Spatial release from masking in children with and without auditory processing disorder in real and virtual auditory environments

Katharina Zenke

A dissertation submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Speech, Hearing and Phonetic Sciences
Division of Psychology and Language Sciences
University College London

May 2021
Declaration

I, Katharina Zenke, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.
Abstract

Auditory Processing Disorder (APD) is a developmental disorder characterised by difficulties in listening to speech-in-noise despite normal audiometric thresholds. It is still poorly understood and much disputed and there is a need for better diagnostic tools.

One promising finding is that some children referred for APD assessment have a reduced spatial release from masking (SRM). Current clinical tests measure SRM in virtual auditory environments created from head-related transfer functions (HRTFs) of a standardised adult head. Adults and children, however, have different head dimensions and mismatched HRTFs are known to affect aspects of binaural hearing like localisation. There has been little research on HRTFs in children and it is unclear whether a large mismatch can impact speech perception, especially for children with APD who have difficulties with accurately processing auditory information.

In this project, we examined the effect of nonindividualised virtual auditory environments on the SRM in adults and children with and without APD. The first study with normal-hearing adults compared environments created from individually measured HRTFs and two nonindividualised sets of HRTFs to a real anechoic environment. Speech reception thresholds (SRTs) were measured for target sentences at 0° and two symmetric speech maskers at 0° or ±90° azimuth. No significant effect of auditory environment on SRTs and SRM could be observed. A larger study was then conducted with APD and typically-developing children aged 7 to 12 years. Individual HRTFs were measured for each child. The SRM was measured in environments created from these individualised HRTFs or artificial head HRTFs and in the real anechoic environment. To assess the influence of spectral cues, SRTs were also measured for HRTFs from a spherical head model that only contains interaural time and level differences. Additionally, the study included an extended high-frequency audiogram, a receptive language test and two parental questionnaires. The SRTs of children with APD were worse than those of typically-developing children in all conditions but SRMs were similar. Only small differences in SRTs were found across environments, mainly for the spherical head HRTFs. SRTs in children were higher than in adults but improved with age. APD children also had higher hearing thresholds and performed worse in the language test.
Acknowledgements

The past years have been quite a journey - not always in the way I had expected. I am thankful for everyone who has accompanied and supported me throughout this experience.

Thank you, Stuart Rosen, for your support and for always being so positive and inspirational and full of new ideas. Talking to you always makes me feel motivated and excited about my work and eager to start the next part. Thanks for always taking the time to discuss my studies in depth and for sharing your passion for hearing science.

Thank you, Shiran Koifman, for the countless discussions we had, the shared challenges of ethics applications and recruitment and the shared joy of testing kids. We have been going through a lot together and it has been great to have you for support and motivation and to share those early morning trips to schools in Kent.

Thank you, Tim Schoof and Laurianne Cabrera, for being so supportive and inspiring and everyone in the SpHeariCal group for all the interesting discussions.

Thank you, Rachel Zheng, for being the best desk neighbour one could ask for, always being up to start the workday with a coffee break and going on fun weekend trips to escape the city. Thanks, Yue, Shego, Gwijde, Faith, Ana, Anqi, Giulia, Jiem, Joaquin, Gise, Emma, Hanin, Kurt, Dan, Max, Anna, Helen, Chengxia, Julie, Bryony, Elise, Clara, Han, Magda and everyone else in the PhD office. It has been amazing to get to know and work with so many people from different cultural and educational backgrounds.

Thank you, Andrew Clark, for helping me with my experiments, the emergency mic-repair service and for making the journey over to the anechoic room countless times. Thank you, Dave Cushing, for the help with the equipment and for my first, best and only caving weekend; and all the other staff at Chandler House for their help. Thank you, Doris-Eva Bamiou, for helping with the recruitment of the APD children and with giving me more insight into clinical practice. Thank you, to everyone who has helped by providing expertise or software for the HRTF measurements, namely Tim Green, Mark Huckvale, Gaston Hilkuysen, Lorenzo Picinali, Jan-Gerrit Richter and Torsten Marquardt.

Thanks also for the support of my two funders: the ETN iCARE which let me be part of a great international community of young researchers and gave me the chance to
experience industrial research, and the Royal National Institute for Deaf People which taught me to communicate my research outside academia.

Thank you, Benjamin, for always being by my side and supporting me through all the ups and downs that come with a PhD project. Thanks for moving to London for me despite the lack of mountains and your difficult relationship with single glazed windows.

Thank you, mum and dad, for always supporting me and for letting me move back in to spend half a pandemic with you. Thanks for making working from home so much better with lunch break swims in the lake and rhubarb cake.

Last but not least, thank you also to all my participants and especially all the children and their families who came into the lab to take part in my study. Without you, this project would not have been possible. Conducting my studies with children, although challenging at times, has been a lot of fun and was one of my favourite part of this project. Talking with parents and children has taught me a lot about the everyday impact of APD.
Impact statement

This project has potential impacts on the clinical diagnosis of auditory processing disorder (APD) in children and on future research studies on speech perception conducted in virtual acoustic environments.

Firstly, the project focused on a test procedure used in the clinical assessment of APD. These listening difficulties in children are highly disputed and more research is needed to better understand what is causing them, how best to assess them and what would be the most effective form of management and intervention. As part of this project, a group of children diagnosed with APD was compared to typically-developing children in a number of tests on speech perception, hearing and language abilities. The main aim was to further evaluate a test that is commonly used in the clinical assessment of APD. We investigated whether the virtual auditory environment used in the test is suitable to measure true performance in speech perception in children with and without APD. The results help to better validate this test which is currently one of the main tools to assess speech-in-noise perception for the APD diagnosis.

On the other hand, measuring speech reception thresholds (SRTs) in different real and virtual auditory environments in children and adults helped to better understand the effect of nonindividualised virtual environments on speech perception. Using individualised environments is important for other aspects of spatial auditory perception, e.g. localisation, but requires complex measurements or calculations. Virtual acoustic environments are generated from features dependent on the anatomy of a person. Since young children are much smaller than adults, it was unclear whether individualisation would be necessary to measure accurate SRTs in children, especially when they have listening difficulties. Our results suggest that it is not necessary to use individualised virtual environments to measure speech perception in terms of SRTs and the spatial release from masking in children or adults, at least for the large spatial separations between acoustic sources tested here. Therefore similar future studies in this field can be conducted in nonindividualised environments which are easy to create.
# Contents

1 Introduction and literature review  
1.1 Auditory processing disorder - current research and clinical practice .......................................................... 16  
1.1.1 Characteristics of APD .................................................. 17  
1.1.2 Assessment of APD .................................................. 18  
1.1.3 APD management and intervention strategies ................. 20  
1.1.4 Controversies in APD research ...................................... 21  
1.1.5 Need for more research on APD ................................. 23  
1.2 Spatial processing abilities in children and adults ............ 24  
1.2.1 Current research on the spatial release from masking .......... 25  
1.2.2 The LiSN-S test and spatial processing disorder ............. 28  
1.3 Virtual acoustic environments created from head-related transfer functions 33  
1.3.1 Generating virtual auditory environments from HRTFs ........ 33  
1.3.2 The effect of HRTF mismatch on auditory and speech perception 36  
1.3.3 Current research on HRTFs in children .......................... 37  
1.4 Overview of this research project ................................. 38  
1.4.1 Motivation .......................................................... 38  
1.4.2 Aims and research questions ...................................... 39  
1.4.3 Structure of the project ............................................ 40  

2 Pilot studies and measurements 42  
2.1 Sentence norming study in children .............................. 42  
2.1.1 Methods .......................................................... 43  
2.1.2 Results .......................................................... 49
2.2 HRTF measurement procedure ........................................ 51
2.2.1 Measurement setup ............................................ 51
2.2.2 Measurement of the impulse responses ....................... 53
2.2.3 Post-processing of impulse responses ......................... 54
2.2.4 Pilot measurements ............................................ 54

3 Spatial release from masking in adults in different auditory environments
3.1 Introduction ...................................................... 57
3.2 Methods .......................................................... 59
  3.2.1 Participants ................................................. 60
  3.2.2 Measurement of individual HRTFs .......................... 60
  3.2.3 Speech stimuli ............................................. 61
  3.2.4 Signal processing ........................................... 62
  3.2.5 Procedure for SRT measurements ......................... 62
3.3 Results ............................................................ 64
  3.3.1 Head related transfer functions ............................. 64
  3.3.2 Speech reception thresholds ............................... 64
3.4 Discussion ........................................................ 69

4 Spatial processing and listening abilities in children with auditory processing disorder
4.1 Introduction ...................................................... 71
4.2 Methods .......................................................... 78
  4.2.1 Participants ................................................. 79
  4.2.2 Measurement of head-related transfer functions ............. 81
  4.2.3 Measurement of speech reception thresholds ............... 82
  4.2.4 Extended high-frequency audiogram ........................ 87
  4.2.5 Language assessment ....................................... 88
  4.2.6 ECLIPS questionnaire and background questionnaire .... 88
4.3 Results ............................................................ 89
  4.3.1 Participant information and background questionnaire .... 90
  4.3.2 Speech reception thresholds ............................... 93
4.3.3 Spatial release from masking ........................................ 101
4.3.4 Hearing thresholds ..................................................... 103
4.3.5 Receptive language: CELF - Recalling sentences ................ 107
4.3.6 ECLIPS questionnaire .................................................. 108
4.3.7 Relationship between speech perception and other results ...... 109
4.3.8 Comparison with results from clinical APD assessment .......... 111
4.4 Discussion ..................................................................... 112

5 Differences between HRTFs of children and adults 114

5.1 Introduction .................................................................. 114
5.2 HRTF measurements ....................................................... 118
5.3 Results ......................................................................... 119
5.3.1 Anthropometric data ..................................................... 120
5.3.2 Binaural cues .............................................................. 122
5.3.3 Spectral cues .............................................................. 127
5.3.4 Effect of HRTF mismatch on speech perception ................. 131
5.4 Discussion .................................................................... 132

6 General discussion ................................................................. 134

6.1 Comparison of speech perception results of children and adults . 134
6.2 Comparison with other studies ............................................. 137
6.3 Discussion of results ........................................................ 141
6.3.1 Effect of auditory environment on speech perception .......... 141
6.3.2 Differences to LiSN-S procedure ..................................... 143
6.3.3 Findings for children with diagnosed APD ....................... 144
6.3.4 Limitations .................................................................. 145
6.4 Summary of the results ..................................................... 146
6.5 Conclusion .................................................................... 147

A Additional figures for child results ........................................ 172

B Documents .................................................................. 175
List of Figures

2.1 Comparison of voice characteristics between ASL sentences and distract-ors. ................................................................. 46
2.2 Speech intelligibility measures. .............................................................. 48
2.3 Psychometric functions of all 240 ASL sentences tested at fixed levels. . . 50
2.4 HRTF measurement setup in the anechoic room. .............................. 52
2.5 HRTFs measured for four adults at the left ear for 270°, 0° and 90° azimuth. 55
2.6 Symmetric HRTFs and repeated measurements for 0° azimuth. .......... 56

3.1 HRTFs for individual participants and large and small head from database. 64
3.2 SRTs for maskers at 0° and ±90° azimuth in the four environments. .... 65
3.3 SRM in the four auditory environments. ........................................ 65
3.4 Correlations of SRTs and SRMs in the four auditory environments. .... 67
3.5 Relationship between ITDs and SRTs in colocated and separated condition. 68

4.1 Age distributions of children in the APD and TD group. ................. 90
4.2 Frequency and severity of middle ear infections in the two participant groups. 93
4.3 SRTs for all participants in the seven test conditions. ...................... 93
4.4 SRTs of individual participants. .................................................. 95
4.5 Difference between repeated SRT measurements. ............................ 96
4.6 SRT performance with age in the seven test conditions. .................... 98
4.7 Age-normed scores for SRTs in the seven test conditions. ............... 100
4.8 SRM results for individual participants and regression over age. .......... 102
4.9 Age-normed SRM for both participant groups. ............................... 103
4.10 Average hearing thresholds for participant groups. ........................ 104
4.11 PTA and EHF hearing thresholds for children related to OME history. . 106
4.12 Scaled scores for receptive language in the CELF-5 recalling sentences. . . 107
4.13 ECLIPS scaled scores for the total scores and the five factors. . . . . . . . 109
4.14 Distributions and correlations between measures in TD and APD children. 110
4.15 SRT and SRM compared with results of the LiSN-S at APD assessment. . 111

5.1 HRIRs and HRTF spectra of the left ear for all three sound incidences. . . 119
5.2 Anthropometric head parameters as a function of age. ......................... 121
5.3 ITDs of children and adults for sound incidences of -90° and +90°. . . . . 123
5.4 ITDs across age calculated from HRIRs low-pass filtered at 3kHz. ........ 124
5.5 ILDs of children and adults: global and SII-weighted ILDs. .................... 125
5.6 Child HRTF with lowest peaks and notches in the spectrum ................. 127
5.7 Spectral peaks P1 and P2 and notches N1 and N2 in HRTFs. ............... 128
5.8 HRTF at 0° azimuth: median and individual results for adults and children.130
5.9 Age-normed SRTs and SRM in children related to circumference and bin-aural cues. .................................................. 132

6.1 SRTs in different AEs in adults and children. ................................. 135
6.2 SRMs in different AEs in adults and children. ................................. 136
6.3 SRTs for TD children in the LiSN-S and the current study. ............... 138
6.4 SRM for TD children in the LiSN-S and the current study. ............... 139

A.1 PTA and EHF hearing thresholds related to history of middle ear infections.172
A.2 Background questionnaire results for all participants. ........................ 173
A.3 Background questionnaire results for APD participants. ...................... 173
A.4 HRTFs for -90°, 0° and 90°: individual and median results for children. . 174
List of Tables

2.1 Examples for ASL and BKB sentences and excerpt from distractor story. . 44
2.2 Correlations between HRTFs from pilot participants. . . . . . . . . . . . . 56
3.1 SRM and SRTs for colocated and separated maskers in adults. . . . . . . 66
4.1 Participant information: age and gender distribution in the two groups. . 91
4.2 Diagnosed and suspected developmental disorders in the participant groups. 92
4.3 Age-normed SRTs and SRM for the two groups. . . . . . . . . . . . . . . . 101
4.4 Hearing thresholds: means and standard deviations in the two groups. . 105
5.1 Anthropometric head measures for children and from literature. . . . . . 122
5.2 Binaural cues of children in each age group and adults. . . . . . . . . . 126
5.3 Monaural cues: lowest spectral peaks and notches in children and adults. . 129
6.1 Studies measuring SRM in children with listening difficulties. . . . . . . 140
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE</td>
<td>auditory environment</td>
</tr>
<tr>
<td>APD</td>
<td>Auditory processing disorder</td>
</tr>
<tr>
<td>ASL</td>
<td>(Audio-visual) adaptive sentence lists</td>
</tr>
<tr>
<td>AU LiSN-S</td>
<td>Australian version of the LiSN-S</td>
</tr>
<tr>
<td>BKB</td>
<td>Bamford-Kowal-Bench (sentences)</td>
</tr>
<tr>
<td>BMLD</td>
<td>Binaural masking level difference</td>
</tr>
<tr>
<td>CELF-5</td>
<td>Clinical evaluation of language fundamentals (Version 5)</td>
</tr>
<tr>
<td>CELF-RS</td>
<td>CELF-5 recalling sentences subtest</td>
</tr>
<tr>
<td>EAS</td>
<td>Environment and auditory sensitivity (ECLIPS factor)</td>
</tr>
<tr>
<td>ECLIPS</td>
<td>Evaluation of children’s listening and processing skills</td>
</tr>
<tr>
<td>EHF</td>
<td>Extended high frequency</td>
</tr>
<tr>
<td>FM</td>
<td>Frequency modulation</td>
</tr>
<tr>
<td>FPL</td>
<td>Forward pressure level</td>
</tr>
<tr>
<td>GOSH</td>
<td>Great Ormond Street Hospital</td>
</tr>
<tr>
<td>HRIR</td>
<td>Head-related impulse response</td>
</tr>
<tr>
<td>HRTF</td>
<td>Head-related transfer function</td>
</tr>
<tr>
<td>ILD</td>
<td>Interaural level difference</td>
</tr>
<tr>
<td>ITD</td>
<td>Interaural time difference</td>
</tr>
<tr>
<td>KEMAR</td>
<td>Knowles Electronics manikin for acoustic research</td>
</tr>
<tr>
<td>LiSN-S</td>
<td>Listening in spatialised noise - sentences (test)</td>
</tr>
<tr>
<td>LLL</td>
<td>Language, literacy and laterality (ECLIPS factor)</td>
</tr>
<tr>
<td>MA</td>
<td>Memory and attention (ECLIPS factor)</td>
</tr>
<tr>
<td>NA LiSN-S</td>
<td>North American version of the LiSN-S</td>
</tr>
<tr>
<td>OME</td>
<td>Otitis media with effusion</td>
</tr>
<tr>
<td>PSS</td>
<td>Pragmatic and social skills (ECLIPS factor)</td>
</tr>
<tr>
<td>PTA</td>
<td>Pure-tone audiogram</td>
</tr>
<tr>
<td>SAP</td>
<td>Speech and auditory processing (ECLIPS factor)</td>
</tr>
<tr>
<td>SII</td>
<td>Speech intelligibility index</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>SPD</td>
<td>Spatial processing disorder</td>
</tr>
<tr>
<td>SRM</td>
<td>Spatial release from masking</td>
</tr>
<tr>
<td>SRT</td>
<td>Speech reception threshold</td>
</tr>
<tr>
<td>susAPD</td>
<td>Suspected auditory processing disorder</td>
</tr>
<tr>
<td>TD</td>
<td>Typically-developing</td>
</tr>
<tr>
<td>UCLH</td>
<td>University College London Hospital</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction and literature review

What if you could hear but not understand? Understanding other talkers is an essential communication skill we all rely on heavily every day. Some people, however, experience listening difficulties that cannot be explained by their peripheral hearing, previous noise exposure and their cognitive and language abilities. In many cases, these difficulties already emerge in early childhood. These children are often diagnosed with auditory processing disorder (APD).

APD is assumed to be a deficit in processing auditory information in the auditory system which results in poor perception of speech and nonspeech sounds (BSA, 2018; de Wit et al., 2016). These children struggle with understanding a speaker in a noisy environment, e.g. the teacher in a busy classroom, and have difficulties comprehending and remembering verbal information. There is also high comorbidity with other developmental disorders of language and cognition. There is a large variety in the characteristics and strength of these listening difficulties and there is no conclusive evidence of their cause. Therefore, it is still unclear whether these deficits can be seen as different symptoms of one distinct disorder or might occur for a variety of different reasons and whether this disorder is purely in the auditory domain or also affects other modalities like cognition and language. The great variation in symptoms and therefore in performance in clinical tests makes it difficult to diagnose APD and to find the best suitable intervention for each child.

Currently, no gold standards exist for clinical assessment of APD and interventions. The assessment typically includes a large test battery of speech and nonspeech tests
as well as questionnaires (AAA, 2010; BSA, 2018; Cameron et al., 2015; Moore et al., 2018). Many tests used in audiological test batteries have originally been developed for other applications and have only limited capability to separate children with APD from typically-developing (TD) children. Often, they are not sufficiently validated and normed. This makes the assessment of APD difficult and highly dependent on the expertise of the audiologist. There is a need for more research on tests specifically targeted at these listening difficulties.

The aim of this project is to investigate difficulties in speech perception in noise of children diagnosed with APD in regards to a specific test currently used in clinical assessment (Cameron & Dillon, 2007). The main outcome of the test is the effect of spatial release from masking (SRM), which is the benefit in speech perception when a target speaker is spatially separated from masking sounds (Bronkhorst, 2015; Freyman et al., 1999). To measure the SRM, the speech reception thresholds (SRT), the level at which speech is 50% intelligible, is measured in two conditions: with target and maskers located at the same position in front of the listener and with the same frontal target but maskers presented from different locations. A subgroup of children with listening difficulties was found to perform below the norms for TD children for the SRM and the SRT in the colocated condition (Cameron & Dillon, 2008). However, it is unclear whether the auditory environment in which the test is presented affects their test performance. The stimuli are presented via headphones in a virtual acoustic environment. The virtual environment is created from so-called head-related transfer functions (HRTFs). HRTFs describe the acoustic filtering of sound due to a person’s body, head and ears and thus vary for every person. They can be measured acoustically or calculated from geometrical data (Xie, 2013). For practical reasons, HRTFs used in the clinical test are from an artificial head that was designed to have head and torso dimensions of an average adult. Nonindividualised HRTFs are known to result in a degradation of the auditory perception in the environment (Møller et al., 1996) which might be large enough to affect the perception of sound sources in adults and even more so for children.

In this project, we aim to investigate whether these differences have a measurable influence on speech perception in these listening conditions and thus influence the test results. Studies on the effect of environment on the SRM are conducted for adults and children with a focus on children with diagnosed APD. This introduction will present
current knowledge on research and clinical practices for APD and on the effect of SRM in children. It will also describe methods to generate virtual auditory environments from HRTFs and their limitations for children.

1.1 Auditory processing disorder - current research and clinical practice

Some children experience listening difficulties despite having normal peripheral hearing and normal intelligence. They are commonly diagnosed with auditory processing disorder (APD), which is also called central auditory processing disorder (CAPD) or described more generally as listening difficulties in recent literature (BSA, 2018; Wilson, 2019). Their difficulties are thought to be caused by disruptions of auditory processes in the central auditory system. APD is often described as an inability to localise, group, discriminate, recognise and comprehend speech and nonspeech sounds. These difficulties are not caused by hearing loss in the peripheral auditory system, as these children have normal hearing thresholds when measured with a pure tone audiogram.

The diagnosis of APD is much disputed in the current literature (e.g. Iliadou et al., 2017; Moore, 2018). The group of children diagnosed with APD is very heterogeneous and their listening problems are likely to have a variety of different causes. Besides auditory processing, other factors such as memory, attention, intelligence and language may contribute to their listening difficulties to varying degrees. Therefore, the term auditory processing disorder has to be used with caution and primarily serves as an umbrella term for any kind of listening difficulty that cannot be explained by current standard methods. At present, it is unclear whether APD is a unimodal disorder in the auditory domain or a multimodal disorder, as many children diagnosed with APD also experience deficits in language or cognitive abilities. There are also suggestions that APD should be considered primarily a cognitive rather than an auditory disorder.

National guidelines on APD from professional societies and official bodies vary considerably in their definition of APD and their recommendations for diagnosis and treatment (e.g. AAA, 2010; BSA, 2018; CISG, 2012; Keith et al., 2019; Wilson, 2018). This section provides a brief overview of the symptoms typically associated with APD, current practices for diagnosis and management and unresolved issues in APD research.
1.1.1 Characteristics of APD

The listening difficulties associated with APD typically emerge once a child enters school and is regularly exposed to a complex and demanding auditory environment. Primarily, higher-order auditory processes are affected, such as speech-in-noise perception, auditory discrimination, auditory pattern recognition and temporal processing (AAA, 2010; BSA, 2018; de Wit et al., 2016). Children with APD experience difficulties in listening in the presence of background noise or when listening to degraded speech. They have a hard time separating similar sound sources and attending to and comprehending a talker. They also struggle to remember complicated verbal information and are easily distracted and quick to fatigue when confronted with complex auditory stimuli due to their higher listening effort (Bamiou et al., 2001). This can lead to reduced processing capacity for other tasks and further secondary deficits in attention and memory as well as language, reading and spelling (Moore et al., 2010). APD can thereby have broader implications on various aspects of learning and can affect performance in school.

Auditory processing difficulties are not specific to children. Processing deficits in the central auditory system can also appear in adulthood due to ageing, trauma, infection or noise exposure. This is often referred to as acquired APD or, in combination with peripheral hearing impairment, as secondary APD (BSA, 2018). These forms, however, have a clear cause or mediator for the deficit, whereas the cause for developmental APD is still very much unknown. Studies investigating possible causes or risk factors for APD in children found evidence that a specific form of APD characterised by an impairment in spatial processing may be related to previous episodes of otitis media with effusion (OME), a severe form of middle ear infection (Graydon et al., 2017; Tomlin & Rance, 2014). Some studies also indicate that there may be predisposing genetic factors for APD (Brewer et al., 2016). Unless stated otherwise, the term APD in this thesis always refers to developmental APD.

A large proportion of children suspected of having APD also have diagnoses of one or more other developmental or learning disorders (Dawes & Bishop, 2009; de Wit et al., 2018). These comorbid disorders include developmental language deficits, like dyslexia or developmental language disorder (Dawes & Bishop, 2010; King et al., 2003; Miller & Wagstaff, 2011), reading disorders (Sharma et al., 2009), attention and memory deficits,
such as attention deficit (hyperactivity) disorder (Gyldenkærne et al., 2014; Riccio et al., 1994), or other cognitive or neurological conditions such as autism spectrum disorder (Kozou et al., 2018; Ocak et al., 2018; O’Connor, 2012; Stevenson et al., 2018), auditory neuropathy spectrum disorder or learning disability (Keller et al., 2006). This comorbidity makes it complicated to separate a child’s auditory difficulties from symptoms of other disorders with similar characteristics. Although there can be common symptoms in different disorders, APD is still viewed as a separate condition to language and attention deficits. Studies found that whilst there was some correlation, deficits in cognitive and language abilities alone do not explain the variance in performance on auditory processing tasks (Brenneman et al., 2017; Gyldenkærne et al., 2014).

Symptoms of APD vary in type and intensity and may not always be correctly identified which makes it difficult to give an accurate estimate of the prevalence of this disorder in the general population. Studies have estimated that between 0.5% to 7% of children are affected by auditory processing difficulties to various degrees (Bamiou et al., 2001; Cameron et al., 2015; de Wit et al., 2016; Hind et al., 2011; Keith et al., 2019).

Identifying APD early on is crucial for raising awareness of a child’s difficulties and for addressing them as interventions in older children are likely to be less effective. Due to the high comorbidity with other developmental disorders and a lack of awareness about APD in the educational system and among paediatricians, listening difficulties are often misdiagnosed or left unattended.

1.1.2 Assessment of APD

Currently, APD is diagnosed in audiology clinics through a combination of clinical tests and questionnaires. Due to the nature of some of these tests, it is often only possible to assess and diagnose children aged 7 years or older and with reasonably normal language abilities and intelligence. Children are typically referred for an APD assessment by their school teacher, speech and language therapist, educational psychologist or paediatrician because of their inexplicable listening difficulties compared to other children their age. In many cases, the child’s cognitive and language abilities have been previously assessed by an educational psychologist or speech and language therapist.

The clinical protocol for assessing APD varies widely between clinics and healthcare
1.1. AUDITORY PROCESSING DISORDER

systems (AAA, 2010; BSA, 2013, 2018; Cameron et al., 2015; CISG, 2012; Keith et al., 2019; Moore et al., 2018). Although there is a growing consensus about the listening difficulties associated with APD, there are still no universally accepted diagnostic criteria nor a standardised set of tests for APD diagnosis. As a consequence, very few clinics in the UK currently assess APD and those that do have developed their own test protocols and diagnosis criteria.

Most commonly, APD assessment involves a structured case history, one or more parental questionnaires, previous professional reports, measures of the peripheral auditory system and an extensive battery of behavioural and objective tests on higher-order auditory skills.

Prior to the appointment, the child’s parents or teacher are asked to fill out questionnaires about the child’s listening abilities. Commonly used questionnaires are the Children’s Auditory Performance Scale, the Evaluation of Children’s Listening and Processing Skills (ECLiPS) or the Fisher’s Auditory Problems Checklist (Barry et al., 2015). Besides questions on hearing difficulties the child is experiencing in everyday life, they also include aspects of the child’s medical history with a focus on middle ear infections and head injuries, other developmental disorders diagnosed in the child and close family members, birth and development and general information about the child’s speech and language development and academic performance (Cameron et al., 2015). This information combined with previous professional reports may also be discussed in more detail in an interview with the audiologists.

To rule out other forms of hearing loss, the function of the peripheral auditory system is tested by measuring a pure tone audiogram, tympanograms and acoustic reflexes. Some clinics also measure otoacoustic emissions with and without suppression as well as auditory brainstem responses to rule out the other deficits such as auditory neuropathy spectrum disorder.

The large behavioural test battery generally consists of various speech and nonspeech tests targeting different auditory and cognitive functions. It may include tests on auditory discrimination, temporal processing, dichotic listening, binaural interaction, monaural low-redundancy speech and speech-in-noise perception (Cameron et al., 2015). Wherever possible, a specific subset of tests is selected for each child, depending on their presentation in the questionnaires and interview, in order to target the child’s specific deficits and
avoid influences of fatigue from an excessive test battery (Chermak et al., 2017). Commonly used nonspeech tests are the gaps-in-noise, random gap-detection or FM detection tests for temporal processing, frequency or duration pattern tests and masking level difference test (Shinn et al., 2009). In terms of speech tests, the test battery will typically include tests on word or sentence perception in quiet and in the presence of noise or speech maskers. Widely used tests are the auditory figure ground test and filtered words test from the SCAN-C for speech-in-noise perception, the dichotic digits test for binaural integration and the LiSN-S test for spatial processing (Cameron & Dillon, 2007; Keith, 2000). Some audiologists also include cognitive tests on nonverbal intelligence, attention and working memory to exclude the possibility of a cognitive disorder (Cameron et al., 2015).

A diagnosis of APD is usually given when a child shows a deficit of more than two standard deviations from the normed mean on two or more speech and nonspeech tests. Depending on the child’s deficits, different subtypes of APD might be diagnosed, such as spatial processing disorder (Cameron et al., 2014; Graydon et al., 2018) or temporal pattern deficit (Tomlin & Vandali, 2019). For children with one or more other developmental disorders that may affect test results, assessment and diagnosis of APD can be more difficult or even impossible.

1.1.3 APD management and intervention strategies

At the current stage, APD interventions are limited and there is no universally accepted gold standard for management and treatment due to the large variation in characteristics and severity of listening deficits. Intervention strategies most commonly used are the amplification of the target speaker, auditory training and modifications of the listening environment (BSA, 2018).

The most straightforward management strategy for APD is to provide a more favourable signal-to-noise ratio, thus increasing the audibility of a target speaker which is beneficial for speech perception overall. Audiologists often recommend the use of a personal FM system to increase the level of the teacher’s voice in the classroom in comparison to the level of background sounds (Johnston et al., 2009; Sharma et al., 2012; Stavrinos et al., 2020). A similar effect can be achieved with a sound reinforcement sys-
1.1. AUDITORY PROCESSING DISORDER

tem or a general improvement of the room acoustics through acoustic treatment. These two latter strategies also benefit other children in the class but are rarely implemented due to their costs. Simple behavioural changes by the child and teacher can also help improve classroom listening. The child can be seated in the front of the room and the teacher can adopt behavioural strategies to ensure the child is able to follow the lessons, such as addressing the child directly, using simple instructions and offering additional reinforcement and clarification.

Another remediation strategy is deficit-specific auditory training that makes use of the neuroplasticity of the brain. This aims to improve auditory skills by improving top-down auditory processing and forming stronger neural connections. Training can be done by using auditory training software for children, such as LiSN & Learn, Earobics and FastForWord, but also through informal methods and musical practise (Fey et al., 2011; Weihing et al., 2015). The LiSN & Learn software (also called Sound Storm) has been developed specifically for children who experience difficulties with spatial processing (Cameron & Dillon, 2011). LiSN & Learn training over several months has been shown to improve performance on the related LiSN-S test (Cameron et al., 2012; Graydon et al., 2018), but it is unclear if and to what extent this improvement translates to listening abilities in other situations. Similar results have been found in a training study using several speech-in-noise games (Loo et al., 2016).

In addition, coping and compensation strategies may also help some children to identify individual listening strengths and weaknesses and develop their own listening strategies. It is important to provide individualised audiological care to find the best management and intervention strategy for each child.

1.1.4 Controversies in APD research

To date, there is a vivid scientific debate about the cause and disease pattern of APD and whether APD can be considered a distinct disorder, a subtype of another developmental disorder or in fact describes a variety of different deficits (e.g. Cacace & McFarland, 2013; Dillon et al., 2012; Iliadou et al., 2017; Moore, 2018; Neijenhuis et al., 2019; Rosen, 2005; Wilson, 2019). This uncertainty is reflected in the guidelines from different countries, which differ substantially in their definition of APD and their strategies for assessment.
CHAPTER 1. INTRODUCTION

and interventions (e.g. AAA, 2010; BSA, 2018).

There is no clear evidence as to which exact part of the auditory system causes APD. Both central and peripheral processes might contribute to the listening difficulties, as the descending auditory pathways are also known to be important for normal auditory function (Moore & Hunter, 2013; Oxenham & Bacon, 2003; Wilson, 2018).

Children with auditory processing difficulties are a very heterogeneous group. The characteristics and degree of their listening difficulties vary greatly and many children have also been diagnosed with other cognitive or language disorders. The potential causality between these deficits in different modalities is not yet sufficiently understood. It is possible that a deficit in language comprehension or cognitive abilities like sustained attention and working memory may lead to difficulties comprehending a speaker. In turn, listening difficulties are likely to lead to an increased listening effort and thus put a strain on cognitive functions. Therefore it is unclear whether APD can be seen as a separate disorder that is specific to the auditory modality or also includes deficits in other modalities.

There have also been suggestions that the deficits associated with APD could be a subtype of another disorder, such as developmental language disorder, rather than a separate entity (Dawes et al., 2008). The listening difficulties could also be part of a more general neurodevelopmental disorder affecting various aspects of communication (Moore et al., 2013). Since there is such high comorbidity and overlap of symptoms between different developmental disorders, the chosen referral route might determine whether a child is (first) diagnosed with a language or attention disorder or APD (Ferguson et al., 2011).

The name ‘auditory processing disorder’ is controversial and several suggestions have been made to rename it to listening difficulties, listening impairment or similar terms (Moore, 2018; Wilson, 2019). Whilst it is unclear whether a unimodal deficit in auditory processing is really the underlying issue, APD is currently used as an umbrella term for these kinds of listening difficulties until causes and relationships to other modalities are better understood.

There is also no universal consensus about the best strategy to diagnose these children. Diagnosis criteria are set arbitrarily and can differ quite strongly from clinic to clinic (Wilson & Arnott, 2013). There have been suggestions to focus on addressing
and treating individual deficits in each patient rather than giving a general diagnosis of APD (e.g. Dillon et al., 2012). A multimodal approach could be used to investigate the underlying reasons for listening problems in each child and address their most important difficulties with an individual management plan that includes auditory, language and cognitive aspects. Unfortunately, an official diagnosis of APD is often needed for families to receive support (e.g. financial support from the local council to purchase an FM system) or recognition of their child’s difficulties in the school (e.g. more awareness from the teachers, quiet space and more time for written tests).

1.1.5 Need for more research on APD

There are still many unanswered questions about APD and the definition and understanding of these listening difficulties are likely to change in the future as new research provides more information. It is a fact, however, that some children experience a reduced ability to listen in complex environments and it is important to find better tools to identify them and to offer targeted interventions. Therefore, interdisciplinary research and collaboration between researchers and clinicians are needed to improve diagnosis and remediation (Neijenhuis et al., 2019).

Firstly, it is important to develop well-validated and standardised screening tools, e.g. parental questionnaires, suitable for very young children. These can help to detect listening difficulties early on in childhood and increase the impact of interventions.

For the clinical assessment, it is important to develop gold standard diagnostic tests for listening abilities and reach a consensus on the best criteria for giving a diagnosis of APD. Since currently clinical practice and diagnosis criteria differ strongly across clinics, it is difficult to compare diagnoses. Many tests currently used in clinical assessment have not been specifically developed to target deficits present in children with APD and have inadequate specificity and sensitivity. Many of them are also not adequately validated and normed for children with APD and do not meet acceptable psychometric standards (Keith, 2009). More complex tests are often not suitable for younger children. Speech-based tests are often only available in a small number of languages and can lead to unreliable results for children with other native languages or dialects (Dawes & Bishop, 2007; Loo et al., 2013). They are also often not suitable for assessing children with
language difficulties.

Therefore, there is a need for high-quality tests for APD that help to reduce the total number of tests and thereby the duration of the test battery and are able to better assess the specific deficits in auditory perception experienced by children with suspected APD. These tests must be valid for real-world listening and need to have reliable age norms for typically-developing children as well as sufficient test-retest reliability. Speech-based tests are more representative of everyday listening situations but are typically also more influenced by other factors such as language, working memory, attention or non-verbal intelligence. Therefore, procedures need to be developed that are more resilient against these effects, either by reducing them or by controlling for them in the outcome measure. Ideally, the tests should be suitable for all children, regardless of their diagnoses of other developmental disorders. It is also important that speech tests are available in a child’s native language. Tests should also be suitable for young children as early identification is critical to the successful management of APD. Good diagnostic tests might also help to find new approaches for auditory training that can be used as APD intervention.

1.2 Spatial processing abilities in children and adults

Children suspected of APD predominantly report problems with understanding talkers in noisy backgrounds. This ability is measured experimentally by the measurement of speech perception in the presence of other sounds. One promising avenue in APD research focuses on speech perception in spatialised noise. This ability is tested by measuring the speech perception thresholds and the spatial release from masking. The SRT is the level difference between target and masker at which a listener is able to understand 50% of the target speech (e.g. Plomp & Mimpen, 1979; Rosen et al., 2013). The SRM describes the benefit in speech perception when stimuli are separated in space versus when they are located at the same position (e.g. Bronkhorst & Plomp, 1988; Freyman et al., 1999; Yost et al., 1996). The SRM is measured as the difference in level between SRTs measured for maskers at the same location as the target and for maskers spatially separated from the target. It is a measure of auditory stream segregation abilities. Research suggests that children with APD perform worse than typically-developing children in these kinds of tests (Cameron & Dillon, 2008; Sharma et al., 2014). In studies by Cameron and Dillon,
around 20% of children had a significant deficit in SRM which is diagnosed as a specific form of APD called spatial processing disorder in some countries (Cameron & Dillon, 2011; Cameron et al., 2015).

### 1.2.1 Current research on the spatial release from masking

The SRM has been researched extensively in adults (Bronkhorst, 2000, 2015; Freyman et al., 1999; Hawley et al., 1999; Peissig & Kollmeier, 1997) and in recent years more research also focused on children (Cameron & Dillon, 2007; Leibold et al., 2019; Litovsky, 2005; Lovett et al., 2012). However, the amount of SRM differs greatly between studies. It is influenced by a variety of factors specific to the stimulus types and locations, the transmission and the listener (Bregman, 1990).

Firstly, the SRM depends on the target and masker stimuli used in the study. The clarity of the target talker’s voice, and therefore its intelligibility, can be influenced by accent, speech rate, pronunciation and many other factors (Smiljanić & Bradlow, 2009; Van Engen et al., 2014). In most studies, the speech material consists of sentences, e.g. the commonly used Coordinate Response Measure sentences or matrix-style sentences (Brungart et al., 2001; Kollmeier et al., 2015). But also syllables or words are used to measure SRTs, especially when testing young children. Degradations of the target stimuli have a negative impact on the SRM, such as using vocoded speech to simulate hearing difficulties which lead to a reduction of the SRM by almost 4 dB in a study by Best et al. (2012). The SRM also depends strongly on the type of masking signal. Maskers can vary greatly in intelligibility - from completely unintelligible noise signals over signals containing some speech characteristics such as speech-shaped noise, vocoded speech or time-reversed speech to fully intelligible speech (Hawley et al., 2004). Speech perception in the presence of a noise masker is only influenced by energetic masking, the information loss at the auditory periphery. The peripheral encoding of the target signal is degraded depending on the amount of spectral overlap of target and masker signals and the temporal density of the masker. This leads to an inaccurate sensory representation and inaccurate neural responses to the target (Arbogast et al., 2002). Speech-in-speech perception is affected by both energetic and informational masking. Speech maskers typically result in a smaller amount of energetic masking since speech
CHAPTER 1. INTRODUCTION

signals are less spectrally dense than noise signals. They, however, lead to additional informational masking due to information loss at higher levels of the auditory system that cannot be explained by the spectro-temporal overlap of the stimuli. Informational masking is dependent on top-down processes like auditory attention and arises from the similarity between masker and target and the degree of uncertainty produced by randomness in the stimuli (Brungart et al., 2001; Shinn-Cunningham, 2008). A strong similarity between the stimuli makes it challenging to group sounds together and to form separate auditory objects for target and masker. It also makes it more difficult to select the target stimulus and focus attention on it. Randomness in the masker can also interfere with selecting the target object because attention is drawn to the salient masker. The overall masking effect is highest for a two-talker masker (Buss et al., 2017; Litovsky, 2005) which combines strong informational and energetic masking. For a larger number of talkers, masking strength decreases since the intelligibility is reduced and the masker is perceived as more noise-like, thus leading mainly to energetic masking (Rosen et al., 2013). Informational masking is weakened when there are other cues, e.g. spatial location, that help to segregate the sound streams and perceive them as distinct auditory objects and thus suppress the stream in the perceptual background. Therefore, informational masking leads to larger SRMs than energetic masking (Arbogast et al., 2002; Jones & Litovsky, 2011). Ewert et al. (2017) found SRMs of 2 - 4 dB for stationary noise, 2 - 7 dB for speech-shaped noise and 4 - 14 dB for speech-like maskers (similar results e.g. in Culling & Mansell, 2013). The SRM is highest for maskers with great similarity to the target, e.g. a speaker of the same gender with similar fundamental frequencies and formants, or even the same voice as the target (Brown et al., 2010; Leibold et al., 2020).

The amount of SRM depends also on the spatial locations of the maskers in the separated condition. The maskers can be presented from one or multiple locations either symmetrically or asymmetrically placed around the listener. A single masker or an asymmetric masker placement enables the listener to listen mainly with the averted ear which has a more favourable signal-to-noise (SNR) ratio. This monaural better-ear listening leads to a lower SRT and therefore higher SRM (Misurelli & Litovsky, 2012). When maskers are symmetrically placed around the listener, no long-term better ear exists but listeners can make use of rapid level fluctuations in both ears leading to short
1.2. SPATIAL PROCESSING ABILITIES

durations of beneficial SNR, so-called better-ear glimpses (Brungart & Iyer, 2012; Glyde et al., 2013a). Whilst this listening-in-the-dips strategy helps with speech perception, SRTs for separated maskers are still higher than in asymmetric configurations in which listeners can benefit from constant better SNR in one ear and don’t need to switch attention between the ears to hear glimpses of the target. Therefore, the SRM is smaller for symmetric maskers. The SRM is often measured for large separations of 60° or 90° but even a small angle between target and masker leads to a substantial SRM since it helps to perceive them as separated auditory objects, thus reducing informational masking. Marrone et al. (2008) showed that in adults already a small separation between target and symmetric speech maskers of ±15° led to an SRM of 8 dB. Further increasing the separation angles led to a maximum SRM of 12 dB. Even a very small separation of ±2° resulted in a measurable SRM for speech maskers in normal-hearing adults in a study by Srinivasan et al. (2016).

The transmission between speaker and listener also affects speech intelligibility. Besides potential degradation of the signals through recording and playback equipment, room acoustics or virtualisation techniques can also affect speech intelligibility. For studies in real auditory environments, higher reverberations lead to increased masking and a reduction of SRM (Ahrens et al., 2020; Biberger & Ewert, 2019; George et al., 2008). Similar effects were found in virtual environments with reverberation (Rychtáriková et al., 2011). Also, the distances of target and maskers to the listener affect SRTs (Westermann & Buchholz, 2015). In virtual environments under headphones, stimuli are spatialised either by convolving them with head-related transfer functions specific to the location and listener or by simpler techniques adjusting only binaural cues like the interaural time or level differences. The presence and precision of these binaural cues can affect the amount of SRM. Presenting only ILDs or ITDs leads to higher SRTs and a reduction of SRM (Edmonds & Culling, 2005; Ewert et al., 2017).

Besides these factors, differences in the test procedure can also affect the SRM and complicate comparison, e.g. different definitions of SRT in percent-correct intelligibility or different scoring methods (e.g. Ahrens et al., 2019; Ozimek et al., 2009).

The measured SRM of course also depends strongly on the abilities of the listener. Sensorineural hearing loss was shown to lead to a smaller SRM in adults and children mainly due to a deficit in the separated condition (Ching et al., 2011; Glyde et al.,
CHAPTER 1. INTRODUCTION

2012; Leibold et al., 2020). Even larger deficits were measured for children with bilateral cochlear implants, who only achieved a small SRM in asymmetric and almost none in symmetric masker conditions (Misurelli & Litovsky, 2012). Cognitive and language skills were also found to influence SRTs and SRM in some studies (MacCutcheon et al., 2019; McCreery et al., 2019).

In children, speech perception in the presence of maskers is generally worse than in adults, thus resulting in higher SRTs. This effect is especially high for intelligible maskers possibly due to difficulties with the allocation of attention (Leibold et al., 2019). SRTs for intelligible maskers are higher and mature much later in childhood than for energetic maskers. This also leads to the SRM being smaller in young children and increasing with age. A benefit from spatial separation has been measured in children from a very young age of 2 - 3 years for asymmetric (Hess et al., 2018; Garadat & Litovsky, 2007) and symmetric maskers (Ching et al., 2011; Misurelli & Litovsky, 2012). The ability to segregate stimuli develops by the age of 2 - 3 years and some studies found a matured SRM at a very young age for single noise maskers (Lovett et al., 2012; Murphy et al., 2011). Other studies measured a reduced SRM in children that only matured in early adolescence (Brown et al., 2010; Cameron et al., 2006c; Vaillancourt et al., 2008; Van Deun et al., 2010; Yuen & Yuan, 2014). The differences are likely to come from differences in the test procedures, with SRM maturing later for intelligible maskers, symmetric masker positions and more complex procedures, e.g. for sentence recognition instead of word recognition (Cameron & Dillon, 2007; Cameron et al., 2011; Vaillancourt et al., 2008).

1.2.2 The LiSN-S test and spatial processing disorder

Children with suspected APD were shown to have a lower SRM for speech maskers, mainly due to reduced performance in the conditions with separated stimuli. This was first found for the LISN (Listening in spatialised noise) test developed in 2006 to measure spatial processing abilities in children (Cameron et al., 2006a). In a study by Cameron et al. (2006b), 9 out of 10 children with suspected APD had SRM performance below the norms of typically-developing children in the LISN test. This suggested that this test procedure may be a suitable tool to detect listening difficulties in children with APD.
A simplified version of the test called the LiSN-S (LiSN-Sentences) was then developed to decrease the linguistic complexity and cognitive load in the test and, hence, reduce the influence of language abilities. In contrast to the original LiSN test, it is also suitable for younger children from the age of 6 years. Cameron & Dillon (2008) showed that also in this version children with suspected APD performed worse than the control group whilst performance in children with a diagnosed language disorder was not significantly different from the control group. The LiSN-S procedure has since been used in clinical practice and a number of studies on children with suspected APD (e.g. Moore et al., 2020; Sharma et al., 2014; Stavrinos et al., 2018) or children with a history of OME (e.g. Graydon et al., 2017; Tomlin & Rance, 2014). Whilst the LiSN-S showed a deficit in APD children, Boothalingam et al. (2019) found no group difference in SRM between children referred to the audiology clinic for suspected APD and typically-developing children for a single speech masker. Thus, this deficit might be specific to the auditory processes necessary for perceiving speech in the presence of intelligible maskers at symmetric locations, thus informational masking and better-ear glimpsing.

**LiSN-S procedure**

In the diagnostic LiSN-S test, the participant listens to a target talker and two simultaneous distractor talkers. The target talker is always presented from the front of the participant at 0° azimuth whilst the two distractor talkers are either presented colocated with the target or symmetrically shifted to both sides of the listener’s head at -90° and 90° azimuth. This symmetric masker presentation only allows participants to hear the target through short better-ear glimpses at both ears (Brungart & Iyer, 2012). For each of the two spatial configurations, two conditions are tested with speech maskers which are either the same voice as the target or different voices of the same gender. The participant is asked to repeat the target sentence to the experimenter. Depending on the number of keywords that were heard correctly, the SNR between target and distractors decreases or increases in the next trial. This adaptive procedure measures the SRT, the level difference for 50% intelligibility in each of the four test conditions.

The LiSN-S is designed to be used with young children from the age of 6 years. The target sentences in the LiSN-S were developed based on the principles of the commonly used BKB sentence lists (Bench et al., 1979), which were specifically developed to use with
CHAPTER 1. INTRODUCTION

young children. The test was developed in an Australian English and American English version (Cameron & Dillon, 2007; Cameron et al., 2009). For each version, recordings were made with a female speaker and normed for equal intelligibility across sentences for children aged 8 to 9 years. The masker signals are children’s stories recorded with the same talker and two other female talkers. The two maskers are played continuously in the background whilst the target sentences are presented separately after a warning signal. The test is conducted with headphones and the three-dimensional auditory environment is created by filtering the signals with HRTFs of an artificial human head. The LiSN-S is targeting skills involved in auditory stream segregation in children and can detect specific difficulties with spatial processing or voice discrimination.

LiSN-S results and norms

The performance in the LiSN-S is evaluated on the 'low-cue' condition (colocated speech from the same talker) and the 'high-cue' condition (spatially separated speech from different talkers) as well as on advantage measures for spatial separation and different talkers. The spatial advantage is defined as the difference between the SRTs of the colocated and the spatially separated condition with the same voice for masker and target. The talker advantage is the SRT difference between the two colocated conditions. The combination of both effects, the total advantage, is the difference between low cue and high cue condition.

Norms for the Australian English version have been determined for normal-hearing listeners between the ages of 6 to 60 years (Cameron & Dillon, 2007; Cameron et al., 2011). For the American English version, norms have been established for an age range of 6 to 30 years (Brown et al., 2010; Cameron et al., 2009). In younger children, the SRT performance is increasing with age, reaching adult-like performance around the age of 14 years. This finding is consistent with similar studies on speech perception in the presence of speech maskers in children (Corbin et al., 2016; Litovsky, 2005). SRT performance is best in young adults and slowly decreases with age in older adults. This effect is higher in the separated conditions leading to a decrease in SRM. To test the reliability of the norms, retests were conducted after a few months, which led to slight improvements in the SRTs due to familiarization with the speech material and voices but similar advantage measures, suggesting that the SRM is a suitable measure, even if sentence material is
1.2. SPATIAL PROCESSING ABILITIES

familiar.

LiSN-S results for children with APD

Children referred for APD assessment are typically between 6 to 11 years old. At this age, typically-developing children get significantly less talker advantage compared to adults but already have a large spatial advantage. The LiSN-S is designed to reveal difficulties in speech perception in challenging situations. Cameron & Dillon (2008) compared a group of TD children with a group of children suspected of having APD (susAPD) and with children with a diagnosed language disorder. Whilst the talker advantage score didn’t yield any differences across groups, the average spatial advantage was significantly smaller for children with suspected APD compared to the other two groups with a mean of only 8 dB instead of 12 dB. This decreased SRM was mostly due to a higher SRT in the spatially separated condition. However, the variability in spatial advantage across all groups was very high and some children in the control group had a similar deficit for SRM as the APD children. There was no significant difference between the performance of TD children and children with a language disorder in any condition, suggesting that the test is also suitable for children with language difficulties.

A subset of children with APD showed a large deficit of more than two standard deviations in this task specifically in the spatially separated conditions. Children with scores in the lowest 2.5th percentile of the norms were labelled as having a so-called spatial processing disorder. They only had 7.6 dB or less average SRM in same voice and different voice condition compared to 10.7 dB ± 1.6 dB in TD children. Spatial processing disorder is assumed to be a subtype of APD and to affect about one in five children suspected of APD (Cameron & Dillon, 2008; Cameron et al., 2015).

Advantages and limitations

The LiSN-S is currently used as part of the behavioural test battery to assess APD in clinics in the UK. A big advantage of the LiSN-S compared to other speech-based tests for APD is that it is based on a difference measure. The influence of higher-order language and learning abilities is greatly reduced since they mainly influence the absolute values of SRTs but have less of an effect on the advantage measures. This test is also feasible for children from the age of 6 years, whereas many other tests can only be conducted
with older children.

Clinical versions of the test have been developed and normed in Australian English and North American English but not in British English. Clinicians in the UK currently use either one of the existing versions and allow for larger deviations from the norm to account for accent differences. A recent study suggests that most norms in the North American version are suitable to be used for typically-developing British English children (Murphy et al., 2019), but it is unclear, whether these accent differences could potentially affect children with language or auditory processing difficulties more and therefore overestimate their listening deficits.

Recently, a language-independent version of the LiSN-S called the LiSN-U test was developed to make the test easily accessible in more countries or regions with different languages or dialects (Cameron et al., 2019). However, it is unclear whether it is dependent on the same auditory processes, as only a low correlation was found between the tests in a small sample of participants (Mealings et al., 2020).

The LiSN-S is performed under headphones since many clinics don’t have a suitable sound-proof room to conduct acoustic tests with loudspeakers. HRTFs are used to generate the virtual environment under headphones, which are taken from a standardised artificial head. One potential problem with this virtualization is that artificial heads are designed for adult head dimensions and likely lead to inaccurate binaural cues for children. This inaccuracy could lead to changes in the perception of the speakers in the test and affect SRTs. It may be possible that children with APD have particular difficulties in adjusting to these different cues and therefore perform worse in the virtual test than they would in a real environment.

In this project, we aim to develop a British English version of the test and examine the influence of virtual environments with nonindividualised spatial cues on speech perception in noise and the SRM. The following section will describe techniques to obtain HRTFs to create virtual auditory environments and discuss the importance of individualised HRTFs for auditory perception.
1.3 Virtual acoustic environments created from head-related transfer functions

In many experimental or clinical procedures, spatialised sound sources are presented in a virtual auditory environment (AE) under headphones instead of in the real acoustic environment. There are many reasons for this, e.g. it reduces the acoustic requirements for the test room, reduces overall costs and makes it easier to ensure equal testing conditions across participants. Virtual acoustic environments were used in many studies that measure the SRM (e.g. Best et al., 2017; Glyde et al., 2013a). There are several techniques to generate virtual AEs but computation from HRTFs is most commonly used because it is hardware efficient and results in a very realistic auditory display for individualised HRTFs (Xie, 2013).

1.3.1 Generating virtual auditory environments from HRTFs

Spatial hearing relies on physical acoustic features, which are the result of the acoustic filtering of sounds from different directions by a person’s pinnae, head and torso (Blauert, 1974). This filtering can be measured as HRTFs that describe the free-field transmission of sound waves from specific directions in space to a person’s ear canals. HRTFs contain both binaural and monaural cues. Binaural cues are direction-dependent differences in timing and level between the two ears. These interaural level differences (ILDs) and interaural time differences (ITDs) result from the fact that the two ears are placed on opposite sides of the head which affects wave propagation. ITDs are particularly important for localizing low frequencies below 1500 Hz (Kuhn, 1977). For higher frequencies, ILDs are the predominant binaural cues. Binaural cues contribute mainly to horizontal localisation of sound sources whilst monaural spectral cues are particularly important for vertical localisation. These spectral cues describe the direction-dependent spectral filtering of sound due to reflections and diffractions of sound waves on the body, head and especially the pinnae (Iida, 2019; Shaw, 1997; Xie, 2013). Peaks in the HRTF spectrum result from resonances in the pinna whilst notches are produced by filtering in cavities in the pinna. Since HRTFs depend on features of the anatomy, they are different for each person.

HRTFs can be measured acoustically and used to realistically reproduce an AE with
headphones (Hammershøi & Møller, 2005; Iida, 2019; Xie, 2013). HRTFs are measured in an anechoic room by presenting test signals through loudspeakers from different azimuth and elevation angles and at a defined radius around the listener. Measurements are typically made for a large number of positions in order to be able to create a full three-dimensional auditory environment. They are measured with excitation signals containing the full audible frequency range, e.g. sine sweeps, MLS signals or golay codes. The listener is required to remain motionless in a standing or seated position throughout the measurement. To reduce measurement time, faster techniques have been explored for continuous measurement (Richter, 2019) and reciprocal measurement for which a small loudspeaker is placed in the ear canal and sound is picked up simultaneously by multiple microphones around the participant (Zotkin et al., 2006). Measurements are the most reliable method to obtain precise HRTFs. However, individual HRTF measurements are rarely incorporated in research studies and are not possible in clinical settings. They are complex and very demanding in terms of equipment for presenting and recording test signals and in terms of the environment, as they require an anechoic room for free-field acoustics. High-resolution HRTF measurements also require the participant to maintain a fixed posture without movement for several minutes, which is not feasible for all participants, especially not for young children. In many cases, therefore, it is not possible to use individually measured HRTFs.

Various attempts were made to establish methods to obtain individualised HRTFs without acoustic measurements. One of them is simulation based on a person’s anatomy. A person’s geometric properties can be obtained by different methods, e.g. measuring anthropometric data by hand, creating a plaster cast for the head and torso, generating a three-dimensional geometric model from laser or MRI scans, or with individual or parametric photogrammetry (e.g. Fels, 2008). HRTFs can then be calculated from the surface of a detailed geometric mesh of a person’s torso, head and pinnae with computational techniques such as the boundary element method or finite element method (Harder et al., 2013; Huttunen et al., 2014; Ziegelwanger et al., 2015). These simulation methods, however, place high demands on the precision of the photographic techniques used for mesh generation and typically require manual post-processing of the geometric data. Therefore, they are only practical to use for a small number of participants. HRTFs can also be modelled by synthesizing them from pinna measures with the use of a principal
1.3. VIRTUAL ACOUSTIC ENVIRONMENTS CREATED FROM HRTFS

In case measurement or numerical calculation from geometric data is not possible, an existing set of HRTFs from an artificial head or another person with similar anatomy can be used. Artificial heads were designed to represent the average adult in torso head and pinna dimensions. Few standardised artificial heads for acoustic measurements have been developed, such as the KEMAR, the Brüel & Kjær 4128 or the Neumann KU100 (Møller et al., 1999). The head most commonly used in experimental procedures is the KEMAR (Knowles Electronics Manikin for Acoustic Research) developed in 1975 by Burkhard & Sachs. It is based on anthropometric data from more than 4000 male and female adults. The head can be fitted with two different pinna sizes and includes a simulation of the ear canal and eardrum based on a Zwislocki coupler. HRTFs from artificial heads provide a good approximation of the head and body dimensions of a typical adult. However, participants show large differences in pinna shapes which are not accounted for by this method.

Another option to find suitable nonindividualised HRTFs is to methodically select the set that most closely resembles the HRTFs of the participant from an existing database. There are a growing number of publicly available HRTF databases for adults such as the ARI\(^1\), IRCAM\(^2\) or CIPIC\(^3\) databases (Algazi et al., 2001). The HRTF set with the best fit can either be identified numerically by comparing pinna measures (Iida et al., 2014) or experimentally by performing a listening test with different sets of HRTFs (Pelzer et al., 2020; Seeber & Fastl, 2003). This procedure, however, can be time-consuming and can only be performed for a small number of HRTF sets. Nonindividualised HRTFs can also be roughly fitted for a participant by expanding or compressing the amplitude spectrum. All of these techniques provide an approximation of an individual’s HRTFs that is sufficient for some applications but will lead to a degradation of the auditory display compared to listener-specific HRTFs. The extent of this degradation is highly listener-specific as it depends on the anatomical similarities of the body dimensions and pinna shapes of the participant and the generic head.

---


\(^2\)http://recherche.ircam.fr/equipes/salles/listen/

\(^3\)https://www.ece.ucdavis.edu/cipic/spatial-sound/hrtf-data/
1.3.2 The effect of HRTF mismatch on auditory and speech perception

Whilst generic HRTFs are easier to obtain than measuring them individually for each person, the use of nonindividualised or inaccurate HRTFs leads to a degradation of spatial audio perception and a loss of immersion and realism of the auditory scene (Sunder & Gan, 2016).

Most noticeably, localisation accuracy is decreased, especially for different elevation angles and frontal directions (Brungart et al., 2017; Møller et al., 1996, 1999; Wenzel et al., 1993). This also leads to a greater number of front-back confusions and up-down reversals (Middlebrooks, 1999b). High-frequency sounds in particular might also be perceived as coming from higher elevations since spectral characteristics at frequencies over 5 kHz are important cues for vertical perception (Iida, 2019). Mismatched HRTFs can also degrade the acuity of localisation, measured as the just noticeable difference between two sound positions. For speech signals in quiet, Begault et al. (2000) found no deficit in localisation accuracy in adults when using nonindividualised HRTFs. However, more demanding listening conditions and a larger mismatch of HRTFs can result in higher localisation uncertainties. A deficit was found for speech-shaped noise in a study by Rychtáriková et al. (2011).

The use of nonindividualised HRTFs can also reduce the effect of externalisation, i.e. the perception of hearing a sound source at a certain distance in space instead of in one’s own head (Best et al., 2020; Iida, 2019). This effect is most perceivable in anechoic virtual environments. In virtual environments that contain reverberations or even just early reflections, the perception of externalisation is more robust and therefore not affected by HRTF mismatch (Begault et al., 2000). For the related perception of distance, studies have found mixed results. Yu & Wang (2018) found no difference in distance perception between individualised and generic HRTFs in adults. However, they used static sound sources in the near-field in an anechoic environment which might reduce overall cues for distance perception and found large variation across participants. Small changes in distance perception are unlikely to affect speech perception. However, different distances of target and masker were found to affect speech-in-noise perception in a reverberant virtual AE (Westermann & Buchholz, 2015).

A large HRTF mismatch also leads to an increase in the apparent source widths.
Sources perceived to be more spread out in space were found to affect speech perception and led to a smaller SRM (Ahrens et al., 2020). For these sources, a larger minimum separation angle between target and masker is necessary to experience any SRM.

Nonindividualised virtual environments are also generally perceived as less vivid and realistic than individualised environments, due to spectral colouring and the reduced externalisation of sources. This leads to listeners being less immersed in the auditory scene.

Overall, the extent of these effects depends on the similarity of the mismatched HRTFs to a listener’s own HRTFs resulting in a slightly different acoustic experience for each person. Studies have shown that the auditory system is able to adapt to mismatched spatial cues and relearn accurate localisation with these modified cues (Hofman et al., 1998; Stitt et al., 2019). However, this auditory learning effect requires the participant to be exposed to the new cues for an extended period of time.

1.3.3 Current research on HRTFs in children

Only very few studies have been conducted on children’s HRTFs. One reason for this is the difficulty of acoustically measuring HRTFs in children. HRTFs are typically measured for a large number of source positions and require the participant to be motionless for a long time, which is not feasible for young children.

Therefore, studies with children have often used generic HRTFs from adult-sized artificial heads to spatialise sound sources (e.g. Brown et al., 2010; Cameron & Dillon, 2007). However, children’s head parameters differ greatly from those of adults. Snyder et al. (1977) measured anthropometric data from a large number of children aged 2 months to 18 years. Whilst the breadth and length of the head increase at a similar rate throughout childhood, the height grows more rapidly, resulting in changes in the overall head shape with age. These changes have implications for the binaural ILD and ITD cues. Moreover, growth also affects the impedances in the ear canals. The pinnae also grow until early adolescence resulting in changes in the HRTF spectra (Iida, 2019).

Efforts have been made to calculate the HRTFs for young children from anthropometric data or three-dimensional meshes generated using photogrammetric techniques (Fels et al., 2004; Harder et al., 2013). Fels (2008) calculated HRTFs for children in a large age
range from digital models generated from photogrammetric data. She compared HRTFs of an adult, and infant and a scaled-down model of the adult with the same head circumference as the infant, as previously used for HRTF adjustment in adults (Middlebrooks, 1999a). There were large differences in the frequencies of the characteristic notches between the child head and the original adult head. Approximating the infant’s head by scaling down the adult head, shifted the notches to higher frequencies but they still did not match those of the child model. This test shows that children’s HRTFs cannot simply be created using scaling algorithms, as the proportions differ strongly from adult heads.

Recently, a first approach has been made to measure complete sets of HRTFs in children. Braren & Fels (2019) measured HRTFs for 26 children aged 5 to 9 years with a fast measurement technique modified to be suitable for children. They evaluated ITDs and predicted localisation performance from these HRTFs but the measurement accuracy has not yet been validated in a study on binaural perception.

Since individualised HRTFs for children are very difficult to obtain, most studies on speech perception in spatial configurations in children were conducted in real environments (Ching et al., 2011; Corbin et al., 2017; Vaillancourt et al., 2008) or in virtual AEs from generic HRTFs of an artificial head (Glyde et al., 2013a) which were only rarely adjusted for children (MacCutcheon et al., 2019). No comparison between these environments was made and thus it is unclear if the generic virtual AE provides the same acoustic display as the real environment and whether differences in spatial cues are large enough to affect speech perception.

1.4 Overview of this research project

In this project, we aim to investigate speech perception in noise in different real and virtual auditory environments in children with and without auditory processing disorder.

1.4.1 Motivation

On the one hand, we hope to be able to determine whether generic virtual environments used in tests like the clinical LiSN-S are suitable to accurately measure speech intelligibility. The LiSN-S test is currently used to assess APD in various countries and is one of the
main tests for speech-in-noise perception. It is important to ensure results are the same as in the real environment and a deficit in the test indicates a genuine deficit in listening abilities and is not just an effect of the mismatched acoustic environment. Currently, APD diagnosis depends much on the expertise and experience of the clinician conducting the assessment and is therefore only assessed in a few specialist audiology clinics in the UK. There is a lack of gold-standard tests that specifically target the different listening deficits common in children who experience auditory processing difficulties and that are suitable to clearly separate children with listening difficulties from typically-developing children. Developing and validating tests that can reliably measure performance in these abilities is therefore essential to facilitate testing in more clinics around the country. The LiSN-S test is currently only available in an Australian and American English version, thus there is a need for a similar test with British English speakers.

On the other hand, we are trying to better understand the general effect of mismatched virtual acoustic environments on speech perception in complex spatial conditions. Whilst we assume that adults will experience no or only a small effect, children might experience a larger effect since their anatomy differs more from those of standardised artificial heads which are typically used to create a nonindividualised virtual auditory environment. HRTFs in children have not been studied well in the past and it is uncertain to which extent the use of adult HRTFs is degrading auditory perception in children. Therefore, it is crucial to further investigate the differences in spatial cues between young children and adults and their change throughout childhood. It is also important to find appropriate ways to obtain accurate HRTFs from young children by measurement or simulation and to develop techniques to better approximate them from generic HRTFs if individualisation is not possible for a more realistic virtual acoustic environment.

1.4.2 Aims and research questions

In this project, we aim to examine whether mismatched spatial cues influence speech perception in the presence of intelligible maskers. We examine the extent to which this affects children and specifically children with APD and, therefore, whether this has implications for the LiSN-S test used in APD assessments.
Individualised spatial cues have been proven to be important for sound localisation and detection in adults (Middlebrooks, 1999b; Wisniewski et al., 2016). Performance is degraded in virtual environments generated with an HRTF set from a different person. Since there are large differences in body and head size between young children and adults, the use of HRTFs from standard artificial adult heads is likely to have an even larger impact on binaural hearing in children. This might also influence the perception of speech in virtual acoustic environments. Since the LiSN-S test for assessing APD is based on a virtual environment generated with HRTFs from an artificial adult head, these inaccurate spatial cues could influence performance, especially for younger children and children with APD.

In our experiments, we want to investigate whether children are performing better on SRM tasks when sources are presented in a real environment or filtered with their own measured HRTFs than with nonindividualised HRTFs. For children with APD, these incorrect spatial cues might be even more disruptive which could potentially lead to worse performance in the LiSN-S under headphones than for the same test performed in a real environment. Another aim was to develop and norm an SRM procedure with British English talkers similar to conditions in the LiSN-S.

In a series of studies, we aim to measure the SRM for adults, normal-hearing children and children with suspected APD in virtual and real conditions and compare the results.

1.4.3 Structure of the project

We first developed, set up and tested a procedure to measure HRTFs for a small set of source positions in young children and adults. We also conducted a first study with children in schools to norm the new sentence material for equal intelligibility in the relevant age group.

After that, a study was conducted with adult participants. It included measuring their HRTFs and using them in a procedure to measure the SRM in different individualised and nonindividualised acoustic environments. This study was done to determine the effect of HRTF mismatch on speech perception in adults.

Finally, a similar experiment was conducted for children with and without a diagnosis of APD. Besides the measurement of their HRTFs and measurement of the SRM in
different listening conditions it also included additional tests and questionnaires on their hearing and language abilities. Speech perception results and measured HRTFs were then compared between adults and children and to similar studies in the literature.
Chapter 2

Pilot studies and measurements

Before the main experiments, preparations had to be made regarding the speech material and the HRTF test procedure. Firstly, the sentences used as target stimuli for SRT measurements had to be recorded and normed to equal intelligibility in young children. This was necessary to ensure that SRTs could be reliably measured with an adaptive procedure in which every change in level would result in a defined relative change in intelligibility regardless of the specific sentence (Plomp & Mimpen, 1979). Therefore, the SRT for each sentence was estimated and the level of the sentence was adjusted accordingly. Secondly, a suitable setup for measuring HRTFs in young children had to be developed. The measurement system had to be planned, set up and tested in adults prior to the main study.

2.1 Sentence norming study in children

Speech intelligibility depends on various factors which are specific to the speech content, the talker, the interference from other simultaneous sounds and the abilities of the listener as described in detail in section 1.2. Speech recognition in young children is affected by both lexical and grammatical parameters of speech (van der Hoek-Snieders et al., 2020). For our experiments, we chose speech material that was specifically designed to suit young children in terms of linguistic complexity with simple sentence structures and vocabulary. But variations in the phonetic content across sentences might still result in variations in intelligibility. Some words may have a higher frequency or lower neighbourhood density than others and are therefore easier to understand (Bradlow & Pisoni, 1999). Also, in
some sentences, it might be easier to predict words based on context. Furthermore, the intelligibility of recorded sentences may be affected by the recording technique and characteristics of the talker, such as accent, clarity of speech and speech rate (Smiljanić & Bradlow, 2009).

For these reasons, a norming study was conducted to estimate the speech intelligibility of all target sentences separately and then adjust sentences in level to match their intelligibility in the following studies. Although we do not expect systematic differences in intelligibility between children and adults for this sentence material, there might be more variation between sentences in young children with developing language abilities. To ensure equal intelligibility for all sentences in both children and adults, the norming was conducted with children in the same age range as in the experiment in Chapter 4. Data collection for this study was conducted in collaboration with another PhD student but analysed separately.

2.1.1 Methods

Participants

The norming study was conducted with 49 children (22 female, 27 male) aged 8.4 to 10.4 years (mean 9.2 years \(\pm\) 0.6 years) recruited from two primary schools in Kent. All participants were native British English speakers and had no known developmental disorder. A hearing screening was conducted at the beginning of the test session, which showed normal audiometric thresholds of less than 20 dB HL between 250 Hz to 8 kHz for all participants. All children provided consent from their parents and received a certificate for participation in the study. The study was approved by the UCL ethics committee (ID 0544/006).

Stimuli

We decided to use sentences from the (audio-visual) adaptive sentence lists (ASL) as target stimuli in all experiments. These lists were developed by Macleod & Summerfield in 1990 and constructed on similar principles to the widely used BKB sentences (Bench et al., 1979). Example sentences are listed in Table 2.1. Both corpora contain sentences with simple syntax and vocabulary and were specifically developed to suit young chil-
dren’s speech and language competence. Unlike the BKB sentences, the ASL sentences have not yet been used in clinical tests and therefore have the advantage of being novel to each participant. The ASL sentence corpus consists of 18 sentence lists with 15 sentences per list, each containing three keywords. Due to the limited number of sentences, BKB sentences were used for training purposes at the beginning of the study. The 270 ASL sentences and 336 BKB sentences were recorded with a male Standard Southern British English speaker in an anechoic room. The talker was an experienced speaker who has been recorded frequently to provide material for speech tests. The sentences were recorded at a sampling frequency of 22.05 kHz and highpass-filtered with a cut-off frequency of 50 Hz.

<table>
<thead>
<tr>
<th>ASL sentences</th>
<th>BKB sentences</th>
<th>Distractor story</th>
</tr>
</thead>
<tbody>
<tr>
<td>The old clothes were dirty.</td>
<td>The clown has a funny face.</td>
<td>“...Amy and her brother Matt had invitations to a Halloween party.</td>
</tr>
<tr>
<td>The knife cut the cake.</td>
<td>Children like strawberries.</td>
<td>Matt cut up cardboard boxes to make a robot costume. Amy glued leaves onto a t-shirt so she could look like a tree...”</td>
</tr>
<tr>
<td>They picked some raspberries.</td>
<td>They’re looking at the clock.</td>
<td></td>
</tr>
<tr>
<td>The sailor stood on the deck.</td>
<td>The little baby sleeps.</td>
<td></td>
</tr>
<tr>
<td>Christmas is coming soon.</td>
<td>Someone’s crossing the road.</td>
<td></td>
</tr>
<tr>
<td>The doctor carries a bag.</td>
<td>The mouse found the cheese.</td>
<td></td>
</tr>
<tr>
<td>They’re living by the sea.</td>
<td>Bananas are yellow fruits.</td>
<td></td>
</tr>
<tr>
<td>The lady was quite cross.</td>
<td>The girl lost her doll.</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Examples for ASL and BKB sentences (keywords underlined) and short excerpt from one of the three distractor stories (full texts attached in section B.1 of the appendix).

Since the SRTs were to be measured in the presence of two speech maskers in the following experiment on SRM, it was important to obtain reliable intelligibility norms for this particular listening condition. Therefore, the same speech maskers were used for the norming procedure. They consisted of three freely available children’s stories with similar linguistic complexity as the target sentences (see Table 2.1 for an excerpt). The stories were recorded with the same male British English speaker. He was instructed to speak with constant pace and intonation throughout the text passages and not to stress keywords. The distractors were recorded at 44.1 kHz sampling frequency and highpass-filtered with a cut-off frequency of 50 Hz. A 3rd order Butterworth filter with forward and backward filtering was used to avoid phase distortions. To eliminate silence in the recordings, breaks between words were truncated to 100 ms. Audio distortions
were removed manually and the recordings were downsampled to 22.05 kHz to match the target sentences. Both stimuli, therefore, contained frequencies up to 11025 Hz.

ASL sentences and distractor stories were recorded several years apart. There is a possibility that some voice characteristics of the speaker may have changed over time (Goy et al., 2013; Stathopoulos et al., 2011). To ensure they are still reasonably similar, we compared the distributions of the fundamental frequencies and the long-term average speech spectra of both sets of recordings.

Fundamental frequency estimates were extracted for the concatenated sentences and the distractors in the SFS software\(^1\). Their distributions are shown in Fig. 2.1a. The distractor recordings had a median fundamental frequency of 126 Hz. The distributions were very similar for the three distractor stories (with median values of 121, 126 and 128 Hz). The distribution of the ASL sentences, however, resembled a bimodal distribution with a median of 144 Hz. A very similar distribution was also found for the BKB sentences recorded at the same time. This difference in the fundamental frequency distribution could be due to the nature of the speech material. Whilst the distractors consist of connected speech and the talker was instructed not to place emphasis on keywords, the sentences are all very similar in structure and length and could therefore lead to very similar intonations. A closer inspection of the change in fundamental frequency over time for all sentences (normalised to equal length) supported this hypothesis. There was a very similar and strong pattern of fundamental frequency change for all sentences, suggesting that the high amount of energy in higher frequencies was due to the similarity in intonation in each recording. Another factor contributing to these differences in the distributions could be the change in voice quality with age. The presence of higher harmonics in the speech depends on the rapid closure of the vocal folds. Since the speech material was recorded several years apart, this closure could have changed, leading to a reduction in high-frequency energy.

Long-term average spectra of the two concatenated recordings are displayed in Fig. 2.1b. There was a more pronounced peak around the fundamental frequency in the distractor recording but the overall shape was very similar. Although there could be slight changes in the talker’s voice characteristics between the recordings, the voice remained

\(^1\)Speech filling system, software for speech research, developed at UCL SHaPS by Mark Huckvale, available from https://www.phon.ucl.ac.uk/resource/sfs/
sufficiently similar to be used as target and masker speaker in this experiment.

![Fundamental frequency distributions](image1.png) ![Long-term average spectra](image2.png)

(a) Fundamental frequency distributions.  
(b) Long-term average spectra.

Figure 2.1: Comparison of voice characteristics between recordings of ASL sentences and distractors.

**Test sessions**

The study was conducted on a laptop in a quiet room at two primary schools. The tests were administered in Matlab and the stimuli were presented diotically through Sennheiser HD 25 headphones via an RME Babyface audio interface. The playback level was calibrated beforehand with a Brüel & Kjær type 4153 artificial ear.

At the beginning of the session, two examples of target sentences were given in quiet and in the presence of two distractor stories and the child was asked to repeat them. Each child was then presented with a short adaptive training consisting of 10 BKB sentences (at an initial SNR of 5 dB and with a step size of 2 dB) in the presence of two distractors. This helped the children to get more familiarised with the listening condition and the nature of the target sentences and to get more confident in their responses. They were encouraged to repeat exactly what they heard to the experimenter and not to question whether it made sense or formed a complete sentence. In case a child had difficulties understanding the task or paying attention to the correct talker, the training was repeated with different sentences.

After the familiarization, an adaptive run was conducted to estimate the participant’s individual SRT for this listening condition. Throughout the adaptive track, two distractor
stories were played at a combined level of 60 dB SPL. The distractors were between 100 to 200 seconds long and looped for continuous playback. The initial target level was 70 dB SPL and changed adaptively with an initial step size of 4 dB which decreased to 3 and 2 dB after the first two reversals, respectively. The responses were scored by keyword in a 1-up 1-down method to track the level of 50% speech intelligibility. Each ASL sentence contains three keywords. Two or three correct keywords resulted in a reduction of target level and zero or one correct keyword in an increase. A total of 30 ASL sentences were presented (ASL sentence lists 1 and 2). The SRT was calculated as the mean target level of the reversals after the second reversal (and shortened to an even number if necessary to balance upper and lower reversals).

The remaining 240 sentences were divided into 12 blocks and presented at fixed levels relative to the estimated SRT. Each block was presented either at the level of the individually measured SRT or at 2 dB above or below it. These three levels were expected to roughly correspond to 30%, 50% and 70% speech intelligibility (Cameron & Dillon, 2007). Levels were counterbalanced between participants to produce an equal number of trials for each sentence and level combination. The presentation order of levels and blocks was also pseudorandomised across participants to avoid learning effects. On average, each child took about 50 minutes to complete the study, split into two sessions on separate days.

The study was initially conducted with 40 children. During the testing, it became clear that for some participants the estimated SRT was very high (see Fig. 2.2a) and represented a higher value than 50% speech intelligibility. A possible explanation for this is that some children were not fully accustomed to the procedure even after the training and hesitant in their responses during the adaptive track. Their performance improved as the test session progressed. This resulted in the average estimate for SRT corresponding to a speech intelligibility of about 65% (see Fig. 2.2b). To ensure a good fit of the psychometric functions in the following analysis, it was important to have data from a wide range of speech intelligibility for each sentence, thus also have data points at levels with low intelligibility. Hence, we tested an additional group of 9 children with a modified version of the procedure. Besides easy level conditions, it also included sentence lists at a fixed absolute level of -4 dB SNR, 2 dB below the estimated average 50%-keyword-correct level across the 40 participants in the first group. Sentences at this
level resulted in an average of 27%-keyword-correct and led to an overall average of 59%. Fig. 2.2c shows the absolute levels and average percent-correct keywords for each sentence list presentation for all 49 participants.

(a) Estimated SRTs (dB SNR) relative to age for 40 participants.

(b) Average speech intelligibility for sentence lists presented at -2, 0 and 2 dB relative to the estimated SRT for each of the 40 participants.

(c) Average speech intelligibility for sentence lists for absolute levels for 49 participant (including additional retest trials at a fixed absolute level of -4 dB SNR).

Figure 2.2: Speech intelligibility measures.

This norming study was designed with similar criteria as the sentence norming of the original LiSN-S test (Cameron & Dillon, 2007) in terms of stimuli and procedure. However, a different scoring method was used. Whilst in the LiSN-S every word is scored, we chose to only score the three main keywords of each sentence. Both versions
are considered to be more precise and effective than scoring complete sentences (Brand & Kollmeier, 2002). Scoring keywords can potentially lead to smaller SRTs than scoring each word but is expected to be more representative of speech intelligibility because it reduces the influence of highly frequent words.

2.1.2 Results

The statistical analysis was conducted in R.

Psychometric functions

The results from the 240 sentences presented at fixed levels were grouped by sentence. A generalised linear model with logit function as link function was used to generate psychometric functions for the speech intelligibility of all sentences across levels as shown in Fig. 2.3. The mean level for 50% speech intelligibility was -2.2 dB SNR with an average maximum slope of 10% per dB. The data was first analysed to remove potential outliers and sentences with abnormally high or low intelligibility. The remaining sentences were then split into balanced sentence lists.

Outlier detection

Influential values for each psychometric function were investigated using Cook’s Distance and Leverage. Values that had a Cook’s D of more than four times the mean were considered potential outliers. When these values were removed, the psychometric functions changed only marginally but had a slightly larger overall spread of SRTs and slopes. Thus, removing them did not improve the fit of the psychometric functions. Most data points identified as potential outliers had zero keywords correct at very low SNR levels indicating that these values were influential points for the fit but could not be considered outliers as they were valid data points. Therefore, no outliers were removed from the data set.

Sentence selection

In a second step, sentences with psychometric functions that deviated strongly from the average were removed from the sentence material. These deviations could have been
caused by artefacts or other distinct features in the recordings or by differences in the complexity of the keywords. These sentences might be unusually easy or difficult to understand or show a very rapid or slow change in intelligibility with level, making them unsuitable to be used in adaptive procedures. Therefore, the fit of each psychometric function was assessed based on the SRT and the slope. Sentences with SRTs of more than 3.4 dB above or below the mean SRT at -2.2 dB were removed. This applied to 22 out of 240 sentences and ensured that level corrections between normed sentences were as small as possible. Sentences with a very slow or fast change in intelligibility and a maximum slope of more than 0.5 or below 0.05 were also removed. This was true for 22 sentences. As an additional criterion to ensure a good psychometric fit, the slope of the fitted model had to be statistically different from zero. Therefore, sentences were removed if the probability of the slope being zero was higher than 0.01, which happened for 26 out of the 240 sentences. In total, 200 out of 240 sentences were found suitable to be used for SRT measurements (black curves in Fig. 2.3). Their levels were adjusted for equal intelligibility.

Figure 2.3: Psychometric functions of all 240 ASL sentences tested at fixed levels: 200 sentences suitable for sentence lists (black lines) and 40 sentences removed due to differences in slope and mean value (red lines).
Sentence list balancing

The selected 200 sentences were then divided into new sentence lists. To avoid any bias, the sentences were distributed into lists depending on four categorical parameters: their SRTs, slopes, keywords and phonemes. This ensured that sentences with similar properties regarding intelligibility and sentences which contain the same words are spread out over different lists as much as possible. Categorical variables for SRT and slope were formed by dividing each parameter into 10 equal-sized ranked groups. Based on the text and audio data, a phonemic transcription was created for each sentence using the Montreal Forced Aligner (developed by McAuliffe et al., 2017).

Custom software was used for the balancing procedure that distributed sentences in lists based on the categories SRT groups, slope groups, keywords and phonemes. The procedure was repeated 1000 times and the run with the lowest dispersion was selected as the best fit for distributing the sentences into balanced lists. Two versions of normed ASL sentences were initially developed to be used in the following study investigating the spatial release from masking. One contained 14 lists with 14 sentences each, the other 20 lists with 10 sentences per list.

2.2 HRTF measurement procedure

To measure head-related transfer functions (HRTFs) in the following studies, we installed a new setup that is suitable for measurements in both adults and children from the age of 7 years upwards. Accurate measurement of HRTFs places high demands on measurement equipment and room (Iida, 2019; Xie, 2013). To reduce background noise and avoid reflections from surrounding surfaces, we built a measurement setup in the anechoic room of the department for Speech, Hearing and Phonetic Sciences at UCL. This anechoic room has a reverberation time of only 20 ms and very low levels of background noise (Nevard & Fourcin, 1995) and is, therefore, suitable for HRTF measurements.

2.2.1 Measurement setup

To obtain HRTFs, impulse responses are measured between loudspeakers positioned at specific locations in space and two microphones mounted in the ear canals of a participant (Carpentier et al., 2014; Xie, 2013; Zhang et al., 2009). HRTF measurements require high
CHAPTER 2. PILOT STUDIES

accuracy of hardware components and great care and precision during the measurement. Therefore, the equipment was carefully selected based on experience and findings from the current literature.

Ideally, the loudspeakers should have the characteristics of point sources. For this reason, Fostex 6301B loudspeakers were chosen, which combine a relatively flat frequency response with a compact size and a single driver to ensure continuity in phase-frequency characteristics. They were positioned at 270°, 0° and 90° azimuth in 1.15 m distance from a chair in the middle of the room at a height of 1.2 m above the grid floor (see Fig. 2.4a). The distance between the chair and loudspeakers was restricted by the dimensions of the anechoic room but was sufficiently large to minimise near-field effects.

For microphones, we chose Knowles FG 23329 omnidirectional electret condenser microphones which have a diameter of only 2.5 mm. They were mounted on silicone domes, manufactured for loudspeakers in hearing aids. This way, we were able to securely insert the microphones into the participant’s ear canals (see Fig. 2.4b) and accommodate for size differences in children and adults. To block the ear canal and avoid resonance effects, closed hearing aid domes were used.

Loudspeakers and microphones were adjusted in level and connected to a laptop via an RME Fireface UC audio interface.

(a) Loudspeaker and chair setup.  (b) Position of microphone in the ear.

Figure 2.4: HRTF measurement setup in the anechoic room.
2.2.2 Measurement of the impulse responses

For the measurement, the participant was seated on a non-rotating chair in the centre of the anechoic room. The chair was adjusted in height to align the ear entrances with the centre of the loudspeaker drivers. The seated position was chosen over a standing position to make the head position more stable and the measurement feasible for young children. Potential reflections introduced by the person’s thighs could be disregarded since the HRTFs were used to compare listening conditions in the virtual environment to the equivalent real environment at the same positions and the reflections would be present in both. Participants were instructed to tie back long hair that could otherwise affect the measurement and remove clothing items like jackets which could lead to noise during the measurements. Unlike most other studies on HRTF measurements, we only conducted impulse response measurements for the three positions that would be used in the following speech intelligibility procedure: at 270°, 0° and 90° azimuth and at 0° elevation from the participant’s head.

To calculate the head-related transfer function for each loudspeaker position, two impulse responses were measured. In the first measurement, impulse responses were measured between the loudspeakers and the microphones in the participant’s ear canal. The participant’s head position was adjusted to face the frontal loudspeaker and the participant was asked to remain motionless and silent throughout the measurement. Due to the short duration of the measurement of only three positions, no head fixture was used. The participant’s position and potential movements were monitored during the measurement via a camera installed above the frontal speaker. In a second measurement, the system response without participant was measured. The microphones were mounted on a thin stand and positioned at the equivalent location of the head centre. This impulse response measurement contained the combined effect of all hardware components in the measurement chain, for example, the transfer function of the loudspeakers, microphones and microphone preamplifiers without filtering through the participant’s body.

The signal generation and playback for the impulse response measurement were realised in Matlab using the ITA Toolbox\(^2\) for acoustic measurements developed at the Institute of Technical Acoustics at RWTH Aachen University (Dietrich et al., 2013).

\(^2\)ITA Toolbox, RWTH Aachen University: http://www.ita-toolbox.org/
tial test measurements were conducted to explore the suitability of various excitation
signals, such as different types of sweeps, white noise and pseudorandom sequences like
MLS and golay codes (Müller & Massarani, 2001; Zhang et al., 2009). Ultimately, ex-
ponential sine sweeps were chosen for future measurements due to their advantages of
avoiding distortion products and achieving a large SNR with a short signal duration,
which is particularly important when testing young children. The exponential sweep had
a duration of 2 s with a sampling rate of 44.1 kHz and a frequency range of 20 to 22500
Hz. The level was set to 75 dB SPL to minimise the influence of low-level noise from
the equipment. To increase the robustness of the measurement, the signal was presented
three times for each loudspeaker and averaged across repetitions.

2.2.3 Post-processing of impulse responses

Both sets of impulse responses were tapered off at 150 samples after the peak to re-
move potential reflections from the metal grid floor and then Fourier transformed into
the frequency domain. To obtain the HRTFs, the transfer functions of the individual
measurements were then divided by the reference transfer functions for each ear:

$$HRTF = \frac{H_{\text{individual}}}{H_{\text{system}}}$$

By this division, all effects introduced by hardware components were erased and did
not need to be accounted for separately. The remaining relative term describes only the
effects introduced by the participant. In the last step, the HRTFs were truncated to 1000
samples (23 ms).

2.2.4 Pilot measurements

Methods

The HRTF measurement was piloted with four young normal-hearing adults (2 female,
2 male, mean age 31 years). To test the robustness of the setup and the replicability of
the measurements, several separate measurements were taken for each participant.
Results

Measurement of the HRTFs at the left ear for all four participants are displayed in Fig. 2.5. The spectra are similar across all participants at low frequencies where they are mainly dependent on overall body dimensions. At higher frequencies, they show greater variations due to listener-specific differences in spectral peaks and notches which are created from filtering at the pinna and thus depend on pinna size and shape.

![Figure 2.5: HRTFs measured for four adults at the left ear for 270°, 0° and 90° azimuth and 0° elevation.](image)

As a measure of similarity between HRTFs, correlation coefficients were calculated between the spectra of repeated HRTF measurements from each participant and between HRTFs from different participants (see Table 2.2). The correlation between repeated measurements was very high for all four participants and three source positions, with an average of 0.97 and a minimum of 0.84. This suggested that the retest reliability of the
procedure was good. As an example, two repeated measurements of one participant can be seen in Fig. 2.6. The graph also shows HRTFs from the left and right ear. They are similar and differ only slightly due to differences in the pinna geometry. Correlations between the HRTF of left and right ear for ipsilateral, frontal and contralateral sound incidences were very high with an average correlation coefficient of 0.84 ± 0.07.

![Figure 2.6: Similarity between symmetric HRTFs (left and right ear) and repeated measurements for one participant for sound incidence from 0° azimuth.](image)

Between different participants, the similarity of HRTF spectra was lower and dependent on the source location. For ipsilateral sound incidence, the correlation coefficient was lowest with an average of 0.27 due to the large effect of listener-specific filtering (see Fig. 2.5). For frontal sound incidence, the correlation was 0.63 and for the contralateral sources 0.78. The results from this short pilot study confirmed that the setup offers good repeatability and precision and is suitable to measure HRTFs in the following studies.

<table>
<thead>
<tr>
<th></th>
<th>Azimuth</th>
<th>270°</th>
<th>0°</th>
<th>90°</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>correlation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left ear</td>
<td>r = 0.98</td>
<td>r = 0.98</td>
<td>r = 0.95</td>
<td></td>
</tr>
<tr>
<td>Right ear</td>
<td>r = 0.95</td>
<td>r = 0.98</td>
<td>r = 0.96</td>
<td></td>
</tr>
<tr>
<td><strong>Between-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>correlation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left ear</td>
<td>r = 0.23</td>
<td>r = 0.73</td>
<td>r = 0.77</td>
<td></td>
</tr>
<tr>
<td>Right ear</td>
<td>r = 0.80</td>
<td>r = 0.52</td>
<td>r = 0.32</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2: Mean correlations between HRTFs for measurements from four adult participants for frequencies from 50 to 20000 Hz.
Chapter 3

Spatial release from masking in adults in different auditory environments

3.1 Introduction

The ability to understand talkers in the presence of background noise is one of our most important skills for everyday listening situations. It is particularly difficult to understand talkers in the presence of other voices. In these situations, we benefit a lot from a spatial separation between the sound sources. This benefit is often called the spatial release from masking (SRM) and is defined as the difference in speech reception threshold (SRT) between conditions with colocated and spatially separated target and masker speakers (e.g. Arborgast et al., 2002; Bronkhorst, 2000; Freyman et al., 1999; Shinn-Cunningham et al., 2001).

SRM has been studied extensively for various target and masker signals, different symmetrical or asymmetrical spatial source distributions and in different groups of listeners (e.g. Best et al., 2017; Culling & Mansell, 2013; Freyman et al., 1999; Kidd et al., 2010; Peissig & Kollmeier, 1997; Shinn-Cunningham et al., 2005). Often these experiments are conducted using a virtual auditory environment (AE) presented with headphones. This realistic AE is created from measurements of head-related transfer functions (HRTFs). HRTFs describe the free-field transmission of sound waves from various locations in space.
to a person’s ear canal, including acoustic filtering by a person’s torso, head and pinnae. Sets of HRTFs are typically measured for a large number of positions in order to be able to reproduce any source location in space. This large number of measurements and the high requirements for equipment and anechoic test environment necessary to obtain precise results make HRTF measurements very demanding. They are also strenuous for the participants who are required to keep their body in a fixed position throughout the measurements.

Therefore, it is rarely possible to measure a participant’s HRTFs as part of a speech intelligibility study. Instead, HRTFs from artificial heads are often used. These dummy heads are constructed to represent an average adult and are therefore assumed to be a good approximation for most adults (Burkhard & Sachs, 1975). However, HRTFs are dependent on anatomical characteristics and vary between individuals. Their binaural components, interaural level differences (ILDs) and interaural time differences (ITDs), depend mainly on the size of the head. The spectral components depend on the characteristic shape of a person’s pinnae. Therefore, HRTFs from artificial heads are not equally well-suited for everybody and may produce a more realistic environment for some persons than for others.

As described in the main introduction, a large mismatch in spatial cues can result in a degraded perception of the auditory environment, e.g. to a reduced localisation accuracy, reduced externalisation and larger apparent source widths. This also leads to the environment being perceived as less realistic and the listener being less immersed in it (e.g. Brungart et al., 2017; Møller et al., 1996). Various methods have been developed to obtain HRTFs without measurements, such as calculation from individual geometric data or methodical selection of the best-fitting set of nonindividualised HRTFs (more detail in section 1.2).

This reduced ability to clearly perceive the position of sound sources and thus separate them could also have implications for speech perception. It is currently unclear whether HRTF mismatch affects accurate speech intelligibility in noise and whether individualised HRTFs are necessary to achieve accurate measurements for the SRM. Most studies investigating the importance of accurate spatial cues for the SRM have focused on the separate contributions of ITDs and ILDs (Culling et al., 2004; Ewert et al., 2017; Edmonds & Culling, 2005; Glyde et al., 2013b). The few previous studies that compared
3.2 METHODS

speech intelligibility in virtual environments generated from individualised and generic HRTFs found conflicting results. Drullman & Bronkhorst (2000) reported no difference in SRTs between generic and individualised HRTFs for bandlimited word and sentence targets. Rychtáříková et al. (2011) measured SRTs in a real anechoic environment and a virtual environment generated from artificial head HRTFs and found no difference in performance. Contrary to that, Orduña-Bustamañe et al. (2018) found better binaural speech intelligibility for individualised HRTFs than for generic HRTFs in most spatial configurations and a similar trend was described in Kondo et al. (2010). These studies all have different limitations, e.g. the filtering of stimuli, the choice of masking signals or the presence of better-ear listening due to asymmetric masker positions and it is still unclear whether their findings would be valid for other conditions. Generic HRTFs are often used in SRM studies or even in clinical applications such as the LiSN-S test (Cameron & Dillon, 2007) where it is impossible to measure individualised HRTFs before the test and it is critical to ensure that generic HRTFs are suitable to accurately measure the SRM in these conditions.

Therefore, in this study, we compared the SRM for sentences in the presence of two symmetric speech maskers in virtual environments generated by individualised and nonindividualised HRTFs and the equivalent real environment.

3.2 Methods

In this experiment, we measured the SRM in different real and virtual auditory environments. To create an individualised virtual environment, we first measured HRTFs for a small number of positions for each participant. In the second part of the experiment, we then measured the SRT, the signal-to-noise ratio (SNR) in dB between target and masker signals at which 50% of speech is intelligible, for each of the following conditions: for frontal target stimuli and colocated or spatially separated maskers in the real anechoic environment or in virtual environments computed from individualised HRTFs or HRTFs from other persons.

We hypothesised that if speech perception in adults is affected by mismatched spatial cues, SRTs would be higher for virtual conditions with nonindividualised HRTFs. Conditions with individualised HRTFs should represent the ‘true environment’ and thereby
lead to the same SRTs as in the real anechoic room.

### 3.2.1 Participants

Seventeen young adults (14 females, 3 males) aged 18 to 35 years (mean 23.6 years ± 5.0 years) participated in this experiment. They were recruited through the university. All participants were native British English speakers with no known listening difficulty or developmental disorder that could affect the outcome of the experiment. They had normal hearing thresholds (below 20 dB HL between 250 to 8000 Hz), which was confirmed by a hearing screening at the beginning of the test session. All participants gave informed consent and were financially compensated for their time. The study was approved by the UCL ethics committee (Project ID 0544/006).

### 3.2.2 Measurement of individual HRTFs

At the beginning of the test sessions, HRTFs were measured for azimuth angles of -90° (= 270°), 0° and 90° for each participant.

Therefore, the participant was seated on a non-rotating chair in the centre of an anechoic room at an equal distance of 1.15 m from three Fostex 6301B loudspeakers at 0° and ±90°. The height of the participant’s ear entrances was aligned with the centre of the loudspeaker drivers. Exponential sine sweeps between 20 and 22500 Hz at 75 dB SPL and with a duration of 2 s were used as excitation signals. The impulse responses were recorded with two Knowles FG 2332 omnidirectional electret condenser microphones placed at the participants’ ear canal entrances. These small microphones, with a diameter of only 2.5 mm, were mounted on closed silicone domes to secure them in the ear and block the ear canal. The posture of the participant was positioned before the measurement and the participant was instructed to remain still during the measurement.

Signal generation and playback were realised in Matlab using the ITA Toolbox\(^1\) for acoustic measurements developed at the Institute of Technical Acoustics at RWTH Aachen University (Dietrich et al., 2013). The combined effect of all hardware components in the measurement chain was accounted for by dividing the measured transfer functions by the equivalent system transfer functions, measured for the system without a participant. The resulting head-related impulse responses were averaged over three

\(^1\)http://www.ita-toolbox.org/
repetitions and shorted to 23 ms to remove possible floor reflections. A more detailed description of the measurement and post-processing can be found in section 2.2. For each participant, multiple measurements were conducted and the best one was selected to be used in the subsequent speech procedure.

3.2.3 Speech stimuli

The sentences used as target stimuli for measuring SRTs were taken from the ASL sentence lists (Macleod & Summerfield, 1990) which have simple vocabulary and syntax and contain three keywords per sentence. As the number of ASL sentences was limited, BKB sentences (Bench et al., 1979) were used for training purposes which are very similar in syntax and vocabulary (for examples see 2.1 in section 2.1). All sentence lists were recorded with a male Standard British English speaker in an anechoic room at a sampling frequency of 22.05 kHz. They were upsampled to 44.1 kHz to match the maskers.

As maskers, three freely available children’s stories were recorded for the same male speaker at 44.1 kHz. They were high-pass filtered at 50 Hz and pauses between words were shortened to 100 ms. For each trial, random segments within the two maskers were selected and started 1 s before the target presentation with a 100 ms ramp at the beginning and end of the signal to avoid onset and offset effect.

The SRM was measured for four virtual and real environments. In each AE the SRT was measured for two conditions: a colocated condition, where the target and the two maskers were presented at 0° azimuth, and a spatially separated condition, where the target was presented from 0° and the maskers from -90° and 90° azimuth. Even though the largest SRM appears to be achieved at slightly larger azimuths of about ±105° (Peissig & Kollmeier, 1997), ±90° was chosen for easy comparability to other studies (e.g. Cameron & Dillon, 2007).

The SRM was measured in a real anechoic room as well as for three different virtual environments using (1) the individualised HRTFs previously measured for the participant, (2) a set of HRTFs from a very large head and (3) a set of HRTFs from a very small head. These two sets of HRTFs were taken from the publicly available HRTF database of the Acoustics Research Institute Vienna², which contains HRTFs for 1150 positions for over 200 participants and also includes anthropometric data for 60 participants. The largest

²https://www.oeaw.ac.at/isf/das-institut/software/hrtf-database
and smallest head of the database were selected based on the sum of their head width, height and depth. The large (male) head had the dimensions 15.7 cm x 23.4 cm x 23.0 cm with a head circumference of 61 cm. The small (female) head had the dimensions 14.6 cm x 18.7 cm x 16.3 cm with a head circumference of 47 cm. Whilst the heads have been chosen for their large difference in head size, differences in pinna shapes were not accounted for.

3.2.4 Signal processing

To create the virtual environment under headphones, target and masker signals were independently filtered with the three different sets of head-related impulse responses, the inverse Fourier transforms of HRTFs, and added together to form two channels. In addition, the signals in both headphone channels were multiplied by the inverse of the headphone transfer functions to account for any spectral alteration introduced by the particular set of Sennheiser HD 25 headphones used in the experiment. These transfer functions were measured multiple times for three adult participants using the same Knowles FG 2332 microphones positioned in the participant’s ear canals as for the HRTF measurement. To erase any strong effects of pinna characteristics the average across all measurements was taken. Measurements on real humans were chosen over measurements with an artificial ear to produce realistic pressure and closure of the headphones on the outer ear. Since very low and high frequencies could not be measured correctly, the inverse transfer function was bandpass filtered with a passband between 70 Hz and 11 kHz.

In conditions in the real environment, the target and masker stimuli were played from three loudspeakers. Again, the loudspeaker transfer functions were measured before the experiment and the stimuli were filtered with the respective inverse functions to avoid the influence of the loudspeaker characteristics.

3.2.5 Procedure for SRT measurements

The participant was seated in the anechoic room at the same position as in the previous HRTF measurement and listened to the stimuli either via the loudspeakers or through Sennheiser HD 25 headphones. The stimuli were created in Matlab and presented via
an RME Fireface UC audio interface. The participant was asked to listen to the target sentences and to repeat them verbally. In the beginning, the participants underwent a short training of 30 sentences under headphones to get familiarised with the six different virtual conditions.

Then, SRTs were measured for the eight test conditions in an adaptive 1-up 1-down procedure (Plomp & Mimpen, 1979). Each adaptive run consisted of 15 target sentences from an ASL list presented in random order. The combined signal level of the two maskers was 65 dB SPL and the target level was adjusted according to the participant’s response (decreased for two or three correct keywords and increased for zero or one correct keyword) to track the SNR at which 50% of the keywords were intelligible, thus the SRT. The step sizes decreased from 4 to 2 dB at the first two reversals. The initial SNR was +5 dB SNR in the colocated conditions and -10 dB SNR in the separated conditions, corresponding to around 1.5 step sizes above the estimated average SRT based on a pilot experiment with four participants and similar studies in the literature (Besser et al., 2015; Cameron et al., 2011). This low initial level was chosen because we were limited to only 15 test sentences per measurement. To ensure that participants were adapted to the new condition at the beginning of the run and did not make careless mistakes due to inattention during the first trials, three BKB practice sentences were presented at the initial SNR before each adaptive run. SRTs were calculated as the mean of the reversals beginning from the third reversal (in case of an odd number of reversals, from the fourth reversal).

The 16 sentence lists and 6 possible masker combinations were pseudo-randomised between conditions across participants using Latin square randomisation. The order of presentation was randomised across participants with the following restrictions: conditions always alternated between colocated and separated maskers and between spatialisation forms, except for the two loudspeaker conditions which were embedded sequentially between the virtual conditions to minimise interruptions due to the switch between presentation forms. After a break, the conditions were repeated in reverse order to obtain a second measure for each SRT. In total, 240 test sentences were presented.

Including the hearing screening and HRTF measurement, the test session took approximately 70 minutes.
3.3 Results

Matlab (R2019a) and R (version 4.0.2)\textsuperscript{3} were used to conduct the statistical analysis.

3.3.1 Head related transfer functions

HRTFs of the left ear for a frontal sound source are displayed in Figure 3.1. They show relatively small differences between the 17 participants and the two nonindividualised HRTFs. The spectral characteristics of the small head (red line) at low and mid frequencies which are affected by head dimensions are shifted, as expected, towards higher frequencies compared to the large head (blue line).

![Figure 3.1: HRTFs for left ear at 0° azimuth measured for individual participants (black dotted lines) as well as for a small (red) and large (blue) head from a database.](image)

3.3.2 Speech reception thresholds

SRTs were averaged across the two repetitions. The SRTs for all eight conditions are shown in Figure 3.2. As expected, participants reached a better SRT of -16.4 dB SNR in conditions with spatially separated targets and maskers than in colocated conditions with a mean of -7.6 dB SNR. In both spatial conditions, the SRTs for the different forms of spatialisation were similar with slightly larger variances in virtual conditions with nonindividualised HRTFs.

\textsuperscript{3}Mixed-effects models were generated with the package lme4.
3.3. RESULTS

Figure 3.2: SRTs for maskers at 0° and ±90° azimuth in the four environments.

Figure 3.3 shows the SRM in the four different forms of spatialisation. The average SRM between colocated and spatially separated conditions was 8.8 dB. The mean SRTs and SRMs and their standard deviations are presented in Table 3.1.

Figure 3.3: SRM in the four auditory environments.

To investigate the effects of different parameters on speech intelligibility, a mixed-effects model was fitted to the SRTs. The model contained three fixed predictors: auditory environment (real, individualised HRTFs, large HRTFs, small HRTFs), masker locations (0°, ±90°) and gender of participant (female, male) as well the interaction of AE and masker locations. To control for other factors that might further influence the outcome, we added random effects for participant, sentence list and order, charac-
CHAPTER 3. SRM IN ADULTS

<table>
<thead>
<tr>
<th></th>
<th>SRT 0° (dB SNR)</th>
<th>SRT ±90° (dB SNR)</th>
<th>SRM (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Room</td>
<td>-7.7 ± 2.6</td>
<td>-15.9 ± 2.4</td>
<td>8.3 ± 2.7</td>
</tr>
<tr>
<td>Individual HRTFs</td>
<td>-7.5 ± 2.7</td>
<td>-16.2 ± 2.6</td>
<td>8.8 ± 2.1</td>
</tr>
<tr>
<td>Large head HRTFs</td>
<td>-7.6 ± 3.1</td>
<td>-17.0 ± 2.6</td>
<td>9.4 ± 2.4</td>
</tr>
<tr>
<td>Small head HRTFs</td>
<td>-7.5 ± 3.3</td>
<td>-16.6 ± 2.7</td>
<td>9.1 ± 3.2</td>
</tr>
</tbody>
</table>

Table 3.1: SRM and SRTs for colocated and separated maskers: mean and standard deviation for adults in the four real and virtual environments.
3.3. RESULTS

Figure 3.4: Relationships between colocated SRTs (SRT00), separated SRTs (SRT90) and SRMs in the four auditory environments (AE) - distributions, regression lines and correlations: Performance in the two SRT conditions correlated strongly, especially in environments based on individual HRTFs and large head HRTFs. The SRM correlated positively with SRTs in the colocated conditions and negatively with performance in the separated conditions. No effect of auditory environment was found.

Even though the group of participants as a whole showed no differences between the AEs, the individual results differed strongly. Two additional models were fitted to investigate whether these differences were systematically related to the participant’s head dimensions.

The head size of each participant (small, medium, or large) was estimated during the session. We compared the results of the four participants with the smallest head sizes (Listener ID: L03, L05, L14, L16) against the results of the four participants with the largest head sizes (Listener ID: L01, L02, L07, L11) in a mixed-effects model. The model included auditory environment (real, individualised HRTFs, large HRTFs, small HRTFs), masker locations (0°, ±90°) and head size (small, large) as well as the interaction between AE and head size as fixed predictors and participant as a random effect. The
interaction was not significant ($\chi^2(3) = 1.31, p = 0.72$), nor was the main effect of head size ($\chi^2(1) = 0.64, p = 0.42$). As in the previous model, there was a large main effect of masker location, the SRM, ($\chi^2(1) = 123.1, p < 0.001$) but none of AE ($\chi^2(3) = 6.22, p = 0.10$).

Binaural cues contained in the measured HRTFs are dependent on the head size, mainly the breadth of the head (Blauert, 1974). The maximum ITD contained in the measured HRTFs can therefore also be an indicator of a person’s head size. The maximum ITD for each participant was calculated as the time shift of the maximum in the cross-correlation function between the head-related impulse response of the left and right ear at lateral sound incidences. To only consider effects from the head but not the pinnae, impulse responses were low-pass filtered at 3 kHz (more details see Chapter 5). A model was fitted with auditory environment, masker locations and ITD as fixed effects and participant as a random effect. Similar to the previous model, masker locations were significantly affecting the SRTs ($\chi^2(1) = 250.4, p < 0.001$) but AE was not ($\chi^2(3) = 1.84, p = 0.61$). ITD, however, was also found to be a significant effect ($\chi^2(1) = 4.44, p = 0.035$) with participants with higher ITDs having lower SRT scores in all conditions as shown in Fig. 3.5.

![Figure 3.5: Relationship between ITDs and SRTs in colocated and separated condition.](image)
3.4 Discussion

This study investigated whether SRTs in adults in colocated and spatially separated conditions are affected by the accuracy of the virtual AE. Speech perception performance was the same in the anechoic environment and the individualised virtual environment. This confirms that the procedure is suitable to measure accurate SRTs in a virtual environment. As shown in Figure 3.2 and in statistical models, there was no difference in the SRTs measured in virtual environments generated from individualised or generic sets of HRTFs. Therefore, also the SRM was very similar in each environment. This result suggests that for normal-hearing adults there is no reduction in speech intelligibility in an environment based on generic HRTFs. Virtual environments using HRTFs from artificial heads seem appropriate for measuring speech perception in spatialised noise. For the investigated conditions of frontal target speakers in anechoic environments, HRTFs from artificial heads, therefore, seem to be suitable for speech perception tests.

Since it is possible that individual participants performed differently in the four environments dependent on their own head dimensions, the effects of head size and ITDs on SRTs were investigated. A comparison of participants with small and large heads did not show any differences in SRTs. A model with ITDs calculated from the individual HRTFs, however, showed a significant effect of ITD on speech perception. Participants with larger ITDs had significantly lower SRTs in both colocated and separated conditions (see Fig. 3.5). This effect was similar in all four environments and therefore could not be explained by a better fit of the generic HRTFs for some participants. Due to the small number of participants, the effect could have been caused by the above-average performance of the three participants with very large ITDs.

Interestingly, SRTs for colocated conditions were about 5 dB higher in comparable studies, e.g. -2.9 dB SNR for young adults in the Australian version of the LiSN-S test (Cameron et al., 2011) or -2.6 dB SNR in the North-American English version (Besser et al., 2015). However, the thresholds in the spatially separated conditions were very similar with -16.3 dB SNR and -16.4 dB SNR, respectively. These differences result in a smaller SRM in the current study of 9 dB compared to 13 dB in the LiSN-S. This effect could be due to differences in the recording or due to differences in the intelligibility of the talkers. As discussed in section 2.1, the stimuli in this study were recorded some
years apart and slight voice differences might lead to a benefit in speech perception particularly in the colocated conditions. The talker also has a voice that’s very clear and easy to understand and there is a possibility that he was more intelligible in the presence of his own voice than other speakers. Additionally, the stimuli in the current study were recorded with a male talker whereas the LiSN-S uses female talkers.

This study showed, that adults are able to achieve the same SRM for virtual environments created with generic sets of HRTFs as for their own HRTFs in the tested conditions. Two sets of HRTFs from very different heads were used in this study, which indicates that this effect is likely to hold for different sets of HRTFs such as commonly used artificial heads which are based on averaged adult head parameters. Whilst this suggests that generic HRTFs are suitable to measure accurate speech perception in adult participants, future research is necessary to show their suitability for children, who could be more affected by this mismatch due to their larger difference in head dimensions and proportions.
Chapter 4

Spatial processing and listening abilities in children with auditory processing disorder

4.1 Introduction

The spatial release from masking (SRM) in adults appears to be the same in different real and virtual auditory environments (AEs). Presenting the speech stimuli in two different virtual AEs generated from nonindividualised HRTFs did not degrade speech perception and lead to similar SRMs as in the anechoic room or a virtual environment based on individually measured HRTFs.

For children, however, the differences in spatialisation accuracy due to nonindividualised adult HRTFs are expected to be larger and might, therefore, affect speech perception. HRTFs describe the sound filtering between a sound source and the ear canal and are therefore dependent on a person’s torso, head and pinna dimensions. Nonindividualised virtual AEs used in SRM experiments in children and adults (e.g. Cameron & Dillon, 2007; Dieudonné & Francart, 2018) are typically generated from HRTFs of artificial heads, which have the anatomy of an average adult. Young children, however, are smaller in size and have different proportions than adults (Snyder et al., 1977). This affects their HRTFs and the binaural and monaural cues contained in them.

Interaural time and level differences arise from the differences in level and travel
time between the two ears. They are the most important cues for localisation in the horizontal plane and therefore also important for the SRM. Binaural cues are the largest for lateral sound incidences. The maximum ITD in adults is around 700 - 800 μs and the maximum ILD around 15 - 20 dB averaged across the hearing range (Blauert, 1974; Kuhn, 1977). ILDs are highly frequency-dependent and largest for high frequencies. ILDs and ITDs are mainly dependent on the breadth of the head and therefore smaller for young children. An average 6-year-old child only has a maximum ITD of about 600 μs and ILD of 12 dB (Fels, 2008). In adults, the just noticeable difference for wide-band signals at lateral directions was found to be around 70 μs for the ITD and 1 dB for the ILD (Iida, 2019). A mismatch in cues larger than that can therefore result in a detectable change in location. In virtual AEs based on adult-size HRTFs, children might therefore perceive sound sources to be more lateralised than in an individualised AE. For supra-ecological ITDs, larger than the maximum ITD of a person, psychoacoustic studies found a strong increase in JNDS and in the diffuseness of the sound image (Blauert, 1974; Mossop & Culling, 1998). This in turn can lead to a reduction in speech intelligibility (Ahrens et al., 2020). Hearing accuracy is also reduced for sounds that are strongly lateralised and perceived as coming from behind the ear.

Monaural cues are the characteristic spectral filtering of the sound due to the acoustic filtering by the body and especially the pinnae. For young children, spectral characteristics based on body dimensions are therefore shifted to higher frequencies and characteristics based on the shape of the pinna are altered in both frequency and magnitude.

A mismatch of HRTFs, therefore, affects different aspects of spatial hearing which are important to perceive the locations of sound sources, distinguish multiple sources and segregate the target from other sounds in the auditory scene. In adults, a mismatch of HRTFs was shown to lead to a decrease of localisation accuracy and acuity, a reduced externalisation, an increase in the apparent source width and spectral colouration as described in detail in section 1.3 (Iida, 2019; Møller et al., 1996). All these effects combined lead to the virtual AE being perceived as less clear and precise and feel less realistic. The size of these effects depends on the similarity of the listener’s anatomy to the anatomy of the person or artificial head used to generate the HRTF set. Whilst for perfectly individualised HRTFs the auditory display is the same for every person, it will differ strongly for nonindividualised ones and lead to a larger degradation of
auditory space perception for listeners who have larger anatomical differences from the generic model. Therefore this reduced quality of the virtual acoustic environment can be expected to be larger for young children. The degradations in sound accuracy make it more difficult for listeners to separate sounds and form acoustic objects. This can potentially lead to an increase in informational masking for speech-in-speech perception.

Children might also be more vulnerable to a degradation of the AE because their spatial processing and speech perception skills are still in development. Measures of basic spatial processing skills like localisation accuracy, spatial acuity and the binaural mask level difference seem to be adult-like from the age of 5 years (Lovett et al., 2012; Van Deun et al., 2009). More complex auditory processing skills, however, are still developing throughout childhood and are assumed to only fully mature in early adolescence, around the age of 10 to 11 years (Moore et al., 2011). Especially for younger children, individual variability is much greater than in adults and different auditory processing skills might follow unique developmental trajectories.

Children perform much poorer than adults in tasks on speech perception in noise (reviewed by Leibold & Buss, 2019). Differences in SRTs between children and adults are large and were found to exceed 20 dB in some test conditions (Litovsky, 2005). SRTs were studied in a large variety of different masking stimuli from purely energetic noise maskers over modulated noise and degraded speech to fully intelligible speech maskers. Children are especially susceptible to the interference of intelligible maskers, meaning the performance difference between children and adults is larger for perceiving speech in the presence of a speech masker than for a noise masker (Hall et al., 2002; Wightman & Kistler, 2005). Whilst noise maskers are spectrally dense and provide better energetic masking, speech maskers lead to higher informational masking due to their similarities with the target and thus increased uncertainty. Speech perception performance for both masker types gradually improves throughout childhood and matures in early adolescence (Brown et al., 2010; Corbin et al., 2016). Corbin et al. (2016) tested children aged 5 to 16 years on an open-set word recognition task in the presence of a speech-shaped noise or a two-talker speech masker. SRTs for noise maskers show adult-like performance around the age of 10 years, whereas SRTs for speech maskers only matured at the age of 13 years. There was no correlation between children’s performance in the two conditions, suggesting that different auditory processes contribute to the two conditions which have
Contradictory results have been found on the age of maturation for the SRM. Some studies showed adult-like performance even in very young children aged 3 years or below (Lovett et al., 2012; Murphy et al., 2011), whilst others found the SRM to only fully mature in early adolescence (Cameron et al., 2006c; Vaillancourt et al., 2008; Van Deun et al., 2010; Yuen & Yuan, 2014). These differences are likely to stem from differences in the measurement of SRM regarding target type, masker type and masker location as described in detail in section 1.2. SRM based on word recognition (Litovsky, 2005; Misurelli & Litovsky, 2012) seems to mature earlier than SRM based on sentence recognition (Cameron et al., 2006c; Vaillancourt et al., 2008).

Also, studies using symmetric maskers seem to find longer developmental trajectories for SRM (Cameron et al., 2006a) than studies with asymmetric maskers. This suggests that better-ear glimpsing might be a more complex effect than better-ear listening and relies on processing skills that develop later in life. Since listeners can use long-term favourable SNRs in one ear, the SRM is generally higher for asymmetric maskers than symmetric maskers (e.g. Culling & Mansell, 2013).

Whilst speech perception in informational maskers is improving throughout childhood, mixed results were found on the development of SRM for informational maskers. Some studies found it to mature later than for a noise masker (Cameron et al., 2006c), whilst other studies already found adult-like performance in young children (Misurelli & Litovsky, 2012). These differences might depend on the complexity of the procedure, e.g. the type of speech target (words or sentences) and the number of speech maskers. Speech maskers generally lead to higher SRMs than noise maskers, since informational masking due to similarities of target and masker is very high in the colocated condition and strongly reduced when the stimuli are spatially separated (Corbin et al., 2017). Speech maskers with very high similarity to the target which are more difficult to discriminate when colocated, e.g. same-gender talkers, lead to a higher amount of masking and higher SRM (Misurelli & Litovsky, 2015). Especially in the presence of intelligible maskers, young children have a large deficit in understanding a speaker compared to adults. Mismatched HRTFs and thus a degradation of the auditory space could therefore be even more disruptive in demanding speech perception tasks.

These disruptions might be even larger for children with a diagnosis of APD. These
children struggle with speech perception in demanding acoustic situations as described in detail in section 1.1 and their speech perception might therefore be especially vulnerable to mismatched spatial cues. The clinical LiSN-S test used in APD assessments is measuring the SRM in four different conditions (see details in section 1.2). Same and different voice maskers are presented colocated and symmetrically separated at ±90° in a virtual AE generated from artificial HRTFs (Cameron & Dillon, 2007). Children with suspected APD were found to perform worse than typically-developing (TD) children in SRT conditions with spatially separated maskers, thus leading to a smaller SRM (Cameron & Dillon, 2008). They also showed a larger variance in these conditions than TD children. Whilst some children achieve similar performance as children in the TD group, low performers can be more than two standard deviations below the norm. This deficit might indicate that these children have particular difficulties making use of binaural cues and difficulties attending to and integrating over short better-ear glimpses across the two ears.

Apart from studies using the LiSN-S paradigm, few other experiments have been conducted measuring speech perception in noise in spatialised conditions for children with auditory processing difficulties. Boothalingam et al. (2019) conducted a study measuring the SRM with a speech-shaped noise masker colocated or separated at 90° azimuth. They found no difference in SRTs between TD children and children suspected of having APD. This also supports the hypothesis that the deficit could lie in more advanced auditory processing skills, which are required for speech perception in the presence of intelligible maskers and for symmetric masker configurations where better-ear listening is reduced to rapid glimpses. The reason no differences were found might also be due to looser selection criteria for children in the suspected APD group. The study was conducted in a sound-treated audiometric booth, therefore there was no mismatch of binaural and spectral cues which could have influenced performance.

Similar abilities were also investigated in studies measuring the binaural masking level difference (BMLD). The BMLD is a measure of binaural interaction, thus on the ability to process interaural phase differences. It measures the difference in detection thresholds for a target in the presence of a noise masker in conditions where one signal is presented diotically and the other either in-phase or 180° out of phase across the two ears. The BMLD is sometimes used as part of a clinical test battery for APD assessment. Different
studies have shown APD children to have a deficit in BMLD tests (Sharma et al., 2014) or perform similarly to TD children (Gyldenkærne et al., 2014; Sharma et al., 2009). Contrasting results have also been found on the age of maturation with some studies suggesting that the BMLD matures during early childhood similar to the underlying binaural processes (Mattsson et al., 2018; Van Deun et al., 2009) whilst other studies found the BMLD to grow until the age of 12 years (Hall et al., 2004). Children with APD seem to perform normally for basic binaural hearing skills but struggle specifically in complex situations with speech understanding rather than detection. More research is needed to understand which factors contribute to these difficulties.

SRT performance in children with APD has not been compared between configurations in a virtual AE based on generic HRTFs and the equivalent real environment generated with loudspeakers. Therefore it is unclear if the differences in separated conditions between APD and TD children found in the LiSN-S are dependent on the virtual AE they were tested in. The inaccuracy of spatial cues due to the use of KEMAR HRTFs could potentially have led to a reduction in the abilities of children with APD to successfully separate maskers and targets in the separated conditions.

Overall, very little research has been done on HRTFs in children. The main reason for this is that measurements of full HRTF sets are lengthy and require the participant to remain motionless during the measurement which is difficult if not impossible for young children. Therefore, approaches have been made to calculate HRTFs from children from geometric head models. Fels et al. have numerically calculated HRTFs from children throughout childhood and adolescence based on geometric data gained from photogrammetry (Fels et al., 2004; Fels, 2008). Harder et al. (2013) have also generated a framework to generate a three-dimensional model from scans (described in more detail in section 5.1). This approach, however, is limited by the accuracy of the geometric mesh, and especially the resolution of the pinna surface, as it is difficult to obtain a geometric model that is precise enough to calculate HRTFs. Especially for young children who are likely to move more during data acquisition, meshes cannot be generated in a fully automatic procedure and require manual post-processing.

Recently, full sets of HRTFs have been measured in young children for the first time by Braren & Fels (2019). A fast HRTF measurement technique was used that made it possible to measure HRTFs in only three minutes using loudspeakers placed on a
4.1. INTRODUCTION

continuously moving circular arc. 26 children aged 5 to 9 years were tested with the procedure and in addition, anthropometric data from the head and torso was captured using a 3D scanner.

Since HRTFs of children are very difficult to obtain, no studies exist yet that use individualised HRTFs from children in experiments on spatial processing. Many studies on binaural hearing in children, however, have been conducted in a real environment using loudspeakers (Misurelli & Litovsky, 2015; Vaillancourt et al., 2008; Van Deun et al., 2010). Some studies have used nonindividualised HRTFs from artificial heads (Cameron & Dillon, 2007; Glyde et al., 2012). However, performance in real and virtual AEs based on artificial HRTFs have never been compared in the same children and it is unclear to what degree the mismatch in HRTFs affects spatial hearing abilities and speech perception. Potential effects might be stronger for young children or children with difficulties processing auditory information.

To address these concerns, an extended study comparing speech perception in different environments was performed with children with and without a diagnosis of APD. In this study, we investigated to what extent a mismatch of virtual acoustic cues influences the SRTs and the SRM in children and whether children with listening difficulties show larger effects. We measured the SRM in an anechoic real environment and equivalent virtual environments based on individual and generic HRTFs. We were aiming to answer the following research questions:

1. Is the SRM the same in real and virtual environments? Is it the same in virtual environments based on individualised and nonindividualised HRTFs?
2. Are there differences between TD children and children diagnosed with APD?
3. Is there an interaction between group and AE, namely a stronger deficit of children with diagnosed APD in some environments?

Since little information is known about child HRTFs, we also aimed to provide a small data set of HRTFs for children at the three stimulus locations which might help to gain insight into systematic differences in transfer functions between adults and children.
4.2 Methods

In this study, speech-in-noise perception was measured in children with and without APD in different acoustic environments. Besides a real anechoic environment, these include virtual environments generated from individualised HRTFs and generic HRTFs. Individual HRTFs were measured for each child at the beginning of the test session. SRTs were then measured in the presence of two colocated or spatially separated speech maskers in each environment, thus obtaining the SRM for that environment. In order to account for any contributions of language deficits or cognitive deficits, the study also included a short receptive language test and two questionnaires for the parents. To ensure all children had normal peripheral hearing, an audiogram was conducted at the beginning of the session. The audiogram included extended high frequencies since studies suggest that some children with APD and history of OME might have increased high-frequency hearing thresholds (Borges et al., 2013; Hunter et al., 1996) and there is evidence that also frequencies over 8 kHz are important for speech perception in complex listening conditions (Hunter et al., 2020; Monson et al., 2014; Motlagh Zadeh et al., 2019).

Since children with APD were recruited from all over the UK and some travelled far to participate, it was necessary to limit the study to a single test session. This affected the choices of test conditions and limited the use of additional screening tests. Overall, the session took around 2.5 hours per child including several breaks.

In general, we predicted that some of the children with auditory processing difficulties would have worse SRT performance than TD children. Based on previous literature (Cameron & Dillon, 2008), this deficit was expected to be stronger in the separated conditions which leads to reduced SRMs. We further hypothesised, that all children, regardless of being diagnosed with APD or not, would get the same results for the virtual environment with their individually measured HRTFs as for the real room. Performance in the condition with nonindividualised HRTFs, however, might be poorer due to the inaccuracy in the spatialisation of the sound sources. Listening with nonindividualised spatial cues requires listeners to adapt to these mismatched cues, which could be more difficult for children with APD who often have problems sustaining attention in complex auditory environments. Therefore, there might be an additional decrease in performance for them compared to TD children of the same age.
4.2. METHODS

4.2.1 Participants

43 children between the ages of 7 to 12 years (mean 10.2 years ± 1.6 years) took part in this study. 18 of them were children who were referred to an audiology clinic for listening difficulties and consequently diagnosed with APD. The other 25 children had no known listening difficulties. The age range was chosen based on the typical age of children being referred for APD assessment. APD is only clinically diagnosed from the age of 7 years onwards due to the complexity of some clinical tests (AAA, 2010).

Recruitment

The children with diagnosed APD were mainly recruited from APD clinics at audiology departments in two local hospitals, Great Ormond Street Hospital (GOSH), and the Nuffield Centre at University College London Hospital (UCLH). These are two of the few clinics in the UK which offer a full diagnosis of APD rather than a short screening assessment. Recruiting through these hospitals ensured that the children underwent a thorough test battery for APD involving both speech and nonspeech tests (see APD test battery in section 1.1) and were found to have difficulties in listening that couldn’t be explained by their peripheral hearing. Processing difficulties associated with APD vary strongly across children, therefore recruiting mainly from these clinics ensured that all children had objectively measurable deficits in listening abilities, thus, resulting in the most homogeneous group possible. Both hospitals incorporate a range of tests concerning different auditory processing skills such as speech perception in adverse conditions (e.g. LiSN-S test, Auditory Figure Ground and Filtered Words tests in the SCAN-3 battery, Speech-in-Babble), temporal processing (e.g. Random Gap Detection Test, Gaps-in-Noise test) and binaural listening (e.g. Dichotic Digits). They also include questionnaires for parents and teachers and take into account findings from previous assessments e.g. from educational psychologists. Additional APD participants were recruited from a private clinic and online support groups.

Typically-developing control participants were recruited through online parent groups and networks, local schools, newsletters at the university and the charity Royal National Institute for Deaf People as well as contacts of past participants. 7 of the 25 TD children had siblings who were diagnosed with APD and who took part in a different study at
the department. Additional tests were performed in the analysis to check that their performance did not differ from the other TD children.

**Inclusion and exclusion criteria**

All children were native English speakers and grew up in an English speaking country. Most children were monolingual British English, but two children were bilingual and three further children were exposed to a second language by one of their parents without actively speaking it. An inclusion criterion for all participants was normal peripheral hearing, which is typically defined as having audiometric hearing thresholds below 20 dB HL for frequencies between 250 Hz to 8 kHz. Therefore an audiogram was conducted at the beginning of the test session.

We aimed to recruit children with no other developmental disorder in both groups. Since APD is highly comorbid with other developmental disorders, this was not always possible. We, therefore, tried to minimise the effect of other disorders. We only included children for whom APD was their most prominent difficulty and who did not have any attentional disorder. Since language disorders were shown not to affect the SRM measure (Cameron & Dillon, 2008), children with a diagnosis of dyslexia were included in the study. None of the children had ear infections at the time of participating or in the weeks before.

**Consent**

At the beginning of the session, the study was explained to the child and the accompanying parent. Both, the parent and the child provided consent for taking part. All children received a young scientist certificate and a voucher as a thank-you for participating in the study. Families were reimbursed for their travel expenses. At the end of the study, families were informed about the general findings of the study and their implications for APD research.

**Ethics**

This study was approved by the UCL ethics committee (Project ID 0544/006). Ethical permission to recruit and test paediatric patients from the audiology departments of NHS sites was obtained (Nr. 18/LO/0250). Due to an agreement with GOSH, for participants
who had been diagnosed with APD there and upon parental consent, we were also able to access the medical records of their previous APD assessment. This applied to 9 children. The results of some of the tests, mostly the LiSN-S, were compared to performance in comparable conditions in our study.

4.2.2 Measurement of head-related transfer functions

At the beginning of the test session, HRTFs for three spatial positions were measured for each participant. They were then used to generate an individualised virtual acoustic environment to measure SRTs with different masker positions. The setup and procedure are described in more detail in section 2.2 and has been verified with adult participants in Chapter 3.

Procedure to measure individual HRTFs

Participants were seated in the middle of an anechoic chamber (Nevard & Fourcin, 1995) at an equal distance of 1.15 m to three Fostex 6301B loudspeakers positioned at 0° and ±90° azimuth. The chair was non-rotating and adjusted for each child to align the height of the ear entrances with the centre of the loudspeaker drivers (thus, attaining 0° elevation). Knowles FG 2332 omnidirectional electret condenser microphones, with a small diameter of only 2.5 mm, were placed at the participant’s ear canal entrances to record the impulse response. They were mounted on closed silicone domes to securely fix them for each child and block the ear canal. Due to the short duration of the measurement for only three spatial positions, no head fixture was used. Each child was positioned before the measurement and instructed to remain motionless and silent throughout the measurement (to “freeze” until the sounds stop).

Exponential sweeps with a frequency range of 20 and 22500 Hz were used as excitation signals. They had a duration of 2 s and were played at 75 dB SPL with a sampling rate of 44.1 kHz. Signal generation and playback were realised in Matlab using the ITA Toolbox for acoustic measurements (Dietrich et al., 2013). To account for the combined effect of all hardware components in the measurement chain, the measured transfer functions for each participant were divided by the equivalent system transfer functions, measured for the same system with the microphones mounted on a thin stand at the equivalent location of the head centre. The resulting head-related impulse responses were averaged across three
repetitions for each loudspeaker and shortened to 23 ms to remove potential reflections from the metal grid floor. Multiple measurements were made for each participant and the set of HRTFs with the fewest artefacts was selected to be used to filter the speech signals in the following SRT measurements.

Special care was taken to ensure all children felt comfortable remaining in the anechoic room by themselves. In some cases, practice measurements in the presence of the parents were conducted first to ease the child into the test. For two participants (N02 and A05), one of the parents remained in the back of the room during the measurement and was instructed to be motionless and silent.

Measurement of head parameters

In addition to measuring the HRTFs, measurements of four head size parameters were taken from each child: the circumference, height, breadth and depth of the head. The circumference was measured at the largest point of the head at the height of the eyebrows. The head height was measured as the distance from below the chin to the highest point of the head. The breadth was measured just above the ears. The depth was measured at the level of the eyebrows. A tape measure and a custom-made distance measure were used to take the measurements. Since the children were in an age range of five years, we expect age-dependent differences for each measure (Snyder et al., 1977). The head parameters were used to investigate differences in HRTFs between participants in Chapter 5.

4.2.3 Measurement of speech reception thresholds

SRTs were measured for seven conditions with different masker locations and auditory environments (AEs) in an adaptive procedure (Plomp & Mimpen, 1979).

Stimuli

ASL sentence lists, developed by Macleod & Summerfield in 1990, were used as target stimuli. They are constructed based on similar principles as the widely used BKB sentences (Bench et al., 1979) but not used in clinical tests, thus novel for all participants. Both sets of lists consist of short sentences with three keywords and simple syntax and vocabulary to specifically suit young children’s speech and language competence (e.g.
"The woman slipped on the ice" or "The yellow bananas are ripe", for more examples see Table 2.1 in section 2.1). The sentences are meaningful and partly predictable. Due to the limited number of ASL sentences, BKB sentences were used in the training phase and as practice at the beginning of each condition. Both sentence lists were recorded with a male Standard Southern British English speaker in an anechoic room at 22.05 kHz sampling frequency.

Three publicly available children’s stories with similar linguistic complexity as the target sentences were used as distractors (for a transcript see section B.1 in the appendix). They were recorded for the same speaker at 44.1 kHz. They were high-pass filtered at 50 Hz with a 3rd order Butterworth filter. Pauses between words were shortened to 100 ms and any audio distortions were removed. The final recordings had durations of 102, 142 and 217 s and were presented repeatedly.

There was a gap of several years between the recording of the target sentences and the distractors. Studies showed that ageing can result in small changes in voice characteristics (Goy et al., 2013; Stathopoulos et al., 2011). Therefore, this difference in the talker’s age and the two voice spectra were discussed in detail in section 2.1 (see Fig. 2.1). The ASL sentences were upsampled to 44.1 kHz to match the sampling frequency of the distractor stories. Due to the differences in original sampling frequencies, the target sentences only contained frequencies up to 11025 Hz whilst the maskers also contained very high frequencies up to 22050 Hz. Whilst high frequencies over 8 kHz can impact speech perception in some listening conditions (Hunter et al., 2020; Motlagh Zadeh et al., 2019), it is unlikely that frequencies over 11 kHz have an effect on speech perception since they are very low in power compared to lower frequencies. Since the higher frequencies are contained in the masker and not in the target, we assumed that they will not affect the SRTs and it is not necessary to limit the frequency range by downsampling to the lower sampling rate.

The levels of the ASL target sentences were normed for equal intelligibility in the presence of two distractors for children aged 8 to 10 years in a previous study (see section 2). The sentences were then regrouped into 14 balanced sentence lists, containing 14 sentences each.
CHAPTER 4. LISTENING ABILITIES IN APD CHILDREN

Spatialisation of speech stimuli

During the test procedure, target and masker signals were either presented from three different loudspeakers or via headphones filtered with HRTFs. In order to present the same stimuli at the listener’s ears and account for potential spectral alterations due to the hardware, the stimuli were filtered with the inverse transfer function of loudspeakers or headphones.

For the loudspeaker presentation, signals were played through three channels and filtered with the inverse transfer functions of the corresponding loudspeaker. These transfer functions were measured before the experiment at the estimated position of a participant’s head.

In the virtual conditions, each speech signal was filtered with the corresponding HRTF and with the inverse of the headphone transfer function. Three different sets of HRTFs were used: the individually measured HRTFs from the participant, generic HRTFs from an artificial head and simplified HRTFs from a spherical head model.

In terms of generic HRTFs, we chose to use a set from a KEMAR artificial head (Knowles Electronics Manikin for Acoustic Research, developed by Burkhard & Sachs, 1975). This head model was designed to represent the average adult dimensions for torso, head and pinna shapes based on anthropometric data of a large group of male and female adults. It can be fitted with two different pinna sizes. The KEMAR head is frequently used in the literature and is also used in the clinical LiSN-S test (Cameron & Dillon, 2007). The specific set of HRTFs used in this study was taken from the publicly available CIPIC\footnote{https://www.ece.ucdavis.edu/cipic/spatial-sound/hrtf-data/} database (Algazi et al., 2001). HRTFs recorded with the small pinna sizes were chosen because they are closer to children’s dimensions and would typically be used in studies involving children.

We also incorporated a condition in which the head was replaced with a rigid sphere. In this simple model, spatial cues are degraded to basic ITDs and ILDs of the head without any pinna or torso effect. This was done in order to investigate whether it is at all necessary to use human HRTFs to measure accurate SRTs in children or whether spatial cues can be approximated by a much simpler model. For adults, Glyde et al. (2013b) found a similar amount of SRM when using only binaural cues instead of full
HRTFs. The calculations were based on the spherical head model from Brown & Duda (1998). It includes a head shadow model to simulate ILDs that computes the diffractions of an acoustic wave at the sphere by Rayleigh’s solution with a single-pole, single-zero filter. ITDs are calculated from a time delay model which is an approximation based on the ray-tracing formula by Woodworth and Schlosberg for high frequencies (Blauert, 1974). In the same way as for the measured HRTFs, the transfer function is free-field normalised by dividing the pressure at the surface of the sphere by the pressure that would be present at the centre of the sphere in free field. The sphere was modelled to be similar to the average adult and, therefore, the KEMAR manikin with a circumference of 57 cm and symmetrically placed ear positions at ±100° azimuth (Blauert, 1974). Due to time constraints and limited ALS sentences, we decided to only measure the SRT for the spherical head in the separated condition. The potential effect of AE was expected to be larger in this condition where target and maskers have large differences in binaural and spectral cues than in the colocated condition where spectral filtering is similar for all stimuli.

For all virtual conditions, target and masker stimuli were filtered with the corresponding HRTFs and the inverse headphone transfer functions to remove any effect of the specific set of headphones used. Headphone transfer functions were calculated from calibration measurements with 4 adult participants before the experiment. A calibration with human participants was chosen over a calibration with an artificial head to measure the transfer functions with realistic pressure and seal of the headphones.

**Adaptive SRT procedure**

The participant was seated in the anechoic room at the same position as for the previous HRTF measurement. Custom Matlab scripts were used to execute the procedure. Stimuli were played either from Sennheiser HD 25 headphones or Fostex 6301B loudspeakers, both connected to the laptop via an RME Fireface UC audio interface.

In the beginning, the task was explained to each child and a short training was conducted. It consisted of three short adaptive runs with 8 BKB sentences to familiarise the child with all different masker locations and spatialisation techniques as well as the test procedure. At first, the easier condition with spatially separated stimuli was played with individual HRTFs, then the colocated condition with KEMAR HRTFs and finally
the separated condition using spherical head HRTFs. In case a child needed longer to understand the task and to establish reliable responses, more training was provided using BKB sentences.

After the training, SRTs were measured adaptively for the seven test conditions: colocated and separated conditions for individual HRTFs, KEMAR HRTFs and the real environment and the separated condition for the spherical head HRTFs. A Latin Square randomisation was used to pseudorandomise the 14 sentence lists and 6 possible combinations from the three maskers for each condition and participant. Conditions of the four different forms of spatialization were presented in pseudorandom order, randomly starting with the separated or colocated condition. Conditions with the same form of spatialization were always conducted consecutively to minimise interruptions due to change between headphones and loudspeaker presentation. All conditions were repeated in reverse order to increase the accuracy of the SRT and test the reliability of the participant.

Each adaptive run consisted of 22 sentences: 8 BKB sentences followed by 14 ASL sentences from one sentence list in random order. The BKB sentences were added to approximate the threshold level at the beginning of the run due to limited ASL sentence material. The two masking stimuli were started at random points within the first 20 seconds of the audio file and looped for continuous playback throughout the run. Their combined level was 65 dB SPL. The first target sentence was presented at a level of 75 dB SPL (+10 dB SNR) in colocated conditions and 70 dB SPL (+5 dB SNR) in spatially separated conditions. Depending on the participant’s response, the next sentence was presented at a higher or lower level. The target level was decreased for two or more correct keywords and increased for one or no correct keywords, thus tracking the SRT at 50% keywords correct. This step size decreased from initially 4 dB to 3 and 2 dB at the first two reversals, respectively. The high initial SNR was chosen intentionally to enable every child to clearly understand the first few sentences and be more motivated. SRTs were calculated as the mean of the levels visited for all trials with ASL sentences after the second reversal (including the calculated subsequent level after the last trial). Including breaks, this test took around 60 minutes.
4.2. METHODS

4.2.4 Extended high-frequency audiogram

Hearing thresholds between 250 Hz and 20 kHz were measured with the Etymotic Research ER-10X extended-bandwidth acoustic probe system. This device is able to generate precise signal levels at high frequencies due to its integrated system to calibrate the sound pressure at the probe microphone in the ear canal. This pressure calibration at the beginning of each test session is necessary since ear canals differ in length and shape among participants and there are slight differences in the position of the measurement probe in the ear canal at every insertion. The system uses forward pressure level (FPL) calibration that measures the sum of all sound waves propagating from the probe towards the eardrum in a person’s ear canal (Lewis et al., 2009; Souza et al., 2014). Compared to SPL calibration, which includes forward-going and reverse sound waves, this has the advantage of avoiding interference from standing waves in the ear canal. The probe system was connected to the computer via an RME Fireface UC audio interface. The execution of calibration and measurement procedure, as well as signal generation, was done with custom code in Matlab.

At the beginning of the test, the participant was seated in a sound-proof room and the probe was carefully inserted in the participant’s left ear. Special care was taken to ensure the probe sealed the ear canal properly and had a deep and stable fit. To avoid any change in probe position during the test, the probe cable was attached to a plastic headband worn by the participant. The cable was draped over a stand to avoid any noise from cable movement over surfaces and any pull on the probe due to movements of the participant. First, the FPL calibration was conducted using the ARLas\textsuperscript{2} Matlab package. Therefore, the calibration signal was simultaneously played and recorded from the probe. The child was instructed to remain quiet and motionless throughout this measurement and the calibration was repeated in case the child moved. The resulting FPL correction was then used to correct the absolute signal levels in the following audiogram measurement.

Three sinusoidal pulses with a duration of 1 s were used as stimuli. These short tones were chosen over a single longer one because they are easier to detect, especially at high frequencies. The child was told to remain quiet, pay attention and raise their hand whenever the sounds were audible. Hearing thresholds were measured at 250, 500, 1000,

\textsuperscript{2}Goodman, S. Auditory Research Lab audio software (ARLas). version 0.20.2, date 2017-11-04. https://github.com/myKungFu/ARLas
2000, 4000, 8000, 11000 and 16000 Hz for every participant. Due to limitations of the maximal signal output level, thresholds at 20 kHz could only be obtained for participants with low hearing thresholds of maximal 10 dB HL. For participants with hearing loss at high frequencies, additional frequencies between 11 and 16 kHz were tested. Thresholds were only measured until a minimum level of 0 dB HL. Calibration and hearing threshold measurements were then repeated for the right ear.

4.2.5 Language assessment

Ideally, all participants would have language abilities within the norms for their age. Since it was difficult to recruit children diagnosed with APD who don’t have any comorbid disorder or difficulties that might affect their language skills, we conducted an additional short test to gain an estimate for their language abilities. The results were used to account for differences in performance based on different language scores and to detect participants that have to be excluded from the final analysis based on their low language scores.

Since we had to keep the study within one session, we were only able to incorporate a very short test. We, therefore, chose the “Recalling sentences” subtest of the Clinical Evaluation of Language Fundamentals (CELF-5 UK) to assess receptive language abilities (Semel et al., 2017). It is regarded as a good indicator for language skills (Conti-Ramsden et al., 2001) and can be administered in a very short time. The test consists of 26 sentences of increasing length and complexity. To ensure the sentence presentation were the same for every child, they were recorded with a female Standard Southern British English speaker and presented to the child in quiet via headphones. The child and experimenter were seated in a sound-proof room and the child was asked to listen to each sentence and repeat it back. A raw score was computed from the child’s responses, which was then transformed into an age-normed scaled score and a percentile rank.

4.2.6 ECLIPS questionnaire and background questionnaire

In addition to the behavioural tests, the study also contained two questionnaires that the accompanying parent filled in during the session.

The first one was the Evaluation of Children’s Listening and Processing Skills (EC-
4.3. RESULTS

LIPS (Lips) questionnaire that was developed at the University of Nottingham specifically to be part of the clinical assessment of APD (Barry & Moore, 2014; Barry et al., 2015). It consists of 38 statements on the child’s auditory, cognitive and language abilities that the parent rates on a 5-point Likert scale. It additionally also contains general questions on the child as well as on deficits in other domains. The results are summarised in one overall score as well as factors in five categories: (1) speech and auditory processing (SAP), (2) environment and auditory sensitivity (EAS), (3) language, literacy and laterality (LLL), (4) memory and attention (MA), and (5) pragmatic and social skills (PSS). All scores are calculated as raw scores as well as age- and gender-normed scaled scores and percentiles.

The second questionnaire addressed different medical and educational factors that could affect the child’s results in the experimental procedures (see document in the appendix section B.2). It contained questions about the child’s language background, history of middle ear infections including any instances of otitis media and grommets and special educational needs. It also included questions on diagnoses or suspicions of hearing loss and other developmental disorders in the child and close family members. For children with diagnosed APD, there were additional questions about the diagnosis and current and past interventions such as FM systems or auditory training software.

4.3 Results

The statistical analysis of this experiment was conducted in Matlab (version R2019a) and R (version 4.0.2).3

Data from three participants had to be removed before the analysis for different reasons. In the typically-developing group, two of the youngest participants (aged 7.0 and 8.1 years) were removed. One of them (N09) was removed because the child spoke another language at home and was only being exposed to English at school and with friends. The other child (N01) was removed because the child only finished the first repetition of the SRT measurements. In the APD group, one child (A18) was removed from the analysis who only had a screening for APD and didn’t undergo a full clinical assessment. The child also displayed a mild hearing loss during the test session. Thus, 3R Packages used for the analysis: tidyverse, lme4, sjstats, GGally, ggbeeswarm, psych, reshape2, patchwork.
the analysis only contains data from 40 participants, 17 in the APD group and 23 in the TD group. All children completed the full study, except for one TD participant (N02) who did not do the CELF-RS language assessment.

4.3.1 Participant information and background questionnaire

Age and Gender distribution

![Bar chart showing age distributions of children in the APD and TD group.](image)

Figure 4.1: Age distributions of children in the APD and TD group.

Fig. 4.1 shows the age distribution for both groups. In our sample, there were fewer young children with diagnosed APD than control children. This age mismatch was just about significant ($t_{(38)} = 2.09, p = 0.043$). It arose due to difficulties recruiting very young children with a diagnosis of APD. Even though APD assessments are offered from the age of 7 years, many children only get diagnosed at a later age. Families might not be interested in participating in research immediately after receiving the diagnosis. Approximately equal numbers of participants were tested for the ages of 9 to 12 years. This age difference was accounted for in the analysis by calculating age-dependent $z$-scores and reduced by removing data from two young TD children as described above.

Age and gender information from the 40 participants who were part of the analysis is displayed in Table 4.1.
4.3. RESULTS

<table>
<thead>
<tr>
<th>Group</th>
<th>Number</th>
<th>Gender</th>
<th>Age (years)</th>
<th>female</th>
<th>male</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>APD group</td>
<td>n = 17</td>
<td>8</td>
<td>9</td>
<td>10.8</td>
<td>1.2</td>
<td>9.0</td>
<td>12.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD group</td>
<td>n = 23</td>
<td>13</td>
<td>10</td>
<td>9.9</td>
<td>1.7</td>
<td>7.2</td>
<td>12.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>n = 40</td>
<td>21</td>
<td>19</td>
<td>10.3</td>
<td>1.6</td>
<td>7.2</td>
<td>12.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Participant information: age and gender distribution in the two groups.

Prevalence of other developmental disorders

For both groups, we tried to recruit children with no other developmental disorder. Due to the nature of APD, which is highly comorbid with several other disorders (see section 1.1), there are few children with purely auditory difficulties. This might also depend on the fact that many children seeking an assessment for APD went through assessments for other developmental disorders before, which only explained parts of their difficulties. For this reason, we also included children with other developmental disorders as long as APD was assumed to be the predominant condition. Any additional diagnoses or cases where families suspected another developmental disorder and were in the process of getting an assessment were recorded and taken into account in the analysis. Three children in the APD group were diagnosed with dyslexia and one child with specific language impairment. Whilst these language disorders have a potential influence on the speech perception score, they are expected to lead to a small degradation in all conditions and thus a stable measurement of the SRM. Two of the children reported learning difficulties and one a sensory processing disorder. We ensured none of the children had a diagnosis of Attention Deficit Hyperactivity Disorder or Autism Spectrum Disorder which could lead to effects on attention that can’t be controlled for in the experiment. Table 4.2 lists the prevalence of disorders in participants of the two groups. It also shows the number of cases in which children were suspected to have tendencies of a specific disorder and were in the process of undergoing assessments or had only a screening test.

History of middle ear infections

In the background questionnaire, parents were asked about their child’s history of middle ear infections. Prolonged and recurring middle ear infections and especially in their more
CHAPTER 4. LISTENING ABILITIES IN APD CHILDREN

Table 4.2: Diagnosed and suspected (in brackets) developmental disorders in the two participant groups.

<table>
<thead>
<tr>
<th>Disorders</th>
<th>APD group (n)</th>
<th>TD group</th>
</tr>
</thead>
<tbody>
<tr>
<td>APD</td>
<td>16 (1)</td>
<td>0</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>3 (2)</td>
<td>0</td>
</tr>
<tr>
<td>Specific language impairment</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Dyspraxia</td>
<td>0 (2)</td>
<td>0</td>
</tr>
<tr>
<td>Autism Spectrum Disorder</td>
<td>0 (2)</td>
<td>0 (1)</td>
</tr>
<tr>
<td>Attention Deficit Disorder</td>
<td>0 (1)</td>
<td>0</td>
</tr>
<tr>
<td>Learning difficulties</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Sensory processing disorder</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Other disorders</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

severe form of otitis media with effusion (OME) have been linked to APD (Borges et al., 2013; Khavarghazalani et al., 2016) and to reduced performance in the clinical LiSN-S test (Graydon et al., 2017; Tomlin & Rance, 2014). Parents were asked about the number of ear infections, their onset time and duration and whether the child has been diagnosed with OME (also called glue ear) and received grommets as treatment. 47% of the APD children and 64% of the TD children reported no middle ear infections (see Fig. 4.2). 13% of the APD children and 32% of the TD children had few middle ear infections and 40% of the APD children and 4% of the TD children reported many infections. 40% of the APD children and 12% of the TD children were diagnosed with OME at least once. Most of them were subsequently treated with grommets in one or both ears: in total, these made up 22% of the children in the APD group and 8% of the TD group. Overall, a similar percentage of children in both groups suffered from middle ear infections. They, however, differed strongly in severity: children diagnosed with APD were more than three times as likely to have experienced OME than children in the control group.
4.3. RESULTS

The background questionnaire included further questions e.g. on the child’s educational needs or the prevalence of disorders in the family. For children with APD, questions were asked about the diagnosis and interventions. Results and group comparisons are listed in the appendix (Fig. A.2 and Fig. A.3).

4.3.2 Speech reception thresholds

The SRTs for the seven test conditions are displayed in Fig. 4.3. Overall, a large difference between SRTs in colocated and separated conditions was obtained in each of the auditory environments.

Figure 4.3: SRTs for all participants in the seven test conditions.
Learning effect

Two SRT measurements were done for each test condition. In the majority of cases, these repeated SRTs were similar. The average absolute difference between the repetitions was 1.8 dB with a low standard deviation of 1.8 dB. Overall, a small improvement in performance in the second repetition of -0.6 dB could be observed. This improvement seemed to be stronger in the colocated conditions with a significant difference of -0.9 dB \( t(222) = 3.36, p = 0.001 \) than in the separated conditions with -0.3 dB \( t(315) = 1.28, p = 0.201 \). When performing t-tests for individual conditions, the improvement was only significant for the two colocated conditions generated from HRTFs (individual HRTFs: \( t(75) = 2.05, p = 0.044 \); KEMAR HRTFs: \( t(63) = 2.20, p = 0.032 \)). This systematic effect was most likely due to ongoing learning for some participants during the first repetition even after the initial training. The colocated condition is more demanding since it has higher informational masking and thus makes it more difficult to separate target and maskers and attend to the target talker. This increase in performance in the second repetition was similar in both participant groups \( t(245) = 0.48, p = 0.629 \). APD children did not need more time to learn the task than TD children. Individual SRT results for each participant are displayed in Fig. 4.4.
Figure 4.4: SRTs of individual participants (APD group in red, TD group in blue): Measurements for first (light shade) and second repetition (dark shade) and averaged value after removing outliers (yellow star).
**Outlier removal**

To ensure only SRT measurements were included in the analysis which represented a listener’s true threshold, we investigated conditions with large differences between the repetitions or with adaptive tracks that did not appear to converge and thus had a very steep slope. These differences were likely to stem from temporary inattention or from a systematic effect of improvement due to learning or degradation due to fatigue. Fig. 4.5 shows the distribution of SRT differences between the first and second repetition. It has a mean of -0.6 dB and a standard deviation of 2.5 dB. Conditions were treated as outliers if they had very extreme differences of more than three standard deviations, thus more than 7.5 dB, between the two repetitions. This condition applied to 5 out of 280 pairs of SRTs which had a Cooks distance larger than 0.02 (see Fig. 4.5). These belonged to two children in the APD group and one child in the TD group and were conditions with separated maskers in different AEs. All pairs showed an improvement from the first to the second repetition, suggesting that these children might have needed more time than others to understand the task and attend to the target speaker. Since we used open sentence lists unfamiliar to the participant, participants were not able to achieve a lower SRT by chance. Hence we regarded the higher SRT values as outliers and remove them from the data set. For the other 275 pairs, the SRT was averaged across the two repetitions to gain a more robust value for the following analysis (see yellow symbols in Fig. 4.4).

\[\text{Figure 4.5: Difference between repeated SRT measurements: (A) distribution of the difference between first and second measurement, (B) spread of differences with outliers which are larger than } \pm 7.5 \text{ dB marked in red.}\]
Model for SRT

A linear mixed-effects model was fitted to describe SRTs as a function of the fixed effects of auditory environment (real, individualised, KEMAR), masker location (0°, ±90°), group (APD, TD), age and repetition (first, second) and their interactions and the random effect of participant and gender. Since for the AE generated with spherical HRTFs we only measured SRTs in the separated condition, we removed it for this analysis to get a full factorial design. Likelihood ratio tests were used to evaluate the influence of fixed effects and their interactions. All main effects were significant, but only one of the interactions. As expected based on previous literature (Cameron & Dillon, 2007), SRTs generally improved with age ($\chi^2(1) = 19.78, p < 0.001$). This improvement was on average 0.6 dB per year and was the same for both groups. SRT performance across age for each condition is shown in Fig. 4.6. There was also a significant difference between the two groups ($\chi^2(1) = 5.61, p = 0.018$) with the APD children on average performing 1.1 dB worse than the control group. Masker location ($\chi^2(1) = 749.76, p < 0.001$) and AE ($\chi^2(2) = 26.42, p < 0.001$) also contributed significantly to the model. However, their interaction did not, suggesting that the benefit from spatial separation was the same in all environments. The main effect of location accounted for SRT differences of 4.6 dB between separated and colocated conditions. Average SRT differences between the environments were up to 1 dB. A significant interaction between masker location and age was found ($\chi^2(1) = 6.11, p = 0.013$), suggesting that the SRM was increasing with age, on average by 0.3 dB per year. None of the other interactions was significant (all $p > 0.48$). A significant difference between the first and second repetition of 0.6 dB on average could be observed ($\chi^2(1) = 14.30, p < 0.001$), supporting the finding of a small learning effect over time. In terms of random effects, variations between participants accounted for 48% of the total variance in the model. Gender did not account for any variance in the data and was dropped from the model.

In a second step, two separate linear mixed-effects models were fitted for the three colocated conditions and four separated conditions to compare SRT differences. They each contained the fixed effects of auditory environment, group and age and a random effect of participant to control for variability between participants. In the model for colocated conditions (three left panels in Fig. 4.6), differences between the AEs ($\chi^2(2) = \ldots$)
17.46, \( p < 0.001 \) and for age (\( \chi^2(1) = 12.28, p < 0.001 \)) were significant, but not between the two groups (\( \chi^2(1) = 3.65, p = 0.056 \)). A post-hoc Tukey HSD test confirmed that the differences were mostly due to differences between the real and the KEMAR conditions (\( p = 0.042 \)), whilst the individualised condition did not differ from any of them. The significant main effect of age showed an improvement in SRT of 0.6 dB per year. SRTs of the two groups differed by 0.9 dB on average.

In the equivalent model for the separated conditions (four right panels in Fig. 4.6), all three factors accounted for significant differences (AE: \( \chi^2(3) = 94.63, p < 0.001 \), age: \( \chi^2(1) = 18.88, p < 0.001 \), group: \( \chi^2(1) = 4.39, p = 0.036 \)). A post-hoc test showed that the differences between AEs in this model were due to large differences between the spherical condition and all other conditions (all \( p < 0.002 \)). No difference was found between the real, individualised and KEMAR condition. SRTs in the spherical condition were on average 2.6 dB worse than in the other environments. Age accounted for an improvement of 0.8 dB per year. The average difference between APD and TD group was 1.2 dB. The results from these two models suggest that differences between the two groups in the previous model are mainly driven by the results of the separated conditions.

![Figure 4.6](image-url)

Figure 4.6: SRT performance with age in the seven test conditions: individual results and regressions for the two participant groups (TD group: square symbols for children with APD siblings, triangle symbols for other children).
Age-normed scores for SRTs

As mentioned above, there was a difference in age between the two groups, with the TD group spanning over a larger age range (7-12 years) than the APD group (9-12 years). Since age has a large influence on the SRT, age-dependent z-scores were calculated based on the performance of the TD participants to allow for comparison across groups.

A linear regression model was fitted for the TD group with age and condition as main effects and their interaction age*condition. Thus, an independent slope and intercept were fitted for each condition. A Levene’s test was conducted before to confirm that the necessary assumption of homogeneity of variance was met for all seven conditions ($F(6) = 0.27, p = 0.94$). To get a reliable model for age-dependent z-scores we excluded extreme values that were more than two standard deviations from the mean and recalculated the model without them. The regression model was then used to predict SRT z-scores for all participants. Positive scores relate to better performance, thus lower SRTs. Fig. 4.7 shows the SRT z-scores for both groups, as a regression over age and as boxplots for each group. Typically, values within two standard deviations of the mean of the norms are considered normal performance.
As detected in the SRT model, children with diagnosed APD performed worse in all seven test conditions. Whilst there was no significant interaction between age and group in the model, when normed by age, the difference in performance between the groups was larger for older children than for younger ones. Across all conditions, age and normed SRT in the APD group had a correlation of -0.31. This effect could be due to different factors: On the one hand, the difference in the age range between the two groups means that we are missing data for very young children in the APD group. Therefore, this trend might be weaker when more young participants would be included.

On the other hand, this result could be caused by a recruitment bias. Children are typically diagnosed with APD from the age of 7 years onwards. In the first years after their diagnosis, many families might be interested in taking part in a research study.
on APD. Later, the incentive to take part might be higher for families of children who continuously have strong listening difficulties and for whom coping strategies and the use of amplification systems haven’t shown any improvement. When comparing performance in each condition separately, significant differences between the groups were found in both conditions, in the real environment (colocated: $t_{(36)} = -2.82, p = 0.008$; separated: $t_{(34)} = -2.18, p = 0.036$) and in the separated condition calculated with individualised HRTFs ($t_{(37)} = -3.18, p = 0.003$).

<table>
<thead>
<tr>
<th>Group</th>
<th>Environment</th>
<th>SRT at 0° (z-score)</th>
<th>SRT at 90° (z-score)</th>
<th>SRM (z-score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD</td>
<td>Real environment</td>
<td>0.00 ± 1.00</td>
<td>0.15 ± 1.20</td>
<td>-0.15 ± 1.21</td>
</tr>
<tr>
<td></td>
<td>Individual HRTFs</td>
<td>0.13 ± 1.18</td>
<td>0.23 ± 1.26</td>
<td>-0.20 ± 1.37</td>
</tr>
<tr>
<td></td>
<td>KEMAR HRTFs</td>
<td>0.30 ± 1.28</td>
<td>0.11 ± 1.20</td>
<td>0.01 ± 1.46</td>
</tr>
<tr>
<td></td>
<td>Spherical HRTFs</td>
<td>/</td>
<td>-0.01 ± 1.39</td>
<td>/</td>
</tr>
<tr>
<td>APD</td>
<td>Real environment</td>
<td>-0.85 ± 0.90</td>
<td>-0.68 ± 1.20</td>
<td>-0.10 ± 1.28</td>
</tr>
<tr>
<td></td>
<td>Individual HRTFs</td>
<td>-0.40 ± 0.96</td>
<td>-0.85 ± 0.90</td>
<td>-0.79 ± 0.83</td>
</tr>
<tr>
<td></td>
<td>KEMAR HRTFs</td>
<td>-0.22 ±0.90</td>
<td>-0.51 ± 1.28</td>
<td>-0.11 ± 1.08</td>
</tr>
<tr>
<td></td>
<td>Spherical HRTFs</td>
<td>/</td>
<td>-0.68 ± 1.45</td>
<td>/</td>
</tr>
</tbody>
</table>

Table 4.3: Age-normed SRTs and SRM: means and standard deviations for the two groups.

7 out of 23 children in the TD group had a sibling with APD who took part in a different study (see squared symbols in Fig. 4.6). These children did not differ significantly from the other children in the TD group in terms of SRTs as verified by a linear mixed-effects model of SRT z-scores with APD sibling (yes, no) as a fixed effect and listener as a random effect ($\chi^2(1) = 2.32, p = 0.127$). However, their performance was slightly worse than that of the other TD children with an average difference in z-scores of 0.7.

### 4.3.3 Spatial release from masking

Overall, a large SRM was reached for all participants in all environments (mean 7.3 dB ± 1.9 dB). Children with diagnosed APD had a very similar SRM to typically-developing children. As expected by the lack of interaction between AE and masker location with
other fixed effects in the main SRT model, no difference in SRM was found between the conditions.

![Figure 4.8: Spatial release from masking results for individual participants and regression over age for both groups in auditory environments in the real room and based on individual and KEMAR HRTFs (TD group: square symbols for children with APD siblings, triangle symbols for other children).](image)

This was supported by a linear mixed-effects model for the SRM with *auditory environment* (real, individualised, KEMAR), *age* and *group* as fixed effects and *participant* as a random effect. No differences were found for AE ($\chi^2(2) = 0.00, p = 0.995$) and group ($\chi^2(1) = 0.59, p = 0.441$). Only age accounted for a small increase in SRM of 0.3 dB per year ($\chi^2(1) = 4.21, p = 0.040$). This age effect is small compared to the direct effect of age on the SRT, which on average improves 0.6 dB per year in the colocated conditions and 0.9 dB per year in the separated conditions (both including and excluding the spherical condition). Therefore this suggests that SRTs overall improve strongly with age, in the separated conditions more so than in the colocated conditions. The SRM across age for individual participants and regression curves for both groups are shown in Fig. 4.8.

Z-scores for the SRM were calculated with the same procedure as for SRTs (see Fig. 4.9 and Table 4.3).
4.3. RESULTS

Recall that the SRM was only measured in AE conditions in the real room and for virtual conditions with individualised and KEMAR HRTFs but not for the spherical head model. Since the SRTs for this head model differed significantly from the other separated conditions, there is a possibility that the overall SRM would also be smaller in this environment.

4.3.4 Hearing thresholds

The hearing thresholds for each participant and the averaged thresholds for the two groups are displayed in Fig. 4.10. Only values at frequencies above 1000 Hz were used in the analysis since measurements of low frequencies were found to be less reliable with the procedure we used leading to inaccurately high thresholds for some participants (see mean results in Tab. 4.4). These low frequencies are only presented at the correct level when the in-ear headphones are sealed correctly. It is possible that the seal was insufficient for children with small ear canals even though the best suitable size of earplug was chosen for each ear. The seal might also have loosened in some children who have moved a lot during the procedure. In these cases, the fit of the headphones was checked and roughly adjusted, however, due to time constraints it was not possible to repeat the full calibration and measurement. Low frequencies were always presented at the end of the procedure, therefore, potential adjustments of the headphones did not affect thresholds at higher frequencies.
In children with higher thresholds at high frequencies, additional frequencies were tested to evaluate the range of this hearing loss. These frequencies, however, were not included in the analysis when comparing across participants. The same was true for thresholds at 20 kHz. They were attempted to be measured for every participant but could only be measured for low hearing sensation levels due to equipment limitations in the maximal signal level that could be provided. This analysis, therefore, focuses mainly on thresholds at 1, 2, 4, and 8 kHz which are included in the standard pure-tone audiogram (PTA), and extended high frequency (EHF) thresholds at 11 and 16 kHz.

Figure 4.10: Average hearing thresholds and standard deviation for participant groups at frequencies in the PTA and EHF range.

Children with a diagnosis of APD showed worse thresholds across the whole frequency range and larger variations within the group (see Tab. 4.4). The average 4FAHL (average threshold for 0.5 - 4 kHz) was 6.0 dB HL in the TD group and 9.7 dB HL in the APD group. Most children were within the limits of normal hearing for all frequencies. 7 children in the APD group and 5 children in the TD group had one or more thresholds higher than 20 dB HL, but the average threshold was below 20 dB HL for all children.

In addition to the difficulties to maintain a correct seal of the headphones in some children, averaged thresholds for the groups might have been slightly overvalued since only levels of 0 dB HL and higher were tested. This was done since we were mainly interested in assessing normal hearing, thus thresholds below 20 dB HL in each participant, but might have resulted in a ceiling effect at group level.

The shapes of the two averaged threshold curves have likely been influenced by the equal-loudness contour used in the calibration software. Since the software was developed
4.3. RESULTS

specifically for the acoustic probe system the calibration function was based on a small number of adult participants and not adjusted for children. However, this systematic effect on all children is small and is not affecting differences between the participant groups, which are the main focus of this analysis.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>TD threshold (dB HL)</th>
<th>APD threshold (dB HL)</th>
<th>Group difference (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>250 Hz</td>
<td>14.9 ± 10.1</td>
<td>18.8 ± 11.0</td>
<td>3.9</td>
</tr>
<tr>
<td>500 Hz</td>
<td>11.9 ± 8.7</td>
<td>15.0 ± 11.4</td>
<td>3.1</td>
</tr>
<tr>
<td>1000 Hz</td>
<td>3.3 ± 3.9</td>
<td>7.9 ± 8.2</td>
<td>4.6</td>
</tr>
<tr>
<td>2000 Hz</td>
<td>2.4 ± 3.3</td>
<td>5.6 ± 6.2</td>
<td>3.2</td>
</tr>
<tr>
<td>4000 Hz</td>
<td>6.2 ± 5.2</td>
<td>10.3 ± 8.5</td>
<td>4.1</td>
</tr>
<tr>
<td>8000 Hz</td>
<td>7.9 ± 7.4</td>
<td>12.9 ± 7.9</td>
<td>5.0</td>
</tr>
<tr>
<td>11000 Hz</td>
<td>5.0 ± 6.8</td>
<td>12.1 ± 10.9</td>
<td>7.1</td>
</tr>
<tr>
<td>16000 Hz</td>
<td>5.5 ± 9.7</td>
<td>10.0 ± 10.5</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Table 4.4: Hearing thresholds: means and standard deviations in the two groups.

A linear mixed-effects model was fitted to explain influences in the hearing thresholds with group (APD, TD), frequency and ear (left, right) as well as the interaction of group and frequency as fixed effects and participant as a random effect. Even though the hearing thresholds were not normally distributed since no values below 0 dB HL were tested, a linear mixed-effects model was assumed to be reasonably robust. As suspected, the two groups differed significantly ($\chi^2(1) = 9.86, p = 0.002$). Thresholds of children in the APD group were on average 4.8 dB higher than those of TD children. There was no significant difference between the left and right ear ($\chi^2(1) = 1.85, p = 0.173$), suggesting that results from the two ears can be averaged for further analysis. Hearing thresholds, however, differed significantly across frequency ($\chi^2(1) = 12.52, p < 0.001$), likely due to the calibration for adult listeners mentioned before. No interaction between group and frequency was found ($\chi^2(1) = 0.86, p = 0.353$). Differences across participants accounted for 26% of the total variance. Post-hoc t-tests showed differences between the groups were present in most frequencies and were significant for 1, 8 and 11 kHz after correcting the significance level with the Bonferroni correction (all $p < 0.007$).
CHAPTER 4. LISTENING ABILITIES IN APD CHILDREN

Figure 4.11: Pure-tone average (PTA) and extended high frequency (EHF) hearing thresholds for participants with a history of OME (reporting a clinical diagnosis of glue ear plus the fitting of grommets in most cases) and participants with less severe middle ear infections (reporting no, few or many ear infections) or severe ear infections. The OME group consisted of 7 APD and 3 TD children.

Some literature suggests that children with a history of OME and/or a diagnosis of APD are likely to experience hearing loss in the extended high frequencies (Hunter et al., 1996, 2020; Margolis et al., 2000). 7 out of the 17 children in the APD group and 3 out of 23 children in the TD group reported a history of OME in the questionnaires. Most of them also reported having received grommets as a form of treatment. Fig. 4.11 shows the influence of group and OME on hearing thresholds in the PTA and EHF range. A more detailed figure splitting the severity of past middle ear infections into five categories is attached in the appendix (Fig. A.1). Participants with diagnosed OME with and without fitted grommets generally had higher hearing thresholds, especially at EHF. This effect was stronger in the APD group.

A mixed-effects model was fitted for hearing thresholds with fixed effects of OME diagnosis (yes, no), frequency range (PTA, EHF) and their interaction and the random effect of participant. Since there was a larger proportion of children with OME in the APD group, group was not included as a factor. The diagnosis of OME had a significant influence on the hearing thresholds ($\chi^2(1) = 7.63, p = 0.006$), but the frequency range did not ($\chi^2(1) = 2.22, p = 0.136$). Their interaction, however, was significant as well ($\chi^2(1) = 5.04, p = 0.02$), supporting the hypothesis that OME leads to increased thresholds at EHF. No statistically significant differences between thresholds in the
4.3. RESULTS

EHF and PTA frequencies were found in the APD group.

Since this study wasn’t primarily intended to examine the effect of otitis media there were large differences across participants in terms of severity, frequency and duration of middle ear infections.

4.3.5 Receptive language: CELF - Recalling sentences

The CELF-5 recalling sentences test (CELF-RS) is evaluated as an age-normed scaled score of 0 to 20 (displayed in Fig. 4.12a). Children with diagnosed APD obtained a significantly lower score than typically-developing children ($t(34) = -3.23$, $p = 0.003$). Their mean score was at $9.5 \pm 2.6$ as opposed to $12.2 \pm 2.5$ for the TD group. Three of the children in the APD group were diagnosed with dyslexia (A01, A11 and A17). Excluding them from the analysis still led to a significant difference between the two groups ($t(29) = -2.86$, $p = 0.008$).

![Figure 4.12: Scaled scores for receptive language in the CELF-5 recalling sentences subtest.](image)

Overall, most children were within one standard deviation of the mean which corresponds to a scaled score of 7 to 13. Whilst the APD group wasn’t significantly different than the normed average score of 10, the TD children were significantly better ($F_{(1,37)} = 10.42$, $p = 0.003$). This might be due to a recruitment bias in the control group that is often found in similar studies. Families are more likely to volunteer to participate in research studies when the parents have a higher educational background.
and the child performs well for their age. For the clinical group of participants, this bias is assumed to be much weaker. Their main motivation for participating was helping to advance research into the disorder and potentially learn more about the child’s abilities. Therefore, children who experience larger listening difficulties in daily life and thus potentially also language problems, are more likely to participate. The 7 TD children with siblings diagnosed with APD were not statistically different from the other TD children ($t_{(10)} = -0.5, p = 0.630$) but had a slightly lower average score of 11.7.

Both participant groups showed a trend for higher scores with age (see Fig. 4.12b), suggesting that older children performed above-average for their age. This effect, however, was not statistically significant ($F_{(1,37)} = 1.31, p = 0.26$).

### 4.3.6 ECLIPS questionnaire

One total score and five factors were calculated from the parents’ answers in the ECLIPS questionnaire on a scale of 0 to 20 (see Fig. 4.13). In terms of the total score, there was a large difference between the groups ($t_{(32)} = -9.09, p < 0.001$). Children with diagnosed APD had an average score of 1.4 whilst TD children scored 8.3. This clear differentiation was to be expected since the questionnaire is based on parental expectations of their child’s performance in listening-related tasks. The 7 TD children with APD siblings were rated even higher than the other TD children, with a mean of 10.1 compared to 7.5 ($t_{(20)} = 2.45, p = 0.023$), most likely due to a perceived better absolute performance in direct comparison to their siblings.

To evaluate differences in the factors, a linear mixed-effects model for the scaled score was fitted with the fixed effects of group (APD, TD) and ECLIPS factor (SAP, LLL, MA, PSS, EAS), their interaction and a random effect for participant. Both main effects were highly significant (group: $\chi^2(1) = 40.28, p < 0.001$, ECLIPS factor: $\chi^2(4) = 19.89, p < 0.001$). The same was true for their interaction ($\chi^2(4) = 18.83, p < 0.001$). These results reinforce that there was a clear differentiation between the groups in each factor. However, there were differences between the factors dependent on the group. Thus, a separate linear mixed-effects model was generated for each group. The models showed that the TD group did not have any differences between the factors ($\chi^2(4) = 8.94, p = 0.06$) whilst the factors differed significantly in the APD group ($\chi^2(4) = 36.47, p <$
0.001). A post-hoc Tukey HSD test showed that these differences were mostly due to the low score in the SAP factor, which was significantly poorer than the scores for MA, PSS and EAS (all \( p < 0.04 \)). This result was expected since children with APD experience the largest difficulties with speech and auditory processing.

Figure 4.13: ECLIPS scaled scores for the total scores and the five factors: speech and auditory processing (SAP), environment and auditory sensitivity (EAS), language, literacy and laterality (LLL), memory and attention (MA), and pragmatic and social skills (PSS).

The ECLIPS questionnaire also contains a language score that’s derived from the factors. Only a weak correlation of 0.26 and a nonsignificant effect in a regression model \( (F_{(1,38)} = 2.86, p = 0.099) \) was found between this score and the CELF-RS score, suggesting that they measure different language abilities.

4.3.7 Relationship between speech perception and other results

This study contained measures of speech perception (SRT and SRM), hearing (PTA and EHF hearing thresholds, history of middle ear infections), language (CELF-5 RS), and reported listening abilities (ECLIPS and background questionnaire). The correlation between the most relevant measures for all participants and the two groups separately are shown in Fig. 4.14. SRTs and SRMs are age-normed and averaged across environments. SRTs in the two spatial configurations correlated strongly. SRTs correlated weakly with the SRM in opposite directions, showing that the variability between participants in both masker conditions contributes to the size of the SRM.
Figure 4.14: Distributions and correlations between measures for the TD children (blue) and APD children (red) and the combined groups (grey) for age, age-normed SRTs for 0° and 90° conditions and age-normed SRM averaged across environments, scaled score of the CELF-5 recalling sentences, scaled ECLIPS total score, average PTA and EHF hearing thresholds and gender: Strong correlations between the two SRT measures were found in both groups. The SRM correlated positively with the separated SRTs and negatively with the colocated SRTs to a similar degree. A weak negative relationship between high frequency thresholds and speech perception measures was found. Hearing thresholds at PTA and EHF frequencies were also correlated. In the APD group, the CELF-5 score had a weak negative correlation with PTA thresholds and a positive correlation with EHF thresholds. Some of the other measures showed clear differences between groups but no notable correlations between measures.

Statistical models were used to investigate the relationship of speech perception with other measures. For models including all participants, age-normed SRTs were influenced by EHF hearing thresholds, a history of OME and the EAS, MA, and PSS factors of the
ECLIPS. Performance in all these measures, however, was related to participant group and effects on the normed SRTs might just describe general differences between the two groups. When investigating both groups separately, no effect of other measures on the age-normed SRTs or SRM could be observed. This might also be due to the small group sizes. Different statistical approaches were explored but none showed significantly large effects.

### 4.3.8 Comparison with results from clinical APD assessment

We were able to gain access to the medical records of the APD assessments for 9 of the 17 APD participants. Whilst the clinical assessment included a variety of different tests (see section 1.1), we were mainly interested in comparing scores of the LiSN-S test and audiometric thresholds. It is important to note that for some participants the clinical assessment was conducted up to 5 years prior to this study. Performance is likely to have changed during that time due to interventions or other factors.

![Comparison with results from clinical APD assessment](image)

Figure 4.15: Raw and normed colocated SRT and SRM in the KEMAR condition compared with previous results of the LiSN-S test during clinical APD assessment (some participants tested twice in the clinic).
The low-cue condition in the LiSN-S is directly comparable to the colocated condition with KEMAR HRTFs and the spatial advantage with the SRM in the KEMAR HRTF conditions. Fig. 4.15 shows the raw and scaled scores for the colocated SRT and the SRM. All children had a lower colocated SRT compared to the LiSN-S results. Most children obtained a slightly smaller SRM. Both effects could be due to differences between the procedures (discussed in more detail Chapter 6) as both TD children and adults reached higher SRM in the LiSN-S than in this study. Z-scores were higher in the current study than in the clinical LiSN-S.

Audiometric thresholds measured in the clinic and the current study were similar. For frequencies between 1 - 8 kHz, they differed on average by 3.8 dB ± 2.8 dB with most participants having slightly worse thresholds in the current study. Only in two participants, average threshold differences exceeded 5 dB.

4.4 Discussion

The present study examined the effect of mismatched spatial cues on SRTs and SRM in children with and without listening difficulties. As expected from previous literature, speech perception performance in the presence of a speech masker improved with age (Cameron et al., 2009). This effect was larger in separated conditions than in colocated conditions. The improvement was on average 0.9 dB per year in the separated conditions and 0.6 dB per year in the colocated conditions. This suggests, that for both conditions mechanisms improving speech-in-noise perception such as better-ear glimpsing only mature in adolescence.

Due to this difference in the age effect on speech perception for colocated and separated maskers, the SRM also increased with age, however only by a much smaller amount of only 0.3 dB per year. Even young children achieved a large SRM. 7-year-old TD children reached an SRM of 6.9 dB, whilst the oldest TD participants aged 12 reached 8.5 dB.

The SRTs were very similar for the anechoic environment and the virtual AEs generated from individualised or KEMAR HRTFs. They differed by up to 1 dB in the colocated conditions and even less in the separated conditions. Therefore, also the SRM was the same for the three environments. These results suggest that the SRM is a stable
measure across the different auditory environments and can be measured well in virtual auditory environments created from nonindividualised HRTFs.

All children showed a degraded performance in the condition based on a spherical head model. Due to restrictions in time and speech material we were only able to include one SRT measure for this condition and therefore did not measure the SRM. We expected any potential differences between this and the other environments to be larger in the separated condition and therefore included this condition. On average, SRTs were 2.6 dB worse in this environment than in the other three AEs. A possible explanation for this are the large differences in spatial cues compared to human HRTFs as discussed in the following chapters. The AE based on spherical head HRTFs only contains the binaural cues from ITDs and ILDs which are approximately similar to the cues for a human head of the same dimensions. It does not contain any monaural spectral filtering since the measurement was done on the surface of a sphere without any pinna-like structure. The deficit in participants from both groups suggests that basic ITD and ILD information is not sufficient for optimal speech perception. Spectral peaks and notches, which depend on sound direction and pinna surface, are necessary cues for the perception of spatialised speakers. Binaural and monaural HRTF characteristics for children, adults and the spherical head model are compared in Chapter 5.

The children with diagnosed APD performed worse than TD children in all SRT measurements, on average by 0.9 dB in the colocated conditions and 1.2 dB in the separated condition. Since the deficit was present in the colocated and separated conditions, the SRM was similar in the two groups. Contrary to expectations, the variability was similar in both groups. In both groups, the variance was smaller in the colocated conditions. These findings were different to the clinical LiSN-S test which only showed differences in the separated condition. The results from this study are compared with results from the adult study and with the clinical LiSN-S in Chapter 6.
Chapter 5

Differences between HRTFs of children and adults

5.1 Introduction

Head-related transfer functions are used to create a virtual auditory display that is a realistic representation of a person’s auditory environment (AE). This virtual AE can be used in situations where it is difficult or impossible to place the person in the original physical environment. Many research studies use this technique for experiments that require spatialised sound sources. There are several techniques for generating virtual AEs, but computation from HRTFs is most commonly used because it is less demanding in terms of computing power and more flexible than room-related techniques such as Ambisonics, sound field synthesis or vector-based amplitude panning (Braren & Fels, 2019; Krebber et al., 2000; Xie, 2013).

HRTFs describe the filtering of a sound wave between a source and a person’s ear. An HRTF can be obtained by measuring the transfer function between a loudspeaker and a microphone in a person’s ear in an anechoic environment (as described in detail in section 2.2). HRTFs are typically measured for a large number of source positions in order to cover the full three-dimensional space. They depend on a person’s anatomical characteristics, mainly the torso, head and pinnae and are therefore highly listener-specific.

HRTFs have been extensively studied in adults (Blauert, 1974; Iida, 2019; Shaw, 1997;
5.1. INTRODUCTION

Xie et al., 2007; Wightman & Kistler, 1989) but very little research has been conducted on children’s HRTFs and the degree to which they differ from adult HRTFs. This is mainly due to difficulties in obtaining HRTFs from children. During HRTF measurements, the participant is required to sit or stand motionless for an extended period of time. This is difficult if not impossible for young children.

Efforts have therefore been made to instead model HRTFs for children based on geometric information of their heads and pinnae. Fels (2008) used photogrammetry to collect geometrical data from 95 children aged between 6 months and 17 years and generated three-dimensional CAD models of their heads. Their HRTFs were then simulated numerically from these models using the Boundary Element Method. Full sets of HRTFs and spatial cues were compared between an average adult, infant and kindergarten child. Harder et al. (2013) also proposed a framework to create three-dimensional models of children’s heads suitable for calculating HRTFs. Multiple 3D surface scans were taken from different angles and then combined using a semi-automatic algorithm. The scans were aligned in an iterative process to optimise the surface. Models were created for 6 children aged 10 months to 9 years.

These studies have shown that it is possible to accurately compute HRTFs from detailed three-dimensional geometric meshes of a head model. The limiting factor for calculating HRTFs from scanned heads is the spatial resolution of current scanning techniques (Harder et al., 2013). In particular, the pinnae have to be scanned with very high precision in order to sufficiently approximate listener-specific HRTF characteristics at high frequencies. This resolution cannot be achieved with standardised scanning methods such as MRI scans. In young children, changes in head position and facial expression between scans are inevitable. Therefore the construction of a high-quality 3D head model from scans cannot be fully automated and requires manual post-processing for each participant. These techniques can therefore only be used for a small number of participants.

In recent years, faster acoustic HRTF measurement techniques have been developed that reduce testing time by using interleaved test signals and fully automated testing procedures (Majdak et al., 2007; Richter, 2019). A first HRTF dataset of children was measured by Braren & Fels (2019). HRTFs were measured in a hemi-anechoic chamber with interleaved sweep signals presented by loudspeakers placed on a continuously moving
circular arc around the participant. Full sets of HRTFs with a resolution of 5° in azimuth and 2.5° in elevation could be measured in only 3 minutes. 26 children aged 5 to 9 years were tested with the procedure and in addition, anthropometric data from the head and torso was acquired with a 3D scanner. ITDs were smaller than for adults. Localisation performance for the artificial head HRTFs was predicted in a model and was similar to performance in adults but varied greatly across individuals. So far, these measurements have not been used in any follow-up experiment on binaural hearing.

The results from these simulations and the recent measurements indicate that the HRTFs of young children are significantly different to adult HRTFs. The smaller dimensions of the torso, head and pinnae lead to a reduction of binaural cues and a shift of monaural spectral cues to higher frequencies. However, results from Fels et al. (2004) showed that a simple approximation of children’s cues from a scaled-down adult model as used previously to adjust HRTFs for adult listeners (Middlebrooks, 1999a) was not sufficient to correctly represent a child’s anatomy and approximate the HRTFs. Besides overall size, the proportions of head and body measures are also different in children. A comprehensive study on children’s body measurements and their growth during childhood was conducted by Snyder et al. in 1977. They measured 87 different body measures from 4127 US-American children aged 2 weeks to 18 years including measurements of head and torso dimensions which are affecting HRTFs.

Binaural cues consist of interaural time and level differences (ITDs and ILDs) between the two ears (Middlebrooks & Green, 1991). They depend mostly on the breadth of the head and the distance between the ear and the shoulder. Both measures increase during childhood, e.g. from 13.4 cm in an average 3-year-old child to 15.2 cm in an 18-year-old adult for the breadth (Snyder et al., 1977). Therefore ILDs and ITDs are expected to be smaller in young children. ITDs are the dominant cues for low frequencies under 1500 Hz. Adults typically have maximum ITDs of 700 - 800 μs for sound sources at ±90° azimuth. Iida (2019) reported variations between 680 μs to 778 μs for a sample of 33 adults. Children were found to have smaller ITDs, e.g. around 500 μs for an average 6-month-old infant or 600 μs for a 6-year-old child (Fels, 2008). The HRTFs measured for children aged 5 to 9 years by Braren & Fels (2019) had an average maximum ITD of 615 μs. ILDs are strongly frequency-dependent and small for low frequencies. They increase with frequency and are the most important cues for frequencies over 1500 Hz. Therefore,
the size of the ILD depends strongly on the frequency content of a sound signal. For the same 33 adults, ILDs were measured for 1/3-octave bands in HRTFs at ±90° azimuth (Iida, 2019). The average ILD increased from 6 dB at 500 Hz to 31 dB at 8 kHz. Large variations were found between participants, e.g. a range of 5 to 32 dB for the 4 kHz band. Fels (2008) found slightly reduced ILDs in infants and young children.

Monaural cues consist of characteristic peaks and notches in the spectrum of each HRTF based on a person’s anatomy (Iida, 2019; Moore, 2012; Shaw, 1997). Below 4 kHz, the spectrum depends on acoustic filtering by the head and body and individual differences between HRTFs are very small. At higher frequencies, the spectrum depends on the highly listener-specific shape and size of the pinna and especially the concha. HRTFs were found to contain six characteristic peaks and notches below 20 kHz which can have level fluctuations of more than 10 dB. The lowest two peaks and notches were found to be most important for localisation (Iida et al., 2007).

Peaks in the spectrum are caused by resonance modes of the pinnae and are approximately constant in frequency for different directions of sound incidence. The first peak arises from the first mode generated in the concha in the horizontal direction, whilst the second peak is the first mode in the vertical direction along the pinna surface. In an analysis of HRTFs of 74 adults, Iida (2019) found an average frequency of 4.1 kHz for the first peak (with a range of 1.8 kHz or 0.6 octaves) and an average of 6.0 kHz for the second peak.

Spectral notches originate from filtering in cavities of the pinnae due to interference between direct and reflected sound waves. They are direction-dependent and most pronounced for the frontal direction. Notch frequencies increase with elevation and azimuth angles of the sound source, therefore notches are particularly important for vertical localisation. The first notch in frontal HRTFs in adults occurs at a mean frequency of 7.7 kHz (range: 4.5 kHz or 0.9 octaves) and the second at 10.3 kHz (range: 5.6 kHz or 0.8 octaves). Whilst the width of the human pinna matures at the age of 3 - 4 years, the pinna depth grows until the age of 9 - 10 years (Fels, 2008; Iida, 2019). Therefore, the frequencies and magnitudes of listener-specific peaks and notches were found to differ between adults and young children. Fels (2008) compared HRTFs for an average adult and a 6-month-old infant. The first peak for the adult was at 3.7 kHz but occurred at a much higher frequency of 6.5 kHz for the infant. However, for the small group of infants
in the study, large variations in peak frequency were found due to differences in body size.

Further studies of HRTFs in children are needed to better understand the effects of head size and head proportions on differences in binaural and spectral characteristics of HRTFs. Even with new, faster techniques, measuring HRTFs requires a highly specialised setup and places high demands on hardware making inclusion rarely possible for research studies and impossible for clinical tests. Finding suitable techniques to approximate a child’s individual HRTFs by simulation or by modifying generic HRTFs based on anthropometric data is therefore important for many tests on the perception of spatialised sound sources in virtual AEs. We, therefore, compared the small sets of HRTFs measured for children and adults in the studies described in Chapters 3 and 4 regarding their head parameters and binaural and monaural cues contained in the transfer functions. These results were also related to speech perception performance in an AE generated from generic HRTFs in the previous study.

5.2 HRTF measurements

As part of the studies described in the previous chapters, HRTFs for three source positions were measured from 43 children aged 7 to 12 years (23 female, 20 male) and 21 young adults (16 female, 5 male). HRTFs were measured at -90°, 0° and 90° azimuth and 0° elevation at 1.15 m distance. The measurement was repeated a minimum of three times for adults and four times for children. For children who had difficulties remaining motionless, more repetitions were measured and the best repetitions were selected. The measurement procedure is described in detail in section 2.2. In addition, four anthropometric parameters were measured for all child participants: circumference, height, breadth and depth of the head (see section 4.2). They were used to estimate the children’s head sizes. In the previous study, generic HRTFs from an artificial KEMAR head and a spherical head model were used. These two sets of HRTFs will therefore be included in the analysis of HRTFs.
5.3 Results

The analysis was performed in Matlab (version R2019a) and R (version 4.0.2). For each participant, the two most accurate HRTF measurements were selected for analysis\(^1\). They were identified by examining the time and frequency signal (both amplitude and phase information) and the group delay of all measurements for artefacts. In general, more artefacts were present in HRTF measurements of children than in measurements of adults because it was more difficult for them to maintain their posture during the measurement. Among children, there were approximately equal numbers of female and male participants at each age. For adults, 16 out of 21 participants in this study were female. Therefore, we expect their heads to be smaller on average than artificial heads.

Figure 5.1: HRIRs and HRTF spectra of the left ear for sound sources at the ipsilateral (-90° azimuth), frontal (0°) and contralateral (90°) positions: example for one adult participant with other adult and child participants in grey as context. Individual waveforms were clipped for better display and the functions were scaled to have equal RMS levels.

\(^1\)With the exception of two children for whom only one valid measurement could be taken (A07 and A13).
like the KEMAR which are based on median values of a large group of adults. Binaural and monaural cues for children were investigated in relation to age and anthropometric data from the head and compared to mature cues in the adult participants.

Fig. 5.1 shows an example of the head-related impulse responses and transfer functions for one listener at the left ear. Binaural cues were calculated from time delays and amplitude differences in the head-related impulse responses (HRIRs) of the left and right ear for lateral sound sources, monaural cues from spectral differences in the HRTFs for frontal sound incidence. Individual results for the children at each location can be found in Fig. A.4 in the appendix.

### 5.3.1 Anthropometric data

Head circumference, height, breadth and depth were measured in the 43 child participants. Measured values and regression lines across age are shown in Fig. 5.2. As expected, all four head parameters increase with age but for each of them, there was a large variability across participants.

Regression models were fitted for each parameter with the effects of *age* and *gender*. No effects of gender were found, suggesting that for all four measures systematic gender differences were far outweighed by variability between individuals. Only the models for circumference and height showed a significant increase of value with age (circumference: $F_{(1,41)} = 4.44$, $p = 0.041$, height: $F_{(1,41)} = 19.13$, $p < 0.001$, breadth and depth $p > 0.45$). The yearly increase was 0.26 cm for circumference and 0.28 cm for height. Growth-dependent changes in breadth and depth were marginal. On average breadth increased by 0.06 cm and depth by 0.04 cm per year.

The same four head parameters were also measured as part of the large study on the growth of body dimensions throughout childhood by Snyder et al. (1977). These growth curves were compared to the results in this study and are included in Fig. 5.2. For height and breadth, these curves were very similar to the average found in this study. Circumference was slightly smaller and depth larger than in this study, but showed very similar growth with age. This offset could result from slight differences in the exact measurement locations between the studies or from systematic differences between participants, as the data from Snyder et al. was from US-American children.
and measured over 40 years ago. More recent data for the circumference from Rollins et al. (2010) showed values similar to the data from Snyder et al. and a small difference between girls and boys (see Fig. 5.2a).

Figure 5.2: Anthropometric head parameters as a function of age for female (orange) and male (blue) participants: regressions over age for male, female and all participants (solid lines), average results from Snyder et al. (1977) (dashed black line with diamonds) and female and male results from Rollins et al. (2010) for circumference (dotted lines).

Table 5.1 displays the mean value of each measure across the age range of the children and reference values for adults. Adult measurements were taken for a large number of female and male adults. These median values were used to create the artificial KEMAR and HATS heads (Burkhard & Sachs, 1975; ITU-T P.58, 2013).

Children were smaller than adults in all four parameters, suggesting that their heads are not fully grown at this age. For example, the head height of a 12-year-old child was on average 1.3 cm larger than that of a 7-year-old, but still 2.1 cm smaller than that of the average adult. Smaller growth effects but a similar ratio can be measured for
CHAPTER 5. HRTF DIFFERENCES

breadth and depth. Fels (2008) found that circumference grows by 10 cm between the age of 5 years and adulthood. In this study, the mean difference between the youngest children aged 7 and adults was 3.8 cm, but variability among children was high and for the smallest participant the difference was 6.3 cm.

To gain a single measure of head size for further analysis, a simplified measure of the head volume was calculated as the multiplication of height, breadth and depth measures. As expected, volume correlated strongly with circumference ($r = 0.75$) but had a slightly stronger relationship with age than circumference ($r = 0.37$ to $r = 0.31$).

<table>
<thead>
<tr>
<th>Participants</th>
<th>Height</th>
<th>Breadth</th>
<th>Depth</th>
<th>Circumference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>n =</td>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 - 6.5 years</td>
<td>31</td>
<td>mixed</td>
<td>19.4 cm</td>
<td>14.5 cm</td>
</tr>
<tr>
<td>7 years</td>
<td>6</td>
<td>3f, 3m</td>
<td>19.0 cm</td>
<td>14.1 cm</td>
</tr>
<tr>
<td>8 years</td>
<td>2</td>
<td>1f, 1m</td>
<td>19.1 cm</td>
<td>14.3 cm</td>
</tr>
<tr>
<td>9 years</td>
<td>10</td>
<td>5f, 5m</td>
<td>18.9 cm</td>
<td>14.6 cm</td>
</tr>
<tr>
<td>10 years</td>
<td>9</td>
<td>4f, 5m</td>
<td>19.7 cm</td>
<td>14.5 cm</td>
</tr>
<tr>
<td>11 years</td>
<td>9</td>
<td>5f, 4m</td>
<td>19.6 cm</td>
<td>14.1 cm</td>
</tr>
<tr>
<td>12 years</td>
<td>7</td>
<td>5f, 2m</td>
<td>20.3 cm</td>
<td>14.7 cm</td>
</tr>
<tr>
<td>Adults &gt; 4000</td>
<td>mixed</td>
<td></td>
<td>22.4 cm</td>
<td>15.2 cm</td>
</tr>
</tbody>
</table>

Table 5.1: Anthropometric head measures for the child participants aged 7 - 12 years and reference values for younger children from Fels (2008) and for adults used to generate the KEMAR and and HATS artificial heads. (Burkhard & Sachs, 1975; ITU-T P.58, 2013)

5.3.2 Binaural cues

Interaural time differences

The maximum ITD for each child was calculated as the time shift of the maximum in the cross-correlation function between the left and right ear HRIR for sound incidences of -90° and +90° azimuth. The ITDs were calculated for two repetitions of the HRTF measurement and averaged over all four values to obtain a more stable measure of ITD. ITDs are plotted as a function of age in Fig. 5.3 and age-dependent averages are listed in Table 5.2. As girls and boys might have different growth trajectories, different colours were used for male and female participants and separate regression lines were fitted. The
5.3. RESULTS

children had ITDs ranging from 578 to 749 μs with mean and standard deviation of 645 μs ± 44 μs for girls and 666 μs ± 35 μs for boys. ITDs in the adults were between 635 - 771 μs, with 672 μs ± 30 μs in female participants and 723 μs ± 36 μs in males. Although the ITDs were spread across a wider range in children, the variance was the same as in adults (Levene’s test: $F_{(1,62)} < 0.001$, $p = 0.98$). The ITD of the KEMAR head was 748 μs which was higher than the average of the five male adults in this study. The spherical head model had an ITD of 635 μs.

![Figure 5.3: ITDs of children aged 7-12 years and adults for sound incidences of -90° and +90°.](image)

A linear mixed-effects model was fitted for the ITDs from the two repeated HRTF measurements with the fixed effects of group (children, adults) and gender (female, male) and a random effect of participant. There was a significant main effect of group ($\chi^2(1) = 12.47, p < 0.001$). Children’s ITDs were significantly smaller than those of adults, by an average of 37 μs. The same was true for gender ($\chi^2(1) = 8.32, p = 0.004$). The absence of an interaction between gender and group suggests that gender differences were present in both groups. However, they were larger in the adults with an average of 51 μs (as shown in a post-hoc t-test: $t_{(5)} = -2.8, p = 0.032$) compared to 22 μs in the children ($t_{(39)} = -1.54, p = 0.132$).

As expected, ITD increased with age in the child group (see Table 5.2), on average by 5 μs per year. However, since individual growth trajectories can vary widely among children, the variability across individuals was much larger than the average effect of
age. Whilst on average, children had ITDs significantly lower than adults, many children already had adult-like ITDs. The ITDs in children were similar to previous findings, e.g. 600 µs in 6-year-old children (Fels, 2008). As expected, ITDs in children correlated with the geometric measures of head circumference ($r = 0.29$) and volume ($r = 0.49$).

For adults, the ITDs from the two repeated measurements were very similar. Children displayed greater differences in ITDs between repetitions, possibly due to small movements during the measurements or posture changes between repeated measurements.

ITDs can be calculated with different methods (Katz & Noisternig, 2014). Instead of calculating the interaural cross-correlation for broadband HRIRs, low-pass filtered HRIRs at 3 kHz can be used that only include low frequencies relevant for ITDs and suppress pinna contributions. This method, however, also has the disadvantage of reducing the sharpness of the onset which can potentially lead to more variation. Calculating ITDs with that technique lead to similar results overall, but to no gender difference in adults (see Fig. 5.4). KEMAR and spherical head ITDs were more similar with 714 µs and 680 µs.

![Figure 5.4: ITDs across age as in Fig. 5.3 calculated from HRIRs low-pass filtered at 3kHz.](image)
Intraural level differences

The ILD is the difference in energy of the HRTFs of the left and right ear. ILDs are strongly frequency-dependent and larger at higher frequencies. The size of the ILDs is therefore strongly dependent on the spectral properties of a sound stimulus, e.g. will be lower for speech than for a broadband signal. The maximum ILD was calculated as the level difference between left and right ear HRTFs for sound incidences at -90° and at +90° and averaged over both locations. To account for the strong effect of frequency, two different values of the ILD were calculated: one for the entire audible frequency range and one for the frequencies that are most relevant for speech perception and thus for the SRM test used in the previous studies. Therefore, ILDs were first calculated for 1/3rd-octave bands with centre frequencies of 31.5 Hz to 20 kHz which roughly resemble human auditory filters. The global ILD was calculated as the average ILD over all 29 bands.

Figure 5.5: ILDs of children aged 7-12 years and adults: A) Global ILD across audible frequency range, B) ILD weighted with SII importance function.
CHAPTER 5. HRTF DIFFERENCES

The second measure of ILD was obtained by weighting the ILDs with an importance function based on the speech intelligibility index (SII). This function includes bands with centre frequencies between 160 and 8000 Hz with the highest weighting for the band at 2000 Hz and was obtained based on different types of speech materials and listening conditions to best reflect speech perception (ANSI-S3.5, 1997).

Same as for the ITDs, a mixed-effects model for the repeated measurements of the global ILD was fitted with fixed effects of group (children, adults) and gender (female, male) and a random effect of participant. No interaction or main effect of gender was found (both \( p > 0.81 \)) but there was a strong effect of group (\( \chi^2(1) = 11.52, p < 0.001 \)).

On average, the global ILD was 6.9 dB ± 0.9 dB for adults and 6.1 dB ± 0.9 dB for children. A similar effect was found for the SII-weighted ILD. Children had a mean ILD of 4.5 dB ± 0.7 dB and adults 5.0 dB ± 0.6 dB. As with ITDs, ILDs varied more in children than in adults, likely due to greater variation in head sizes and a higher likelihood of measurement inaccuracies. Global ILDs were correlated with head measures, mainly with circumference (\( r = 0.41 \)). ILDs calculated for the spherical head model were smaller since the spectra are flat at higher frequencies relevant for ILDs.

<table>
<thead>
<tr>
<th>Age</th>
<th>ITD</th>
<th>ILD (global)</th>
<th>ILD (SII)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>female</td>
<td>male</td>
<td>female</td>
</tr>
<tr>
<td>7 years</td>
<td>627 μs</td>
<td>646 μs</td>
<td>6.0 dB</td>
</tr>
<tr>
<td>8 years</td>
<td>623 μs</td>
<td>737 μs</td>
<td>6.1 dB</td>
</tr>
<tr>
<td>9 years</td>
<td>621 μs</td>
<td>669 μs</td>
<td>5.8 dB</td>
</tr>
<tr>
<td>10 years</td>
<td>660 μs</td>
<td>678 μs</td>
<td>6.1 dB</td>
</tr>
<tr>
<td>11 years</td>
<td>673 μs</td>
<td>655 μs</td>
<td>6.0 dB</td>
</tr>
<tr>
<td>12 years</td>
<td>649 μs</td>
<td>663 μs</td>
<td>6.8 dB</td>
</tr>
<tr>
<td>Adults</td>
<td>672 μs</td>
<td>723 μs</td>
<td>6.9 dB</td>
</tr>
<tr>
<td>KEMAR</td>
<td>745 μs</td>
<td></td>
<td>5.3 dB</td>
</tr>
<tr>
<td>Sphere</td>
<td>635 μs</td>
<td></td>
<td>3.7 dB</td>
</tr>
</tbody>
</table>

Table 5.2: Binaural cues of children in each age group and adults.
5.3.3 Spectral cues

Listener-specific spectral cues were examined for HRTFs for frontal sound incidence. Values were averaged across the left and right ear.

Figure 5.6: Child HRTF with lowest peaks and notches in the spectrum

**Peaks and notches**

The two lowest frequency peaks and notches in the HRTF spectrum, displayed in Fig. 5.6, are the most relevant spectral cues for spatial perception (Iida, 2019). The frequencies of these four extreme values were extracted for each participant and are displayed in Fig. 5.7 as a function of age for both genders (see Table 5.3 for age-dependent averages).

The first two peaks stem from the first horizontal and vertical modes in the concha. The first peak P1 was between 3.9 to 5.3 kHz for the adult participants (0.4 octaves, 4.4 kHz ± 0.3 kHz) and between 3.5 and 5.5 kHz for children (0.6 octaves, 4.4 kHz ± 0.4 kHz). The second peak P2 was between 7.5 and 10.0 kHz in adults (0.4 octaves, 8.9 kHz ± 0.7 kHz) and 7.1 to 11.5 kHz in children (0.7 octaves, 8.8 kHz ± 1.2 kHz). Previous literature suggests that the peaks are shifted to higher frequencies in young children due to their smaller pinna dimensions. However, large individual differences were observed even in infants (Fels, 2008). In this study, there was no systematic difference between the peak frequencies of children and adults (P1: \( t_{(50)} = 0.28, p = 0.778 \); P2: \( t_{(59)} = -0.11, p = 0.913 \)) and no change with age or head size in children. However, children showed significantly more variability than adults for both peak frequencies (Levene’s test
CHAPTER 5. HRTF DIFFERENCES

for P1: \( F_{(1,62)} = 5.04, p = 0.028 \) and P2: \( F_{(1,62)} = 6.29, p = 0.015 \). This could be due to larger differences in pinna size or a lower precision of the HRTF measurements in younger children.

Figure 5.7: Lowest two spectral peaks P1 and P2 and notches N1 and N2 in HRTFs for sound location at 0° azimuth and elevation for female (orange) and male (blue) child and adult participants.

Notches in the spectrum arise from interference of sound waves in the pinna. The first notch N1 was between 6.5 and 8.9 kHz in adults (0.4 octaves, 7.8 kHz ±0.7 kHz) and 6.1 and 10.2 kHz in children (0.7 octaves, 7.7 kHz ± 0.8 kHz) and the second notch N2 between 8.4 to 13.5 kHz in adults (0.7 octaves, 10.7 kHz ± 1.3 kHz) and 8.2 to 14.7 kHz in children (0.6 octaves, 10.4 kHz ± 1.3 kHz). There was no difference in frequency between children and adults (N1: \( t_{(53)} = -0.54, p = 0.591; \) N2: \( t_{(41)} = -0.37, p = 0.716 \) and the variability was equal (Levene’s test for N1 and N2 both \( p > 0.09 \)), likely due to notches mainly depending on the shape of the pinna rather than the size.

The level of P1 was found to correlate strongly with head measures and thus with age in children. A larger circumference resulted in a higher peak level (\( F_{(1,41)} = 9.41, p = \))
None of the levels of the other extreme values correlated with the circumference ($r < 0.21, p > 0.32$) or other growth measures.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Peak P1</th>
<th>Notch N1</th>
<th>Peak P2</th>
<th>Notch N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 years</td>
<td>4.5 kHz</td>
<td>7.8 kHz</td>
<td>8.9 kHz</td>
<td>10.1 kHz</td>
</tr>
<tr>
<td>8 years</td>
<td>3.9 kHz</td>
<td>7.1 kHz</td>
<td>9.4 kHz</td>
<td>11.0 kHz</td>
</tr>
<tr>
<td>9 years</td>
<td>4.5 kHz</td>
<td>7.5 kHz</td>
<td>8.6 kHz</td>
<td>10.3 kHz</td>
</tr>
<tr>
<td>10 years</td>
<td>4.4 kHz</td>
<td>8.2 kHz</td>
<td>9.3 kHz</td>
<td>10.9 kHz</td>
</tr>
<tr>
<td>11 years</td>
<td>4.5 kHz</td>
<td>7.8 kHz</td>
<td>8.7 kHz</td>
<td>10.3 kHz</td>
</tr>
<tr>
<td>12 years</td>
<td>4.3 kHz</td>
<td>7.6 kHz</td>
<td>8.7 kHz</td>
<td>11.0 kHz</td>
</tr>
<tr>
<td>Adults (female)</td>
<td>4.5 kHz</td>
<td>7.9 kHz</td>
<td>9.0 kHz</td>
<td>10.8 kHz</td>
</tr>
<tr>
<td>Adults (male)</td>
<td>4.1 kHz</td>
<td>7.6 kHz</td>
<td>8.6 kHz</td>
<td>10.2 kHz</td>
</tr>
<tr>
<td>Adults (KEMAR)</td>
<td>4.0 kHz</td>
<td>7.9 kHz</td>
<td>8.4 kHz</td>
<td>9.5 kHz</td>
</tr>
</tbody>
</table>

Table 5.3: Monaural cues: lowest spectral peaks and notches in children (averaged across gender) and female adults, male adults and the median adult used for the KEMAR head.

**Spectral envelope**

Correlation coefficients between HRTF spectra for frequencies between 50 Hz and 20 kHz were calculated as an indicator of similarity for HRTFs at the same ear and source position. Repeated measurements for the same participant had an average correlation of 0.96 in adults and 0.89 in children (similar to the results for pilot participants in Chapter 2). The lower correlation in children is likely due to an increased chance of movement during the measurements or change of head position between measurements. Symmetric HRTFs (left and right ear HRTFs for ipsilateral, frontal and contralateral sound sources) had an average correlation of 0.72 in adults and 0.62 in children.

Between-subject correlations were on average 0.49 in adults and 0.56 in children. The average correlation between a child and an adult participant was similar ($r = 0.51$). No effect of growth could be detected for children, i.e. the correlation was not higher for children with similar head circumferences. This is in line with the results from the spectral peaks and notches, suggesting that differences in HRTF spectra were mostly
due to individual variability rather than due to growth during childhood. Spectral cues mainly depend on the shape of the pinna which is not changing much with age.

Fig. 5.8 displays HRTFs in 0° azimuth direction for adults and children as well as the symmetric HRTFs from the KEMAR head and spherical head model. The median spectra for children and adults are representative of the overall shape of the spectra but remove distinct peaks and notches present in individual spectra. The two curves are very similar for high frequencies but at lower frequencies, the envelope of children is shifted towards higher frequencies due to their smaller dimensions.

![HRTF in frontal direction (0° azimuth and elevation): median and individual results for 21 adults and 43 children and artificial HRTFs from the KEMAR head and the spherical head model.](image)

Figure 5.8: HRTF in frontal direction (0° azimuth and elevation): median and individual results for 21 adults and 43 children and artificial HRTFs from the KEMAR head and the spherical head model.
5.3.4 Effect of HRTF mismatch on speech perception

In the study described in Chapter 4, SRTs were measured for children in different HRTF conditions. One central question of this study was whether children whose binaural and monaural cues differed more from those of an artificial head like the KEMAR would experience more difficulties in speech perception in the environment generated from it. Therefore we investigated the relationship of head dimensions with age-normed SRTs and SRM. Since SRT performance in children diagnosed with APD is affected by their deficit in listening, this analysis only included the 25 typically-developing children.

Overall, the head parameters of circumference and volume of TD children were found to correlate only very weakly with age-normed SRTs and SRM. There was a trend for SRT z-scores to be higher for children with larger circumferences (see Fig. 5.9). This relationship, however, was nonsignificant in a mixed-effects model for all conditions that accounts for clustering from repeated measures with a random effect of listener ($\chi^2(1) = 2.74, p = 0.097$). Statistical significance was only found in the colocated loudspeaker condition R0° ($F_{(1,22)} = 5.78, p = 0.025$). This trend is likely to be by chance due to the small sample size. But it could potentially also hint at differences in the development of same-aged children with some children with smaller heads lagging behind.

None of the binaural cues extracted from the HRTFs had a strong correlation with age-normed SRTs (ITD: $r = 0.07$, SII-weighted ILD: $r = 0.11$, Global ILD: $r = 0.10$) as shown in Fig. 5.9. The same was true for the frequencies of spatial peaks and notches (all $r < 0.25$).

The relationship between the variables was similar in the KEMAR conditions and the individualised and real conditions. These results support the findings from the previous chapter, that speech perception in virtual auditory environments is not affected by small differences in binaural and monaural cues between the listener’s HRTFs and nonindividualised HRTFs, at least not for the large separation angles used in this procedure.
5.4 Discussion

HRTFs at 0°, 90° and -90° azimuth and 0° elevation were measured for children aged 7 to 12 years and adults. Listener-specific differences were investigated for binaural cues, the maximum ITD and ILD, as well as monaural cues in the spectrum, especially characteristic peaks and notches at lower frequencies.

As expected, all head measures for children were smaller than measurements for adults reported in the literature (Fels, 2008; Burkhard & Sachs, 1975; ITU-T P.58, 2013). Circumference and height showed a significant increase with age within the 6-year age
range of the children in the study. Both binaural cues were significantly smaller for children than for adults. There was only a small increase with age within the children suggesting that even for the oldest children binaural cues are not yet matured. On average, ITDs were 31 μs smaller and ILDs 0.8 dB smaller in children. Gender differences were weaker in children than in adults. The frequencies of spectral peaks and notches did not differ between children and adults but children had significantly more variability in the peak frequencies. This variability might come from larger differences in body dimensions between the children than between grown adults. Another reason for larger differences and also for a slightly lower correlation of repeated HRTFs in children is that the HRTF measurements might have been less precise than in adults. Even though we developed the measurement procedure to be as fast and easy as possible, children were more likely to move during the measurement or change their position before the measurement started.

Few other studies have published measured or simulated results for HRTFs in children. Fels (2008) simulated HRTFs from geometric data for children in a large age range. Results were reported for a typical 6-month-old infant, 6-year-old child and 16-year-old teenager. The maximum ITD in the 6-year-old child was approximately 625 μs, comparable to the youngest children in the current study aged 7 years with 636 μs. ILDs were only reported in visual form and were fluctuating greatly across source positions. As expected, ILDs were generally larger for the teenager than the younger children. Braren & Fels (2019) measured HRTFs for children aged 5 to 9 years. They had an average maximum ITD of 615 μs, slightly lower than the most similar group of 6 to 9-year-olds in this study who had 644 μs.

These findings suggest that children differ from adults in some spatial cues but also exhibit larger variability across individuals. A larger sample size that also includes younger children would be needed to better investigate growth effects.
6.1 Comparison of speech perception results of children and adults

The SRM was measured in children and adults in virtual auditory environments created from individualised HRTFs and generic HRTFs and in the equivalent real environment. For all participants, SRTs were measured twice and averaged to obtain a more stable result. In the study with children, the generic HRTFs came from an artificial KEMAR head. For the adults, two sets of nonindividualised HRTFs were used, from an adult with large head dimensions and an adult with small head dimensions. Since there was no difference between the SRTs for these nonindividualised AEs, mean SRT values of the two colocated and separated conditions were used for this comparison. Fig. 6.1 shows the SRTs measured in colocated and separated masker conditions from adults and children with and without APD. For comparison with the adults, only the results of the TD children were used, as they are representative of normal childhood development.

For children, SRTs decreased strongly with age. The youngest TD children, aged 7 years, had mean SRTs of -2.1 dB SNR in the colocated conditions and -9.0 dB SNR in the separated conditions. The oldest TD children, aged 12 years, had -5.6 dB SNR and -14.2 dB SNR, respectively. Thresholds improved on average by 0.6 dB per year in the colocated conditions and by 0.9 dB per year in the separated conditions. Adults had an average SRT of -7.6 dB SNR in the colocated conditions and -16.4 dB SNR in the separated conditions. As can be seen in Fig. 6.1, there was a clear difference between
children and adults in all environments. Even the oldest children still had higher SRTs than adults. This is consistent with previous findings in the literature suggesting that SRTs for symmetric speech maskers decrease until adolescence, with children reaching adult-like levels only at around 14 years of age (Brown et al., 2010; Corbin et al., 2016).

Figure 6.1: SRTs in generic and individualised virtual AEs and the equivalent real AE: results for adults (boxplots), TD children (points and solid regression lines) and APD children (triangles and dashed lines). Regression lines were fitted separately for each AE, age and group of children, therefore, not all differences seen are necessarily statistically significant.

To better understand the extent of these differences, a mixed-effects model for SRT was fitted with the fixed factors of group (TD children, adults), age, auditory environment (real, individualised HRTFs, generic HRTFs) and masker locations (0°, ±90°), their two-way interactions and the random effect of participant. As expected, there was a main effect of group ($\chi^2(1) = 8.77$, $p = 0.003$) and age ($\chi^2(1) = 5.65$, $p = 0.018$) as well as an interaction between the two ($\chi^2(1) = 7.63$, $p = 0.006$), due to the change with age in the child group but not in the adults. There was also a strong main effect of masker location ($\chi^2(1) = 136.7$, $p < 0.001$), thus SRM, and interaction of age with masker locations ($\chi^2(1) = 20.67$, $p < 0.001$) due to the increase in SRM with age in children, albeit small. Also, a small main effect of AE ($\chi^2(2) = 7.74$, $p = 0.021$) was revealed, mainly due to
differences between the real and the generic environment, with SRTs being 0.7 dB lower in the generic environment. The random effect of participant accounted for 53% of the total variance in the model.

The SRM in children also increased with age but only by an average of 0.3 dB per year, much less than the SRTs. Even the youngest TD children at age 7 had a large SRM of 6.9 dB. The oldest TD children aged 12 years had an SRM of 8.6 dB, almost the same as adult performance, which was on average 8.9 dB.

A mixed-effects model for SRM was fitted with the fixed effects of group (TD children, adults), age and auditory environment (real, individualised HRTFs, generic HRTFs) and the random effect of participant. As there was neither a main effect of group nor an interaction with age (both $p > 0.20$), group was removed from the model. For the reduced model, a main effect of age was found ($\chi^2(1) = 12.03, p < 0.001$) but none of environment ($\chi^2(2) = 1.45, p = 0.482$). Only a small effect of participant was found, accounting for 16% of the variance. This suggests that individual performance in the three environments and, therefore, the size of the SRMs, differed quite strongly (for an example of this variability see individual results for SRTs in children in Fig. 4.4 in section

---

Figure 6.2: SRMs in generic and individualised virtual AEs and the equivalent real AE: results for adults (boxplots), TD children (points and solid regression line) and APD children (triangles and dashed line).

---

136
These results showed that although SRTs in children decreased strongly with age, children had poorer abilities than adults to perceive speech in the presence of intelligible maskers even at age 12. In terms of SRM, a large effect was found even in the youngest participants but the SRM also increased with age and was not yet mature in the oldest children. These findings are consistent with other studies that have found improvements in SRM into early adolescence (Cameron et al., 2011; Corbin et al., 2016; Misurelli & Litovsky, 2012).

The age of maturation likely depends on the type of masker signal. Corbin et al. (2016) found improvements in SRTs in colocated conditions until the age of 10 years for speech-shaped noise and 13 years for a two-talker masker. Speech maskers lead to a combination of energetic masking due to the spectro-temporal overlap of the stimuli and informational masking due to similarities between the stimuli and their uncertainty. Masking was found to be highest for two-talker maskers (Buss et al., 2017). Another reason for this late maturation is the number and placement of the masker signals. Maskers placed asymmetrically on one side of the head allow the participant to listen with the averted ear and take advantage of the consistently better SNR at that ear. This better-ear listening is greatly reduced for symmetrical maskers for which participants can only take advantage of brief glimpses of better SNR in both ears (Brungart & Iyer, 2012).

6.2 Comparison with other studies

Most other studies that tested sentence perception in the presence of colocated and symmetrically separated speech maskers in children were conducted using the clinical LiSN-S procedure by Cameron & Dillon (2007) which is described in more detail in section 1.2. The conditions in the current procedure were designed to be similar to the same voice conditions of the LiSN-S procedure. Both procedures have a frontal target speaker and two maskers that are either colocated with the target or symmetrically separated at ±90° azimuth so that the listener has to rely on better-ear glimpses to understand the target. The same talker is used as target and maskers resulting in very high informational
CHAPTER 6. GENERAL DISCUSSION

masking, especially in the colocated condition. However, the procedures differ in the gender of the speaker with female speakers in the LiSN-S and a male speaker in the current study. Another difference is the scoring method. In the LiSN-S procedure, every word is scored whilst in the current procedure, we used a keyword scoring method, which was expected to be more representative of speech intelligibility.

Results from the two procedures were compared for two measures: the colocated condition, corresponding to the low-cue or SV0 condition in the LiSN-S, and the SRM which is equivalent to the spatial advantage measure. The separated condition in the LiSN-S is rarely reported but was calculated by subtracting the SRM from the colocated SRT. The LiSN-S procedure uses a virtual AE created from KEMAR HRTFs. Besides comparing it to the mean results across environments, the LiSN-S will therefore also be compared directly to the KEMAR condition in the child study.

![Graph showing SRTs in the colocated (SV0) and separated same voice conditions (SV90) for TD children: results from the current study (mean SRTs for all three environments and SRTs for the KEMAR environment) and age norms for the NA LiSN-S (Brown et al., 2010; Cameron et al., 2009) and the AU LISN-S (Cameron & Dillon, 2007; Cameron et al., 2011).](image)

Figure 6.3: SRTs in the colocated (SV0) and separated same voice conditions (SV90) for TD children: results from the current study (mean SRTs for all three environments and SRTs for the KEMAR environment) and age norms for the NA LiSN-S (Brown et al., 2010; Cameron et al., 2009) and the AU LISN-S (Cameron & Dillon, 2007; Cameron et al., 2011).

The LiSN-S has been developed in an Australian (AU LiSN-S) and a North American version (NA LiSN-S). The Australian version (Cameron & Dillon, 2007) was normed for 70 typically-developing children, gender-balanced and equally distributed across the age range of 6 to 11 years. They had mean SRTs of -0.8 dB in the colocated and -12.6 dB in the separated condition, thus obtaining 11.8 dB SRM. The NA LiSN-S was normed for

138
a similar participant group of 72 gender-balanced children aged 6 to 11 years (Cameron et al., 2009). Their mean SRT in the colocated condition was -0.3 dB and their mean SRT for the separated condition -9.6 dB, leading to an average SRM of 9.3 dB. The NA LiSN-S norms were found to be representative of the performance of British English children for all measures of the same voice conditions (Murphy et al., 2019).

**Typically-developing children**

Fig. 6.3 shows the averaged SRTs in each age group for TD children in the current study averaged across all environments and specifically in the KEMAR environment compared to the age norms for the two LiSN-S versions. In the KEMAR environment, the mean SRT for the age group 7 to 11 years was -4.1 dB for the colocated and -11.1 dB for the separated condition. Average values across all environments were -3.7 dB and -10.1 dB, respectively. This resulted in an SRM of 7.1 dB which was smaller than in the two versions of the clinical LiSN-S with 11.8 dB and 9.3 dB (see Fig. 6.4). This difference is likely to depend on the fact that SRTs in the colocated conditions in the current study were about 3 dB lower, thus better, than in the LiSN-S norms whilst separated SRTs were similar. This effect might be due to differences in the speech stimuli of the LiSN-S and the current study, as discussed further in section 6.3.

![Figure 6.4: Comparison of SRM (called spatial advantage in the LiSN-S) for TD children: results for the current study and LiSN-S normes as in Fig. 6.3.](image)
CHAPTER 6. GENERAL DISCUSSION

Children with diagnosed APD

The LiSN-S test is used in the diagnosis of APD but few studies have published results for children with suspected or diagnosed APD or for potentially related groups of children with a history of otitis media. Typically, age-normed scores are reported for the SRT in the colocated condition and the SRM. Table 6.1 lists their results in comparison to the age-normed results of the current study. For reference, the table also includes two studies with groups diagnosed with spatial processing disorder (SPD), thus children who were specifically selected because they were more than two standard deviations below the norm in the LiSN-S test at their APD assessment. In all studies, children with listening difficulties had negative z-scores, thus performed worse than TD children of the same age.

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>Age (years)</th>
<th>LiSN-S version</th>
<th>SRT 0° (z-score)</th>
<th>SRM (z-score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cameron &amp; Dillon, 2008</td>
<td>9 susAPD</td>
<td>6 - 11</td>
<td>AU</td>
<td>-0.2</td>
<td>-2</td>
</tr>
<tr>
<td>Cameron et al., 2012</td>
<td>10 SPD</td>
<td>6 - 9</td>
<td>AU</td>
<td>-0.56 ± 0.96</td>
<td>-2.6 ± 0.70</td>
</tr>
<tr>
<td>Tomlin &amp; Rance, 2014</td>
<td>35 OME</td>
<td>6 - 12</td>
<td>AU</td>
<td>-0.13 ± 0.14</td>
<td>-1.25 ± 0.18</td>
</tr>
<tr>
<td>Sharma et al., 2014</td>
<td>21 susAPD</td>
<td>10 - 15</td>
<td>AU</td>
<td>-0.12 ± 0.77</td>
<td>-0.67 ± 2.44</td>
</tr>
<tr>
<td>Graydon et al., 2017</td>
<td>82 OME</td>
<td>6 - 13</td>
<td>AU</td>
<td>-0.39 ± 0.81</td>
<td>-0.84 ± 1.08</td>
</tr>
<tr>
<td>Graydon et al., 2018</td>
<td>16 SPD</td>
<td>6 - 10</td>
<td>AU</td>
<td>-1.03 ± 1.17</td>
<td>-2.16 ± 0.53</td>
</tr>
<tr>
<td>Stavrinos et al., 2018</td>
<td>27 susAPD</td>
<td>7 - 11</td>
<td>AU</td>
<td>-0.86</td>
<td>-1.24</td>
</tr>
<tr>
<td>Moore et al., 2020</td>
<td>35 susAPD</td>
<td>6 - 13</td>
<td>NA</td>
<td>-0.8</td>
<td>-0.5</td>
</tr>
<tr>
<td>Current study</td>
<td>16 APD</td>
<td>7 - 12</td>
<td>custom</td>
<td>-0.22 ± 0.90</td>
<td>-0.11 ± 1.08</td>
</tr>
</tbody>
</table>

Table 6.1: Studies measuring SRM in children with listening difficulties including results from the current study: mean and standard deviation (z-scores) for colocated SRT and SRM for participant groups with diagnosed APD with and without spatial processing disorder (SPD), suspected APD (susAPD), or a history of OME.

Many studies using the Australian version found the deficit to be greater for the SRM than for the SRT in the colocated condition, suggesting that it mainly affects performance in the separated condition. However, Moore et al. (2020) measured a larger deficit in the SRT of the colocated condition than for the SRM for American children as found in
the current study, suggesting that performance is lower in both SRT conditions. A large
deficit in the colocated condition was also found in a study by Stavrinos et al. (2018) that used the AU version for British children. In the current study, children with APD performed worse than the control group in all SRT conditions. This suggests that this group of children has an overall deficit in processing speech-in-noise rather than a deficit specific to conditions with spatially separated sound sources. Due to this general SRT deficit, the reduction in SRM compared to TD children was smaller than in studies using the clinical LiSN-S procedure.

Age norms in the current study were calculated from the 23 TD participants. Due to this small group size compared to the norming studies for the LiSN-S with 70 or more participants, age norms were calculated from regression over age instead of independently for each age group. This is likely to result in less precise norms and thereby to smaller differences between age-normed scores of TD and APD children.

**Adults**

For adults, the LiSN-S procedure led to a much larger SRM of 13.9 dB compared to 8.9 dB in the current study (norms reported in Brown et al., 2010). Whilst SRTs in the separated condition were similar (-15.5 dB SNR and -16.4 dB SNR), there were large differences in the colocated condition with -1.6 dB SNR in the LiSN-S and -7.6 dB SNR in the current procedure. For the LiSN-S, SRTs changed very little with age in the colocated condition. The difference between the youngest children aged 6 years old and adults was less than 2 dB, whilst it was more than 5 dB in the current procedure.

### 6.3 Discussion of results

#### 6.3.1 Effect of auditory environment on speech perception

SRTs and SRM in the virtual AEs calculated from HRTFs were very similar to those measured in the real environment for adults and children. This result suggests that even generic HRTFs from artificial heads are sufficiently similar to individual HRTFs to be used for speech perception tasks with spatialised stimuli. This means that the LiSN-S procedure in its current form is a suitable tool to accurately measure the SRM even in young children. In the current procedure, large separation angles of $\pm 90^\circ$ were used for
CHAPTER 6. GENERAL DISCUSSION

the SRM. There is a possibility that smaller separation angles, which lead to smaller and potentially less robust effects of SRM, could be influenced by a mismatch in HRTFs.

The study in Chapter 4 also showed that approximating HRTFs with a simple spherical head model was not sufficient to accurately measure SRTs in children. Unfortunately, only the condition with separated sound sources could be included for this AE. Both children with and without APD had increased SRTs in this condition, on average by 2.6 dB. Whilst the spherical head model is able to provide binaural cues similar to a human head, it does not contain any spectral filtering which is part of real HRTFs due to pinna filtering. Studies have examined the contributions of ITDs, ILDs and spectral cues on the SRM in adults (e.g. Kidd et al., 2010). For the LiSN-S procedure, stimuli only containing ILD cues lead to an SRM of around 13 dB, similar to the norms for full KE-MAR HRTFs with 13.9 dB (Glyde et al., 2013b). Only providing ITDs resulted in an SRM of 8 dB, suggesting that the rapid changes in level at the left and right ear due to head diffraction are the main contributors to the SRM. However, Ellinger et al. (2017) found the opposite effect in a study with Coordinate Response Measure sentences. SRTs for separated speech maskers were increased by 2 dB when only ITDs were presented and by 3 dB for only ILDs. Other studies found that both cues contributed roughly equal to the SRM for speech and noise maskers (Edmonds & Culling, 2005; Ewert et al., 2017). This decreased performance for separated stimuli and the related reduction in SRM might depend on the lack of spectral filtering leading to a flat frequency response at mid frequencies which are important for speech perception (see Fig. 5.8).

Another factor contributing to the higher SRTs could be that ITDs and ILDs calculated from the spherical head model were smaller than for an average adult (see Table 5.2) even though the spherical head was modelled with the same circumference as the KEMAR head. This might have been due to the difference in shape between a sphere and a real head. ITDs were similar to those of younger children but ILDs were smaller which could have had a negative effect on speech perception.
6.3.2 Differences to LiSN-S procedure

In the current procedure, SRTs in the colocated conditions were much lower than in the LiSN-S test for all three participant groups. Some studies have found deviations from the original LiSN-S norms for normal-hearing participants, e.g., Glyde et al. (2013b) measured SRTs in adults which were 3.5 dB lower than the norms in the colocated condition and 2.5 dB in the separated condition in an environment only containing binaural cues. In our procedure, however, the difference was over 5 dB in the colocated condition but there was no difference in the separated condition, leading to a smaller SRM in adults of only 9 dB.

Some of this may be attributed to the variability across target talkers and their potential differences in voice intelligibility (Smiljanić & Bradlow, 2009). The current study used a talker who is experienced with speech recordings and has a very clear and intelligible voice. It is possible that this talker might also be more intelligible in the presence of his own voice(s), thus leading to less informational masking. The gender of the talkers was also different, with a female talker in the LiSN-S and a male talker in the current procedure. It is also important to recollect that the recordings for target and masker were done several years apart. This led to some differences in voice characteristics as described in detail in section 2.1, which might have been large enough to reduce the similarity between the voices and thus informational masking. This effect would be most apparent in the colocated condition. It is also likely that speech rate and intonation differ between single sentences and longer texts, and therefore between target and masker, which might have reduced informational masking further.

Slight differences in SRTs could also be due to differences in the scoring methods. In the current procedure, only three keywords were scored whilst in the LiSN-S every word in the sentence is used for scoring (typically 4 - 6 words). On the one hand, scoring only keywords could result in lower SRTs since fewer words need to be understood to mark a sentence as correct. On the other hand, the additional words in the sentence are likely to be words with very high frequency, therefore are easier to fill in from context and might also result in lower SRTs. The two procedures also use different sentence material. Whilst they were both developed on very similar principles based on the BKB sentence lists, there could be subtle differences in intelligibility.
6.3.3 Findings for children with diagnosed APD

The study also found a significant deficit of children with APD for SRTs in both spatial configurations, whilst most studies with the original LiSN-S procedure only found a deficit in the separated condition and therefore also a reduced SRM (e.g. Cameron & Dillon, 2008; Sharma et al., 2014). This effect might partly be connected to the lower SRTs for colocated stimuli in all participant groups. In the LiSN-S, SRTs in the colocated same-voice conditions are -1.6 dB SNR for adults and even higher for children, thus, almost at equal levels for target and masker. For positive SNR levels, it is likely that masking is greatly reduced in all participants, leading to a high speech intelligibility also for children with APD. This ceiling effect might lead to rapid changes in intelligibility with level and therefore steep psychometric functions. Potential differences in speech perception between participants could be reduced. This could also be a factor that is responsible for the smaller variability in the colocated condition. Overall, the SRT results in the current study suggest that the children diagnosed with APD have a broad deficit in speech-in-speech perception rather than a deficit specific to the spatial processing of sounds.

Children with APD were found to also perform worse than TD children in hearing and language tasks. They had slightly higher hearing thresholds than TD children, on average by 3.8 dB for frequencies up to 8 kHz and 5.8 dB for extended high frequencies. Apart from a few exceptions mainly at high frequencies, thresholds for APD children were still within the limits of normal hearing. There was also a higher likelihood for past episodes of OME in children with APD supporting previous findings that this might be a risk factor for APD. 41% of APD children reported a history of OME in the questionnaires whilst only 13% did so in the TD group. In the CELF-5 recalling sentences subtest for receptive language, children with APD again performed lower than the control group but still within the norm. This finding supports the suggestion that these children might have more general difficulties that affect different abilities instead of purely auditory deficits or even deficits specific to spatial or temporal processing. These deficits might be low for each ability but in summary, lead to difficulties in perceiving acoustic information in challenging environments.
6.3.4 Limitations

As mentioned before, it is difficult to measure HRTFs in children as they are more likely to move during the measurement. This was also the case in this project, resulting in HRTFs from some children being less precise than for adults. We were aware of these difficulties and tried to address them through multiple adjustments to the measurement procedure. The duration of each measurement was very short (20 s) since only three sound locations were measured. Measurements were repeated at least three times for each child and more often if the child showed difficulties in keeping a fixed body posture. Children were closely monitored during the measurement to detect movements. Unfortunately, the measurement had to be done at the beginning of the test session when children may have still been a bit insecure regarding the room and the testing. We ensured all children had a chance to get sufficiently accustomed to the anechoic room before the HRTF measurement started and were feeling comfortable throughout the measurements. They also had a chance to listen to the stimuli before the measurement. We were unable to use a motion tracking system and decided against constraining the children with a head fixture during the measurement since we did not want to frighten them or to put them under additional stress by constraining their movements.

In the two main studies, we used two intelligible speech maskers with the same voice as the target speaker just as in the LiSN-S procedure. These stimuli lead to a large amount of masking but results can also be influenced by other factors, such as attention, especially in the colocated condition. Due to time constraints, it was not possible to include more test conditions. Ideally, more conditions would have been tested with additional masker types which are less intelligible but still lead to high informational masking and conditions measuring the separate contributions of ILDs and ITDs on the SRM, as has been studied in adults. Difficulties in recruiting children with diagnosed APD limited the size of the participant groups and additional testing. Especially for the APD group, a larger sample size would have been helpful for the analysis since these children are very heterogeneous and can differ greatly in their abilities.
6.4 Summary of the results

In two studies with adults and children, we investigated the effect of the auditory environment on the perception of a talker in the presence of speech maskers at the same or different locations in space with a focus on children with APD.

Overall, speech intelligibility in the different AEs was similar. For adults, SRTs in nonindividualised environments based on HRTFs from other adults were the same as in the individualised virtual environment and the equivalent real environment. For children, there were small differences between the results, but no overall degradation of speech perception could be detected for AEs from generic HRTFs. Only the use of a simplified spherical head model instead of a human head led to a significant drop in performance. SRTs in children decreased with age across the age range of 7 - 12 years. Even in the oldest children, SRT performance was not yet matured. The SRM also increased with age, although, to a lesser degree. The oldest children had an SRM comparable to adults. Children with APD had a significant deficit in all conditions of the speech test compared to same-aged TD children. The derived measure of SRM, however, was similar to TD children of the same age. This finding suggests that children with APD have an overall deficit in speech-in-noise perception for this type of masker rather than a difficulty with spatial processing.

The children with APD also showed slight deficits in other tests on hearing and language abilities. Their hearing thresholds were slightly higher than those of the control group, on average by 5 dB, but still within limits for normal hearing in most participants. They also performed worse than the control group in a test on receptive language skills but were still within normal limits.

To gain more knowledge about the extent of the mismatch between spatial cues of adults and young children, the measured HRTFs were compared based on the size of their binaural cues and the characteristics of their monaural spectral cues. ITDs and ILDs were significantly lower in children than in adults, as expected from their differences in head size. The frequencies of spectral peaks and notches, which depend on the shape and size of the pinnae, varied greatly in children and did not differ statistically from those of adults.
6.5 Conclusion

The project showed that artificial head HRTFs, or more generally adult HRTFs, are suitable for experiments on speech perception with intelligible maskers in children. This was found for target and masker stimuli that were either colocated or had large spatial separations. There is still a possibility that for small separation angles the accuracy of spatial cues might be more important thus speech-in-noise perception might potentially be worse for mismatched HRTFs. But for the large separation angles typically used for SRM measurements, individualised and nonindividualised HRTFs resulted in the same SRTs even in young children and children with APD. The auditory environment used in the clinical LiSN-S procedure is, therefore, suitable to correctly measure the SRM in children.

This research has also supported previous findings that HRTFs from children are more varied and differ from adult HRTFs. There is a need for more research on HRTFs of children and on suitable techniques to obtain them for all ages. So far, most research into virtual auditory environments has focused on adults, since potential research and consumer applications are typically targeted at adults. HRTFs however have many more promising applications beyond that, such as new opportunities for clinical tests and simplifying research studies by conducting them in a virtual rather than a real environment. This could also simplify the recruitment of children, as research studies could be conducted in quiet rooms in schools or children could be tested online under remote supervision. Research teams are currently working on better techniques to extract precise geometric information from lower quality photogrammetric data (such as videos or pictures from a smartphone) and to make the computation of HRTFs more efficient and faster. These less demanding techniques would make individualised HRTFs more accessible and easier to obtain for a large number of participants. It could also be used for very young children or infants who cannot get their HRTFs measured acoustically.

Results from children with diagnosed APD showed that, as a group, they performed worse in all tests of the study. They had significantly higher SRTs in all environments, both in the colocated and the spatially separated conditions. They also had scores in the receptive language test that were worse than for TD children but still within the limits for normal performance. The same was true for hearing thresholds across the full
audible range, likely connected to the higher prevalence of OME during childhood. These deficits in auditory processing, hearing and receptive language might be too small to be picked up separately in a normal assessment. Their effect is likely to be small when other cues are available but can have a considerable impact on auditory perception in demanding situations with less redundancy. The cause and effect of these difficulties are still unclear, but the results support the suggestion that these children might have more general difficulties that affect a variety of abilities instead of a mainly auditory deficit or even deficit specific to spatial or temporal processing.

This project showed that SRTs and SRM can be measured accurately in a nonindividualised virtual environment in both children and adults. This suggests that further studies on speech perception with spatialised sound sources in children can be conducted in a generic virtual environment instead of needing to place the child in the real acoustic environment. The results also validate the LiSN-S as a suitable tool to test listening difficulties in children with APD. These children showed an overall deficit in speech perception in the presence of multiple speech maskers. However, further research with modified stimuli is necessary to understand whether the cause for this is in the auditory domain or more dependent on language, attention and/or working memory difficulties. Speech perception tasks in different realistic listening situations created in virtual environments could be used as part of a larger intervention tool.

There are still many unanswered questions on APD, about its causes and the interplay of auditory, cognitive and language difficulties. More targeted diagnosis tools and individualised intervention strategies are needed to help these children and reduce the impact of their listening difficulties on their life.
Bibliography


Appendix A

Additional figures for child results

Figure A.1: Pure-tone average (PTA) and extended high frequency (EHF) hearing thresholds for participants in both groups with different severities of past middle ear infections: no infections, few infections, many infections, diagnosed with OME (glue ear) and diagnosed with OME and fitted with grommets as treatment.
Figure A.2: Background questionnaire results for the two participant groups: language background (monolingual, passively exposed to second language, bilingual), musical practice, diagnosis of an additional developmental disorder apart from APD, special educational need.

Figure A.3: Background questionnaire results for APD participants: clinic that diagnosed participant with APD (GOSH, UCLH, other clinic, screening without full diagnosis), age at diagnosis and intervention with FM system (no intervention with training software reported).
Figure A.4: HRTF measurements for the left ear for -90°, 0° and 90° azimuth at 0° elevation: individual results for 43 children and median child and adult functions.
Appendix B

Documents

B.1 Full text of distractor stories

Princess Pea  There was once a prince who wished to marry a princess; but then she must be a real princess. He travelled all over the world in hopes of finding such a wife; but there was always something wrong. He found plenty of princesses, but whether they were real princesses or not it was impossible for him to decide. Now one thing, now another, seemed to him not quite right about the ladies. At last he returned to his palace quite cast down, because he so much wanted to have a real princess as his wife. One evening a terrible storm came up. There was thunder and lightning and the rain poured down from the sky. It was as dark as pitch. Suddenly they heard a knocking at the palace gate, and the old king went out himself to open it. It was a princess who was standing outside the door. But what a sight the rain and wind have made her look. The water trickled down from her hair, and her clothes clung to her body. And yet she said she was a real princess. “Well, we will soon find out!” thought the old queen-mother. But she said no word of what she was going to do. She went quietly into the bedroom, took all the bed-clothes off the bed, and put three little peas on the bottom. She then laid twenty mattresses on top of the three peas, and put twenty feather beds over the mattresses. In this bed the princess lied all night. The next morning she was asked how she had slept. “Oh, very badly!” she replied. “I have hardly closed my eyes the whole night. I do not know what was in my bed, but I had something hard under me, and am all over black and blue. It has hurt me so much!” Now it was clear that the lady must be a real princess, since she had been able to feel the three little peas through the twenty mattresses and twenty feather beds. None but a real princess could have had such a sensitive sense of feeling. The prince made her his wife as he was now convinced that he had found a real princess. The three peas were put into the cabinet of curiosities, where they can still be seen today.
Lara and Bernie Lara spent a lot of time alone. It wasn’t that Lara was an unfriendly girl, just that her face was naturally serious. Unless she made an effort it sometimes looked as if she wanted to be left alone, even though she really wouldn’t have minded talking to someone. However, just because she spent time alone didn’t mean she was always lonely. She was reading a lot and in her mind she went on great adventures. Whenever she got sad, which happened a little more often than it should, she talked to Bernie, her patchwork blue bear. Some would say she should talk to her family or friends but she always found what helped the most was talking to her little blue Bernie, who always listened with the same interested look on his face. They spent a lot of time together, Lara holding his little round paws. Lara had had Bernie ever since her previous birthday, when her aunt had taken her present shopping. Lara had noticed Bernie right away, sitting amongst the other blue bears on the shelf. All the bears were made of blue patchwork with big buttons connecting their arms and legs. They all had little round ears sticking out the top of their head and black beady eyes above their cute black noses. But while in many ways Bernie looked exactly like all the other bears in the shop he also looked completely different. His ears were rounder, his eyes were glossier, his patchwork brighter, the gold in it holding an extra shine. Although Lara did not know it at that time this was because Bernie was a special bear. All Lara knew was that Bernie was the one for her. What she didn’t know was that she hadn’t really chosen Bernie. Bernie had chosen her. One particularly rainy and lonely Thursday evening Lara was in her bedroom reading a book about spiders. She didn’t particularly like spiders and in all honesty was a little afraid of them but she felt safe enough reading about them in a book and she always felt it was good to get to know and understand the things that scared you. The more she knew, the less there was to be scared of. However, as Lara read she had the feeling that somebody was watching her. She got that strange little tickle on the back of the neck. She turned and looked at her room but all she could see was her own bed, her shelf of toys, the poster on her wall and Bernie, sitting in his favorite place on the yellow chair in the corner. The door was closed and her curtains drawn. She was alone. Lara went back to her book but soon enough she felt it again, that feeling that someone was watching her, like a light feather on her neck. She turned quickly this time, hoping to catch whoever it was but there was still no one there. Just her room as it always was. She went back to her book again but she couldn’t really concentrate anymore. She was waiting for the feeling again. And when it came, as she had known it would, she did not turn around. Instead she said, “Hello.”

Spider Halloween Tilly the spider had spun her web in the corner of the kitchen by the broom closet, where nobody could see her. Tilly was afraid of people. People scream when they see spiders. The loud noise makes the tiny hairs on her body vibrate. When her hairs vibrate, Tilly gets scared. Tilly liked to watch the family from her corner in the kitchen. Amy and her brother Matt had invitations to a Halloween party. Matt cut up cardboard boxes to make a robot costume, and glued aluminum foil on top. Amy glued
leaves onto one of her dad’s old t-shirts so she could look like a tree. Tilly wanted to go to the party, too. She wanted to be a ghost. She spun herself a snug little white ghost costume leaving room for all eight of her spindly spider legs to stick out the bottom. She thought – I can spin a thread over the doorway and drop down onto Amy’s costume when she goes out the door. I’ll get to go to the party, and nobody will scream at me. Tilly waddled along the ceiling. It was hard to walk gracefully in her ghost costume. She perched over the doorway, and spun her thread ready to drop onto Amy when she walked underneath. Matt came to the doorway. The aluminum foil on his cardboard robot costume crackled and he walked stiff-legged. The door handle clicked and a breeze pushed the door open. Matt stomped out onto the sidewalk. The breeze rustled Amy’s leaves. She waved her arms like branches in the wind. Tilly attached her sticky white thread to the lintel, ready to push-off with all eight of her spindly spider legs. Her thread was stronger than steel and more flexible than silk. It could hold her weight even though she was extra heavy with her ghost costume. Amy put one foot over the threshold. Tilly let go. She glided down, down on her lovely silken thread. A gust of wind blew a leaf off Amy’s costume. Amy dashed back into the living room to catch the leaf. Tilly stopped gliding down, and started climbing back up her silken thread to wait for Amy. Just then, Amy’s mother, who always screamed at spiders, stepped through the doorway. Amy’s mother’s ear caught on Tilly’s web. Tilly found herself dangling from Amy’s mother’s left ear, like an earring. Tilly was so startled, she hung onto her thread for dear life, all the way to the party. At the party, Matt and Amy showed off their costumes. Tilly thought “look at me! Look at me I’m wearing a costume, too!” She wiggled in her ghost costume as she dangled from Amy’s mother’s ear. Then she remembered: if people saw her somebody would scream, but it was too late. Amy looked up at her mother. She saw Tilly wiggling and dangling from her mother’s ear. “You are wearing a costume.” said Amy. “You’re wearing an earring that looks like a spider wearing a ghost costume.” “I didn’t put on any earrings,” said Amy’s mom. Tilly tried to pull her legs up inside her costume. But they wouldn’t fit. Someone was going to see her. Then Amy’s mother screamed. The hairs on Tilly’s body vibrated so fast she shook all over. She looked like a dancing ghost hanging from Amy’s mother’s ear. Tilly had never heard anybody scream that loudly before. Tilly felt sorry for her – she sounded even more scared than Tilly felt. Amy’s mother batted at her ear. She screamed again. Tilly’s hairs vibrated. She hung onto her ghost costume as tightly as she could with all eight of her spindly legs. Tilly’s silk thread came loose from Amy’s mother’s ear. Her ghost costume acted like a tiny parachute. She flew through the air. Tilly had never been so scared in her life. She landed on Amy’s shoulder. Amy picked Tilly up gently in her hand. She saw the hairs vibrating all over Tilly’s body. “You scared the spider,” she told her mother. “Fair is fair. She scared me, too,” said Amy’s mother. “Everybody wants to be scared on Halloween,” said Matt. “But your spider isn’t scary. It’s cute,” said Matt. “That spider is wearing a costume. It must have really wanted to come to this party.” “Put it back on my ear. Maybe we can scare some more people,” said Amy’s mom. “It’s fun to be scared on Halloween.” Tilly looked around at the party and she agreed.
B.2 Background questionnaire

Kids Cues Study – Background questionnaire

Participant ID:

1. What is your child’s native language?

2. Is your child bilingual?

3. Has your child have any
   □ Ear inflections: ………… (number) from age ………………… until age …………………
   □ Glue ear: from age ………………… until age …………………
   □ Grommets: at age …………………

4. Does/Did your child have any other form of hearing loss? Please specify.

5. Is there anyone with hearing loss in the family (siblings/parents of child)? Please specify.

6. Has your child been diagnosed with any developmental disorder?
   □ Auditory Processing Disorder (APD) □ Attention Deficit Hyperactivity Disorder (ADHD)
   □ Dyslexia □ Autistic spectrum disorder (ASD)
   □ Dyspraxia □ Developmental Language Disorder/Specific Language Impairment
   □ Learning problems □ Other: …………………….

If not, is your child currently undergoing assessment for any of the above OR thought to have some tendencies for any of the above but not fully meeting criteria? Please specify.

7. Is there anyone with a developmental disorder in the family (siblings/parents of child)? Please specify.

8. Does your child have special educational needs? □ Yes □ No

9. Do you think your child’s speech and language production are normal? □ Yes □ No

10. Do you think your child’s understanding of spoken language is normal? □ Yes □ No

11. Has your child received speech and language therapy input? □ Yes □ No

12. Does your child play any instrument? □ Yes □ No

In case your child is diagnosed with APD please specify:

13. At what age was your child diagnosed with APD?

14. At which clinic was your child diagnosed with APD?

15. Is your child currently using a remote microphone system, a training software or other forms of APD intervention? Please specify. If not, has your child previously been using any of them?