 1 and 2 were not included in this experiment. No significant correlations were found between the results of the additional tasks and localisation of speech in babble performance (except for

the LocalQ$_{5CS}$ task), so it was decided that they were not the best use of testing time.

6.3 Methods

6.3.1 Participants

Many of the participants for this experiment had also taken part in Experiments 1 or 2. Parents of all participants whose data were not excluded from analysis were contacted and asked if their child would like to participate in another experiment. Many of them responded positively and participated in Experiment 4, some bringing friends or siblings to participate who had not been in the previous experiments. More participants were recruited via Facebook.

As in Experiments 1 and 2, participants with a hearing loss or learning or language disorders were not recruited for this experiment as it was thought these conditions could affect their performance on the task. As the purpose of this experiment was to investigate performance in typically developing children the presence of any learning/language disorders or hearing impairment would have introduced confounds.

6.3.1.1 Adults

Five adults (four female, one male) aged 18 to 21 years (mean 19.6, s.d. 1.0), who were all English speakers from birth with normal hearing and no language or learning difficulties took part in the experiment. They were

recruited through posters placed around the University of Nottingham and posts on the University of Nottingham Students Facebook pages.

6.3.1.2 Children

45 children (18 female, 27 male) aged 6 to 17 years (mean age 11.0 years, s.d. 3.3) completed the experiment. The number of participants of each age is shown in Table 6.1. All of the children spoke English since birth, had normal hearing and no confirmed language or learning difficulties. An additional six year old child was recruited for the experiment but no data was collected as they were unable to complete the tasks (they are not included in the total of 45 children).

Age (years)	Number of participants
6	5
7	4
8	4
9	4
10	4
11	3
12	4
13	4
14	4
15	4
16	4
17	1
18-21	5

Table 6.1: Table showing number of participants of each age

6.3.2 Equipment

The same test room and equipment was used as in Experiment 3. The MATLAB code used to run the experiment was written by Michael Akeroyd as part of the overall programming of the laboratory.

6.3.3 Stimuli

6.3.3.1 Accuracy of the Panning

For this experiment pairs of stimuli need to be presented with separation angles of less than 15°, which is less than the separation of each loudspeaker. To do this, the laboratory software used a panning method to present sounds at intermediate angles. It is important to check that the presentation of such angles is accurate.

The panning method used was an "equal power pan", implemented in software written by Owen Brimijoin and Michael Akeroyd as part of the overall laboratory software. To generate an intermediate direction, a signal was played simultaneously from adjacent loudspeakers, but with the individual amplitudes set by the functions shown in the left panel of Figure 6.1. Their powers (i.e. amplitude squared) then sum to the same value across all intermediate angles (right panel off Figure 6.1).

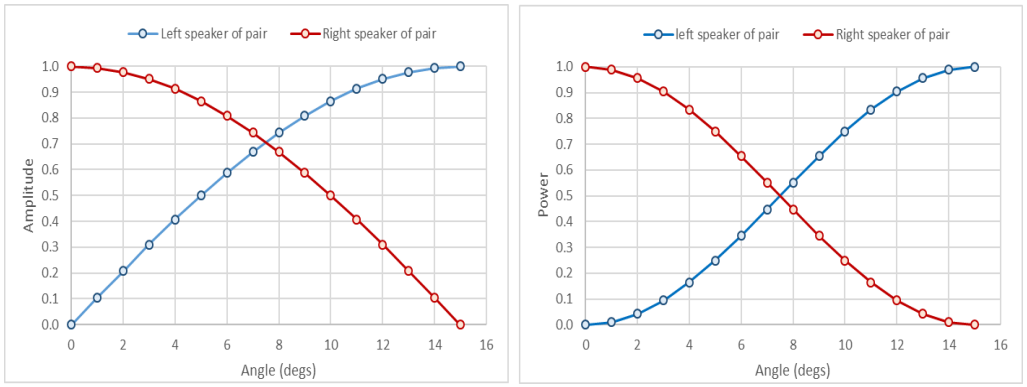


Figure 6.1: Graphs showing amplitude (left) and power (right) of speakers

To check the method's accuracy, I recorded signals at the position of a listener's head, then analysed their binaural characteristics. A manikin/head and torso simulator (Bruel & Kjaer, Type 4100D), which has microphones

placed to take measurements which are representative of how a person hears binaurally, was used. This was placed on a chair in the middle of the speaker ring, connected to a multi-track recorder (TASCAM DR-680MKII). Signals were played using MATLAB. The stimulus was pink noise, presented for a duration of 20 s with intervals of 1 s between each presentation. A sample frequency of 192 kHz was used for recording. I made measurements with and without the table and computer monitor (which was positioned in the same way for participant testing) present, shown in Figure 6.2 and Figure 6.3 respectively. Signals were presented at azimuths of -60° to + 60° at 1° intervals (so -60°, -59°, -58° etc).

I then transferred the recordings from the multi-track recorder, split them using Audacity and analysed them using MATLAB (with code written by Michael Akeroyd as part of the overall programming of the laboratory). This first applied an A-weighting filter, then worked out binaural parameters across a sequence of windows (100-ms “Hanning”). The ITD was defined as the internal delay that gave the largest value of the left-right cross-correlation. The ILD was defined as the difference between the left and right signal powers in decibels.

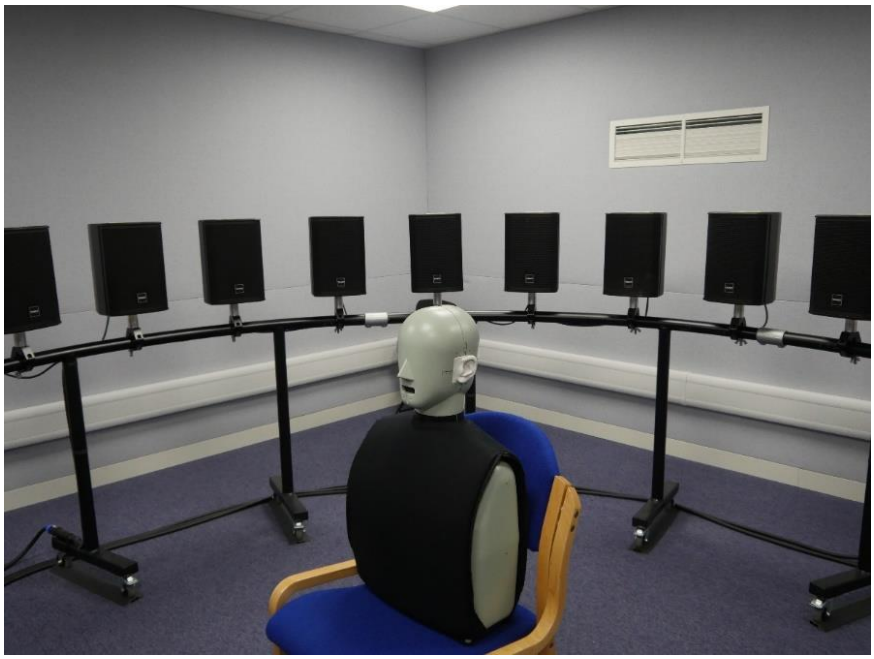


Figure 6.2: Photograph showing testing set up without table and monitor

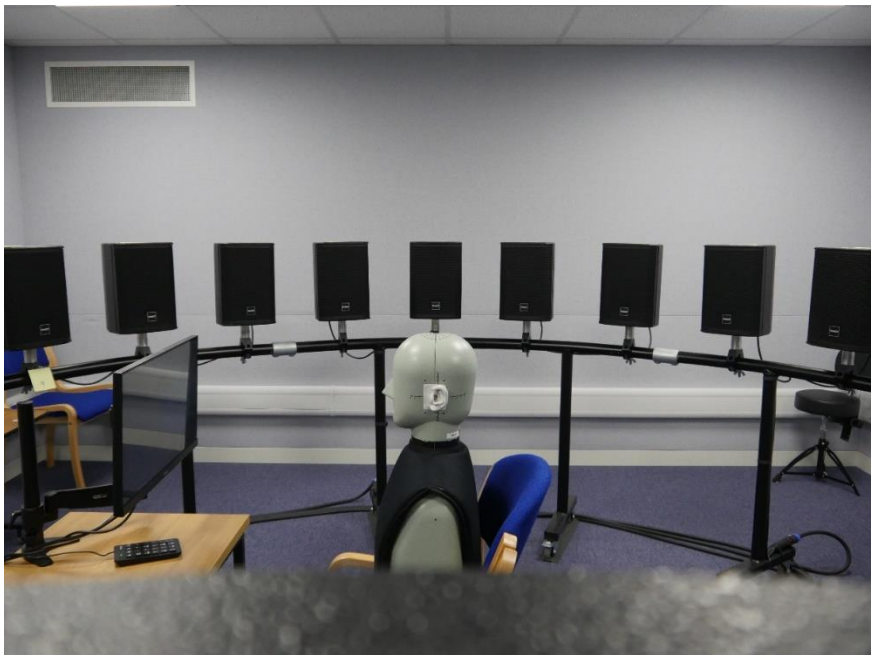


Figure 6.3: Photograph showing testing set up with table and monitor

If the panning worked perfectly, then it would give ITDs and ILDs indistinguishable from what would have been found had there actually been loudspeakers physically at the intermediate angles. However, this could not be tested directly as the loudspeakers were not in these positions. Instead an

indirect method was chosen to check that the ITDs and ILDs acted as would be expected: that the magnitude of the ITDs would be largest at ±60° and, if the panning was operating properly, the ITDs would decrease essentially linearly to a minimum at 0° (Feddersen et al., 1957). A similar trend would be expected for the ILDs, however results are likely to be influenced by factors such as the head shadow effect and other related filtering effects. It was also expected that for both the ITDs and ILDs further influences would likely come from room reflections, as the measurements were not made in an anechoic chamber.

The measurements from the recordings are shown in Figure 6.4 (ITDs) and Figure 6.5 (ILDs). It can be seen that both ILD graphs are very similar to each other and both ITD graphs are also similar to each other. This suggests that the impact of the computer monitor and desk are minimal, therefore further analysis will only discuss the recordings where the table and monitor were present. Overall, the ITD graphs show that the function is fairly linear (R^2 values of 99.92% with table and monitor, 99.88% without table and monitor). This is as expected and shows that, overall, the panning is working well. It is noteworthy that no point is out-of-order, so the rank ordering of each direction is correct. The ILD graphs have an overall positive trend (R^2 values of 97.42% with table and monitor, 97.54% without table and monitor), but the functions contain a visible “ripple”. This cycles every 15°, which is the separation of the loudspeakers, and so suggests that the panning software may not have been working quite as expected.

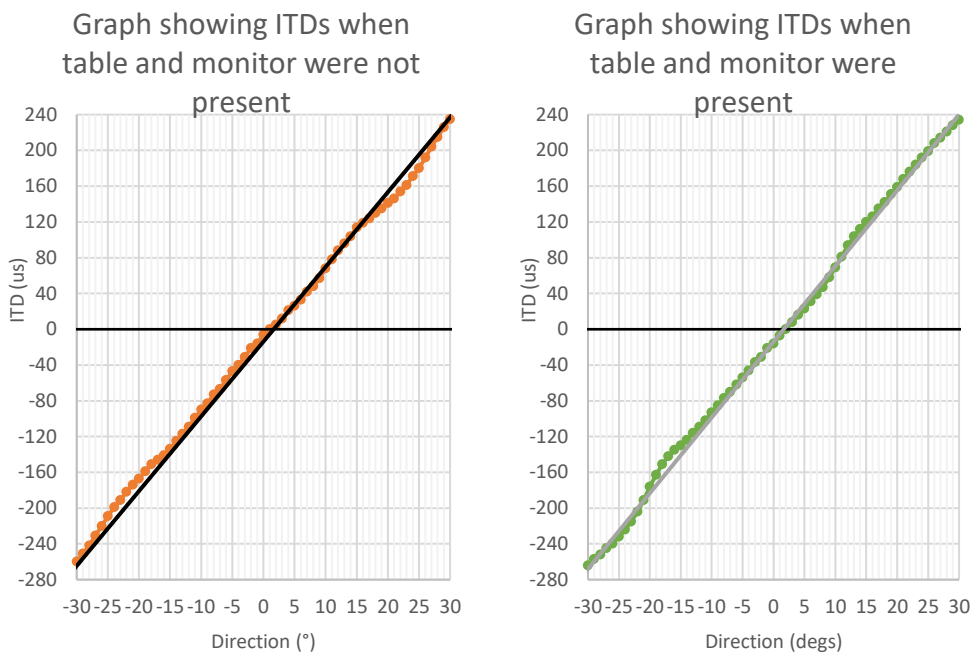


Figure 6.4: Figure showing ITDs when table and monitor were present and when they were not present

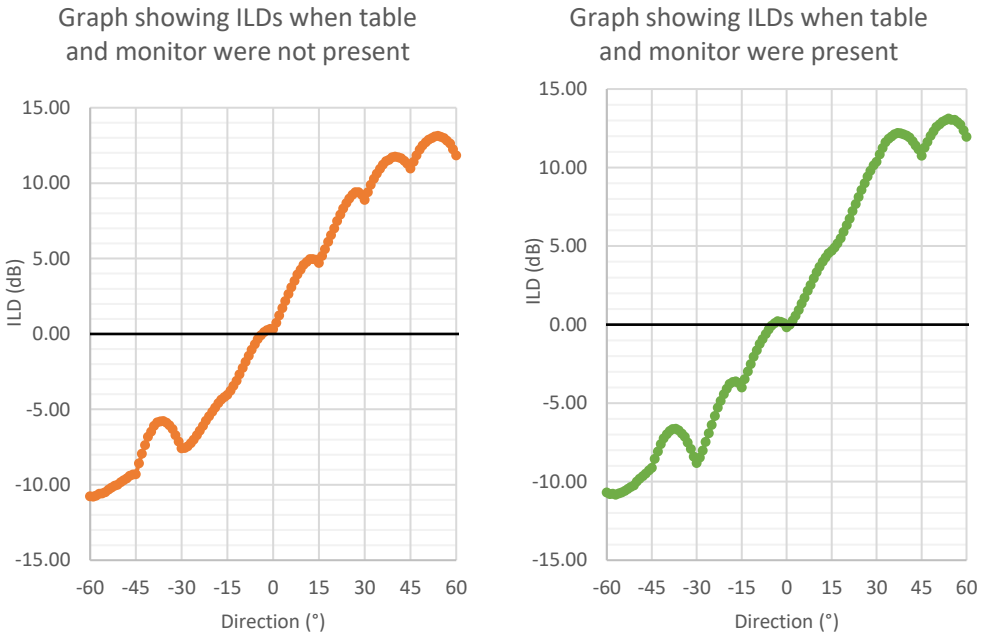


Figure 6.5: Figure showing ILDs when table and monitor were present and when they were not present

To look at these in further detail, Figure 6.6 shows the ITD and ILD graphs for the range of -15° to 15°. This range is the most important as this is where the majority of MAAs are found. These graphs show that in this region the

panning is very accurate in terms of the ITDs (R^2 values of 99.62%). The ITD increased linearly as the angle increased, giving a unique ITD for each direction presented. To determine if the ripple was likely to be noticeable, I compared them to the just noticeable difference (JND) for ITDs. For broadband noise, the JND for ITDs are at least 10μs (Klumpp and Eady, 1956). In my measurements the greatest distance of a data point from the line of best fit was 9.1μs (in the -15° to 15° range, shown in Figure 6.7). Thus, for ITD, the panning is accurate to within a just-noticeable difference.

However, for ILDs, although the ILD graphs for this range of directions (-15° to 15°, shown in Figure 6.6) shows a similar overall trend to that of the ITD graphs, the ILD values around 0° are not as accurate as those on the ITD graphs (R^2 values of 95.68%). Nevertheless, there are no data points further than 1.0 dB from the line of best fit (shown in Figure 6.7). For ILDs the just noticeable difference is about 0.5 to 1 dB (Mills, 1960), and so again the panning appears accurate to within a JND.

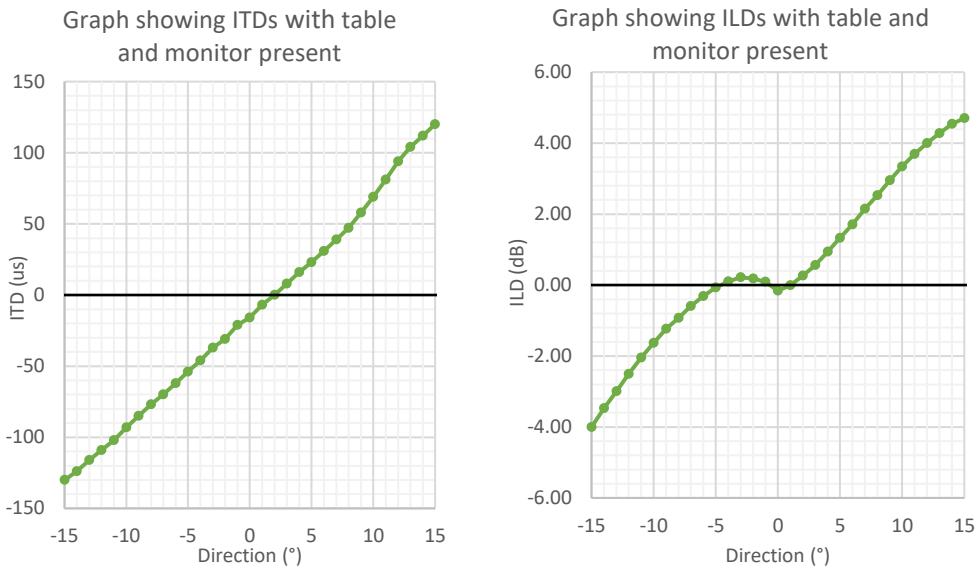


Figure 6.6: Figure showing ITDs and ILDs for the directions -15° to 15°

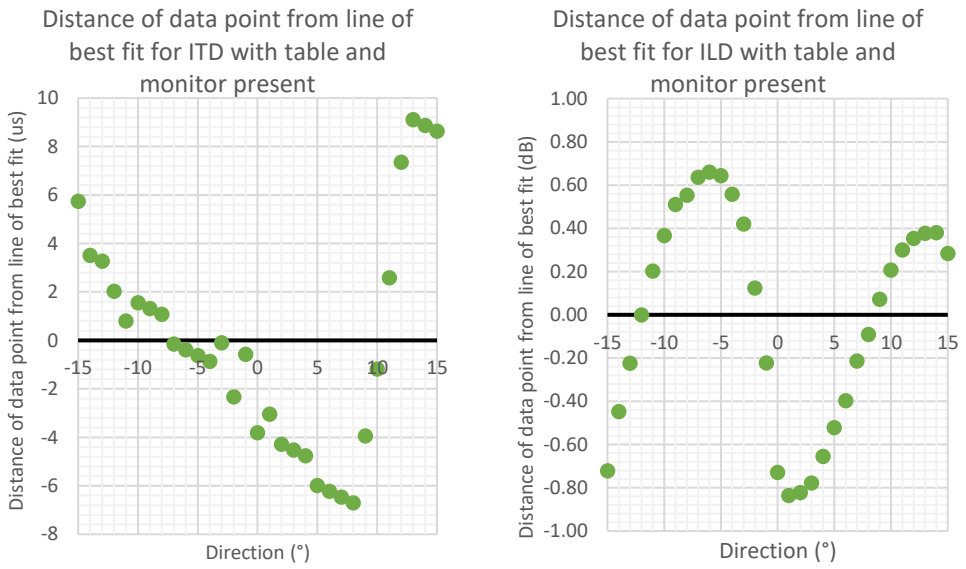


Figure 6.7: Distance of data point from line of best fit for ITDs and ILDs with table and monitor present

Subsequently, further recordings were made by Michael Akeroyd and Jennifer Firth. Measurements of signal power were made using a single microphone in the centre of the speaker ring, therefore avoiding the complications introduced by head shadow and ears. They found that the ripple was also present in these new recordings. It appears that the ripple may be a result of

the calculation of amplitude of the signals from the adjacent loudspeakers. A likely explanation is that the signals were assumed to be incoherent and so their powers add together. However, this is not the case as the signals are the same and in phase when equidistant to both loudspeakers. As the signals are the same, their amplitudes add and thus their combined power is greater, causing a ripple. An illustration of this can be seen in recordings made by Michael Akeroyd and Jennifer Firth, using a single microphone in the middle of the speaker ring. Figure 6.8 shows the total power received at the microphone. The ripple is there, so it is an effect in overall level, not ILD on its own.

Though not perfect, it is interesting that this data shows that the ripple in overall level could not be used to distinguish directions (left versus right). Within the range of -15° to 15°, the level at a left direction (blue in Figure 6.8) was within 1 dB of the level from the right direction (red in Figure 6.8). This means that they are within a JND of each other, and therefore unlikely to be noticeable, meaning that participants could not have used the ripple to help them in the task.

Overall, these measurements have shown that the panning is not perfect, but it appears to be accurate enough as its errors are generally less than a JND. The panning is good in the important range of -15° to 15° for the ITDs and is fairly good in this region for the ILDs. It is unlikely that they have affected the results of the experiments in any substantial way.

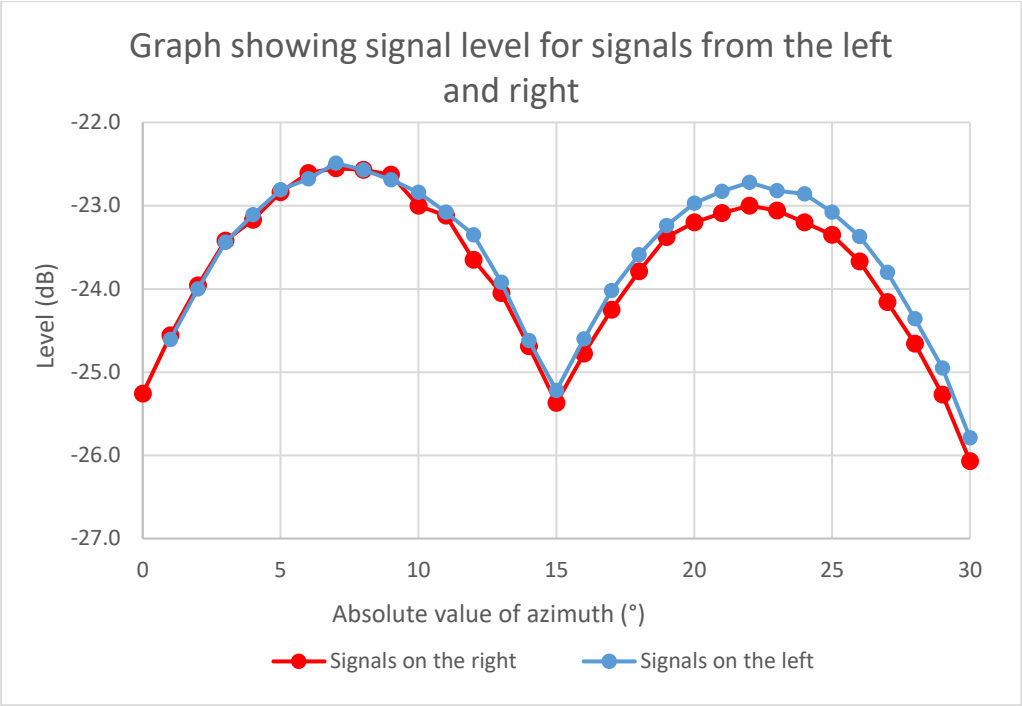


Figure 6.8: Graph showing power sum for signals from the left and right

6.3.3.2 Signal

The signals used were the same single speech words from the CRM sentence corpus as used in Experiments 1, 2 and 3. As in Experiments 2 and 3, only female voiced words were used for the signal.

6.3.3.3 Masker

As in Experiments 1, 2 and 3 the stimuli used were six IEEE sentences and each one was played from a different location in the ring (locations chosen as described previously). As in Experiments 2 and 3, only male voiced sentences were used for the masker.

6.3.4 Procedure

All participants attended a single session in the Hearing Sciences building on the University of Nottingham campus. Sessions lasted approximately 60 minutes for adults and 90 minutes for children, largely due to children taking

more (and longer) breaks than adults. Adult participants were given £7.50 inconvenience allowance and any travel expenses were reimbursed. Children were given £15 in Amazon vouchers to thank them for their time and their parents were reimbursed for any travel expenses. Prior to the experiment beginning consent was obtained.

6.3.4.1 Study tasks

Prior to the study tasks all participants underwent pure tone audiometry, to check that they had normal hearing. The practice and main tasks are explained below and shown in Table 6.2.

Practice task	1 track of SNR 20 dB
Main task	3 tracks of SNR 20 dB
	3 tracks of SNR 10 dB
	3 tracks of SNR 0 dB
	3 tracks of SNR -10 dB

Table 6.2: Table showing study tasks in order

Practice task

Participants completed a single track of the first (and easiest) part of the study. This ensured that they understood the task. After they had completed this, the researcher checked verbally that the participant was happy with the task.

They were instructed that they would hear babble from all around them, which they should ignore. They were told that they would hear two colour words in succession and that one would be from the right hand side and the other from the left hand side of the room. They were asked to think about the order in which the words occurred and to respond using the touchscreen to

indicate if the order was 'right then left' or 'left then right.' A two-down, one-up adaptive track procedure (Levitt, 1971), which is the same used in Experiments 1 and 2), working in log(angle) was used to find the threshold at which the participant achieved 70.7% correct. The angle between the target words was initially about 60°, but then changed depending on the participants responses. When the listener got two correct in a row then the angle between the target words decreased, when they got one wrong the angle increased. The amount by which the angle increased or decreased was dependent on how far through the adaptive track the participant was and the size of the most recently presented angle.

For all presentations prior to the completion of two reversals the step size was two (this was referred to as part one). Following the completion of the first two reversals there were four reversals with a step size of one (referred to as part two). The equations used to calculate the angles and examples for these calculations are shown in Table 6.3. A logarithmic scale was used for these equations because it provided a way of quickly reaching an appropriate difficulty level for the participant (without presenting an excessive number of easy trials which would bore the participant). When the angle chosen was between two speakers (i.e. not a multiple of 15°) then the signal was panned to sound as if it came from that angle. MAA was defined as the angular distance between the two target stimuli (see Figure 6.9).

Part	Equation	Example
Part 1	$Next\ angle = 10^{(\log(previous\ angle) \pm 0.2)}$	Previous angle of 60°, participant gives correct answer twice $Next\ angle = 10^{(\log(60°) - 0.2)}$ $Next\ angle = 37.8°$
Part 2	$Next\ angle = 10^{(\log(previous\ angle) \pm 0.1)}$	Previous angle of 9.5°, participant gives incorrect answer $Next\ angle = 10^{(\log(9.5°) + 0.1)}$ $Next\ angle = 12°$

Table 6.3: Table showing how separation angles were calculated

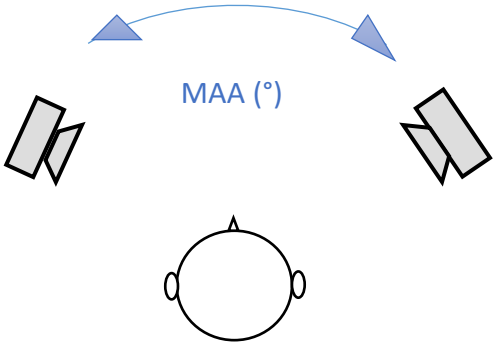


Figure 6.9: Figure showing how MAA was measured

Main tasks

After it had been confirmed that the participant understood the practice task, the main task commenced. This task consisted of the participant completing three interleaved tracks for each of four SNRs (20, 10, 0 and -10 dB). These angles were selected as it was thought that they would provide a good range of difficulty for most of the child participants. By looking at the data from Experiments 1 and 2, it was thought that most children would find the -10 dB condition challenging and the 20 dB condition very easy.

The first set of interleaved presentations was always at the SNR of 20 dB (the same as the practice task). Participants were instructed that the task was the same as the practice, but that it would last longer.

After the participant had completed a set of three interleaved tracks they were offered a break. Before the next set of three interleaved tracks (at a different SNR) they were instructed that the task was the same but the target words would be quieter. For the final and hardest SNR (-10 dB) participants were told they might not always hear the colour words due to the masking babble. Analysis in Chapter 2 found that four to six tracks were ideal for data collection. However, due to time constraints it was decided to make a compromise on the number of tracks for each SNR, in order to test four different SNRs.

Adult participants were offered the opportunity to take a break between tracks but generally did not. Children were strongly encouraged to take a break between each track and generally took breaks ranging from two to ten minutes.

6.4 Results

All adult participants were able to complete the whole experiment, returning a total of 12 thresholds each over the four different conditions.

35 children completed all 12 thresholds over four different conditions. For 10 children between 1 and 11 thresholds were returned, the details of these children are shown in Table 6.4. Sufficient thresholds were returned for 40 children to provide a score for all four conditions. For five children, not enough thresholds were returned for all four conditions, the details of these children are shown in Table 6.5. These two tables show that when the full

number of thresholds were not returned the child was usually one of the youngest of the children tested.

In this chapter, analysis is carried out on each condition, and on the mean of all conditions for each participant. This mean analysis excludes the participants who did not complete at least one threshold for each condition. The data from these participants must be excluded for the mean data as they did not complete the hardest conditions, so their mean score would be high compared to other children of a similar age/ability who did complete the harder conditions.

Number of thresholds returned (maximum of 12)	Frequency	Age of children (years)
1	2	6, 6
6	1	7
7	1	7
8	1	6
9	1	7
10	1	8
11	3	6, 9, 11
12	35	6-17

Table 6.4: Table showing number of thresholds returned in children

Number of conditions with data (maximum of 4)	Frequency	Age of children (years)
1	2	6, 6
2	0	-
3	3	6, 7, 7
4	40	6-17

Table 6.5: Table showing number of conditions with data completed/partially completed by children

The researcher ensured that each child understood the task well, sometimes going over it multiple times or even doing a practice task together if necessary. It is likely that all of the children understood the task but due to the way it was coded were sometimes unable to complete a track and

therefore unable to return a threshold. This was because when a participant got an answer wrong the separation angle increased. If the separation angle reached 90°, was then presented at this angle five times and the participant answered incorrectly then the track stopped. This was done because if the separation angle was greater than 90° then the sounds could be coming from behind the participants so front-back confusions could be introduced. Therefore, if a child had an MAA in babble of 90° or greater then a threshold would not be returned. When children struggled to return thresholds the researcher asked them to complete two tasks which were used in Experiment 2. These were a localisation in quiet and localisation in low background sound tasks. This allowed the researcher to assess if the child had localisation in quiet abilities and to see the impact of a small amount of babble, the results of which are discussed below in section 6.4.9.

6.4.1 Main summary of results

Graphs showing MAA and age for each SNR can be seen in Figure 6.10, Figure 6.11, Figure 6.12, Figure 6.13 and Figure 6.14 (for 20, 10, 0, -10 dB SNR and mean MAA over all conditions respectively). Exponential functions have been fitted (direct fit), using the Solver function in Excel, to each of the graphs. Figure 6.15 shows all of these exponential functions on one graph. From this it can be seen that the performance predicted by the exponential functions is worst in the -10 dB condition (as would be expected), followed by the 0, 10 and 20 dB conditions (which were very similar to each other). This suggests that participants found the -10 dB conditions very challenging compared to

the other three conditions, which they found to be of fairly similar difficulty levels to each other.

The equations for the fitted exponential curves for each SNR are shown in Table 6.6 and then broken down in Table 6.7, which shows the M, T and C values from the equations. The C values are the SNR if age equals infinity and represents the adult asymptote. The M values are the slope. The T values are the time constants in years.

SNR (dB)	Equation for exponential curve fitted to data
20	Predicted $MAA(°) = 12.6 * Exp\left(\frac{age\ (years)}{-2.6}\right) + 0.2$
10	Predicted $MAA(°) = 4.5 * Exp\left(\frac{age\ (years)}{-4.6}\right) + 0.06$
0	Predicted $MAA(°) = 4.7 * Exp\left(\frac{age\ (years)}{-4.5}\right) + 0.2$
-10	Predicted $MAA(°) = 2.6 * Exp\left(\frac{age\ (years)}{-7.7}\right) + 0.4$
Mean of all conditions	Predicted $MAA(°) = 3.1 * Exp\left(\frac{age\ (years)}{-5.8}\right) + 0.2$

Table 6.6: Table showing equations and time constants for exponential curve for each SNR

SNR (dB)	M value	T value	C value
20	12.6	2.6	0.2
10	4.5	4.6	0.06
0	4.7	4.5	0.2
-10	2.6	7.7	0.4
Mean of all conditions	3.1	5.8	0.2

Table 6.7: Table showing values from exponential curve fits

In all of the different SNR conditions there is a gradual improvement with age, but the steepness of the slope and the age at which performance levels off is different in each condition. In the SNR 20 dB condition there is a gentle slope and then performance appears to level off around 12 years of age. The SNR 10 and 0 dB conditions appear quite similar, with less of a visible slope than the SNR 20 dB condition, they both appear to level off around 16 years of age. In the SNR -10 dB condition the data appears a lot flatter than in the other

conditions, although the end of the curve appears to level off in a similar way to those in the other conditions. The mean MAA graph appears to be fairly similar to the graph for the 0 dB condition.

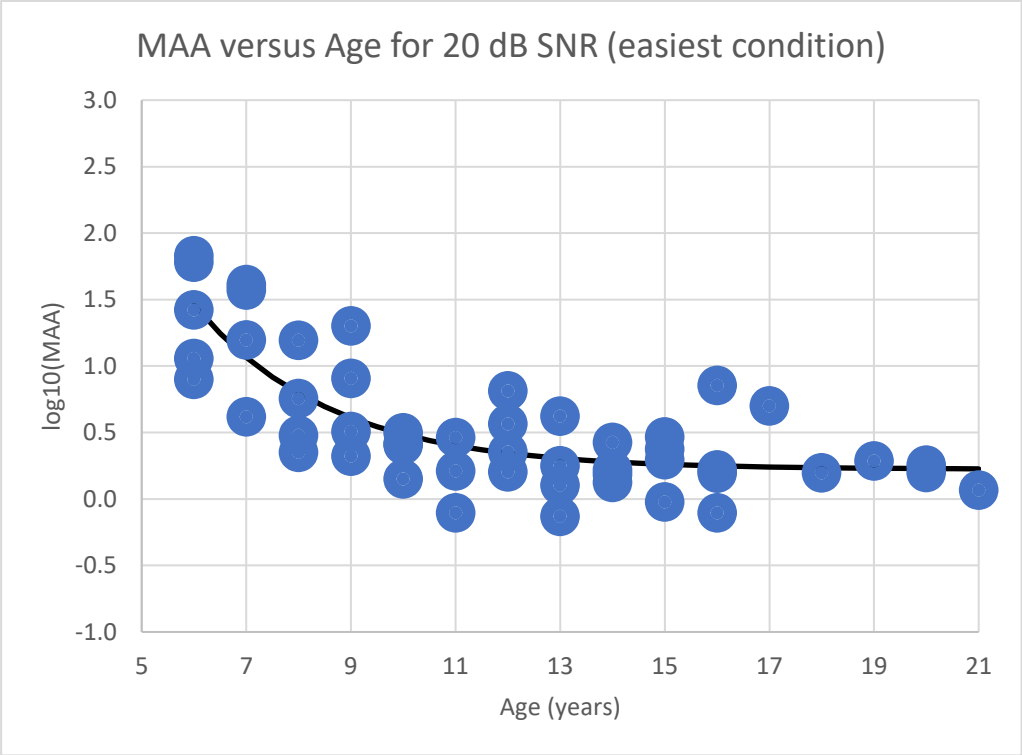


Figure 6.10: Graph showing MAA and Age for 20 dB SNR

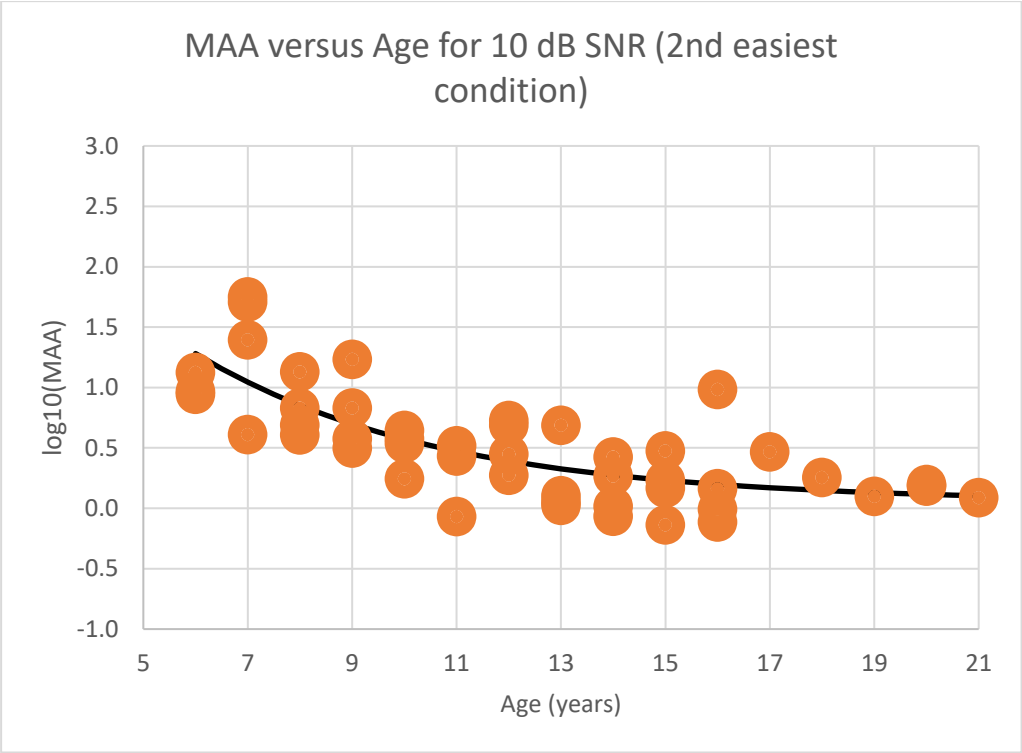


Figure 6.11: Graph showing MAA and Age for 10 dB SNR

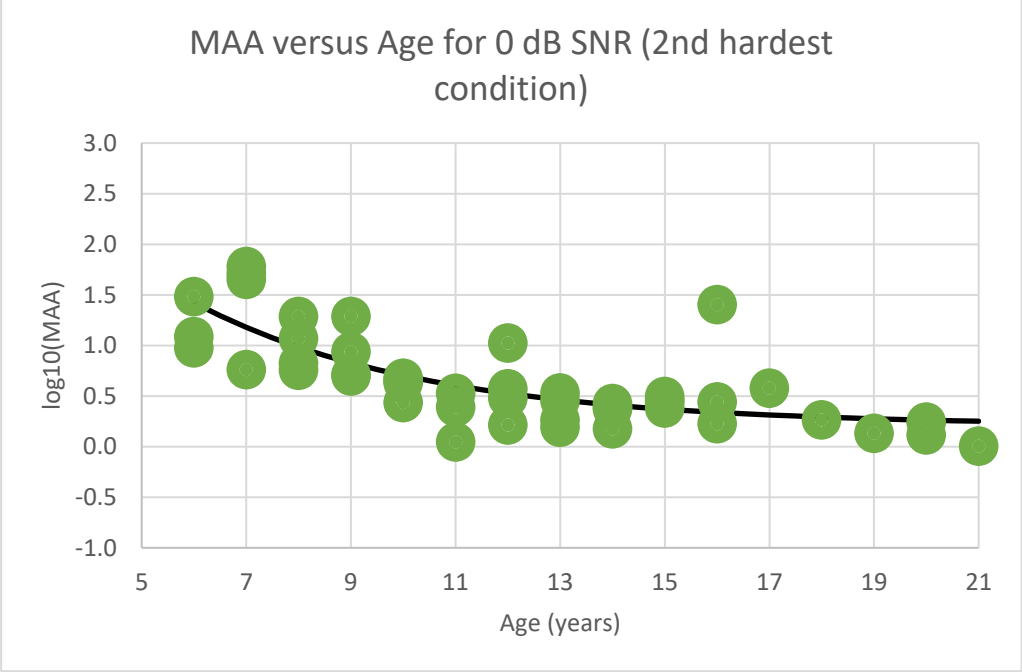


Figure 6.12: Graph showing MAA and Age for 0 dB SNR

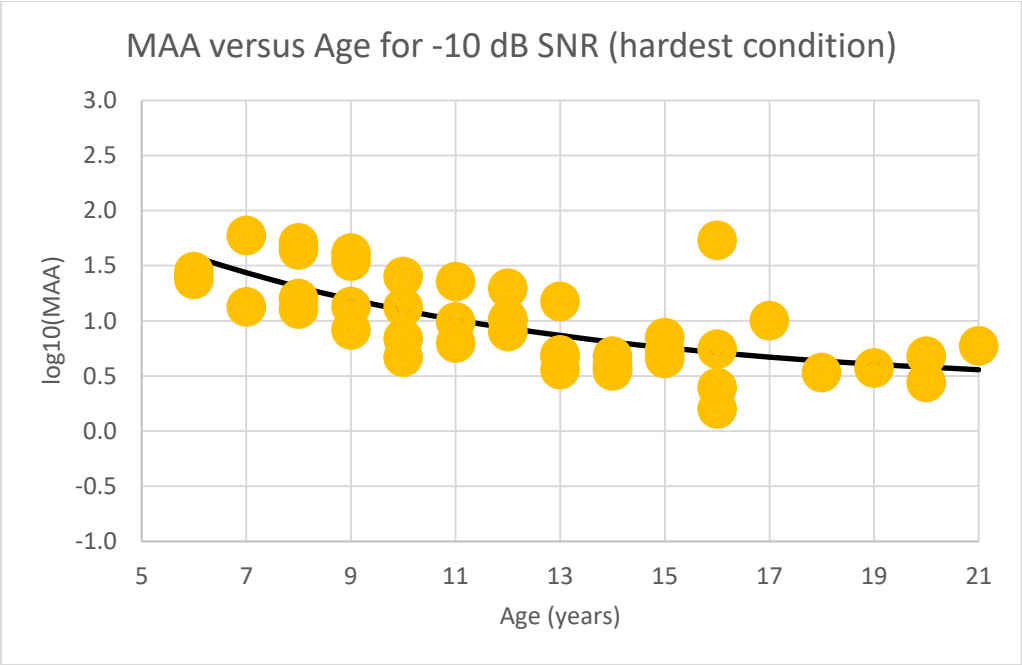


Figure 6.13: Graph showing MAA and Age for -10 dB SNR

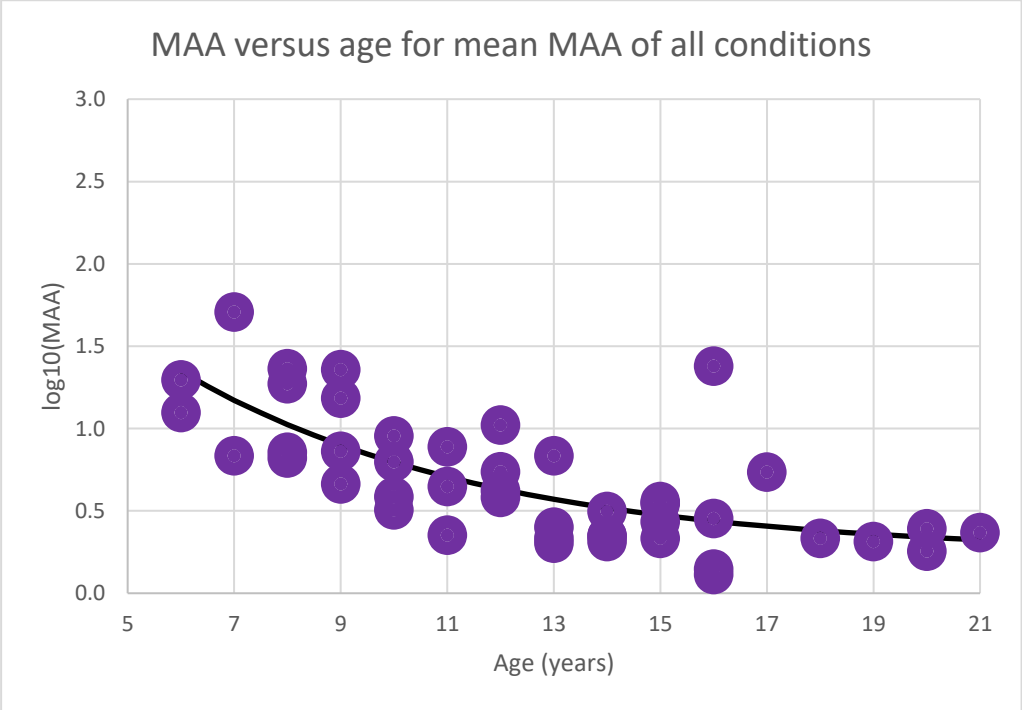


Figure 6.14: Graph showing MAA and Age for the mean MAA of all conditions (for participants who returned at least one threshold for each condition, n=40)

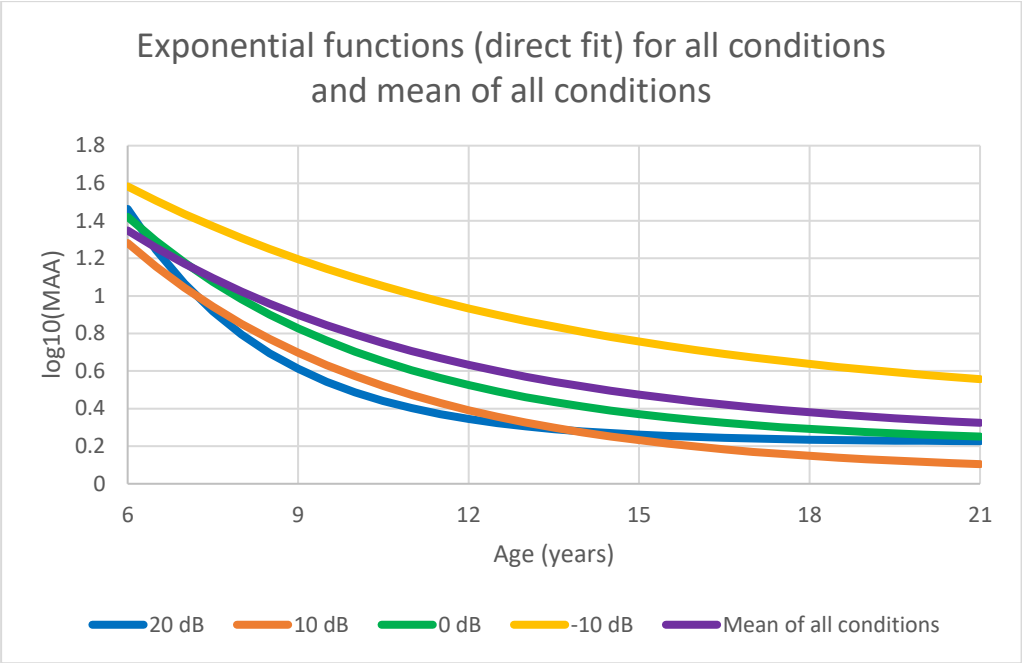


Figure 6.15: Graph showing exponential functions for MAA and age for all conditions and mean of all conditions

The relationship between age and MAA are further confirmed by Spearman's rank correlations. This test was chosen as the data for all conditions (and the average of all conditions) are monotonic and non-linear. There were strong and significant correlations between MAA and age for all SNRs and the mean of all the conditions, the values of which are shown in Table 6.8.

SNR (dB)	Correlation coefficient	Significance value
20	-0.663	$p < 0.0005$
10	-0.725	$p < 0.0005$
0	-0.754	$p < 0.0005$
-10	-0.700	$p < 0.0005$
Mean of all conditions	-0.714	$p < 0.0005$

Table 6.8: Table showing Spearman's rank correlations

Table 6.9 shows the mean and standard deviation of MAA for each SNR (for adults and children combined). From this it can be seen that the mean values are fairly similar for the three easiest conditions (20, 10 and 0 dB SNR) and then larger for the hardest condition (-10 dB SNR). Standard deviations are

slightly higher for the 20 and -10 dB SNR conditions than for the 10 and 0 dB SNR conditions.

SNR (dB)	Mean MAA (°)	Standard deviation of mean MAA (°)
20	8.2	14.2
10	6.4	10.8
0	8.5	12.9
-10	14.3	14.7
Mean of all conditions	7.5	9.0

Table 6.9: Table showing means and standard deviations for each SNR

6.4.2 Comparing results between conditions

Analysis was carried out comparing the participants mean data on each of the different conditions (SNRs) to each other. Scatterplots for each of these comparisons are shown in Figure 6.16. They are all shown with the easier SNR on the x-axis and the harder SNR on the y-axis. From these graphs it can be seen that performance on the different conditions appear to be correlated with each other. These correlations were confirmed using Pearson's correlations (as the data is linear), which are shown in Table 6.10. This shows that there are large and significant correlations for all comparisons. This suggests that participants were performing consistently and that the results were not random.

	SNR 10 dB	SNR 0 dB	SNR -10 dB
SNR 20 dB	*r* = 0.9 *p* < 0.0005 N = 48	*r* = 0.9 *p* < 0.0005 N = 48	*r* = 0.8 *p* < 0.005 N = 45
SNR 10 dB	x	*r* = 0.9 *p* < 0.0005 N = 48	*r* = 0.7 *p* < 0.0005 N = 45
SNR 0 dB	x	x	*r* = 0.8 *p* < 0.0005 N = 45

Table 6.10: Table showing Pearson's correlations for different MAA conditions

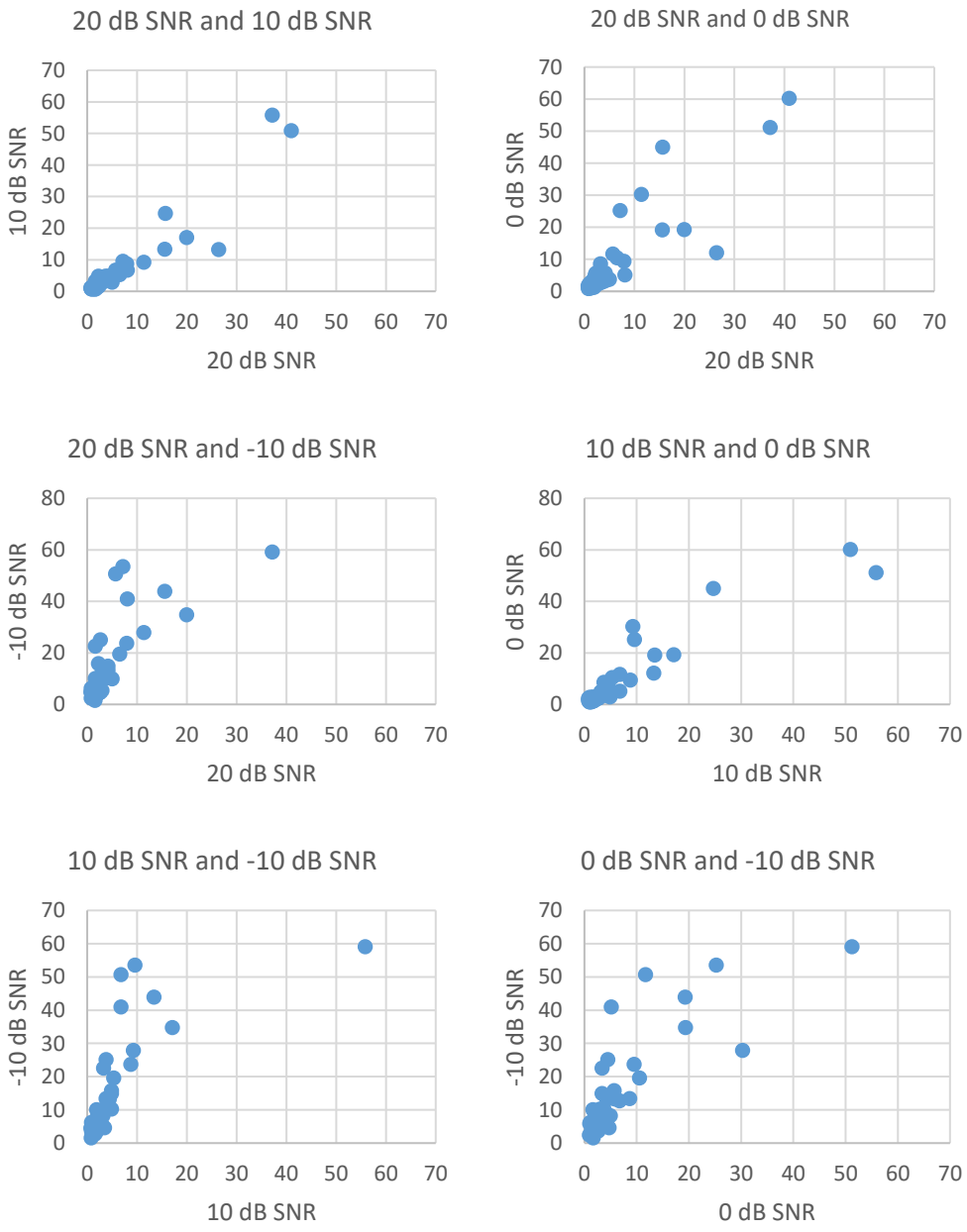


Figure 6.16: Graphs showing comparisons between all conditions

6.4.3 Differences between adults and children

A Mann-Whitney U test was chosen to compare adult and child performance on the MAA tasks. Children were classed as being participants aged 6 to 17 years and adults being participants aged 18 years and older. This test was chosen because the data was not all normally distributed (Shapiro-Wilk test of normality showed $p<0.05$ for children in all conditions) and although there were no outliers in the adult data, there were many in the child data (as shown in the boxplots in Figure 6.17 and Figure 6.18, note the y-axis in Figure 6.18 is 10x that of Figure 6.17). For all SNRs and the mean of all conditions, distributions of the MAAs for adults and children were similar, as assessed by visual inspection of population pyramids.

- **SNR 20 dB:** Median MAAs for children (2.9°, mean rank=26.9) were significantly higher than for adults (1.6°, mean rank=13),
  $U = 50$, $z = -2.021$, $p = 0.042$, r = -0.29
- **SNR 10 dB:** Median MAAs for children (3.3°, mean rank=25.8) were not significantly higher than for adults (1.5°, mean rank=13.2),
  $U = 51$, $z = -1.907$, $p = 0.057$, r = -0.28
- **SNR 0 dB:** Median MAAs for children (3.7°, mean rank=26.6) were significantly higher than for adults (1.3°, mean rank=6.6),
  $U = 18$, $z = -3.021$, $p = 0.001$, r = -0.43
- **SNR -10 dB:** Median MAAs for children (9.9°, mean rank=24.7) were statistically significantly higher than for adults (3.7°, mean rank=9.5),
  $U = 32$, $z = -2.456$, $p = 0.011$, r = -0.37

- **Mean of all conditions:** Mean MAAs for children (4.5°, mean rank=24.8) were statistically significantly higher than for adults (2.2°, mean rank=8.9), U = 29.5, z=-2.548, p = 0.008, r = -0.36)

These results are summarised in Table 6.11. Overall, adults had significantly better MAAs than children in all conditions except the SNR 10 dB condition, with medium effect sizes.

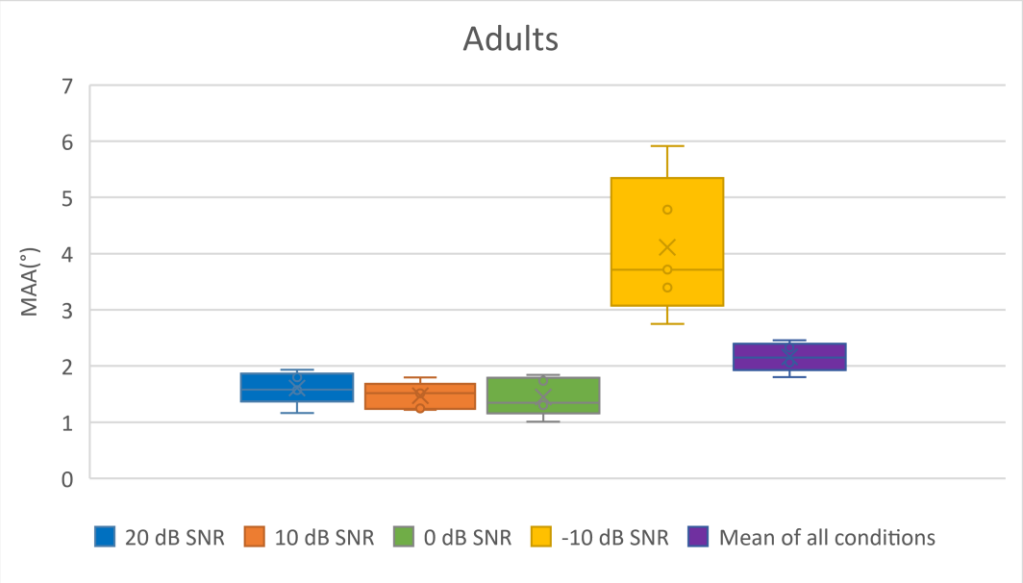


Figure 6.17: Boxplots for adult MAAs

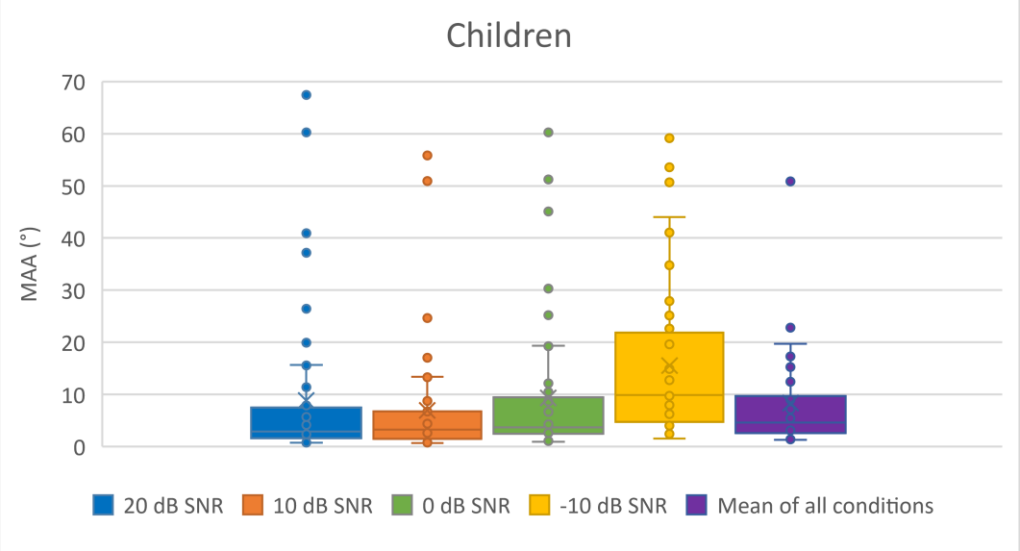


Figure 6.18: Boxplot for child MAAs

		20 dB SNR	10 dB SNR	0 dB SNR	-10 dB SNR	Mean of all conditions
Median values	**Children**	2.9°	3.3°	3.7°	9.9°	4.5°
	Adults	1.6°	1.5°	1.3°	3.7°	2.2°
Mean rank values	**Children**	26.9	25.8	26.6	24.7	24.8
	Adults	13.0	13.2	6.6	9.4	8.9
Effect sizes (*r*)		-0.29	-0.28	-0.43	-0.37	-0.36

Table 6.11: Table showing median and mean rank values for adults and children for each SNR

It was not possible to look at interactions between the factors (SNR and adult/child) using a test such as the two-way mixed ANOVA due to lack of homogeneity of variances in the groups (using the MAA data and log10(MAA) data). A Welch ANOVA was carried out on the log10(MAA) data (this test was chosen due to the lack of homogeneity of variances, as assessed by Levene's test for equality of variances, $p < 0.05$. Table 6.12 shows that significant differences were found between adults and children for all conditions (and for the mean of all conditions), with medium to large effect sizes. These results are fairly similar to the Mann-Whitney U test results, where a significant difference was found between adults and children for all conditions except 0 and 10 dB SNR) with medium effect sizes.

Condition	Welch ANOVA results	Partial η^2
20 dB SNR	$F(1, 41.1) = 21.0$, $p < 0.0005$	0.056
10 dB SNR	$F(1, 42.8) = 24.7$, $p < 0.0005$	0.068
0 dB SNR	$F(1, 27.4) = 42.3$, $p < 0.0005$	0.133
-10 dB SNR	$F(1, 16.6) = 23.2$, $p < 0.0005$	0.108
Mean of all conditions	$F(1, 43.2) = 40.0$, $p < 0.0005$	0.111

Table 6.12: Table showing results from Welch ANOVA

In order to further understand the difference between adults and children on this task, data from all four conditions was pooled and the amount by which participant's performance differed from the predicted performance values was plotted (predicted values were taken from exponential curve fits

described in section 6.4.4). Figure 6.19 shows the combined data from the adults and children, which shows a fairly normally shaped distribution around 0 (no difference between actual and predicted values). Figure 6.20 shows just the data from the adults and Figure 6.21 shows just the data from the children. From these graphs it is clear that the adult's differences from the predicted values are all much closer to 0 than the children's (i.e. no difference between actual and predicted values).

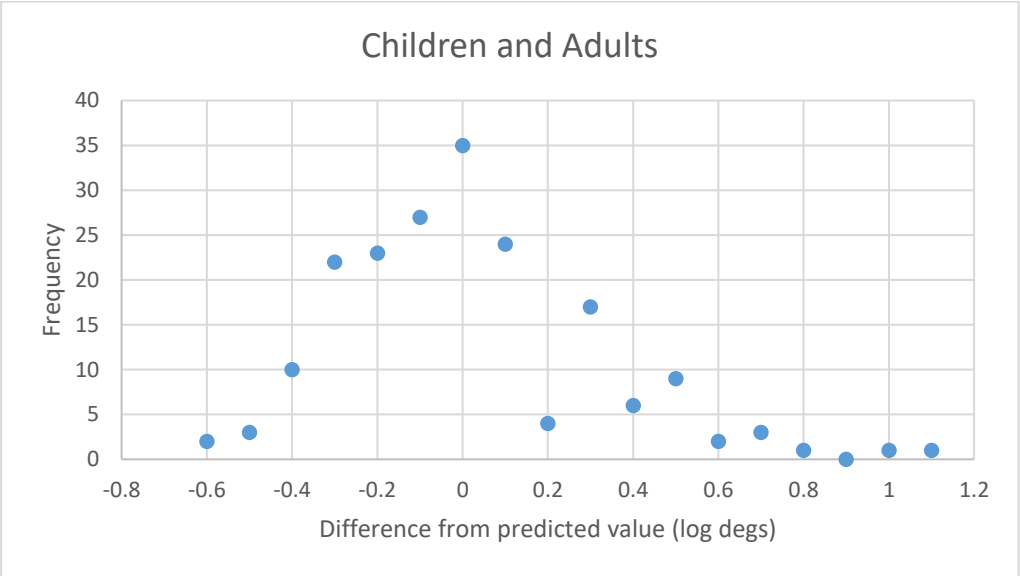


Figure 6.19: Graph showing how all participants performance varied from predicted values

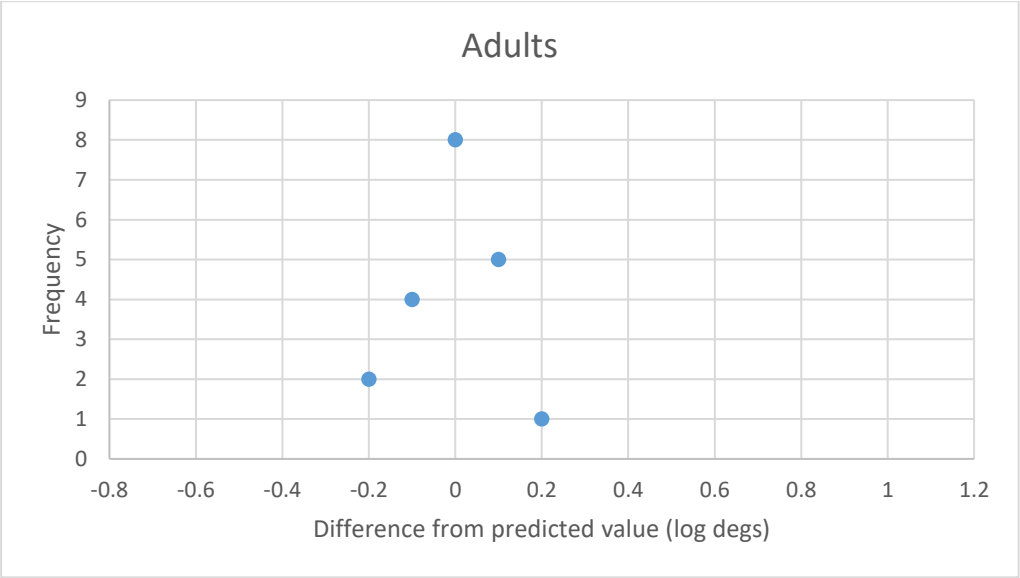


Figure 6.20: Graph showing how adult participant's performance varied from predicted values

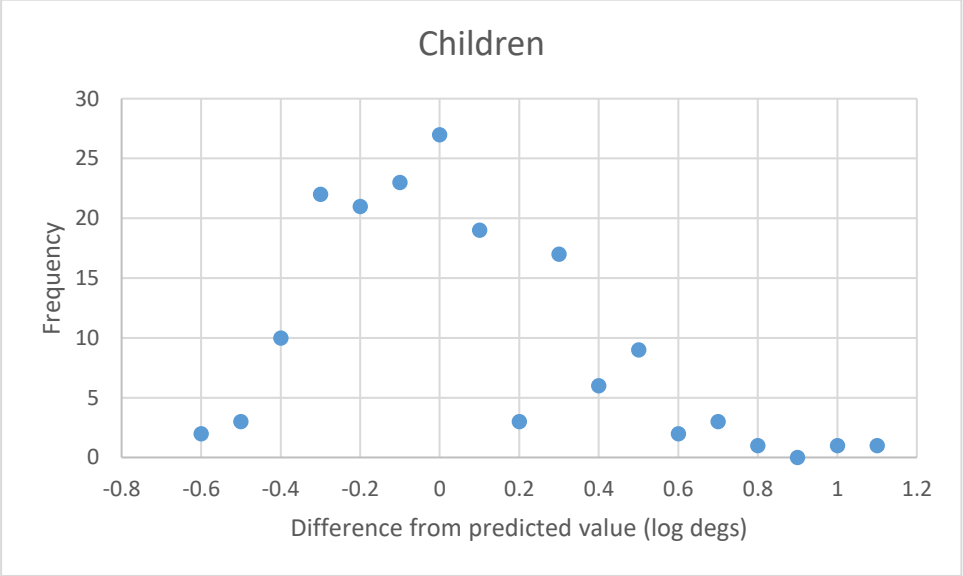


Figure 6.21: Graph showing how child participants performance varied from predicted values

6.4.4 Adult-like performance: Exponential function (direct fit)

The age at which adult-like performance was reached was defined using exponential functions fitted to the data and RMS errors (RMS error was 0.3 log(dB) for all conditions). This is the same method which was used and fully described in Chapter 4. Results from this analysis are shown in Table 6.13. From this, it can be seen that generally as SNR decreases the age range at which adult-like performance is achieved increases, although the developmental time courses seem to be roughly the same length for all conditions (they do not increase with difficulty of the SNR). However, the age ranges for SNR 10 dB and 0 dB are very similar.

SNR (dB)	Age at which adult-like performance is reached (years, nearest 0.5 years)
20	8.0-12.0
10	9.5-13.0
0	9.5-12.5
-10	12.0-17.0
Mean of all conditions	10.5-14.5
Overall range	8.0-17.0

Table 6.13: Table showing age at which adult-like performance is reached for each SNR using the exponential function (direct fit)

6.4.5 Adult-like performance: Broken stick analysis

Broken stick analysis was also carried out (as described in Chapter 4), where the slope after the change point was zero (using R code from Schwarz (2015)). The R analysis fitted a piecewise linear function to the data, found the change point and used bootstrapping to find the slope of the functions and the confidence intervals for the change point. This type of broken stick function was chosen over the type where there are different (non zero) slopes either

side of the change point because adult-like performance should be stable in young adults.

The graph showing the broken stick functions are plotted in Figure 6.22, Figure 6.23, Figure 6.24, Figure 6.25 and Figure 6.26 (for conditions SNR 20, 10, 0 and -10 dB and the mean of all conditions) and the data from the analysis is shown in Table 6.14. All of the broken stick functions are shown on the same graph in Figure 6.27. From this it can be seen that performance predicted by the broken stick functions is worst in the -10 dB condition (as would be expected), is then better in the 20 dB and 0 dB conditions, and is best in the 10 dB conditions. The change points are different for each condition, being 10.6 years in 20 dB, 13.1 years in 10 dB, 11.1 years in 0 dB and 14.0 years in -10 dB. Although these values are all different, they have quite wide 95% confidence intervals (shown in Figure 6.28). This suggests that these change points are not accurate as point estimates, so it is better to use the confidence intervals as a measure of when adult-like performance is reached. This gives ranges of 10.5-16.0 years for the 20 dB SNR condition, 10.5-15.0 years for the 10 dB SNR condition, 10.5-16.5 years for the 0 dB SNR condition, 11.5-18.0 years for the -10 dB SNR condition and 10.5-16.0 years for the mean of all SNRs (and an overall range of 10.0-18.0 years). Figure 6.29 shows the values for initial slope with 95% confidence intervals for each of the conditions. From this graph it can be seen that the values are fairly similar for each condition and all have quite wide confidence intervals, similar to the

graphs for the change point values. This suggests that there are probably not significant differences between the slope values for each condition.

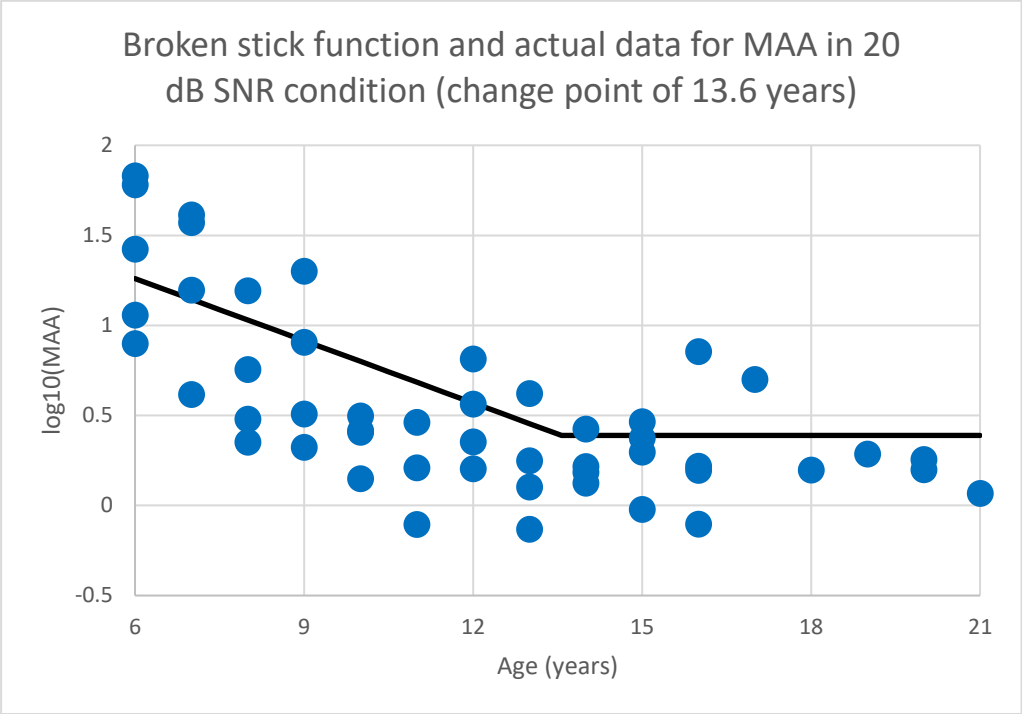


Figure 6.22: Graph showing broken stick function for MAA of speech in babble 20 dB condition

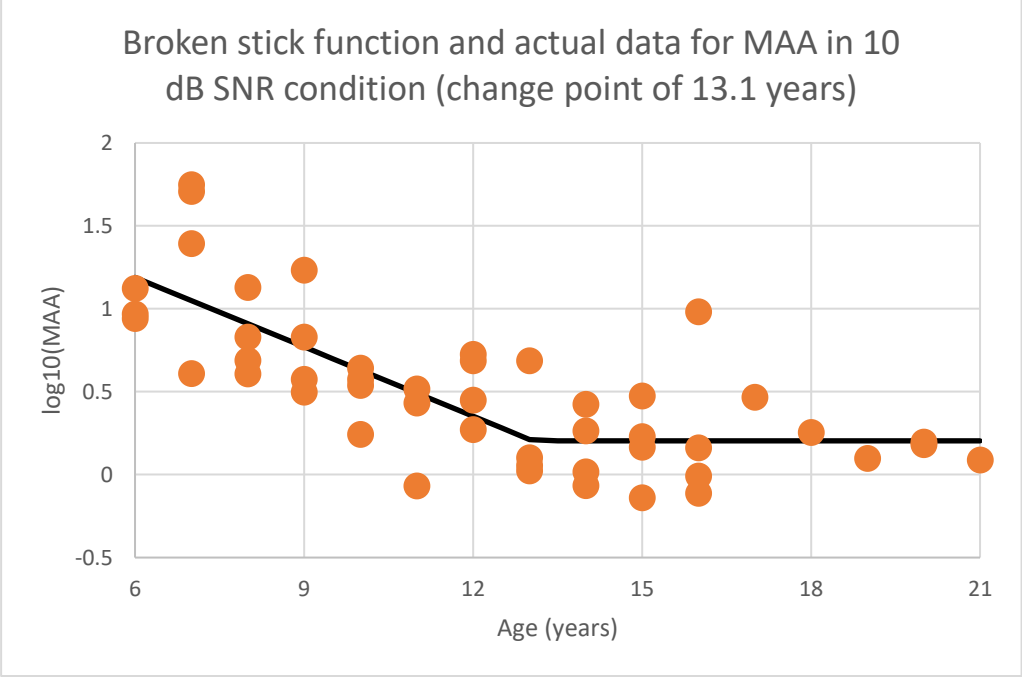


Figure 6.23: Graph showing broken stick function for MAA of speech in babble 10 dB condition

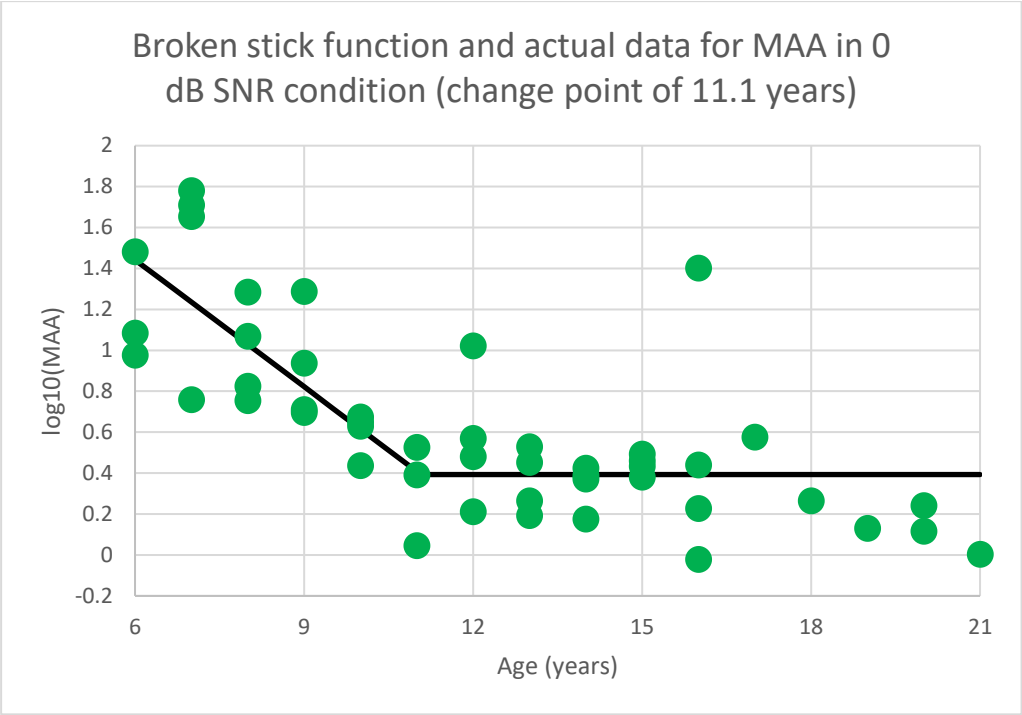


Figure 6.24: Graph showing broken stick function for MAA of speech in babble 0 dB condition

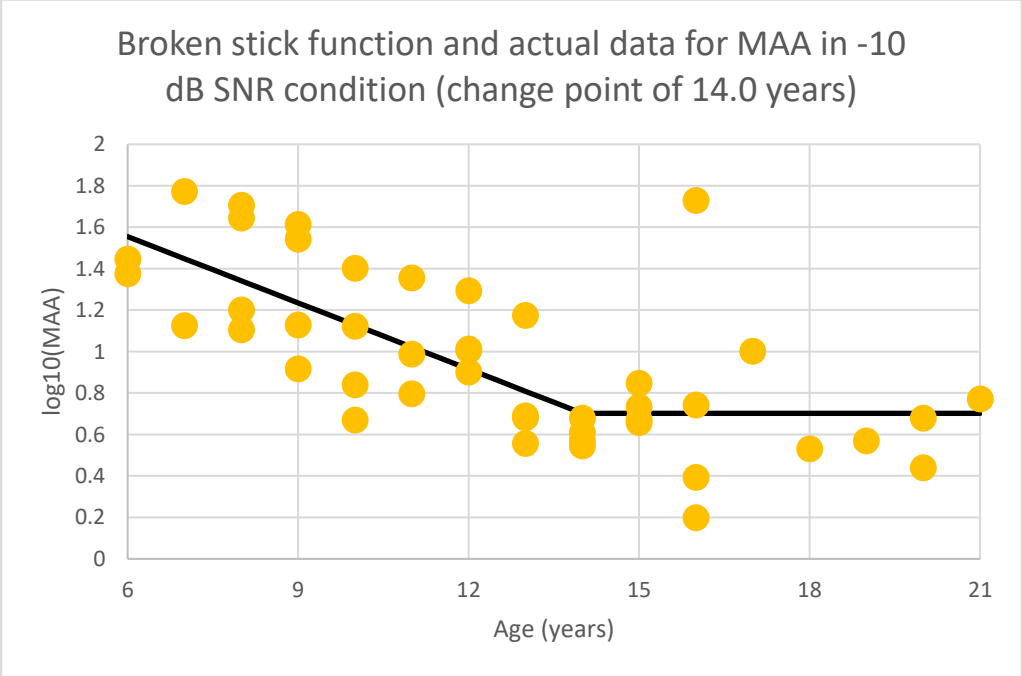


Figure 6.25: Graph showing broken stick function for MAA of speech in babble -10 dB condition

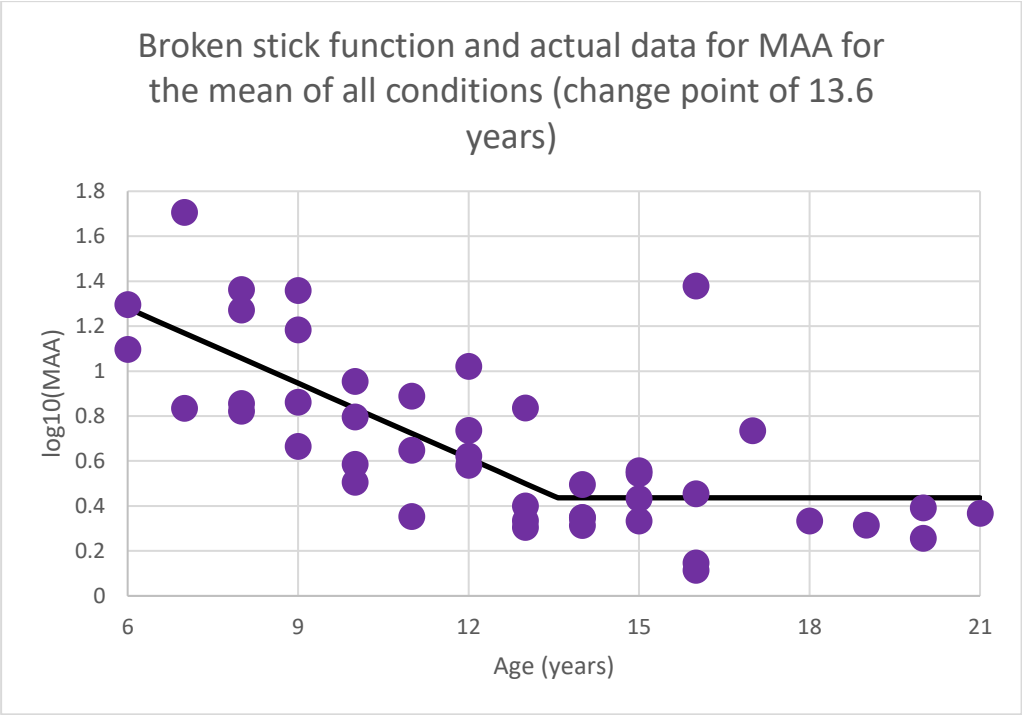


Figure 6.26: Graph showing broken stick function for MAA of the mean of all conditions

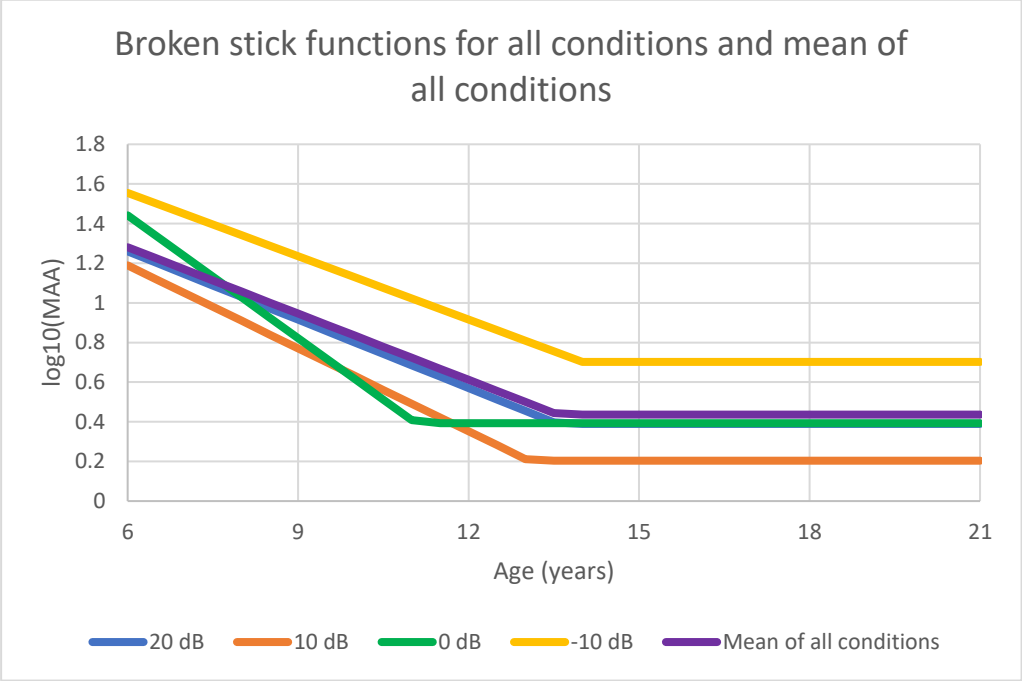


Figure 6.27: Graph showing all broken stick functions together

		SNR 20 dB	SNR 10 dB	SNR 0 dB	SNR -10 dB	Mean of all conditions
Change point (years)	**Value**	13.57	13.06	11.08	14.00	13.57
	2.5 % confidence interval	10.45	10.38	10.39	11.42	10.55
	97.5% confidence interval	16.00	15.00	16.50	18.00	16.00
Slope before change point (dB/year)	**Value**	-0.12	-0.14	-0.21	-0.11	-0.11
	2.5 % confidence interval	-0.26	-0.30	-0.31	-0.16	-0.24
	97.5% confidence interval	-0.08	-0.10	-0.09	-0.07	-0.074

Table 6.14: Table showing data from R broken stick analysis

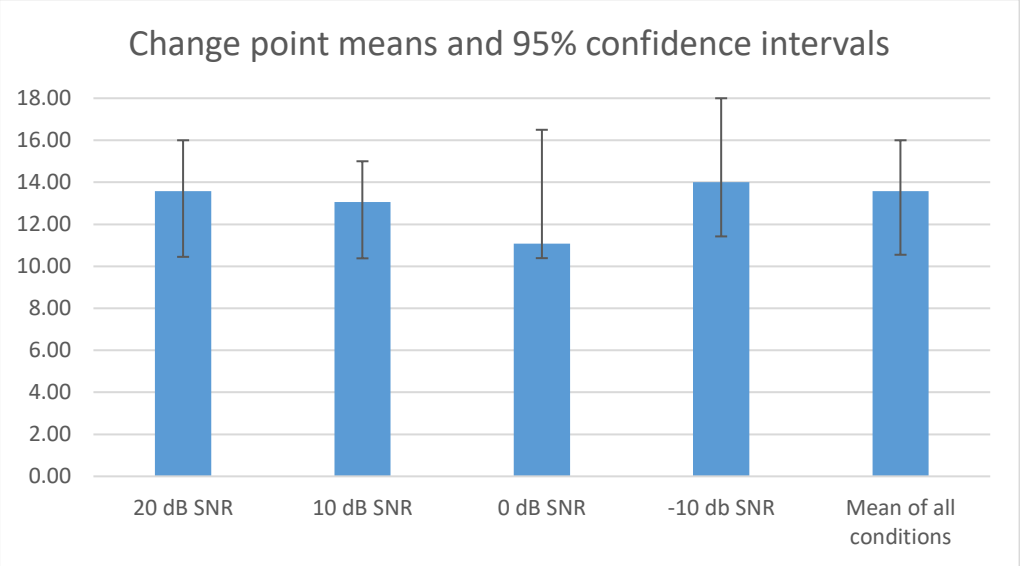


Figure 6.28: Graph showing mean values and 95% CIs for change points from broken stick function for each condition

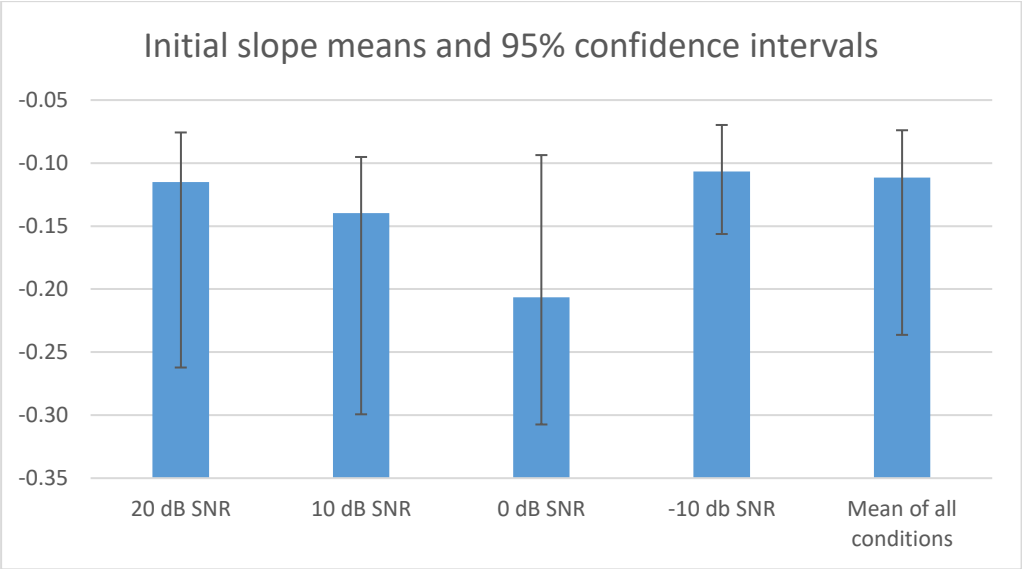


Figure 6.29: Graph showing mean values and 95% CIs for initial slopes from broken stick function for each condition

6.4.6 Adult-like performance: Exponential function (bootstrap)

A Bootstrap analysis was carried out using MATLAB code written by Michael Akeroyd as part of the laboratory statistical software. The full dataset was input and the values used to plot exponential functions similar to those previously fitted to the data (slope, time constant and asymptote) were output following 1000 bootstraps. This was done for the mean MAA over all conditions (as before, excluding participants who did not achieve at least one threshold for each condition) and for each condition (20, 10, 0 and -10 dB). The results are shown in Table 6.15 and were used to plot the exponential functions shown in Figure 6.30. From the table it can be seen that some of the results are larger than might be expected. For example, the standard deviation of the slope values for the 20 dB and mean SNR conditions are 203 dB/year and 232 dB/year respectively, and the upper confidence interval for the crossover in the -10 dB condition is 39.9 years. This suggests that the bootstrap analysis is not completely reliable.

SNR	Parameter	Median	Standard deviation	Lower Confidence Interval	Upper Confidence Interval
20 dB	**Slope (dB/year)**	13.0	203.0	4.3	68.6
	Time constant (years)	2.6	0.7	1.5	4.5
	Asymptote (dB)	0.2	0.07	0.07	0.3
	Crossover (years)	7.9	0.7	7.0	9.4
10 dB	**Slope (dB/year)**	4.4	42.7	2.6	18.3
	Time constant (years)	4.6	1.4	2.5	7.9
	Asymptote (dB)	0.06	0.1	-0.2	0.2
	Crossover (years)	9.5	1.5	7.6	13.1
0 dB	**Slope (dB/year)**	4.9	32.3	2.5	34.7
	Time constant (years)	4.3	1.9	2.0	9.8
	Asymptote (dB)	0.2	0.2	-0.2	0.5
	Crossover (years)	9.5	2.2	6.8	15.3
-10 dB	**Slope (dB/year)**	2.8	2.9	1.9	8.1
	Time constant (years)	7.3	19.1	3.5	24.0
	Asymptote (dB)	0.4	1.3	-0.9	0.7
	Crossover (years)	12.0	10.0	7.2	39.9
Mean of all conditions	**Slope (dB/year)**	3.3	232.0	2.0	12.6
	Time constant (years)	5.7	4.5	2.7	11.7
	Asymptote (dB)	0.2	0.4	-0.2	0.5
	Crossover (years)	10.7	4.1	7.2	18.8

Table 6.15: Table showing results from bootstrap analysis

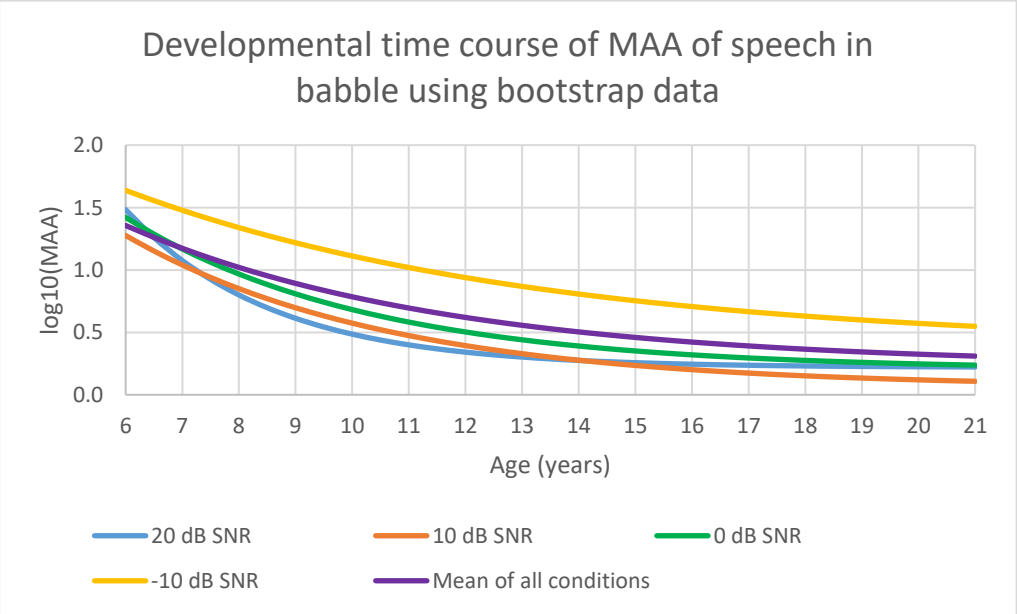


Figure 6.30: Graph showing exponential functions of MAA of speech in babble from bootstrap analysis

The equivalent graph plotted using the actual data from the experiment is shown earlier in Figure 6.15. In Figure 6.15 the data is as might be predicted, with MAAs being highest for the hardest condition (-10 dB) then decreasing as the SNR of the conditions increased and became easier (although 20 and 10 dB SNR are very similar). However, in Figure 6.30 it appears that MAA is highest for the -10 dB SNR condition (the hardest condition), followed by 20 dB SNR, then 0 dB SNR and finally 10 dB SNR. This may suggest that the bootstrapping analysis is unreliable for these datasets.

Confidence intervals were also obtained from the bootstrap analysis for each of the variables (slope, time constant and asymptote) and crossover (the point at which adults and children differed by two RMS errors). Figure 6.31 shows the mean values and confidence intervals (95%) for the slope of the exponential functions. From this, it would appear that the slope values are fairly similar for all conditions. However, the confidence intervals are very variable and are especially large for the 0 dB condition. This means that it is possible that the slope values are actually quite different to each other.

Figure 6.32 shows the mean values and confidence intervals (95%) for the time constants of the exponential functions. From this, it can be seen that the time constants are fairly similar to each other and the -10 dB condition has a very large confidence interval. This implies that the time constants for all conditions except -10 dB are likely to be fairly similar to each other.

Figure 6.33 shows the mean values and confidence intervals (95%) for the asymptote values of the exponential functions. From this it can be seen that

most of the mean asymptotes values are very similar and the confidence interval is much larger for the -10 dB SNR condition then the other three conditions. Therefore, it is likely that these values are not significantly different to each other (except at -10 dB SNR).

Figure 6.34 shows the mean values and confidence intervals (95%) for the crossover values of the exponential functions, which is when the adults and children were separated by two RMS errors. From this it can be seen that these values were fairly similar to each other, and again the confidence interval for the -10 dB condition is much larger than the others. As with the other plots, this suggests that there is unlikely to be a significant difference at the point when adult-like performance is achieved between the results from the different conditions, except for in the -10 dB condition. Many of the 95% confidence interval values for the -10 dB condition (but especially for the crossover value) are quite large. This may be due to an issue with the bootstrap analysis code, possibly linked to a fairly high value from one of the 16 year old participants on this conditions (which can be seen in Figure 6.13).

As the mean values of the crossover point, slopes, time constants and asymptotes for each condition were generally fairly similar and the confidence intervals so large, the mean averages for slope, time constant and asymptote were calculated over all four conditions. Figure 6.35 shows the exponential functions plotted using the means of all four conditions for the actual data and the bootstrap data. From this it can be seen that the bootstrap data is essentially the same as the actual data. Although the bootstrap analysis may

not be highly accurate it still shows the overall trend of MAAs decreasing with age so supports the conclusion that MAAs of speech in babble improves with age and likely reaches an adult-like level in the early to mid-teenage years.

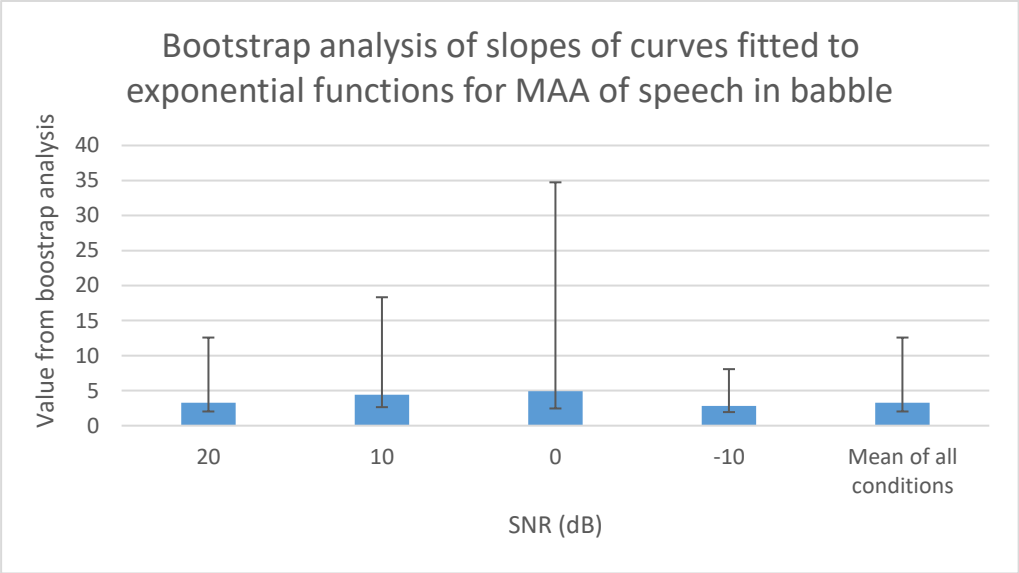


Figure 6.31: Graph showing bootstrap analysis of slope for MAA of speech in babble exponential functions (error bars showing 95% confidence intervals)

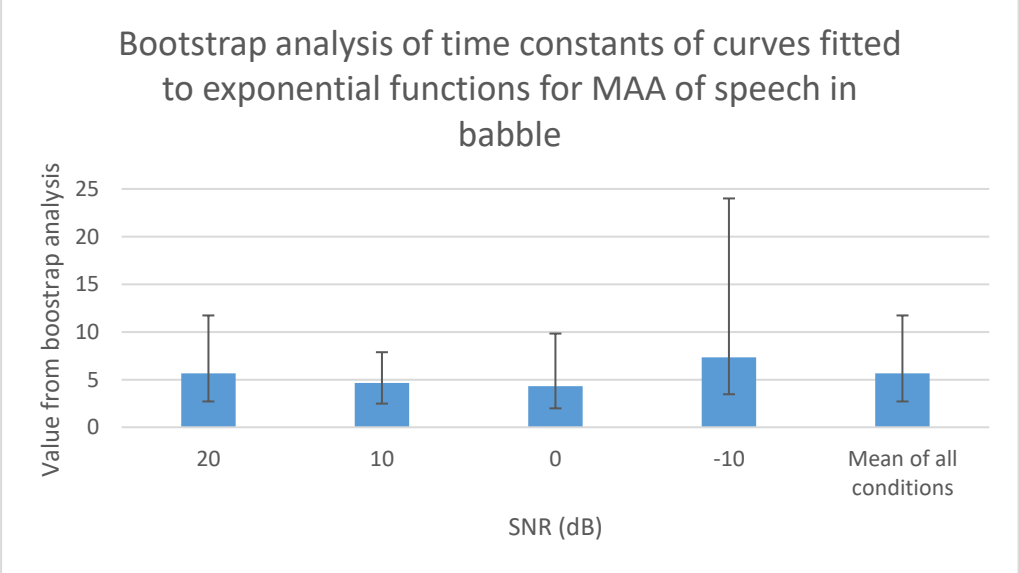


Figure 6.32: Graph showing bootstrap analysis of time constant for MAA of speech in babble exponential functions (error bars showing 95% confidence intervals)

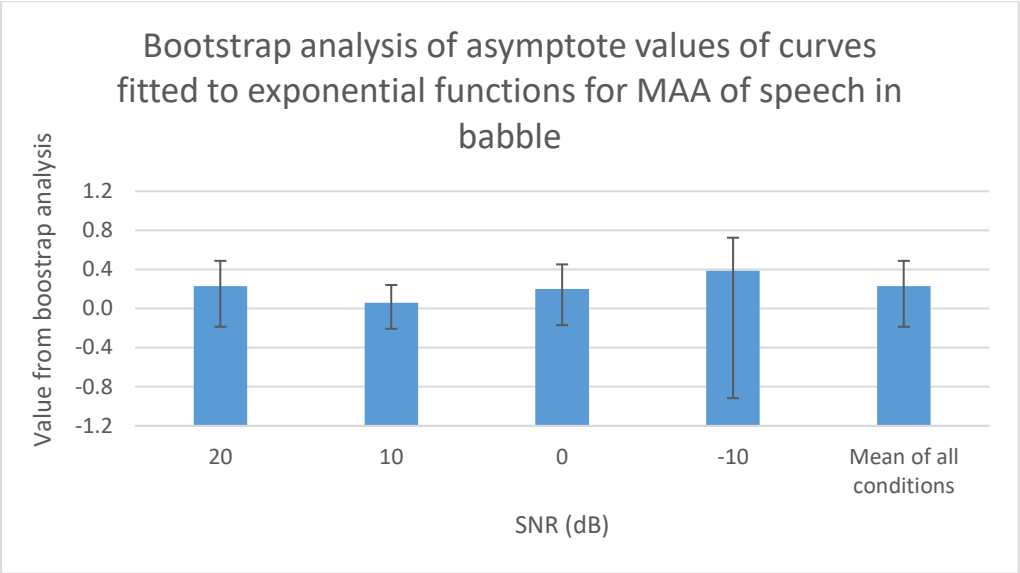


Figure 6.33: Graph showing bootstrap analysis of asymptote for MAA of speech in babble exponential functions (error bars showing 95% confidence intervals)

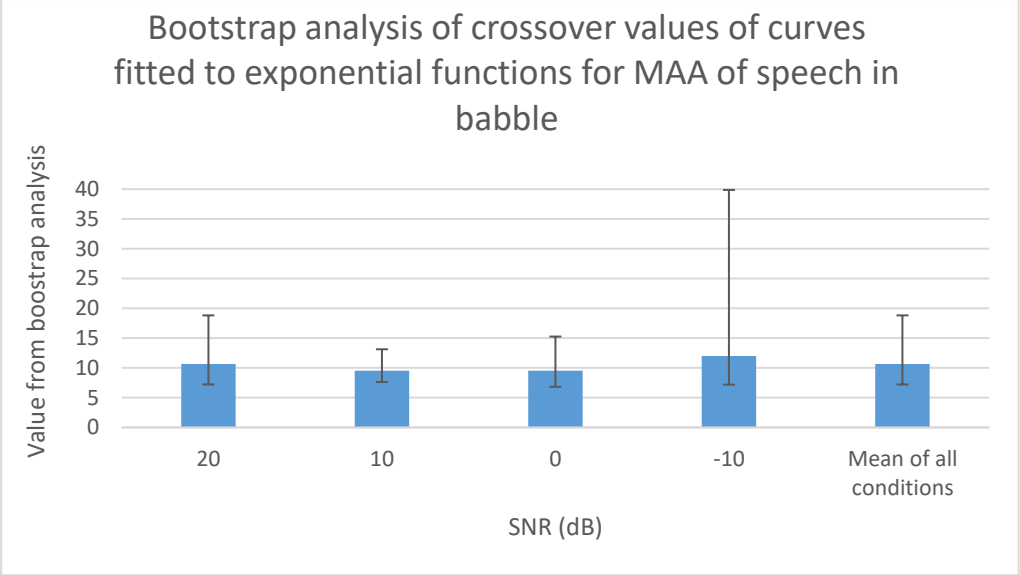


Figure 6.34: Graph showing bootstrap analysis of crossover for MAA of speech in babble exponential functions (error bars showing 95% confidence intervals)

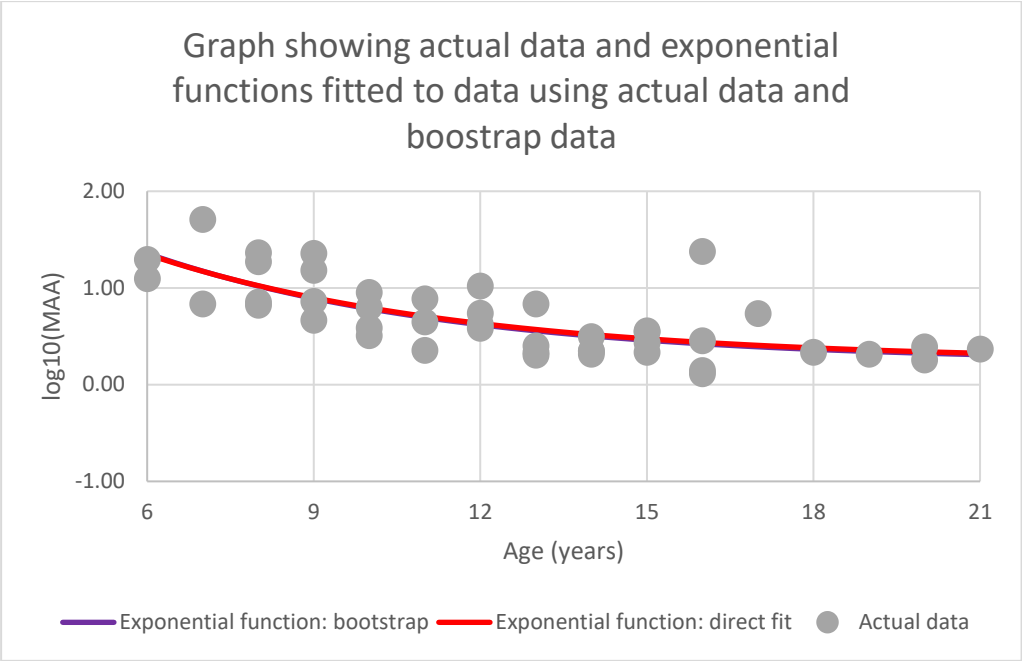


Figure 6.35: Graph comparing bootstrap and actual data for MAA of speech in babble (mean of all conditions)

6.4.7 Comparison of different analysis methods

Figure 6.36 shows the exponential functions (direct and bootstrap) and the broken stick function on the same graph (along with the participant data) for the mean MAA over all conditions. From this graph, it can be seen that the exponential curve fits (direct and bootstrap) are very similar. The second (flat) part of the broken stick function is higher than the asymptotes for both curve fits. This means that the adult-like performance level estimated using this method (0.44 log(dB)) is higher (worse) than that estimated by the curve fit (0.2 log(dB)) and bootstrap curve fit (0.23 log(dB)). The change point of the broken stick function (13.6 years, with 95% confidence intervals of 10.5 to 16 years) is older than adult-like performance estimated by the exponential function (bootstrap) (10.7 years, 95% confidence intervals of 7 to 18 years). The exponential function (direct fit) gave an estimate of 10.5 to 14.5 years,

which is very similar to that given by the broken stick function. This is the same pattern of results seen in the LocalB_{5AT} task in Chapter 4 and again it seems most likely that adult-like performance is reached in the early to mid-teenage years.

R^2 values for all analysis methods and conditions are shown in Table 6.16. From this it can be seen that R^2 values are very similar for the exponential function (direct) fit and exponential function (bootstrap) methods, ranging from 47-62% and 47-63% respectively. The R^2 values from the broken stick method are fairly similar with values ranging from 43-59%.

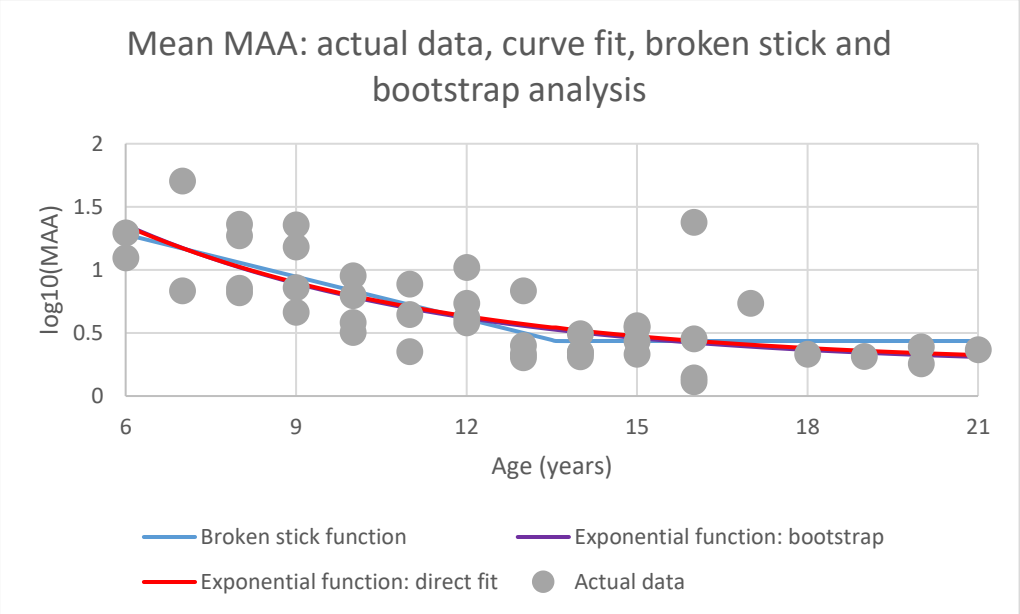


Figure 6.36: Graph comparing different methods for estimating adult-like performance

	R^2 values from different analysis methods (%)		
SNR	**Exponential function (direct fit)**	**Exponential function (bootstrap)**	**Broken stick**
20 dB	62	62	55
10 dB	57	57	58
0 dB	57	57	59
-10 dB	47	47	43
Mean of all SNRs	60	63	53

Table 6.16: Table showing R^2 values from different analysis methods

6.4.8 Comparison to results from LocalB$_{5AT}$ task (Experiments 1 and 2)

A total of 22 participants completed the MAA task (and returned at least one threshold at each SNR) and the LocalB$_{5AT}$ task in either Experiment 1 or 2. Their results from the LocalB$_{5AT}$ and mean MAA score over all conditions are shown in Figure 6.37. From this it can be seen that there is a correlation between good performance on both of the tasks. This was confirmed by a Spearman's rank correlation (chosen as the data is non-linear but monotonic), which gave $r_s = 0.7$ and $p < 0.0005$. The fact that performance on these two tasks is strongly correlated is not surprising, given that they were both aiming to measure localisation of speech in babble abilities but in different ways.

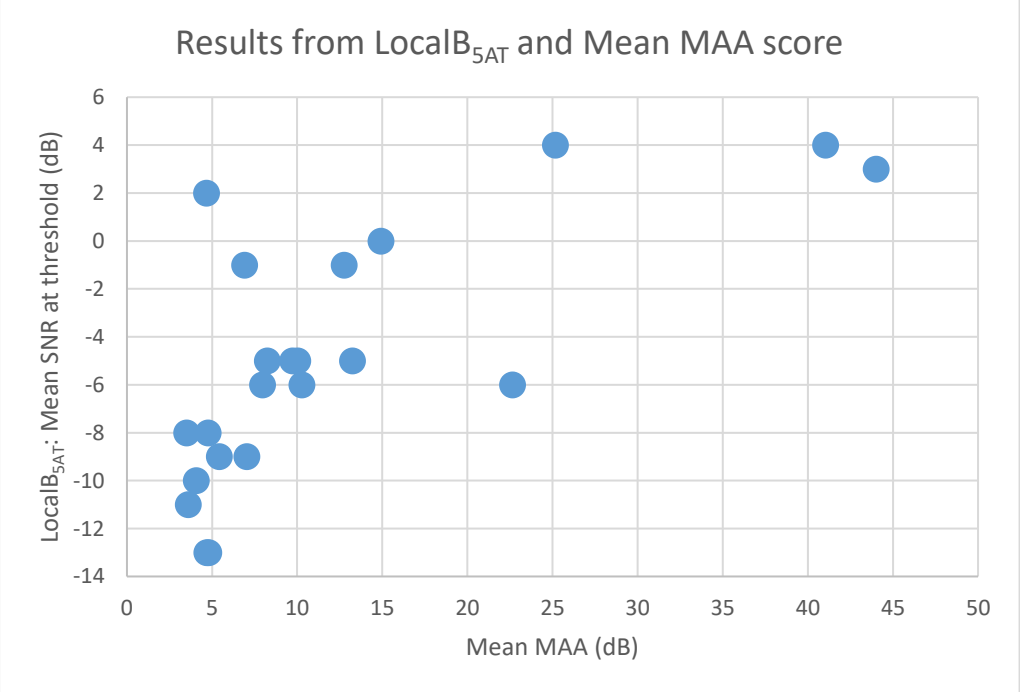


Figure 6.37: Graph showing results from Experiment 1/2 and Experiment 4

6.4.9 Additional results of children who had difficulty with the MAA task

Three participants had great difficulty with the MAA task. Two participants, RB328 and RB346 only managed to complete one track (at the easiest SNR and none at the more challenging ones), and the third, RB345 did not complete any tracks. All of these participants were six years of age. The details of these participants are shown in Table 6.17. Three other six year olds participated in the experiment and successfully completed most or all of the conditions.

ID	Age	Number of tracks completed at SNR 1	Localisation in quiet (% correct)	Localisation in babble (easy SNR, % correct)	Reason for poor MAA
RB328	6	1 (MAA of 60°)	73	60	Still developing, more affected by babble than normal
RB345	6	0	93	80	MAA masker too much for participant
RB346	6	1 (MAA of 67°)	23	30	Very poor localisation abilities

Table 6.17: Table showing details of participants who had difficulty with MAA tasks

In more detail, one participant (RB346, aged six years) struggled with both the localisation in quiet and localisation in low background sound tasks, scoring 23% and 30% respectively. Both scores are not far from chance for these tasks (20%). These results suggest that this participant has not yet developed the adult-like localisation in quiet abilities normally expected by approximately six years of age, which is why they struggled with the MAA in babble task. One of the other participants (RB345, aged six years) scored 73% on the localisation in quiet task and 60% on the localisation in low background sound task. This

suggests that localisation abilities were still developing in this child and perhaps that they were more significantly affected by the presence of babble than would be expected for a child of their age (although there was only a small number of presentations in the low background sound task compared to the quiet task). The third participant (RB345, aged six years) scored fairly well in both the localisation in quiet task and the localisation in low background sound task, scoring 93% and 80% respectively. This suggests that their localisation in quiet abilities were developed, but perhaps the more challenging SNR used for the MAA task made the MAA task too difficult for them.

Overall, it seems that one participant was very poor at localising generally, another was still developing localisation abilities and the third one had developed localisation in quiet abilities but struggled greatly when the level of the masker became challenging. This information would be useful for developing a future experiment with a similar age range. If it was important that six year olds could complete the experiment then it would be necessary to include conditions with less challenging SNRs. It would also be beneficial to have a good measure of localisation in quiet for all participants, so as to identify whether this is a problem for any participants who struggle with the main tasks. Three other six year olds participated in the experiment and completed most or all of the conditions, therefore this is a successful completion rate of 50% for six year old participants. This is a large proportion

of the six year old participants so any future similar studies should consider these recommendations.

6.5 Discussion

This experiment showed that performance on an MAA task with SNRs of 20, 10, 0 and -10 dB improved with age, with small to medium effect sizes in all conditions except 10 dB SNR. Issues were present with the panning code (detailed in section 6.3.3.1) but It is unlikely that they have affected the results of the experiments in any substantial way.

As discussed in previous chapters, there are many different ways to define adult-like performance. In this experiment exponential functions (direct fit and bootstrap) and broken stick functions were used. Table 6.18 summarises the age ranges at which adult-like performance was reached using these methods. For the mean of all conditions, analysis showed that the age ranges at which adult-like performance was reached was 10.5-14.5 years (for the exponential function (direct fit)), 10.5-16.0 years (for the broken stick function method) and 7.0 to 19.0 (for the exponential function (bootstrap)). These values are fairly similar but the range from the bootstrap analysis is likely to be an overestimate due to the influence of some unexpectedly large points, which are likely to be erroneous. It seems most likely that the age-range at which adult-like performance was reached on this task is between 10 and 16 years of age.

	Age at which adult-like performance was achieved with different methods (to the nearest 0.5 years)		
SNR (dB)	**Exponential function (direct fit)**	**Exponential function (bootstrap)**	**Broken stick function**
20	8.0-12.0	7-9.5	10.5-16.0
10	9.5-13.0	7.5-13.0	10.0-15.0
0	9.5-12.5	7.0-15.5	10.5-16.5
-10	12.0-17.0	7.0-40.0	11.5-18.0
Mean of all conditions	**10.5-14.5**	**7.0-19.0**	**10.5-16.0**
Range over all SNRs	8.0-17.0	7.0-15.5 (ignoring the 40.0 result from SNR -10 dB)	10.0-18.0

Table 6.18: Table summarising different methods of analysis for finding age at which adult-like performance was reached

6.5.1 Comparison to other developmental time courses (MAA in quiet and masked speech perception)

As summarised in the Introduction, there are a few studies with conflicting results about the development of MAAs in quiet in children. These results will now be compared to the findings of this experiment. The development of masked speech perception abilities will also be discussed in relation to the results of this experiment.

Litovsky (1997) found that some five year olds had MAAs in quiet which were adult-like and some did not. The mean MAA for adults was 0.78° and for five year olds was 1.55°. The youngest children tested in this experiment (Experiment 4) were six years of age. Some of them had great difficulty with the task, which may support Litovsky's conclusions that around five years of age children have a significant improvement in their MAA abilities (as some of the youngest children may have not yet developed MAA in quiet abilities). However, this difficulty could also be due to the masker present. The presence

of the masker most likely also explains why in this experiment none of the youngest children performed at adult-like levels, unlike in Litovsky's study.

Kühnle et al. (2013) showed age effects between children aged 6-7 and 13-18 years with their MAA in quiet study. These findings are not dissimilar to those of this experiment, which also showed age effects between these ages. The MAAs for each reference angle and age group for low frequency stimuli and high frequency are shown Table 6.19 and Table 6.20 respectively. From these tables it can be seen that for both low and high frequency stimuli the MAAs decrease with increasing age.

	6-7 years	**8-12 years**	**13-18 years**
0°	7.6°	5.0°	3.2°
±30°	11.4°	6.7°	3.3°
±60°	15.1°	11.6°	4.0°
±90°	16.0°	9.4°	5.9°

Table 6.19: Table showing MAAs for each age group and reference angle for low frequency stimuli (0.3-1.2 kHz Gaussian noise bursts) (Kühnle et al., 2013)

	6-7 years	**8-12 years**	**13-18 years**
0°	10.3°	6.0°	3.8°
±30°	14.7°	9.0°	4.5°
±60°	21.1°	15.2°	6.6°
±90°	21.1°	14.6°	6.9°

Table 6.20: Table showing MAAs for each age group and reference angle for high frequency stimuli (2-8 kHz Gaussian noise bursts) (Kühnle et al., 2013)

Overall, the findings of Litovsky (1997) and Kühnle et al. (2013) suggest development of adult-like MAA in quiet abilities in children vary from the age of five years old and upwards, potentially even as late as 13 years of age. The experiment described in this chapter showed that the development of MAAs in babble varied from 10 to 16 years of age (for the mean MAA of all conditions). This range is higher than those found by Litovsky (1997) and

Kühnle et al. (2013), which is not surprising given that their studies looked at MAAs in quiet. It would be expected that MAAs in quiet would reach adult-like levels before MAAs in babble, as was found to be the case for the localisation of speech in babble task in Chapter 4.

The later development of MAAs of speech in babble compared to MAAs in quiet may be linked to the development of speech reception/recognition in background sound abilities, which have been shown to develop throughout childhood into the teenage years although there are some inconsistencies in the literature. Sobon et al. (2019) showed age-related changes in a sentence recognition task using speech as a masker. They tested children aged 8 to 10 years and young adults and attribute the age-related improvements to maturation in segregation and selective attention. Corbin et al. (2016) showed improvement between the ages of 8-12 years and 13-16 years but not between 13-16 years and adults using a speech recognition task with a two-talker masker. They demonstrated a gradual improvement in 5-13 year olds, followed by a sharp improvement to adult-like levels in 13-14 year olds. However, Jacobi et al. (2017) showed a gradual improvement in speech reception thresholds using broadband masking noise in children aged from 12-17 years of age, with 17 year olds performing at an adult like level. In addition, Rashid, Leensen and Dreschler (2016) showed that speech reception thresholds using broadband masking noise improved gradually in 12-18 year olds. Moore et al. (2019) predict adult-like performance at 14-15 years of age based on their data from 5-13 year olds on a speech in four-talker babble task.

Leibold and Buss (2019) carried out a review into masked speech recognition in children, considering the reasons for age-related improvements seen throughout childhood and into adolescence. They write that the difficulty children experience with masked speech in not due to immaturity in the sensory representation of speech stimuli or in the neural transmission through the pathways of the auditory brainstem. They write that immature abilities to group and/or selectively attend to stimuli may limit how well children perceptually isolate target and masker speech, but that this area is under-studied potentially due to lack of suitable behavioural testing to isolate immature selective attention from poor formation of auditory objects (both of which are detrimental to recognising speech-in-speech and could occur in combination with each other).

Overall, it seems most likely that speech reception/recognition in background sound abilities continue developing into the mid teens. The experiment in this chapter found that MAA of speech in babble abilities reached adult-like levels at approximately 10.5-16.0 years of age. As these age ranges are similar/overlapping, it is possible that spatial hearing in babble and speech reception/recognition in babble abilities are related, however the nature of this relationship is not known. On one hand, it is possible that better speech recognition in background sound abilities contribute to better spatial hearing in babble abilities, as if a listener can use speech recognition abilities to separate out targets and maskers then they should be able to localise better. On the other hand, the relationship might be reversed, where improved

localisation abilities may mean that listeners are better able to localise a sound in babble then they may be better able to focus on it and therefore understand it. Alternatively, it could be that both mechanisms are driven by a common factor.

Cabrera et al. (2019) carried out a study looking at temporal auditory processing, using amplitude modulation in children aged 5-11 years old and in young adults. They found that sensory factors were mature in the youngest children. They used computational modelling to show that age-related improvements in the task were likely due to processing efficiency and so better use of the temporal information of speech. Although this study only used a relatively small age-range it may be that these abilities continue to develop throughout the teenage years, as seen in the localisation of speech in babble tasks in this thesis. If so, then age-related improvements for localisation of speech in babble could also be due to better temporal processing caused by improved processing efficiency.

In Chapter 4, other abilities were highlighted which may develop at similar times to localisation of speech in babble (and also speech identification in noise), potentially due to being driven by a common factor. These were glimpsing, working memory, sustained attention and selective attention. If all of these abilities develop over a similar/overlapping time period then it should be considered what could be causing this. Possible causes could be biological changes in the auditory system and/or behavioural changes.

Litovsky (2015) wrote a review on the development of the auditory system, stating that although many auditory abilities are present at birth many other abilities develop into the teenage years and involve non-auditory pathways such as memory, attention and cognition. In the review, Litovsky states that auditory informational masking (a factor in both localisation in babble and speech and many masked speech perception/identification studies) continues development into the teenage years. It is suggested that the late age-related improvement in informational masking may be due to maturation of the central auditory pathways/cortical maturation. Coch, Sanders and Neville (2005) carried out a complex dichotic listening task looking at sustained auditory selective spatial attention, which showed that auditory event-related potential waveforms are immature in 8 year olds children, meaning that attentional network used for processing linguistic content matures beyond the age of 8 years. Litovsky (2015) writes that this is not surprising, as cortical maturation continues into the teenage years. This view is supported by the work of Karns et al. (2015), who carried out an event-related potential study looking at auditory spatial selective attention in children aged 3-16 years and young adults. They found that auditory spatial selective attention develops throughout the teenage years and state that this is a complex and potentially non-linear developmental trajectory, much like other domains of cognitive function and brain development. They suggest that this prolonged auditory development may be linked to the neural maturation in the frontal and parietal cortices, including increased myelination which leads to faster processing and potentially greater influence of top-down control. They write

that during late childhood and adolescence great changes are seen in brain structure and that improvements in selective attention may be the driver for maturation in many other areas, including cognition and working memory. Therefore, it also seems likely that the wider neurodevelopmental changes which cause improvements in selective attention in late childhood and adolescence also leads to maturation of localisation of speech in babble abilities.

6.5.2 Comparison to results from Experiments 1 and 2

Next, the results from this experiment were compared to the combined results from Experiments 1 and 2. Previously, results from these two studies were shown to have a large correlation of $r_s = 0.7$ and $p < 0.0005$ (Spearman's rank correlation, shown in section 6.4.8).

As the $LocalB_{5AT}$ tasks in Experiments 1 and 2 estimated the SNR at which participants achieved 70.7% correct and this experiment (Experiment 4) estimated the angle at which participants achieved 70.7% correct, they cannot be compared directly. However, the ages at which adult-like performance was achieved on the different tasks can be compared. These values are shown in Table 6.21.

Experiment	Task	Method of finding age at which adult-like performance occurred	Age at which adult-like performance was achieved (to the nearest 0.5 years)
1 and 2	LocalB$_{5AT}$	Exponential function (direct fit)	**13.0-17.5**
		Broken stick function (R analysis)	**13.5-19.0**
		One-way Welch ANOVA	**15+**
		Exponential function (bootstrap)	**10.5-14.0**
		Range over all methods	**10.5-19.0 years**
4	MAA of speech in babble (adaptive track, 2AFC)	Exponential function (direct fit)	SNR 20 dB: 8.0-12.0 SNR 10 dB 9.5-13.0 SNR 0 dB: 9.5-12.5 SNR -10 dB: 12.0-17.0 **Mean of all conditions: 10.5-14.5**
		Broken stick function (R analysis)	SNR 20 dB: 10.5-16.0 SNR 10 dB: 10.0-15.0 SNR 0 dB: 10.5-16.5 SNR -10 dB: 11.5-18.0 **Mean of all conditions: 10.5-16.0**
		Exponential function (bootstrap)	SNR 20 dB: 7.0-9.5 SNR 10 dB: 7.5-13.0 SNR 0 dB: 7.0-15.5 SNR -10 dB: 7.0-40.0 **Mean of all conditions: 7.0-19.0**
		Range over all methods	**10.5-19.0 years**

Table 6.21: Table showing age at which adult-like performance is achieved in different experiments

From this is can be seen that the age ranges when adult-like performance was reached in Experiments 1 and 2 overlaps with the age range within which adult-like performance was reached in this experiment (Experiment 4) for the mean of all conditions of the MAA task. However, the age-ranges for the LocalB$_{5AT}$ task generally start at and extend to older ages than for the mean of all conditions MAA task. This suggests that the LocalB$_{5AT}$ task is generally more difficult than the MAA task. This may be due to the greater flexibility of SNRs used in the LocalB$_{5AT}$ task.

It must be noted that there were differences between the two tasks which must be considered when comparing them. The task in Experiments 1 and 2

only presented one target signal at a time and listeners had to identify the location of it, whereas Experiment 4 presented two signals and listeners had to hear the change in location. Also, Experiments 1 and 2 used SNRs which varied with signal level and Experiment 4 used the same fixed SNRs for all participants. These differences were necessary in order to have both source identification and source discrimination methods for measuring localisation. Overall, the time courses are similar but are not identical. This should not be surprising, given the differences between the experiments which meant that they were measuring localisation in different ways and to different extents.

6.5.3 Summary of results and conclusions

This experiment aimed to further understand the developmental time course of spatial hearing in babble by testing MAAs in children and young adults. It was expected that MAAs would improve following a similar developmental time course as the localisation abilities improved in Experiments 1 and 2. It was also expected that the effect of decreasing the SNR would be reduced with age.

It was shown that MAAs improve with age and a significant difference was found between adults and children in all conditions except 10 dB SNR. A similar developmental time course was shown as in Experiments 1 and 2 (shown in Chapter 4), with adult-like performance generally occurring in the early to mid-teenage years. This is likely linked to the neurodevelopmental changes during this age range, which leads to improvements in many abilities, especially selective attention.

6.5.4 Limitations and future work

The MAA task in this experiment differed from the localisation task used in previous experiments and therefore listeners may have adopted different strategies for completing this task. In the previous task listeners were required to hear the word in the background of babble and then identify the location it came from (out of a choice of 5). In the MAA task listeners heard two words one after the other, each from a different location but one from the right hand side of the speaker ring and the other from the left hand side of the speaker ring. They were then asked which order the sounds occurred in, whether it was right then left or left then right. This task therefore required listeners to hear the two words in the background of babble and be able to tell which side each of them came from. However, listeners could adopt other strategies for this task. For example, instead of comparing the locations of the sounds in order to tell which one came from the right and which one from the left, they could ignore the first word and then judge whether the second word came from the right or left of the room. If a listener adopted this strategy it may have been less effort for them but could have reduced their chances of getting each one correct as they may have benefited from comparing the two words to each other when they were very close together and there is a chance that they might not hear the second word (and if they had heard and paid attention to the first word then they could have potentially known which side it came from). Therefore, this strategy, where listener would not have to rely on their localisation abilities as much as they did in the previous localisation task, may have worked well in the early stages of the adaptive

track, but towards the end when the words were much closer together listeners would need to utilise their localisation abilities more. It cannot be known for sure which strategies listeners chose to use and whether they switched to the most efficient ones throughout the adaptive track as it became more challenging (although tracks were interleaved overall the words would get closer together as time went on). Therefore, when considering the results of this experiment it should be noted that it may not be as good a measure of pure localisation in babble abilities as the previously used localisation in babble task.

The main limitation of this experiment was the limited number of thresholds for each condition. If more thresholds had been returned for each condition then the results would have been more accurate. It may have been better to have collected six thresholds at two different SNRs (e.g. 10 dB and -10 dB) rather than three thresholds at four different SNRs.

As discussed in Chapter 1 and in Chapter 4, there are limitations to using a cross-sectional design as you cannot be certain that the differences seen are due to the factor of interest (in this case age) rather than other factors. Due to the recruitment methods used, it is likely that the groups of children are comparable to each other but the young adults tested (who were all university students or graduates) may not be well matched for comparing to the children. Therefore there is a stronger case for arguing that there are age effects within the groups of children than there are between the children and the young adults. As discussed in Chapter 4, it is possible that due to

recruitment methods the age at which adult-like performance is achieved is over-estimated.

As discussed in section 6.3.3.1, issues were present in the panning code. Although it seems most likely that they have not substantially impacted the results of the current experiment, if a similar experiment was to be carried out it would be beneficial to fix the issues with the code in order to improve the robustness of the results.

In the future it could be useful if this experiment was replicated either using only a few speakers or over headphones, using HRTFs. If this was successful then it could be possible for health professionals to use this task in clinics in order to assess spatial hearing in quiet and in the presence of a masker in patients with hearing loss and/or auditory processing disorder. It could also be used as a diagnostic tool and as a way to measure hearing aid or cochlear implant benefit.

Chapter 7

Conclusions

7.1 Overview

The main aim of this thesis was to measure as accurately as possible the development of localisation of speech in babble in typically developing children with normal hearing. The age at which adult-like performance was reached and reasons behind the development of localisation of speech in babble abilities were also investigated.

This main aim was achieved by mapping the developmental time course of localisation of speech in babble in children aged six years and older, using both source identification and source discrimination methods. Previously, only one published experiment has investigated localisation of speech in babble in children (Boothalingam et al., 2016) and no age effects were found, making the experiments in this thesis the first to successfully demonstrate age effects, map the developmental time course and find the age at which adult-like performance is reached for localisation of speech in babble.

Localisation of speech in babble abilities were further investigated in relation to detection and identification of speech in babble abilities in children aged six years and older and in young adults. This is the first experiment to investigate

these abilities all together, and the first to successfully compare localisation to identification and detection.

The LocalB$_{5AT}$ task, which was developed for use in the experiments in this thesis, is a novel approach to testing localisation of speech in babble in children. It provides a way of testing participants with a variety of ages and abilities on a localisation task and meaningfully comparing their results.

7.2 Key Findings

7.2.1 Development of localisation of speech in babble

The source identification task (LocalB$_{5AT}$), described in Chapters 2, 3 and 4, used an adaptive track that varied the SNR in order to estimate the SNR at threshold for 68 participants aged 6 to 29 years. The data was fitted with an exponential function, which had a time constant of 6.6 years. Adult-like performance was estimated using an ANOVA, exponential functions (direct fit and bootstrap) and a broken stick analysis. It was concluded that an age-range was a better way of defining adult-like performance (rather than a point estimate) and it was most likely that adult-like performance is reached in the early to mid-teenage years.

The source discrimination task (MAA), described in Chapter 6, used an adaptive track that varied separation angle in order to find the smallest difference in locations that a listener could detect. This task was carried out at four different SNRS (20, 10, 0 and -10 dB) for 50 participants aged 6 to 21 years. The data was fitted with an exponential function, the time constants of

which varied with SNR. The age at which adult-like performance occurred also varied with SNR. Adult-like performance was estimated using the exponential functions (direct fit and bootstrap) and a broken stick analysis. As with the source identification task, adult-like performance was reached in the early to mid-teenage years. Measurements showed that the panning code did not work perfectly for this experiment, however the effect of this issue was small so is unlikely to have impacted the results of the experiment.

In Chapters 4 and 6 potential reasons for the age-related improvement of speech in babble abilities are discussed. It seems most likely that the substantial changes in brain structure and neural maturation during late childhood and adolescence, which lead to improved selective attention (Karns et al., 2015), drive the age-related improvements of speech in babble abilities presented in this thesis.

Both of these tasks demonstrated clear improvements in localisation of speech in babble abilities with increasing age for children, although it must be remembered that as the experiments were cross-sectional other factors may have influenced the results (although due to the recruitment methods for children it is likely that they are representative of the general population of this age group). Following this likely age-related improvement plateauing occurred for adults (and children who had reached adult-like levels). When considering the age at which adult-like performance is reached, it is essential to define what adult-like performance is. However, there is no clear way to do this. As adult-like is not a solid concept it is difficult to accurately measure and

define it. Therefore, in these experiments, adult-like performance was defined using age ranges estimated using a variety of methods. Using age ranges, rather than cut-off ages, also seemed more appropriate given the nature of the data where age-related improvement appeared gradual (rather than suddenly improving at a specific age). In the literature there have been a variety of approaches to defining adult-like performance and in this thesis several different methods are used and discussed. In this thesis, the methods used were: ANOVAs, t-tests, exponential functions (direct fit and bootstrap) and broken stick analysis. Although ANOVAs and t-tests are useful for showing age effects between different groups of participants, they are not ideal for identifying the age at which adult-like performance is reached on a task. If a non-significant difference is found between a group of children and adults, this does not mean that they are the same and that the group of children is actually adult-like. The exponential function (bootstrap) analysis was also not the optimum method for estimating when adult-like performance occurred, as due to the way the analysis worked sometimes very large values were given for the parameters of the curve fits. The exponential function (direct fit) method was useful in giving a range of ages at which adult-like performance occurred. However, the broken stick function has some advantages over the exponential function (direct fit) method. The broken stick function fits a flat line after the change point. Although the change point of the broken stick function is not reliable on its own, its 95% confidence intervals provide a good estimate of adult-like performance. For Experiments 1 and 2 the R^2 values were similar for all different analysis methods (broken stick and exponential

function: direct fit and bootstrap). For Experiment 4 the R^2 values were generally similar for both of the exponential function (direct fit and bootstrap) methods. The broken stick method gave lower R^2 values than the exponential functions in two out of four of the conditions (the values were similar in the other two conditions). Despite this, the broken stick function is a good choice for this data as the change point with confidence intervals is a useful measure of the age at which adult-like performance is reached and the flat line following the change point fits the data well as performance did not improve once it reached an adult-like level

For the source identification task the broken stick function gave a change point of 14.8 years with 95% confidence intervals from 13.3 to 18.9 years. For the source discrimination task the broken stick function gave a change point of 13.5 years with 95% confidence intervals from 10.5 to 16.0 years (looking at the mean of all conditions). However, due to the recruitment methods used for all experiments in this thesis (especially for young adult participants) it is possible that these are over-estimates.

7.2.2 Detection, identification and localisation of speech in babble

Experiment 2 investigated detection, identification and localisation of speech in babble in participants aged 10 to 29 years and is described in Chapter 3. This experiment used personalised SNRs for each participant, based on their performance in the localisation of speech in babble task (which was a source identification task). Generally participants scored highest in identification,

followed by detection and localisation. This contrasted with the findings of Experiment 3, which tested adults aged 18 to 30 years on the same tasks but with set SNRs. Experiment 3, which tested young adults only (using harder SNRs), found that participants scored highest in detection, followed by identification and one interval-four choice localisation (which were the same) and then one interval-five choice localisation. These results suggest that even if a listener can detect and identify a sound they cannot necessarily localise it. Although previous studies (Hawkins and Stevens, 1950; Liu and Jin, 2019) have found that identification can be better than detection in background sound, these studies are quite different to the ones in this thesis. Differences in findings may be due to the different methods, i.e. different choice of maskers (the experiments in this thesis used informational maskers, whereas the previous studies used energetic maskers) and the use of more challenging SNRs, especially in the adult experiment.

7.3 Limitations and future work

7.3.1 Experiment specific limitations

In the source identification experiment (Chapters 2, 3 and 4) the developmental time course of localisation of speech in babble abilities were mapped for participants aged 6 to 29 years of age. However, there were only four participants in the age range of 17 to 22 years (with no participants aged 20 or 21 years). The time course would be improved if more participants of these ages were included and the estimates of the age at which adult-like performance is reached may be made more accurate.

In the source discrimination experiment (MAA, Chapter 6) only three tracks were carried out for each of the four SNRs (so a total of 12 tracks). The results would be more accurate if a greater number of tracks had been carried out for each participant. The decision to only have this number was made due to difficulties with recruiting participants for long studies (or even studies with two visits). However, if more time had been available then investing more time in recruitment and testing in order to run more tracks would have improved this experiment.

The detection in babble task used in Experiment 2 may have been too long to maintain the children's attention. It lasted 15 minutes, which was twice as long as the other tasks they completed (detection and identification of speech in babble). It may have been better to have split it into two tasks and given the children a break between them. This could have helped them to stay focused and potentially improve their performance, making results more accurate and comparable to those of adults (who have better attentional abilities).

The detection, identification and localisation of speech in babble tasks with adult participants (Chapter 5) used challenging SNRs (-33, -30, -27, -24 and -21 dB). Although piloting was carried out in order to guide the choice of these SNRs, it would have been better to include an easier SNR at -18 dB.

Experiments 1, 2 and 4 used cross-sectional designs and due to recruitment methods, it may be that the young adults are not as comparable to the children as the different age groups of children are to each other. The main

impact of this is that the age at which adult-like performance occurs may have been overestimated. However, there are clear age effects between the different age groups of children. The ideal method for this type of experiment would be a longitudinal experiment, but this was not possible due to the limited time available.

7.3.2 General limitations

One of the main ways in which the experiments described in this thesis could potentially be improved is by increasing the number of tracks. However, this has to be balanced with ensuring that child participants are still engaged with the task otherwise they will lose attention and their performance will get worse. Even if more tracks were carried out over multiple sessions (rather than just having longer sessions with lots of breaks), participants may become bored with the tasks and their motivation could decrease. Therefore, it is important to balance the number of tracks with maintaining engagement of the participant. Overall, in most of the experiments the results infer that this balance was maintained quite well. However, the bootstrap analysis suggested that the MAA of speech in babble task (described in Chapter 4) probably would have benefitted from having a greater number of tracks to improve accuracy (or more tracks for each condition, but fewer conditions).

There are a couple of practical elements of the experiments which could possibly be improved. Firstly, the experiments were carried out in a soundproof test room, similar to those used for clinical hearing tests. Although this is a very good environment for carrying out hearing studies, it is

not as good as an anechoic chamber. Secondly, more loudspeakers could have been used. Although the panning between speakers worked well, having a greater number of speakers (for example having a speaker every 5°, rather than every 15°) would give more accurate presentation of signals. However, it is important to remember that children do not often sit still, especially when they have been doing a task for a while and may be starting to get bored. Therefore, any improvements gained by using an anechoic chamber or more speakers may be insignificant when accounting for the loss in accuracy caused by children moving their heads from the ideal positions.

A limitation of testing speech in babble became clear in the early stages of Experiment 1. Children (and some adults) who were fluent English speakers but had not learned English since birth had great difficulty with the localisation of speech in babble task (even if they had normal localisation in quiet abilities). This meant that it was very important to ensure participants for all the experiments described in this thesis had learned English since birth, otherwise confounds would be introduced into the data.

7.3.3 Future work

The main outcome of these experiments was the mapping of the developmental time course of localisation of speech in babble in typically developing English-speaking children with normal hearing. This will hopefully provide a benchmark for other researchers and clinicians working with populations who may have difficulties with spatial hearing in babble, such as children with Auditory Processing Disorder or hearing loss. It could be

beneficial for these patient groups to also be tested on these tasks, to understand more about their development of localisation of speech in babble in relation to normal development for their age. Knowing how much these abilities are delayed or deficient in these groups could help to guide rehabilitation. The tasks used in this thesis could be adapted for use over headphones, to make them suitable for use in clinics or research facilities without the space for a ring of loudspeakers.

It could also be beneficial to measure localisation of speech in babble in children using different types of localisation tasks. In this thesis two different methods were used, which gave slightly different results. This helped to provide a broader picture of the development of localisation in babble. Therefore, using different methods again may give an improved estimate of when localisation of speech in babble reaches adult-like levels. An example of another type of task would be carrying out a source identification task with RMS error as the performance measure.

In these experiments the youngest participants were six years of age. This age was chosen as it has previously been found that children of this age can localise to an adult-like level in quiet (Van Deun et al., 2009; Lovett et al., 2012; Kühlne et al., 2013; Otte et al., 2013). However, it would be interesting to further investigate the localisation of speech in babble abilities of even younger children. In these experiments a large proportion of six year olds were unable to complete the tasks, so any experiment seeking to do this would need to employ different methods which would allow larger angles of

separation and easier SNRs to be used. Tasks would also need to be shorter, have even more breaks and be more engaging in order to maintain the attention of younger children. A possible way to do this may be to gamify the experiment tasks.

It is likely that localisation of speech in babble abilities underpin the use of head orientations to support listening to speech in babble. It is known that head orientations are only utilised by normally hearing adults approximately half of the time and even then only half of these orientations provide optimum listening conditions (Grange and Culling, 2016a; 2016b). The underutilisation of head orientations is also shown by Brimijoin, McShefferty and Akeroyd (2012), and can most likely be attributed to the challenges of auditory processing presented by listening to masked speech, the necessity to align with social norms such as facing a speaker and not wanting to lose the benefits gained by utilising audiovisual cues. However, we have little understanding about whether or how children use orientation to support listening in background sound. Studies by Ricketts and Galster (2009) and Ching et al (2009) suggest that children use head orientations in a variety of environments. However these results do not directly tell us how children use head orientations in background sound. Therefore, it would be interesting to carry out a experiment investigating whether children use head orientation to improve listening to masked speech and how beneficial the head orientations they choose are.

7.4 Summary

This thesis aimed to measure the developmental time course of localisation of speech in babble abilities and to estimate the age at which adult-like performance is reached. Localisation was measured in the horizontal plane and the frontal hemifield, using both source identification and discrimination methods. Clear improvement with age was seen throughout childhood and into the early-mid teenage years. Multiple analysis techniques were used, but a broken stick function was found to be the most appropriate method. Adult-like performance has been estimated to occur in the early to mid-teenage years, although this may be an over-estimate.

The relationships between detection, identification and localisation of speech in babble abilities were also investigated. It was found that even if a listener was able to detect and/or identify a sound they could not necessarily localise it.

References

Akeroyd, M.A. and Whitmer, W.M. (2016) Spatial Hearing and Hearing Aids. In: G.R. Popelka et al., ed. 2016. *Hearing Aids*. Switzerland: Springer. pp. 181-215.

Altman, D. G. and Bland, J. M. (1995) 'Absence of evidence is not evidence of absence', *BMJ*, 311(7003), pp. 485.

Barry J. G. and Moore D. R. (2014) Evaluation of Children's Listening and Processing Skills (ECLiPS) London, United Kingdom: MRC-T.

Barry, J. G., Tomlin, D., Moore, D. R. and Dillon, H. (2015) 'Use of Questionnaire-Based Measures in the Assessment of Listening Difficulties in School-Aged Children', *Ear and Hearing*, 36(6), pp. e300-13.

Batteau, D. W. (1967) 'The role of the pinna in human localization', *Proceedings of the Royal Society B Biological Sciences*, 168(1011) pp. 158-80.

Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M. and Rzeczkowski, C. (1984) 'Standardization of a test of speech perception in noise', *Journal of Speech and Hearing Research*, 27, pp. 32-48.

Blauert, J. (1997) *Spatial Hearing: The Psychophysics of Human Sound Localization*. 2nd ed. Cambridge, Massachussets: The MIT Press.

Bolia, R. S., Nelson, W. T., Ericson, M. A. and Simpson, B. D. (2000) 'A speech corpus for multitalker communications research', *Journal of the Acoustical Society of America*, 107(2), pp. 1065-6.

Boothalingam, S., Macpherson, E., Allan, C., Allen, P. and Purcell, D. (2016) 'Localization-in-noise and binaural medial olivocochlear functioning in children and young adults', *Journal of the Acoustical Society of America,* 139(1), pp. 247-62.

Boothalingam, S., Purcell, D. W., Allan, C., Allen, P. and Macpherson, E. (2019) 'Auditory Localization and Spatial Release From Masking in Children With Suspected Auditory Processing Disorder', *Ear and Hearing*, 40(5), pp. 1187-96.

Brimijoin, W. O. and Akeroyd, M. A. (2012) 'The role of head movements and signal spectrum in an auditory front/back illusion', *i-Perception*, 3(3), pp. 179-82.

Brimijoin, W. O. and Akeroyd, M. A. (2016) 'The Effects of Hearing Impairment, Age, and Hearing Aids on the Use of Self-Motion for Determining Front/Back Location', *Journal of the American Academy of Audiology*, 27(7), pp. 588-600.

Brimijoin, W. O., McShefferty, D. and Akeroyd, M. A. (2012) 'Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task', *Hearing Research*, 283(1-2), pp. 162-8.

Brungart, D. S., Simpson, B. D., Ericson, M. A. and Scott, K. R. (2001) 'Informational and energetic masking effects in the perception of multiple

simultaneous talkers', *Journal of the Acoustical Society of America*, 110(5), pp. 2527-38.

Buss, E., Leibold, L. J., Porter, H. L. and Grose, J. H. (2017) 'Speech recognition in one- and two-talker maskers in school-age children and adults: Development of perceptual masking and glimpsing', *Journal of the Acoustical Society of America*, 141(4), pp. 2650.

Cameron, S., Brown, D., Keith, R., Martin, J., Watson, C. and Dillon, H. (2009) 'Development of the North American Listening in Spatialized Noise-Sentences Test (NA LiSN-S): Sentence Equivalence, Normative Data, and Test-Retest Reliability Studies', *Journal of the American Academy of Audiology*, 20(2), pp. 128-46.

Cabrera, L., Varnet, L., Rosen, S. and Lorenzi, C. (2019) 'Development of temporal auditory processing in childhood: Changes in efficiency rather than temporal-modulation selectivity', *Journal of the Acoustical Society of America*, 146(4), pp. 2415-29.

Ching, T. Y. C., O'Brien, A., Dillon, H., Chalupper, J., Hartley, L., Hartley, D., Raicevich, G. and Hains, J. (2009) 'Directional Effects on Infants and Young Children in Real Life: Implications for Amplification', *Journal of Speech, Language, and Hearing Research*, 52(5), pp. 1241-54.

Ching, T. Y., van Wanrooy, E., Dillon, H. and Carter, L. (2011) 'Spatial release from masking in normal-hearing children and children who use hearing aids', *Journal of the Acoustical Society of America*, 129(1), pp. 368-75.

Clifton, R. K., Gwiazda, J., Bauer, J. A., Clarkson, M. G., & Held, R. M. (1988). Growth in head size during infancy: Implications for sound localization. *Developmental Psychology*, 24(4), pp. 477–83.

Coch, D., Sanders, L.D. and Neville, H.J. (2005) 'An event-related potential study of selective auditory attention in children and adults', *Journal of Cognitive Neuroscience*, 14(4), pp. 605-22.

Conners, C. K., Sitarenios, G., Parker, J. D. A., and Epstein, J. N. (1998) 'Revision and restandardization of the Conners Teacher Rating Scale: Factor structure, reliability, and criterion validity', *Journal of Abnormal Child Psychology,* 26(4), pp. 279-91.

Corbin, N. E., Bonino, A. Y., Buss, E. and Leibold, L. J. (2016) 'Development of Open-Set Word Recognition in Children: Speech-Shaped Noise and Two-Talker Speech Maskers', *Ear and Hearing*, 37(1), pp. 55-63.

Crone, E. A., Wendelken, C., Donohue, S., van Leijenhorst, L. and Bunge, S. A. (2006) 'Neurocognitive development of the ability to manipulate information in working memory', *Proceedings of the National Academy of Sciences of the United States of America*, 103(24), pp. 9315-20.

Drullman, R. and Bronkhorst, A. W. (2000) 'Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation', *Journal of the Acoustical Society of America*, 107(4), pp. 2224-35.

Efron, B. and Tibshirani, R. (1993) *An introduction to the bootstrap*. New York: Chapman & Hall.

Egami, C., Yamashita, Y., Tada, Y., Anai, C., Mukasa, A., Yuge, K., Nagamitsu, S. and Matsuishi, T. (2015) 'Developmental trajectories for attention and working memory in healthy Japanese school-aged children', *Brain & Development*, 37(9), pp. 840-8.

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J. and Boothroyd, A. (2000) 'Speech recognition with reduced spectral cues as a function of age', *Journal of the Acoustical Society of America*, 107(5 Pt 1), pp. 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', *Journal of the Acoustical Society of America*, 66(3), pp. 651-3.

Feddersen, W.E., Sandel, T.T., Teas, D.C., Jeffress, L.A. (1957) 'Localization of High-Frequency Tones', *Journal of the Acoustical Society of America*, 29(9), pp. 988-91.

Field, A. (2012) *Discovering Statistics Using IBM SPSS Statistics*. 4th ed. London: Sage.

Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K. and Esterman, M. (2015) 'Sustained Attention Across the Life Span in a Sample of 10,000: Dissociating Ability and Strategy', *Psychological Science*, 26(9), pp. 1497-510.

Freyman, R.L., Balakrishnane, U. and Helfer, K.S. (2001) ‘Spatial release from informational masking in speech recognition’, *Journal of the Acoustical Society of America*, 109(5), pp. 2112-22.

Games, P.A. and Howell, J.F. (1976) ‘Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: A Monte Carlo Study’, *Journal of Educational Statistics*, 1(2), pp. 113-25.

Garadat, S. N. and Litovsky, R. Y. (2007) 'Speech intelligibility in free field: Spatial unmasking in preschool children', *Journal of the Acoustical Society of America*, 121(2), pp. 1047-55.

Graber (2019). *MSD MANUAL Professional Version: Physical Growth of Infants and Children*. [online] Available at: https://www.msdmanuals.com/en-gb/professional/pediatrics/growth-and-development/physical-growth-of-infants-and-children [Accessed 16th February 2020].

Grange, J. A. and Culling, J. F. (2016a) 'Head orientation benefit to speech intelligibility in noise for cochlear implant users and in realistic listening conditions', *Journal of the Acoustical Society of America*, 140(6), pp. 4061-72.

Grange, J. A. and Culling, J. F. (2016b) 'The benefit of head orientation to speech intelligibility in noise', *Journal of the Acoustical Society of America*, 139(2), pp. 703-12.

Greenhouse, S. W., and Geisser, S. (1959). On the methods in the analysis of profile data. *Psychometrika*, 24, pp. 95-112.

Haggard, M. (1993) *Research in the development of effective services for hearing-impaired people*. London: Nuffield Provinicial Hospitals Trust.

Hartmann, W.M., Rakerd, B. and Gaalaas, J.B. (1998) 'On the source-identification method', *Journal of the Acoustical Society of America*, 104(6), pp. 3546-57.

Hawkins, J. E. and Stevens, S. S. (1950) 'The Masking of Pure Tones and of Speech by White Noise', *Journal of the Acoustical Society of America*, 22(1), pp. 6-13.

Hawley, M. L., Litovsky, R. Y. and Colburn, H. S. (1999) 'Speech intelligibility and localization in a multi-source environment', *Journal of the Acoustical Society of America, 105*(6), pp. 3436-3448.

Hirsh, I. J. (1948) 'Binaural summation and interaural inhibition as a function of the level of masking noise', *American Journal of Psychology*, 61(2), pp. 205-13.

Huelke, D.F., 1998. An Overview of Anatomical Considerations of Infants and Children in the Adult World of Automobile Safety Design. *Annual Proceedings/Association for the Advancement of Automotive Medicine*, 42, pp. 93-113.

IEEE (1969) 'IEEE Recommended Practices for Speech Quality Measurements', *IEEE Transactions on Audio and Electroacoustics*, 17(3), pp. 225-46.

Jacobi, I., Sheikh Rashid, M., de Laat, J. and Dreschler, W. A. (2017) 'Age Dependence of Thresholds for Speech in Noise in Normal-Hearing Adolescents', *Trends in Hearing*, 21, pp. 2331216517743641.

Johnson, C. E. (2000) 'Children's phoneme identification in reverberation and noise', *Journal of Speech, Language, and Hearing Research*, 43(1), pp. 144-57.

Jones, P. R. (2018) 'The development of perceptual averaging: Efficiency metrics in children and adults using a multiple-observation sound-localization task', *Journal of the Acoustical Society of America*, 144(1), pp. 228-41.

Kalikow, D. N., Stevens, K. N. and Elliott, L. L. (1977) 'Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability', *Journal of the Acoustical Society of America*, 61(5), pp. 1337-51.

Karns, C. M., Isbell, E., Giuliano, R. J. and Neville, H. J. (2015) 'Auditory attention in childhood and adolescence: An event-related potential study of spatial selective attention to one of two simultaneous stories', *Developmental Cognitive Neuroscience*, 13, pp. 53-67.

Kitterick, P. T., Bailey, P. J. and Summerfield, A. Q. (2010) 'Benefits of knowing who, where, and when in multi-talker listening', *Journal of the Acoustical Society of America*, 127(4), pp. 2498-508.

Klatte, M., Lachmann, T. and Meis, M. (2010) 'Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting', *Noise and Health*, 12(49), pp. 270-82.

Klenberg, L., Korkman, M. and Lahti-Nuuttila, P. (2001) 'Differential development of attention and executive functions in 3- to 12-year-old Finnish children', *Developmental Neuropsychology*, 20(1), pp. 407-28.

Klumpp, R.G. and Eady, H.R. (1956) 'Some measurements of interaural time difference thresholds', *Journal of the Acoustical Society of America*, 28(5), pp. 859-60.

Kolarik, A. J., Moore, B. C., Zahorik, P., Cirstea, S. and Pardhan, S. (2016) 'Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss'*, Attention, Perception & Psychophysics*, 78(2), pp. 373-95.

Kopčo, N., Best, V. and Carlile, S. (2010) 'Speech localization in a multitalker mixture', *Journal of the Acoustical Society of America*, 127(3), pp. 1450-7.

Kühnle, S., Ludwig, A. A., Meuret, S., Kuttner, C., Witte, C., Scholbach, J., Fuchs, M. and Rubsamen, R. (2013) 'Development of auditory localization accuracy and auditory spatial discrimination in children and adolescents', *Audiology and Neurotology*, 18(1), pp. 48-62.

Leibold, L. J. (2017) 'Speech Perception in Complex Acoustic Environments: Developmental Effects', *Journal of Speech, Language, and Hearing Research*, 60(10), pp. 3001-8.

Leibold, L. and Buss, E. (2019) 'Masked Speech Recognition in School-Age Children', *Frontiers in Psychology*, 10:1981. doi: 10.3389/fpsyg.2019.01981 [Accessed 8th March 2020].

Lentz, J. (2018) *Psychoacoustics: Perception of Normal and Impaired Hearing with Audiology Applications.* San Diego: Plural Publishing.

Levitt, H. (1971) 'Transformed up-down methods in psychoacoustics', *Journal of the Acoustical Society of America*, 49(2B), pp. 467-77.

Litovsky, R. (1997) 'Developmental changes in the precedence effect: Estimates of minimum audible angle', *Journal of the Acoustical Society of America*, 102(3), pp. 1739-45

Litovsky, R. (2005) 'Speech intelligibility and spatial release from masking in young children', *Journal of the Acoustical Society of America*, 117(5), pp. 3091-9.

Litovsky, R. (2015) 'Development of the auditory system', *Handbook of Clinical Neurology*, *129*, pp. 55–72.

Liu, C. and Jin, S. H. (2019) 'Psychometric Functions of Vowel Detection and Identification in Long-Term Speech-Shaped Noise', *Journal of Speech Language and Hearing Research*, 62(5), pp. 1473-85.

Lovett, R. E. S., Kitterick, P. T., Huang, S. and Summerfield, A. Q. (2012) 'The Developmental Trajectory of Spatial Listening Skills in Normal-Hearing Children', *Journal of Speech Language and Hearing Research*, 55(3), pp. 865-78.

Lufi, D. and Haimov, I. (2018) 'Effects of age on attention level: changes in performance between the ages of 12 and 90', *Aging, Neuropsychology, and Cognition*, 26(6), pp. 904-19.

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A. and Sweeney, J. A. (2004) 'Maturation of cognitive processes from late childhood to adulthood', *Child Development*, 75(5), pp. 1357-72.

Macmillan, N.A. and Creelman, C.D. 2008. *Detection Theory: A User's Guide*. Second edition. Hove: Psychology Press.

Martin, K., Johnstone, P. and Hedrick, M. (2015) 'Auditory and visual localization accuracy in young children and adults', *International Journal of Pediatric Otorhinolaryngology*, 79(6), pp. 844-51.

Martin, K., Johnstone, P. and Hedrick, M. (2015). Auditory and visual localization accuracy in young children and adults. *International Journal of Pediatric Otorhinolaryngology*, 79, pp. 844-51.

McFadden, D. (1966) 'Masking-level differences with continuous and with burst masking noise', *Journal of the Acoustical Society of America*, 40(6), pp. 1414-9.

Middlebrooks, J. C. and Green, D. M. (1991) 'Sound localization by human listeners', *Annual Reviews of Psychology*, 42, pp. 135-59.

Mills, A. W. (1958) 'On the Minimum Audible Angle', *Journal of the Acoustical Society of America*, 30(4), pp. 237-46.

Mills, A.W. (1960) 'Lateralization of High-Frequency Tones', *Journal of the Acoustical Society of America*, 32(1), pp. 132-4.

Misurelli, S. M. and Litovsky, R. Y. (2012) 'Spatial release from masking in children with normal hearing and with bilateral cochlear implants: effect of

interferer asymmetry', *Journal of the Acoustical Society of America*, 132(1), pp. 380-91.

Moore, B.C.J. (2012) *An Introduction to the Psychology of Hearing.* 6th ed. Bingley: Emerald.

Moore, D. R., Cowan, J. A., Riley, A., Edmondson-Jones, A. M. and Ferguson, M. A. (2011) 'Development of auditory processing in 6- to 11-yr-old children', *Ear and Hearing*, 32(3), pp. 269-85.

Moore, D.R, Whiston, H. Lough, M., Marsden, A., Dillon, H., Munro, K.J. and Stone, A. (2019) 'FreeHear: A new sound-field speech-in-babble hearing assessment tool', *Trends in Hearing*, 23. doi:10.1177/2331216519872378. [Accessed 14th March 2020].

Morrongiello, B. A. (1987) 'Infants' localization of sounds in the median sagittal plane: Effects of signal frequency', *Journal of the Acoustical Society of America*, 82(3), pp. 900-5.

Morrongiello, B. A. and Rocca, P. T. (1987) 'Infants' Localization of Sounds in the Horizontal Plane: Effects of Auditory and Visual Cues', *Child Development*, 58(4), pp. 918-27.

Morrongiello, B. A., Fenwick, K. D., Hiller, L. and Chance, G. (1994) 'Sound Localization in Newborn Human Infants', *Developmental Psychobiology*, 27(8), pp. 519-38.

Muir, D. and Field, J. (1979) 'Newborn Infants Orient to Sounds', *Child Development*, 50(2), pp. 431-6.

Muir, D. and Hains, S. (2004) 'The U-Shaped Developmental Function for Auditory Localization', *Journal of Cognition and Development*, 5(1), pp. 123-130.

Muir, D. W., Clifton, R. K. and Clarkson, M. G. (1989) 'The Development of a Human Auditory Localization Response: A U-Shaped Function', *Canadian Journal of Psychology*, 43(2), pp.199-216.

Murphy, J., Summerfield, A. Q., O'Donoghue, G. M. and Moore, D. R. (2011) 'Spatial hearing of normally hearing and cochlear implanted children', *International Journal of Pediatric Otorhinolaryngology,* 75(4), pp. 489-94.

Musicant, A. D. and Butler, R. A. (1984) 'The influence of pinnae-based spectral cues on sound localization', *Journal of the Acoustical Society of America*, 75(4), pp. 1195-200.

Musiek, F.E. (1983) ‘Assessment of central auditory dysfunction: the dichotic digit test revisited’, *Ear and Hearing*, 4(2), pp. 79-83.

Németh, R., Haden, G. P., Török, M., and Winkler, I. (2015) 'Processing of Horizontal Sound Localization Cues in Newborn Infants', *Ear and Hearing*, 36(5), pp. 550-6.

Otte, R. J., Agterberg, M. J. H., Van Wanrooij, M.M., Snik, A. F. M. and Van Opstal, A. J. (2013) 'Age-related Hearing Loss and Ear Morphology Affect Vertical but not Horizontal Sound-Localization Performance', *Journal of the Association for Research in Otolaryngology*, 14(2), pp. 261-73.

Perrott, D.R. and Saberi, K. (1990) ‘Minimum audible angle thresholds for sources varying in both elevation and azimuth’, *Journal of the Acoustical Society of America*, 87(4), pp. 1728-31.

Rashid, M. S., Leensen, M. C. and Dreschler, W. A. (2016) 'Application of the online hearing screening test "Earcheck": Speech intelligibility in noise in teenagers and young adults', *Noise and Health*, 18(85), pp. 312-8.

Rebok, G. W., Smith, C. B., Pascualvaca, D. M., Mirsky, A. F., Anthony, B. J. and Kellam, S. G. (1997) 'Developmental changes in attentional performance in urban children from eight to thirteen years', *Child Neuropsychology*, 3(1), pp. 28-46.

Ricketts, T. A. and Galster, J. (2008) 'Head Angle and Elevation in Classroom Environments: Implications for Amplification', *Journal of Speech, Language, and Hearing Research*, 51(2), pp. 516-25.

Schafer E.C., Beeler, S., Ramos H., Morais, M., Monzingo J. and Algier, K. (2012) 'Developmental Effects and Spatial Hearing in Young Children With Normal-Hearing Sensitivity', *Ear and Hearing*, 33(6).

Schwarz, C. J. (2015). Regression - hockey sticks, broken sticks, piecewise, change points. In Course Notes for Beginning and Intermediate Statistics. Available at http://www.stat.sfu.ca/~cschwarz/CourseNotes. Retrieved 2019-09-19.

Shaw, E. A. G. (1974a) 'Acoustic Response of External-Ear Replica at Various Angles of Incidence', *Journal of the Acoustical Society of America*, 55(2), pp. 432.

Shaw, E. A. G. (1974b) 'Wave properties of the human external ear and various physical models of the ear', *Journal of the Acoustical Society of America*, 56(S3).

Shield, B. M. and Dockrell, J. E. (2008) 'The effects of environmental and classroom noise on the academic attainments of primary school children', *Journal of the Acoustical Society of America*, 123(1), pp. 133-44.

Simpson, B.D., Brungart, D.S., Iyer, N., Gilkey, R.H. and Hamil, J.T. (2006) 'Detection and localization of speech in the presence of competing speech signals,' *Proceedings of the 12th International Conference on Auditory Display.* London, UK, 20-23rd May 2006, pp. 129-33.

Simpson, S. A. and Cooke, M. (2005) 'Consonant identification in N-talker babble is a nonmonotonic function of N', *Journal of the Acoustical Society of America*, 118(5), pp. 2775-2778.

Sobon, K., Taleb, N.M., Buss, E. and Grose, J.H. (2019) 'Psychometric function slope for speech-in-noise and speech-in-speech: Effects of development and aging', *Journal of the Acoustical Society of America*, 145(4), pp. EL284-90.

Spear, L.P. (2000) 'The adolescent brain and age-related behavioral manifestations', *Neuroscience & Biobehavioral Reviews*, 24(4), pp. 417-63.

Stollman, M. H., van Velzen, E. C., Simkens, H. M., Snik, A. F. and van den Broek, P. (2004) 'Development of auditory processing in 6-12-year-old children: a longitudinal study', *International Journal of Audiology*, 43(1), pp. 34-44.

Strutt, J.W. (1909). 'On our perception of sound direction', *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13(74), pp. 214-32.

Stuart, A. (2005) 'Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise', *Ear and Hearing*, 26(1), pp. 78-88.

Tabri, D., Abou Chacra, K.M. and Pring, T. (2011) 'Speech perception in noise by monolingual, bilingual and trilingual listeners', *International Journal of Language & Communication Disorders*, 46(4), pp. 411-22.

Trehub, S. E., Bull, D. and Schneider, B. A. (1981) 'Infants' detection of speech in noise', *Journal of Speech and Hearing Research*, 24(2), pp. 202-6.

Vaillancourt, V. C. L., Giguère C. and Soli, S.D. (2008) 'Establishment of Age-Specific Normative Data for the Canadian French Version of the Hearing in Noise Test for Children', *Ear and Hearing*, 29(3), pp. 453-66.

Van Deun, L., Van Wieringen A., Van den Bogaert, T., Scherf, F., Offeciers, F.E., Van de Heyning, P.H., Desloovere, C., Dhooge, I.J., Deggouj, N., De Raeve, L., and Wouters, J. (2009) 'Sound Localization, Sound Lateralization, and Binaural

Masking Level Differences in Young Children with Normal Hearing', *Ear and Hearing*, 30(2), pp. 178-90.

von Hapsburg, D., Champlin, C.A., Shetty, S.R. (2004) 'Reception thresholds for sentences in bilingual (Spanish/English) and monolingual (English) listeners', *Journal of the American Academy of Audiology*, 15(1), pp. 88-98.

Wallach, H. (1940) 'The role of head movements and vestibular and visual cues in sound localization', *Journal of Experimental Psychology*, 27(4), pp. 339-68.

Wechsler D. (1991) Wechsler Intelligence Scale for Children. 3rd ed. The Psychological Corporation; San Antonio, Tex.

Wechsler, D. (1999) Wechsler Abbreviated Scale of Intelligence. The Psychological Corporation: Harcourt Brace & Company. New York, NY.

Wightman, F. L. and Kistler, D. J. (1999) 'Resolution of front–back ambiguity in spatial hearing by listener and source movement', *Journal of the Acoustical Society of America*, 105(5), pp. 2841-53.

Wightman, F. L. and Kistler, D. J. (2005) 'Informational masking of speech in children: effects of ipsilateral and contralateral distracters', *Journal of the Acoustical Society of America*, 118(5), pp. 3164-76.

Yuen, K. C. and Yuan, M. (2014) 'Development of spatial release from masking in mandarin-speaking children with normal hearing', *Journal of Speech, Language, and Hearing Research*, 57(5), pp. 2005-23.