My thesis title
Title second row

Shiran Koifman

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at University College London

January 15 2021

For Yihui Xie

Acknowledgements

This is where you will normally thank your advisor, colleagues, family and friends, as well as funding and institutional support. In our case, we will give our praises to the people who developed the ideas and tools that allow us to push open science a little step forward by writing plain-text, transparent, and reproducible theses in R Markdown.

We must be grateful to John Gruber for inventing the original version of Markdown, to John MacFarlane for creating Pandoc (http://pandoc.org) which converts Markdown to a large number of output formats, and to Yihui Xie for creating `knitr` which introduced R Markdown as a way of embedding code in Markdown documents, and `bookdown` which added tools for technical and longer-form writing.

Special thanks to Chester Ismay, who created the `thesisdown` package that helped many a PhD student write their theses in R Markdown. And a very special tahnks to John McManigle, whose adaption of Sam Evans' adaptation of Keith Gillow's original maths template for writing an Oxford University DPhil thesis in LaTeX provided the template that I adapted for R Markdown.

Finally, profuse thanks to JJ Allaire, the founder and CEO of RStudio, and Hadley Wickham, the mastermind of the tidyverse without whom we'd all just given up and done data science in Python instead. Thanks for making data science easier, more accessible, and more fun for us all.

Ulrik Lyngs
Linacre College, Oxford
2 December 2018

Abstract

This *R Markdown* template is for writing an Oxford University thesis. The template is built using Yihui Xie's `bookdown` package, with heavy inspiration from Chester Ismay's `thesisdown` and the `OxThesis` LaTeX template (most recently adapted by John McManigle).

This template's sample content include illustrations of how to write a thesis in R Markdown, and largely follows the structure from this R Markdown workshop.

Congratulations for taking a step further into the lands of open, reproducible science by writing your thesis using a tool that allows you to transparently include tables and dynamically generated plots directly from the underlying data. Hip hooray!

Contents

List of Figures **xi**

List of Tables **xiii**

List of Abbreviations **xv**

Introduction **1**

Speech-in-noise in children 1

APD definition 2

Diagnosis 2

Binaural and spatial listening in APD 2

Summary 3

1 Binaural listening: interripted and alternated speech-in-noise in adults **5**

1.1 Influence of distractor type on IM 6

1.1.1 Introduction 6

1.1.2 Experiment I: speech vs. non-speech distractors 6

1.1.3 Experiment II: speech distractors spoken in a familiar vs. unfamiliar language 6

1.1.4 General discussion and conclusion 8

1.2 Dichotic vs. monotic presentation and the influence of speech material 8

1.2.1 Introduction 9

1.2.2 Methods 9

1.2.3 Results 9

1.2.4 Discussion 10

1.2.5 Conclusion 10

2 Spatial listening: development and normalisation of a children's spatialised speech-in-noise test 11
2.1 Introduction . . . 12
2.2 Methods . . . 12
2.3 Discussion . . . 12
2.4 Conclusion . . . 12

3 APD study 13
3.1 Introduction . . . 14
3.2 Methods . . . 14
3.2.1 Participants . . . 14
3.2.2 Auditory evaluation . . . 15
3.2.3 Questionnaires . . . 22
3.3 Results . . . 23
3.3.1 Standard audiometry . . . 23
3.3.2 Extended high-frequency audiometry (EHFA) . . . 24
3.3.3 Switching task (ST) . . . 26
3.3.4 Spatialised speech-in-noise (LiSNS-UK) . . . 35
3.3.5 The Environmental Auditory Scene Analysis task (ENVASA) 42
3.3.6 CELF-RS . . . 45
3.3.7 Questionnaires . . . 46
3.4 Discussion . . . 47
3.5 Conclusion . . . 47

General discussion 49
Summary of main findings . . . 49
Conclusion . . . 49

Appendices

A The First Appendix 53

B The Second Appendix, for Fun 55

Works Cited 57

List of Figures

2.1 Code chunk syntax 12

3.1 APD participants pure-tone audiogram thresholds for standard frequencies plotted for the left and the right ear (black). The shaded grey area represents the TD group range of audiometric thresholds and the white line represents the mean at each frequency. The dashed line represents the threshold criteria of hearing level $\leq$ 25 dB HL. . 15

3.2 Schematic of the ENVASA experimental paradigm (taken from Leech et al., 2009) 21

3.3 Add caption here. 23

3.4 Add caption here. 25

3.5 APD participants pure-tone thresholds for extended high-frequencies plotted for the left and the right ear (black). The shaded grey area represents the TD group range of audiometric thresholds and the white line represents the mean at each frequency. 26

3.6 Scatterplot and linear regression lines for the listeners SRdTs measured with the switching task with the ASL (A) and CCRM speech corpus (B) as a function of age. Red indicates data from the APD group and green indicates data from the TD control group. The grey area represents the regression line 95% confidence interval. Reference data for normal hearing adults taken from Chapter 2 is given in the boxplots. 30

3.7 Boxplots of the listeners age-independent standardised residuals for data measured with the ASL (A) and the CCRM corpus (B). Residuals were calculated seperately for each condition and are based on a model predicton for TD group only. The grey area represents the deviance cut-off for abnormal score (SD $\pm$ 1.65 below/above the TD mean), where about 90% of the normal population is e expected to lay within. The dashed line represents the theorethical TD group mean (z = 0). 32

3.8 Add caption here. 36

3.9 Add caption here. 37
3.10 Scatterplot and linear regression lines for the LiSN-S UK and SPIN SRTs (A) and the derived measure SRM (B) as a function of the listeners age. Red indicates data from the APD group and green indicates data from the TD control group. The grey area represents the regression line 95 % confidence interval. 38
3.11 Add caption here. 38
3.12 Add caption here. 41
3.13 Scatterplot and linear regression lines for the ENVASA task %-correct as a function of age for single background, dual background and combined measure. Red indicates data from the APD group and green indicates data from the TD control group. The grey area represents the regression line 95 % confidence interval. 42
3.14 ENVASA standardised residuals for the %-correct in single background, dual background and combined measure. The dashed red line (z = -1.64) indicate the cut-off value for normal score, where only about 5 % of the normal population would be expected to score below it; the dashed green line (z = +1.64) indicate scores above which only 5 % of the normal population is expected to exceed. . . 44
3.15 Boxplots for CELF-5 UK Recall Sentences subtest scaled scores by groups. The dashed line represents the norms mean and the grey area indicate the upper and lower limit of an average performance in the normal population (± 1 SD). 46
3.16 Add caption here. 47
3.17 Add caption here. 47

List of Tables

3.1 Standard audiometry descriptive and test statistics for group differences tested using Wilcoxon rank-sum test for unpaired samples*. PTA_{Right} and PTA_{Left} were calculated by taking the individual mean for thresholds at 0.5, 1, 2 and 4 kHz in the respective ear. PTA denotes the listeners grand-mean for PTAs in both ears and BE represents the listeners PTA at the better-ear. 24

3.2 Extended high frequency hearing screening descriptive statistics and group difference analysis by frequency and ear or PTA. 25

3.3 Add caption here 29

3.4 Switching task descriptives for standardised residuals (z-scores) calculated for data measured with the ASL and CCRM speech material. 33

3.5 LiSNS standard residuals (z-scores) descriptives by group. 36

3.6 Add caption here. 40

3.7 Add caption here. 41

3.8 ENVASA standard residuals (z-scores) descriptive statistics and group difference analysis by test measures. 44

List of Abbreviations

1-D, 2-D . . . One- or two-dimensional, referring in this thesis to spatial dimensions in an image.

Otter . . . . . One of the finest of water mammals.

Hedgehog . . . Quite a nice prickly friend.

Introduction

Welcome to the *R Markdown* Oxford University thesis template. This sample content is adapted from `thesisdown` and the formatting of PDF output is adapted from the OxThesis LaTeX template. Hopefully, writing your thesis in R Markdown will provide a nicer interface to the OxThesis template if you haven't used TeX or LaTeX before. More importantly, using *R Markdown* allows you to embed chunks of code directly into your thesis and generate plots and tables directly from the underlying data, avoiding copy-paste steps. This will get you into the habit of doing reproducible research, which benefits you long-term as a researcher, but also will greatly help anyone that is trying to reproduce or build upon your results down the road.

Using LaTeX together with *Markdown* is more consistent than the output of a word processor, much less prone to corruption or crashing, and the resulting file is smaller than a Word file. While you may never have had problems using Word in the past, your thesis is likely going to be about twice as large and complex as anything you've written before, taxing Word's capabilities.

Speech-in-noise in children

R Markdown creates a simple and straightforward way to interface with the beauty of LaTeX. Packages have been written in **R** to work directly with LaTeX to produce nicely formatting tables and paragraphs. In addition to creating a user friendly interface to LaTeX, *R Markdown* allows you to read in your data, analyze it and to visualize it using **R**, **Python** or other languages, and provide documentation and commentary on the results of your project.

Further, it allows for results of code output to be passed inline to the commentary of your results. You'll see more on this later, focusing on **R**. If you are more into

Python or something else, you can still use *R Markdown* - see 'Other language engines' in Yihui Xie's *R Markdown: The Definitive Guide*.

APD definition

Anyone who needs to use data analysis, math, tables, a lot of figures, complex cross-references, or who just cares about reproducibility in research can benefit from using *R Markdown*. If you are working in 'softer' fields, the user-friendly nature of the *Markdown* syntax and its ability to keep track of and easily include figures, automatically generate a table of contents, index, references, table of figures, etc. should still make it of great benefit to your thesis project.

Diagnosis

Anyone who needs to use data analysis, math, tables, a lot of figures, complex cross-references, or who just cares about reproducibility in research can benefit from using *R Markdown*. If you are working in 'softer' fields, the user-friendly nature of the *Markdown* syntax and its ability to keep track of and easily include figures, automatically generate a table of contents, index, references, table of figures, etc. should still make it of great benefit to your thesis project.

Binaural and spatial listening in APD

Anyone who needs to use data analysis, math, tables, a lot of figures, complex cross-references, or who just cares about reproducibility in research can benefit from using *R Markdown*. If you are working in 'softer' fields, the user-friendly nature of the *Markdown* syntax and its ability to keep track of and easily include figures, automatically generate a table of contents, index, references, table of figures, etc. should still make it of great benefit to your thesis project.

Summary

Anyone who needs to use data analysis, math, tables, a lot of figures, complex cross-references, or who just cares about reproducibility in research can benefit from using *R Markdown*. If you are working in 'softer' fields, the user-friendly nature of the *Markdown* syntax and its ability to keep track of and easily include figures, automatically generate a table of contents, index, references, table of figures, etc. should still make it of great benefit to your thesis project.

Neque porro quisquam est qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit...

There is no one who loves pain itself, who seeks after it and wants to have it, simply because it is pain...

— Cicero's *de Finibus Bonorum et Malorum.*

1

Binaural listening: interripted and alternated speech-in-noise in adults

Contents

1.1 Influence of distractor type on IM 6
1.1.1 Introduction 6
1.1.2 Experiment I: speech vs. non-speech distractors 6
1.1.3 Experiment II: speech distractors spoken in a familiar vs. unfamiliar language 6
1.1.4 General discussion and conclusion 8
1.2 Dichotic vs. monotic presentation and the influence of speech material 8
1.2.1 Introduction 9
1.2.2 Methods 9
1.2.3 Results 9
1.2.4 Discussion 10
1.2.5 Conclusion 10

Here is a brief introduction to using *R Markdown*. *Markdown* is a simple formatting syntax for authoring HTML, PDF, and MS Word documents and much, much more. *R Markdown* provides the flexibility of *Markdown* with the implementation of **R** input and output. For more details on using *R Markdown* see http://rmarkdown.rstudio.com.

Be careful with your spacing in *Markdown* documents. While whitespace largely

is ignored, it does at times give *Markdown* signals as to how to proceed. As a habit, try to keep everything left aligned whenever possible, especially as you type a new paragraph. In other words, there is no need to indent basic text in the Rmd document (in fact, it might cause your text to do funny things if you do).

1.1 Influence of distractor type on IM

1.1.1 Introduction

- *Italics* are done like *this* or _this_
- **Bold** is done like **this** or ___this___
- ***Bold and italics*** is done like ***this***, _____this_____, or (the most transparent solution, in my opinion) **_this_**

1.1.2 Experiment I: speech vs. non-speech distractors

- `Inline code` is created with backticks like `` `this` ``

Methods

Sub_2 and $super^2$ script is created like this~2~ and this^2^

Results

- ~~Strikethrough~~ is done ~~like this~~

Discussion

- To include an actual *, _ or \, add another \ in front of them: *, _, \\

1.1.3 Experiment II: speech distractors spoken in a familiar vs. unfamiliar language

- – and — with -- and ---

Methods

Do like this:

> Put a > in front of the line.

Results

- are done with #'s of increasing number, i.e.
  - # First-level heading
  - ## Second-level heading
  - ### Etc.

In PDF output, a level-five heading will turn into a paragraph heading, i.e. `\paragraph{My level-five heading}`, which appears as bold text on the same line as the subsequent paragraph.

Discussion

Unordered list by starting a line with an * or a -:

- Item 1
- Item 2

Ordered lists by starting a line with a number:

1. Item 1
2. Item 2

Notice that you can mislabel the numbers and *Markdown* will still make the order right in the output.

To create a sublist, indent the values a bit (at least four spaces or a tab):

1. Item 1
2. Item 2

3. Item 3

    - Item 3a
    - Item 3b

1.1.4 General discussion and conclusion

The official *Markdown* way to create line breaks is by ending a line with more than two spaces.

Roses are red. Violets are blue.

This appears on the same line in the output, because we didn't add spaces after red.

Roses are red.
Violets are blue.

This appears with a line break because I added spaces after red.

I find this is confusing, so I recommend the alternative way: Ending a line with a backslash will also create a linebreak:

Roses are red.\
Violets are blue.

To create a new paragraph, you put a blank line.

Therefore, this line starts its own paragraph.

1.2 Dichotic vs. monotic presentation and the influence of speech material

- This is a hyperlink created by writing the text you want turned into a clickable link in `[square brackets followed by a](https://hyperlink-in-parentheses)`

1.2.1 Introduction

- Are created[1] by writing either ^[my footnote text] for supplying the footnote content inline, or something like `[^a-random-footnote-label]` and supplying the text elsewhere in the format shown below [2]:

```
[^a-random-footnote-label]: This is a random test.
```

1.2.2 Methods

To write comments within your text that won't actually be included in the output, use the same syntax as for writing comments in HTML. That is, <!-- this will not be included in the output -->.

1.2.3 Results

The syntax for writing math is stolen from LaTeX. To write a math expression that will be shown **inline**, enclose it in dollar signs. - This: $A = \pi*r^{2}$ Becomes: $A = \pi * r^2$

To write a math expression that will be shown in a block, enclose it in two dollar signs.

This: $$A = \pi*r^{2}$$

Becomes:

$$A = \pi * r^2$$

To create numbered equations, put them in an 'equation' environment and give them a label with the syntax `(\#eq:label)`, like this:

```
\begin{equation}
  f\left(k\right) = \binom{n}{k} p^k\left(1-p\right)^{n-k}
  (\#eq:binom)
\end{equation}
```

[1] my footnote text
[2] This is a random test.

Becomes:

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{1.1}$$

For more (e.g. how to theorems), see e.g. the documentation on bookdown.org

1.2.4 Discussion

- *R Markdown: The Definitive Guide* - https://bookdown.org/yihui/rmarkdown/
- *R for Data Science* - https://r4ds.had.co.nz

1.2.5 Conclusion

2

Spatial listening: development and normalisation of a children's spatialised speech-in-noise test

Contents

2.1 Introduction 12
2.2 Methods 12
2.3 Discussion 12
2.4 Conclusion 12

The magic of R Markdown is that we can add code within our document to make it dynamic.

We do this either as *code chunks* (generally used for loading libraries and data, performing calculations, and adding images, plots, and tables), or *inline code* (generally used for dynamically reporting results within our text).

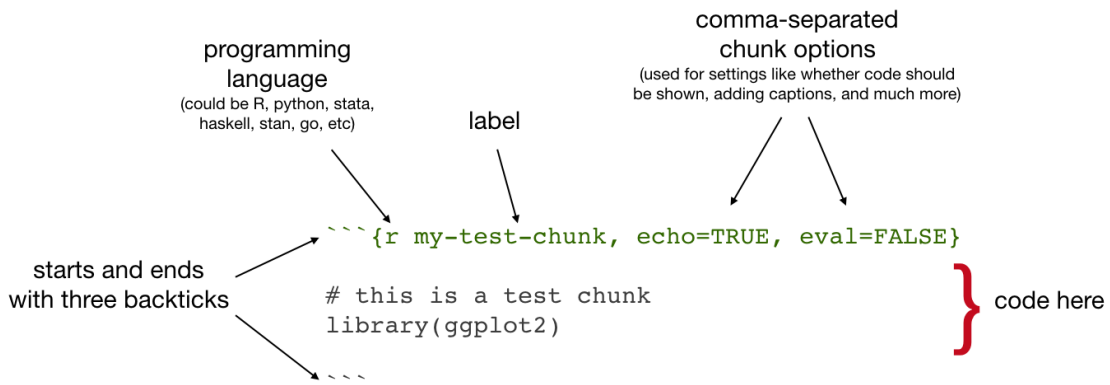

Figure 2.1: Code chunk syntax

2.1 Introduction

2.2 Methods

2.3 Discussion

2.4 Conclusion

The syntax of a code chunk is shown in Figure 2.1.

Common chunk options include (see e.g. bookdown.org):

- `echo`: whether or not to display code in knitted output
- `eval`: whether or to to run the code in the chunk when knitting
- `include`: wheter to include anything from the from a code chunk in the output document
- `fig.cap`: figure caption
- `fig.scap`: short figure caption, which will be used in the 'List of Figures' in the PDF front matter

IMPORTANT: Do *not* use underscoores in your chunk labels - if you do, you are likely to get an error in PDF output saying something like "! Package caption Error: \caption outside float".

3

APD study

Contents

3.1 Introduction 14
3.2 Methods 14
3.2.1 Participants 14
3.2.2 Auditory evaluation 15
3.2.3 Questionnaires 22
3.3 Results 23
3.3.1 Standard audiometry 23
3.3.2 Extended high-frequency audiometry (EHFA) 24
3.3.3 Switching task (ST) 26
3.3.4 Spatialised speech-in-noise (LiSNS-UK) 35
3.3.5 The Environmental Auditory Scene Analysis task (ENVASA) 42
3.3.6 CELF-RS 45
3.3.7 Questionnaires 46
3.4 Discussion 47
3.5 Conclusion 47

3.1 Introduction

3.2 Methods

3.2.1 Participants

Forty-four primary school children native British English speakers with normal hearing acuity participated in the study. Amongst them 20 belonged to the APD clinical group (5 females) with an average age of 11.04 ± 1.42 years (range: 7.8 - 12.9 years). One APD child was excluded from the analysis due to raised thresholds (PTA>25 dB HL). APD children were recruited in two ways. Children diagnosed with APD at Great Ormond Street Hospital (GOSH) and at the London Hearing and Balance Centre (LHBC), London, UK, and fulfilled the recruitment criteria were identified and contacted by a clinical team member. The parents/caregivers were provided with information about the study and means of contact to express interest in participation. Others were recruited by advertisements in social networks, where parents were requested to fill-out an interest form that included screening questions to ensure they fulfil the participation requirements. To add percentage for clinics, diagnosed/LiD/susAPD and SPD pattern? The remaining 23 (12 females) comprised of typically developing control children (TD) with no reported concerns or diagnosis of a language or other cognitive developmental disorders. The TD group average age was 9.47 ± 1.58 years and ranged between 7 to 12.1 years (A detailed description of the groups is shown in Tab. ??).

Difference in variance for age between the two groups was tested using t-test with Welch degrees of freedom correction for uneven sample-size, showing a significant difference in age between the groups [t(40.95)=3.43, p=0.001].

The project was approved by the UCL Research Ethics Committee (Project ID Number 0544/006) and the NHS Health Research Authority HRA (REC reference: 18/LO/0250). The testing commenced once an informed consent was given by both the parent/caregiver and the child.

- Background questionnaire

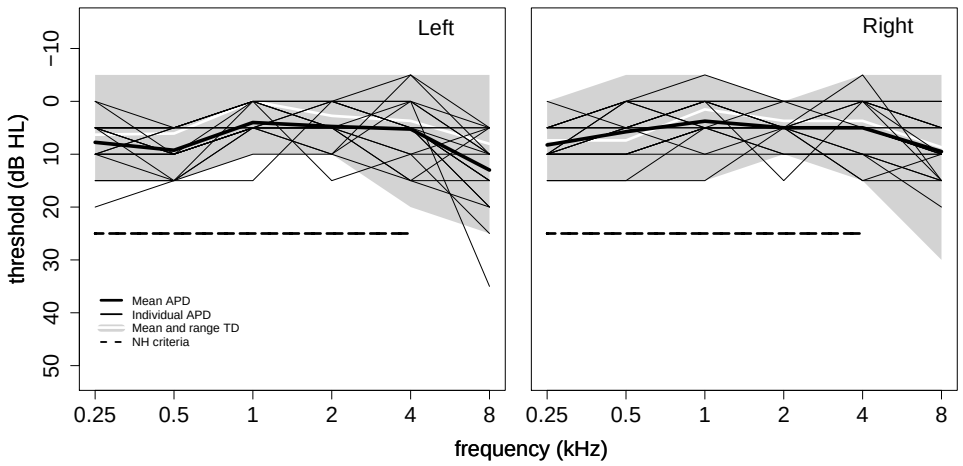

Figure 3.1: APD participants pure-tone audiogram thresholds for standard frequencies plotted for the left and the right ear (black). The shaded grey area represents the TD group range of audiometric thresholds and the white line represents the mean at each frequency. The dashed line represents the threshold criteria of hearing level $\leq$ 25 dB HL.

- Otoscopic examination was carried out to ensure the eardrum is visible, healthy and intact.
- Location of the testing
- duration of the session

Participants from both TD and APD group completed the same battery of tests listed below

3.2.2 Auditory evaluation

Standard audiometry

A standard air conduction pure-tone audiometry was carried out at six octave frequency bands ranging between 0.25 to 8 kHz using ???? audiometer and ??? headphones. Normal hearing acuity was defined as thresholds $\leq$ 25 dB HL for the octave frequency bands between 0.25 to 4 kHz. Thresholds at 8 kHz were $\leq$ 25 dB HL for all the participants, excluding two participants with thresholds at 35 and 30 dB HL in one ear, respectively. One participant from the clinical group (APD) was excluded from the analysis due to raised thresholds predominantly in the right ear (PTA_{Right} = 36.25 dB HL; PTA_{Left} = 13.75 dB HL). Otoscopy inspection of the child's ear canal revealed a large accumulation of cerumen whereby the right

eardrum was not visible. The listeners' thresholds for the left and the right ear are plotted in Figure 3.1. The shaded grey area represents the TD group thresholds range and the white line represents their mean at each frequency. The black lines represents the individual thresholds in the APD group and the group mean is marked by the bold black line. The dashed line represents the maximal thresholds criteria of $\leq$ 25 dB HL for participation in the study. Results belong here??

Extended high-frequency audiometry (EHFA)

Extended high-frequency pure-tone audiometry was carried out at four $\frac{1}{3}$ octave band frequencies 8, 11, 16, & 20 kHz using a locally developed MATLAB based software which generated and collected the data. Measurements took place at SHaPS, UCL laboratory in an electromagnetically shielded sound proof booth which is typically used for EEG measurements. A Windows PC situated outside the booth was connected via USB to an RME ???? sound card (Audio AG, Haimhausen Germany) and an ER10X Extended-Bandwidth Acoustic Probe System (Etym$\bar{o}$tic Research, Elk Grove Village, IL) which was located in the testing booth. Once the ear probe was placed in the child's ear, an in-situ sound pressure level calibration was performed (chirp noise) using a MATLAB code provided by ????.

Speak with KZ about the measurements

Switching task (ST)

The switching task (ST) is a novel speech-on-speech listening task that involves perception of interrupted and periodically segmented speech that is switched between the two ears out-of-phase with an interrupted distractor. Since segments of the target and of the distractor are never presented in the same ear at the same time, it enables to eliminate peripheral (EM) masking, while maintaining high IM for speech distractors. The task assesses the ability to switch attention and integration of binaural information.

Refer to Chapter 2 and briefly describe the stimuli and difference in the methods.

As described in Chapter 2 Section ???, two test versions were used with varying in sentence structure and complexity: 1. ASL 2. CCRM

Masker Types..

Spatialised speech-in-noise (LiSNS-UK)

The Listening in Spatialised Noise Sentences UK (LiSNS-UK) assesses the ability to use binaural cues in speech-on-speech listening conditions. The test development, speech material normalisation, and norms standardisation followed Cameron and Dillon (2007) development steps and are described in detail in Chapter ???. The test uses virtualisation techniques to create spatial distribution of sound sources in space for headphones presentation where target sentences (ASL; MacLeod & Summerfield, 1990) are presented in two simultaneous speech distractors (unrelated children's stories spoken by the target talker). It comprises of two main listening conditions, differing in their availability of spatial cues. The target sentences are configured to always appear in front of the listener's head, at 0° azimuth on the horizontal plane, with the two streams of speech distractors either co-located in space with the target (S0N0), resulting in relatively poor speech perception, or offset in space, with one distractor to either side of the target at ± 90°, resulting in an improvement in speech perception of circa 13 dB (Cameron et al., 2011), typically termed as spatial release from masking (SRM). This SRM advantage is calculated by taking the difference between performance in the co-located condition and the separated condition. The speech distractors were presented continuously throughout a run at a fixed 65? dB SPL output level and comprised of a combination of two out of three different passages children stories. A 1-up/1-down adaptive procedure was used, varying the level of the target talker relative to the distractors depending on listener's correct/incorrect response to measure the listeners' speech reception threshold (SRT), i.e., the signal-to-noise-ratio (SNR) yielding 50% speech intelligibility. A 2 ms long 1 kHz pure-tone was presented 500 ms before the target sentence onset at 65? dB SPL (0 dB SNR) as a reference cue signalling the listener to attend the coming target sentence. The initial target output level was ?? dB SPL

with an initial step-size of 4 dB SNR. The step-size was reduced after every reversal, reaching a minimum step-size of 2 dB SNR after three practice reversals. A stopping rule was introduced in case the maximal SNR was reached more than three times and the procedure was considered to be successfully completed in case test reversals were obtained. The SRT was then calculated by averaging test reversals SNRs (i.e., following three practice reversals). Each run consisted of 25 sentences taken from 8? phonemically-balanced test list which were constructed following the normalisation of the speech. In addition, a sentence-specific level correction was applied to the target signal (see Chapter ?? for more information). The order of the listening condition, test lists, target sentences and distractors combinations was fixed across all the participants and started with the collocated condition. Spatialisation was applied by convolving each stimuli with head-related transfer functions (HRTFs) at the corresponding azimuthal direction separately for the left and the right channel. The HRTFs were measured with a Knowles Electronics Manikin for Acoustic Research (KEMAR, REF) manikin with a small pinnae taken from the CIPIC HRTF database[1] [Algazi et al. (2001); see "special" HRTF data]. A post-equalisation step was applied in order to flatten the magnitude of the headphones frequency response. Headphone-to-ear Transfer Functions (HpTFs) measured with KEMAR manikin for HD-25 supraaural headphones were extracted from Wierstorf et al. (2011) HRTF database. The final mixed stimulus was filtered with the inverse HpTFs separately for the left and the right channel before being combined together as a final step. Every participant was presented with two runs, one for each listening condition (collocated/separated). Testing started following a practice phase of two runs for each of the test conditions with five BKB sentences each (Bench et al., 1979). Listeners were instructed to verbally repeat the target sentences to the experimenter who was situated alongside the participant in a sound treated chamber. The experimenter scored the response by selecting the correctly repeated keywords on the screen. Listeners were encouraged to guess if unsure while no feedback was given at any time. A loose keyword scoring method was

[1]The database is available online in: https://www.ece.ucdavis.edu/cipic/spatial-sound/hrtf-data/

used, whereby errors of case or declension were considered as correct responses. For example, a repetition of the keywords '<clown*s*> <funny> <face*s*>' to the stimulus 'The <clown> had a <funny> <face>'.

Speech-in-noise (SPIN)

The speech-in-noise test was used as a more realistic listening situation that is widely used in the clinic as opposed to more complex listening tasks as listed above. The normalised ASL sentences were presented in a speech-shaped-noise (SSN) with spectrum matched to the ASL corpus. The SSN onset was 500 ms before the target sentence begin. The exact same adaptive procedure as for the LiSNS-UK was used with the same stop-rules. Each listener was presented with a single run of 25 sentences following a practice phase with seven BKB sentences. The same test list and sentences order was used across all the listeners.

The Environmental Auditory Scene Analysis task (ENVASA)

In analogy to the classic 'cocktail-party' scenario, ENVASA is a non-linguistic paradigm (Leech et al., 2009) that measures detection of everyday environmental sounds presented in naturalistic auditory scenes and can be used to asses IM effects as well as sustained selective auditory attention skills. In the task, short environmental target sounds (e.g., a "dog's bark", "door knock" or "bouncing ball") were presented in a dichotic background scene (i.e., the target sound is presented only in one ear) consisting of either a single background scene,presented in both ears, or two background scenes, each presented in a different ear. The number of targets, the onset time and presentation ear varied across trials. Four target/background SNRs were employed split into two categories 'low' (-6 and -3 dB) and 'high' (0 and +3 dB) by varying the target level. Target/background contextual agreement was manipulated by embedding the target sound in a *congruent* background scene that is in agreement with the listener's expectations (e.g., a cow's 'moo' in a farmyard scene) or in an *incongruent* background scene which violate these expectations (e.g., a cow's 'moo' in a traffic scene).

Procedure:

The experiment was carried out using the original code and laptop as used and described by Leech et al. (2009). Sounds were presented via Sennheiser HD-25 headphones (REF) and the participants response was recorded using ???? gamepad. The output level was adjusted to a comfortable level before the test started. The participants were situated in front of the laptop placed on a desk and were instructed to hold the gamepad. Prior to the test begin the listeners were presented with a short child-friendly video covering the task's instructions and demonstrated two test trials. Following the instruction video, the examiner gave the child a short recap of the task's instructions and simulated with the child an exemplary trial to make sure the child is familiarised with the task. The task began with three practice trials with provided feedback, while no further feedback was given in the testing phase.

Every trial was made of two parts, starting with a target audio and visual familiarisation phase before the main target detection phase. Target identification was recorded by pressing one of the three buttons on the gamepad which corresponded to the location of the target objects on the screen. A response was counted as correct only if the participants pushed the corresponding button within 2 s time interval, 300 ms following the target onset. The outcome measure was calculated as the percentage of target sounds correctly identified within a condition (%-correct).

In total there were 92 target sounds presented over 40 trials, with half of the target sounds presented in a single- and half in a dual-background condition. In Occasional foil target items were played at 0 dB SNR without a corresponding picture on the screen and were used to estimate the quality of the participants performance. Each target item was served once as a foil item and their order was randomised.

CELF-RS

The Recalling Sentences (RS) sub-test of the Clinical Evaluation of Language Fundamentals Fifth UK edition (CELF-5-UK H Wiig et al., 2017) was administered to assess the listeners expressive language ability and has been shown to be a good indicator of the listeners general language skills (REF). In the task the child is

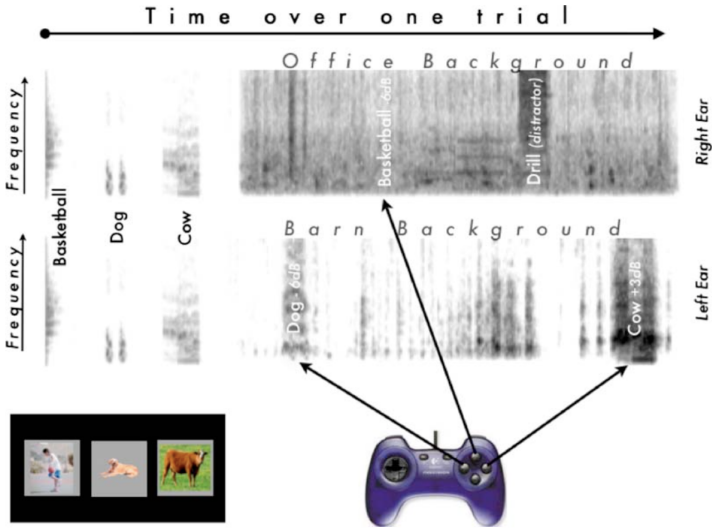

Figure 3.2: Schematic of the ENVASA experimental paradigm (taken from Leech et al., 2009)

presented with pre-recorded sentences of increasing length and complexity and required to repeat sentences without any changes. Scoring were marked by hand by the examiner as instructed by the test manual. The sentences were spoken by a standard southern British English female and were recorded in a sound-treated recording booth at the SHaPS UCL laboratory, London. The sentences were presented using a MATLAB program via headphones using the same experimental equipment as listed above at a comfortable output level of 70 dB HL. The task began with two practice sentences while the number of test items varied depending on the child's age performance. No repetitions or feedback was given during the testing and the test was discontinued in case the child failed to score any points for four consecutive items. Age-scaled score were calculated based on the test norms with a mean score of 10 and SD of 3. Scaled scores within $\pm$ 1 SD from the norm mean (between 8 to 12) are classified as average scores, whereas performance beyond $\pm$ 1 SD are classified as above/below the average score. Thus, the cut-off for abnormally poor performance is a scaled score $\leq$ 7.

3.2.3 Questionnaires

Medical, Neurological, and Pysiological History

The Evaluation pf Children's Listening and Processing Skills (ECLiPS)

The ECLiPS questionnaire (Barry & Moore, 2014) comprises of 38 items where the users are asked to express their agreement simple statements about the child's listening and other related skills or behaviours using a five-point Likert scale (from "strongly agree" to "strongly disagree"). The ECLiPS was design to identify listening and communication difficulties in children aged 6 to 11 years. Nonetheless, the UK standardisation study (REF) found little to no age effect on scores in many of the scale items, suggesting the testing age could be extended below and beyond the population used for the development. Based on factor analysis the items are grouped into five subcategories: 1. Speech & Auditory Processing (SAP), 2. Environmental & Auditory Sensitivity (EAS), 3. Language, literacy & laterality (L/L/L), 4. Memory & Attention (M&A), 5. Pragmatic & Social skills (PSS). Age- and sex-scaled scores were computed using the test excel scorer.

A score below the 10^{th} percentile (corresponding to a scale score of circa 6) is generally considered clinically significant.

The Children's Communication Checklist 2^{nd} edition (CCC-2)

Communication abilities were assessed using the Children's Communication Checklist second edition questionnaire (CCC-2; D. V. M., 2003) was completed by the child's parent/guardian. The CCC-2 was designed to screen communication problems in children aged 4 to 16 years and comprises of 70 checklist items each comprising of a behaviour statement like "Mixes up words of similar meaning". The respondents are asked to judge how often the behaviours occur using a four-point Likert rating scale: 0. *less than once a week (or never)*, 1. *at least once a week, but not every day*, 2. *once or twice a day*, 3. *several times (more than twice) a day (or always)*. The items are grouped into ten sub-scales of behaviours tapping into different skills (A. Speech, B. Syntax, C. Semantics, D. Coherence, E.

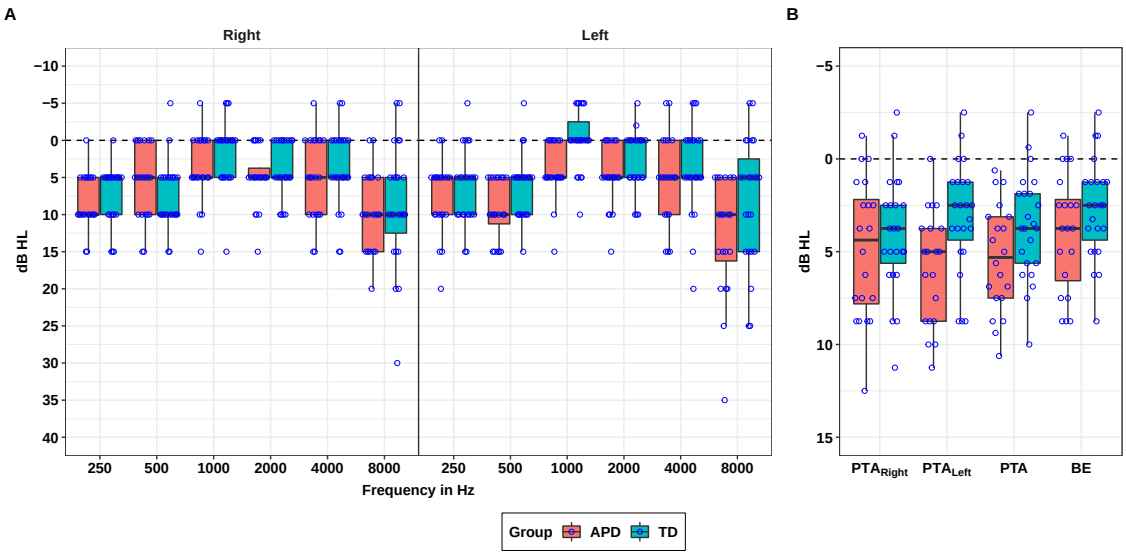

Figure 3.3: Add caption here.

Inappropriate initiation, F. Stereotyped language, G. Use of context, H. Non-verbal communication, I. Social relations, J. Interests). Taking the sum of scores for the sub-scales A to H are used to derive the General Communication Composite (GCC) which is used to identify clinically abnormal communication competence. A GCC score < 55 was found by Norbury and Bishop (2005) to well separate between control and clinical groups, identifying children with scores at the bottom 10%. Another composite (Social-Interaction Deviance Composite, SIDC) was taken by taking the difference in sum of scales E, H, I, and J from the sum of scales of A to D. Abnormal GCC (< 55) combined with a negative SIDC score has been shown to be indicative of an autistic spectrum disorder profile (D. V. M., 2003). The CCC-2 scaled and composite scores were computed using the test excel scorer. Data from one TD child out of forty four children was excluded from the analysis due to inconsistent reports flagged by the test scorer.

3.3 Results

3.3.1 Standard audiometry

Boxplots of listeners thresholds measured with a standard audiometer at the six octave frequency bands (0.25 to 8 kHz) and their corresponding PTAs are shown in

Table 3.1: Standard audiometry descriptive and test statistics for group differences tested using Wilcoxon rank-sum test for unpaired samples*. PTA_{Right} and PTA_{Left} were calculated by taking the individual mean for thresholds at 0.5, 1, 2 and 4 kHz in the respective ear. PTA denotes the listeners grand-mean for PTAs in both ears and BE represents the listeners PTA at the better-ear.

		APD					TD					Wilcoxon rank-sum test			
	Ear	N	median	sd	min	max	N	median	sd	min	max	CI	p	effect size (r)	magnitude
octave frequency bands															
250	R	20	10.00	3.73	0.00	15.00	23	5	3.95	0	15	-0.00 - 5.00	*0.35*	*0.14*	small
500	R	20	5.00	4.94	0.00	15.00	23	10	4.49	-5	15	-0.00 - 5.00	*0.51*	*0.10*	small
1000	R	20	5.00	4.55	-5.00	15.00	23	0	4.38	-5	15	-5.00 - 0.00	*0.19*	*0.20*	small
2000	R	20	5.00	3.97	0.00	15.00	23	5	3.76	0	10	0.00 - 5.00	***0.03***	*0.34*	moderate
4000	R	20	5.00	5.62	-5.00	15.00	23	5	5.27	-5	15	-0.00 - 5.00	*0.08*	*0.27*	small
8000	R	20	10.00	5.36	0.00	20.00	23	10	8.29	-5	30	0.00 - 5.00	***0.00***	*0.48*	moderate
250	L	20	5.00	4.99	0.00	20.00	23	5	5.05	-5	15	-0.00 - 5.00	*0.29*	*0.16*	small
500	L	20	10.00	4.06	5.00	15.00	23	5	4.25	-5	15	-0.00 - 5.00	*0.12*	*0.24*	small
1000	L	20	5.00	3.84	0.00	15.00	23	0	3.99	-5	10	-0.00 - 5.00	*0.46*	*0.11*	small
2000	L	20	5.00	4.13	0.00	15.00	23	0	4.03	-5	10	-0.00 - 5.00	*0.32*	*0.15*	small
4000	L	20	5.00	5.95	-5.00	15.00	23	5	6.07	-5	20	-5.00 - 5.00	*0.58*	*0.09*	small
8000	L	20	10.00	8.01	5.00	35.00	23	5	8.49	-5	25	-0.00 - 10.00	*0.05*	*0.30*	small
PTAs and better-ear															
PTA_{Right}	R	20	4.38	3.78	-1.25	12.50	20	5	5.62	-5	15	-1.25 - 3.75	*0.56*	*0.09*	small
PTA_{Left}	L	20	5.00	3.04	0.00	11.25	20	10	5.36	0	20	1.25 - 5.00	***0.01***	*0.42*	moderate
PTA		20	5.31	2.92	0.62	10.62	20	5	4.99	0	20	-0.00 - 3.75	*0.08*	*0.27*	small
BE		20	3.75	3.17	-1.25	8.75	20	10	4.06	5	15	-1.25 - 3.75	*0.23*	*0.18*	small

* significance level = 0.05

Figure 3.3 A-B. The listeners PTAs were calculated separately for the right and left ear (PTA_{Right}, PTA_{Left}) by taking the mean of thresholds at the frequency bands 0.5, 1, 2 and 4 kHz which are known to be important for speech perception (REF WHO). PTA denotes the listeners grand-mean for PTAs in both ears, form a single measure for the listeners hearing ability, and BE represents the listeners PTA at the better-ear. Descriptive statistics is given in Table 3.1, split by groups.

Do the groups differ in their hearing abilities?

- To report Wilcoxon or LMEM test? - Wilcoxon test and LMEM found a significant difference between the groups for thresholds in the left ear. Nonetheless, the differences are rather small and clinically negligible, and there is no reason to assume this difference occurred by random.

3.3.2 Extended high-frequency audiometry (EHFA)

The listeners' thresholds measured with the ER10X at the EHFs 11 and 16 kHz are plotted in Figure 3.5 for the left and the right ear. The shaded grey area represents the TD group thresholds range and the white line represents their mean at each frequency. The black lines represents the individual thresholds in the APD

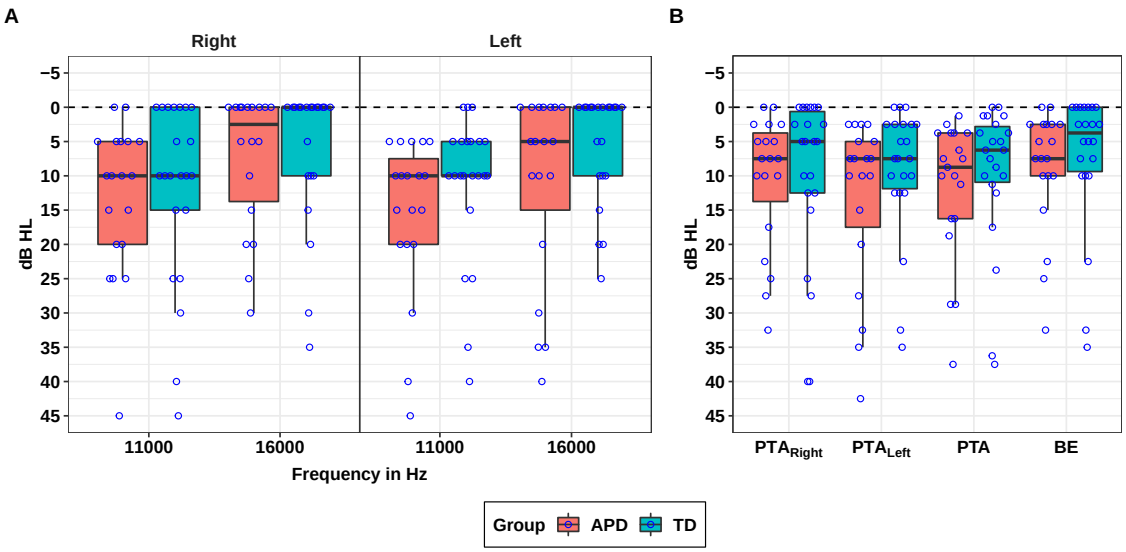

Figure 3.4: Add caption here.

Table 3.2: Extended high frequency hearing screening descriptive statistics and group difference analysis by frequency and ear or PTA.

		APD					TD					Wilcoxon rank-sum test			
	Ear	N	median	sd	min	max	N	median	sd	min	max	CI	p	effect size (r)	magnitude
octave frequency bands															
11000	R	19	10.00	11.19	0.00	45.0	22	10.00	13.24	0	45.0	-5.00 - 10.00	*0.47*	*0.11*	small
16000	R	18	2.50	10.04	0.00	30.0	21	0.00	10.51	0	35.0	-0.00 - 10.00	*0.14*	*0.23*	small
11000	L	19	10.00	11.65	5.00	45.0	22	10.00	10.71	0	40.0	-0.00 - 5.00	*0.57*	*0.09*	small
16000	L	19	5.00	13.80	0.00	40.0	21	0.00	8.11	0	25.0	-0.00 - 10.00	*0.21*	*0.20*	small
PTAs and better-ear															
PTA_{Right}	R	19	7.50	9.74	0.00	32.5	22	5.00	12.42	0	40.0	-2.50 - 5.00	*0.48*	*0.11*	small
PTA_{Left}	L	19	7.50	12.28	2.50	42.5	22	7.50	9.62	0	35.0	-2.50 - 7.50	*0.27*	*0.18*	small
PTA		19	8.75	10.18	1.25	37.5	22	6.25	10.57	0	37.5	-2.50 - 6.25	*0.29*	*0.17*	small
BE		19	7.50	8.95	0.00	32.5	22	3.75	10.09	0	35.0	-2.50 - 5.00	*0.24*	*0.18*	small

group and the group mean is marked by the bold black line. A comparison of the group means indicates that the differences in thresholds are small. Again, boxplots of the listeners thresholds by frequency and ear as well as PTAs are shown in Figure 3.4 A-B. Descriptive statistics and Wilcoxon rank-sum test outcomes for group comparison (unpaired samples) is given in Table 3.2, split by groups.

Difference between groups Difference in thresholds between the groups across frequencies (11 & 16 kHz) and ears (left/right) as well as for the calculated PTA and BE measures were tested using a Wilcoxon rank-sum test for unpaired samples ('rstatix::wilcox_test' with bonferroni adjustment; REF). No significant difference was found between the groups for all thresholds (all p>.05; see Table 3.2). A LMEM model with **frequency** and **ear** as fixed factors showed similar results

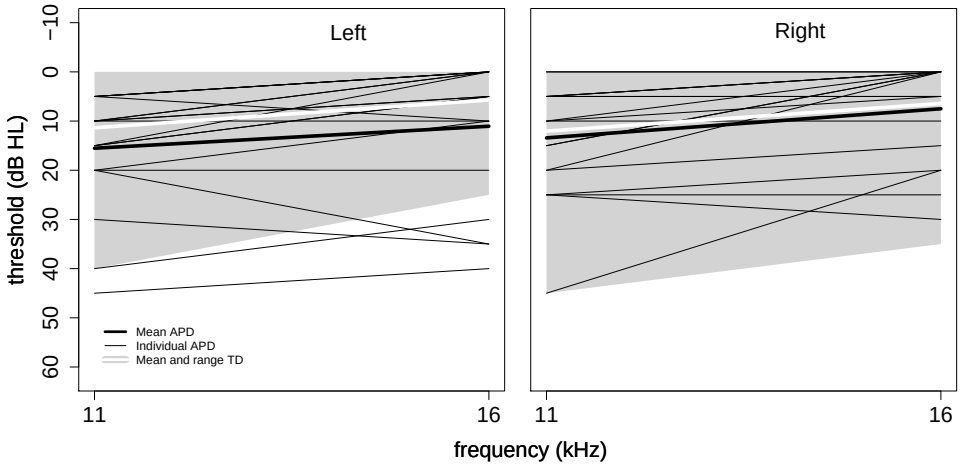

Figure 3.5: APD participants pure-tone thresholds for extended high-frequencies plotted for the left and the right ear (black). The shaded grey area represents the TD group range of audiometric thresholds and the white line represents the mean at each frequency.

and thus was not reported here.

3.3.3 Switching task (ST)

Data Analysis

Outliers & missing data

As a first step, the listeners adaptive track and psychometric functions (PF) were manually inspected for abnormalities. The proportion of correct keywords within the final test trials (LevsPC) was calculated as a measure describing the success of the adaptive procedure. Since the adaptive procedure was set to yield 50%-correct, a successful procedure is expected to be within the 50% range. A binomial statistical test was applied to identify observations that significantly differ from 50%. Observations with LevsPC $\leq$ 35% were flagged as possible outliers and were further inspected (see Figure ??). Interestingly, the majority of the flagged cases belonged to the CCRM test condition where targets were presented with a competing CCRM-type sentences (CCRM_F). Three observations out of 215 (5 conditions x 43 listeners) were flagged for data measured with the ASL corpus, and 29 observations out of 258 (6 conditions x 43 listeners) were flagged for data measured with the CCRM corpus. As expected, most of the identified cases were for observations measured with the more demanding conditions with speech distractors

(see Figure ??). In five cases (2 ASL; 3 CCRM) we were able to confidently determine that the listener's true score was near to ceiling, and thus these observations were set to the maximum DC (0.97). In other cases it was not possible to confidently determine the true SRdT, either because the maximum number or trials were presented before a minimum number of test reversals were obtained (CCRM = 1; ASL = 2), or due to aberrant adaptive tracks (CCRM = 5). Since all these cases belonged to more challenging test conditions with speech distractors, it is very likely that the children's true score is at celling. Thus, to account that, rather than removing these observations, which will consequently reduce the statistical power, they were set to a DC of 1, which is above the task's upper DC limit of 0.97.

Regression lines + z-scores

Age-independent scores were estimated using a linear regression model. The model was fitted condition-wise separately for each test version (ASL/CCRM) and was based on the control group data only with %-correct as the dependent variable and age as a predictor. A two-steps model comparison was performed to test the assumption that performance displays a monotonic linear relationship with age versus a non-monotonic (segmented) linear relationship. Extreme outliers were initially trimmed from the TD group to reduce noise in the data and to improve the models fit. In the first step, both models were computed and the best model was selected based on a model comparison using analysis of variance **anova()** test. Standard residuals were next calculated for each TD listener, based on the selected model prediction. The standardised residuals are age-independent and are comparable to z-scores for data with normal distribution, with a mean and SD of approximately 0 and 1, respectively. Since the main goal of the study was to find a measure that is able to well separate between the APD group and the typically developed control group, individual differences and group differences were explored using a deviance analysis procedure proposed by Ramus et al. (2003). Abnormal scores were defined by a two-tailed deviance cut-off of ± 1.65 SD from the TD group mean. Thus, circa 90% of the normal population residuals are expected to be within the deviance range of ± 1.65. Occasional occurrence of abnormal scores

in the normal population is not unusual in behavioural measures. Therefore, since the prediction of the residuals is based on the control data, such outliers may skew the TD group true mean or SD and thus may introduce an error in the model prediction. Therefore, in the second step, TD outliers (with standardised residuals below/above TD mean ± 1.65) were trimmed from the data and the two models were re-fitte and compared again. Finally, the selected model with the best fit was used to calculate the standardised residuals for all the listeners, including the trimmed TD observations and the APD group.

parametric data assumptions Inspection of the data revealed that the assumption of normal distribution (Shapiro Wilk test) and the assumption of homogeneity of variance (Levene's test) were not met. Thus, the relationship between the listeners SRdTs and their age was analysed separately for each group per condition using Kendall's tau (τ) which is a non-parametric correlation test (Kendall's τ and p values are displayed in Figure 3.6 A-B). Similar to Spearman's rho coefficient, Kendall's τ ranges between +1 and -1, denoting an optimal correlation at both extremes. The test is recommended for small sample size and a large number of tied ranks (REF Andy fields book p181), which fits to the restricted range possible outcomes in the ST task due to the nature of the adaptive procedure.

ADD HERE: Description about nParLD Rank-based ANOVA type statistic test (ATS) was used f2 ld f1 type test (between/within...) –> nparLD offers different test statistics, whereas due to the small sample-size the ANOVA-type test statistic was preferred (Noguchi al., 2012).

other tests were considered and delievered the same results!!

SRdTs by age

Since the age of the TD/APD children that participated in the present study spanned between circa 7 to 13 years, a developmental age effect was expected, where performance was expected to improve with an increasing age. To inspect this effect across the different test conditions and speech material (ASL/CCRM), scatterplots and linear regression lines for the listeners SRdTs as a function of age are shown in

Table 3.3: Add caption here

	Statistic	df	p-value
Group	13.555	1.000	***0.000***
Condition	3.730	3.450	***0.008***
Material	0.246	1.000	*0.62*
Group:Condition	1.957	3.450	*0.109*
Material:Condition	1.473	3.472	*0.214*
Group:Material	0.181	1.000	*0.67*
Group:Condition:Material	0.688	3.472	*0.58*

* significance level = 0.05

Figure 3.6 A-B. The effect of age was tested against the TD group only since the variability in scores in the TD group is expected to be relatively smaller than in the clinical APD group. Therefore, any interpretation based on the APD regression lines should be carried out cautiously. The TD regression lines were determined based on model comparison and outliers trimming procedure to improve the model's fit (see section ??). Regular regression lines were found to be the most parsimonious in describing the relationship between the TD children performance and their age. This was the case in all test conditions but the MDR_F distractor in the ASL material, where a segmented line was found to give the best fit. Based on the segmented line, DC improved with age by circa 0.1 per year until reaching a plateau at the age of 9.5 years. Inspection of the regression lines revealed a clear developmental effect in both speech material, where SRdTs improved with age (smaller SRdT $\rightarrow$ better performance), with a larger improvement in performance when additional distractor is presented. While the trend in performance was similar in both speech materials, as expected, CCRM sentences were more intelligible, with performance for the CCRM corpus shifted towards lower (i.e., more difficult) DC relative to performance for the ASL corpus. This is due to the more simple speech material and the restricted alternative responses of the CCRM matrix-based sentences.

The effect of age on performance of the trimmed TD group was analysed using Kendall's τ non-parametric correlation test, with Kendall's τ and p-values displayed in Figure 3.6 A-B, split by speech material and condition. The test results showed

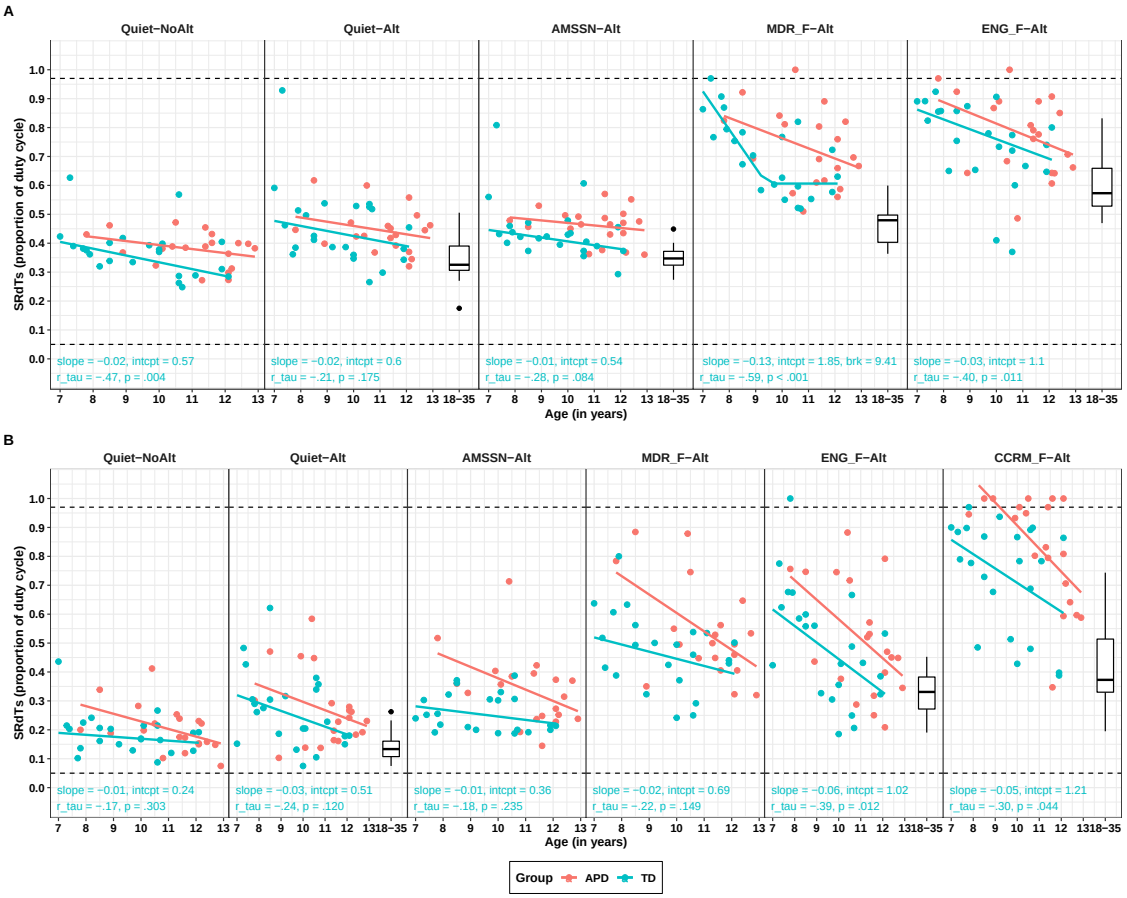

Figure 3.6: Scatterplot and linear regression lines for the listeners SRdTs measured with the switching task with the ASL (A) and CCRM speech corpus (B) as a function of age. Red indicates data from the APD group and green indicates data from the TD control group. The grey area represents the regression line 95% confidence interval. Reference data for normal hearing adults taken from Chapter 2 is given in the boxplots.

a mixed developmental trend across conditions and test material. There was a significant correlation (moderate effect size) between age and SRdTs for the test conditions Quiet-NoAlt, MDR_F, and ENG_F when measured with the ASL material, and for ENG_F and CCRM_F for the CCRM speech material, whereas no significant correlation was found for the remaining conditions (*all p's* > 0.05).

A closer look at the linear lines shows several interesting trends. Firstly, the improvement in performance by age was more prominent for speech distractors, with relatively steeper slopes than for the non-speech distractor (AMSSN) or for conditions without a distractor. Furthermore, alternation on it's own had only a small impact on performance when comparing performance for the target sentences

in quiet with and without alternations. Moreover, only speech distractors resulted noticeable decrement in performance (i.e., increased SRdTs), whereas performance for AMSSN was fairly similar to the Quiet conditions. Interestingly, when comparing the regression lines, there appears to be a relatively larger separation between the groups for data measured with the CCRM speech material, especially for AMSSN, but also for the speech distractors. However, it is possible that the APD reggression lines do not reflect the true population due to due to the large spread in performance and the small sample size and thus any interpertation should be taken with a pinch of salt. Another interesting observation is that the children showed little-to-no *masking-release* for speech spoken in an unfamiliar language (MDR_F) when compared with a distractor spoken in English (ENG_F). This is in agreement with findings in the adults study in Chapter 2. Lastly, it is apparent from the figure that performance for CCRM_F distractor was near-to-ceiling for some children, mostly among the APD group.

An exploratory comparison between between the children's data measured in the present study with data measured across young NH adults collected in Chapter 2 further highlight the strong developmental trend, with SRdTs still not entirely "adult-like" even at the age of 13 years, especially for speech distractos (see boxplots in Figure 3.6 A-B). The children in both groups seems to be markedly susceptible to competing CCRM sentences or familiar/unfamiliar speech presented with ASL sentences, with performance at the age of 12 years still largely differing from those obtained by the adults. On the other hand, by the age of 12 years the TD children reached near to "adult-like" performance when CCRM target sentences were presented with with ENG_F speech distractor or when ASL sentences were presented with AMSSN distractor.

Age-independent z-scores

Boxplots & abnormal scores by conditions Age-independent (standardised residuals) z-scores were calculated based on the TD group data using a multiple-case study approach (Ramus et al. (2003); see section ??? for more details). Boxplots

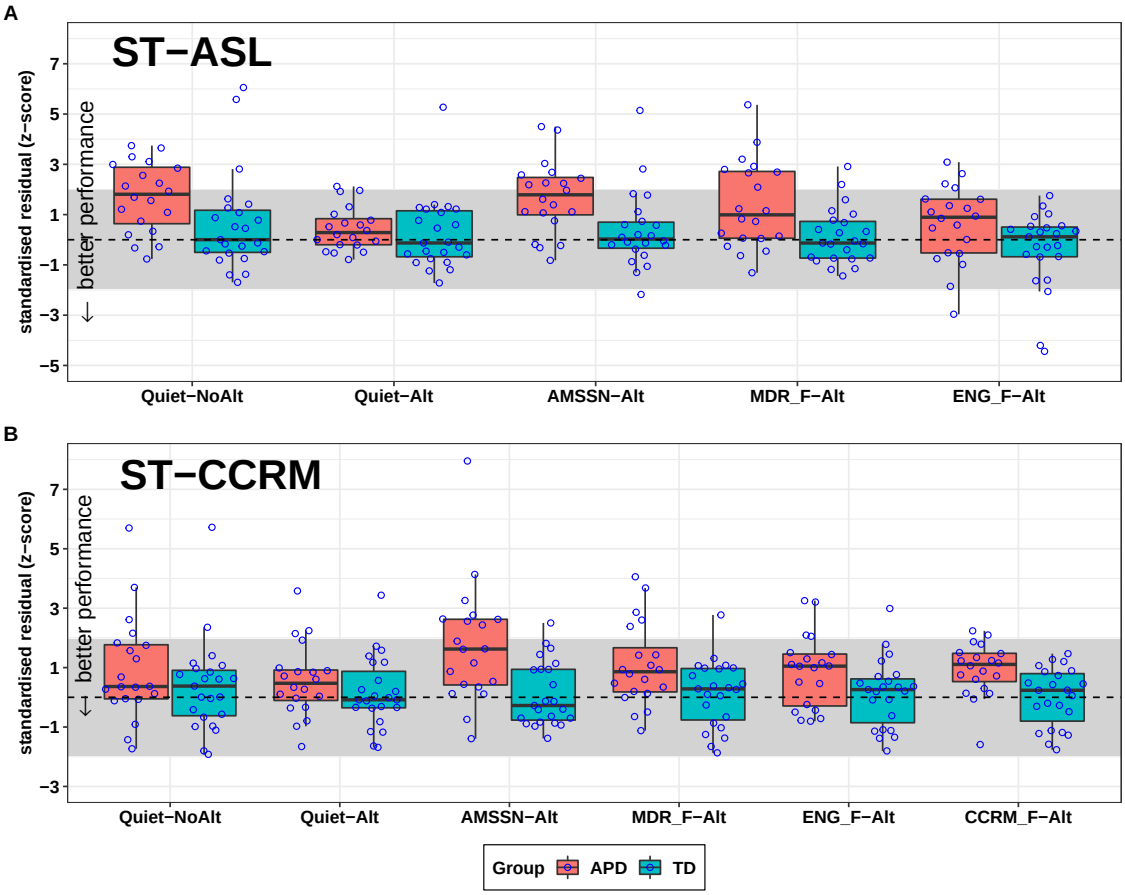

Figure 3.7: Boxplots of the listeners age-independent standardised residuals for data measured with the ASL (A) and the CCRM corpus (B). Residuals were calculated seperately for each condition and are based on a model predicton for TD group only. The grey area represents the deviance cut-off for abnormal score (SD $\pm$ 1.65 below/above the TD mean), where about 90% of the normal population is e expected to lay within. The dashed line represents the theorethical TD group mean (z = 0).

of the listeners age-independent z-scores is shown in Figure 3.7 A-B, for the ASL and CCRM speech material respectively. Scores were calculated separately for each test condition, with better performance indicated by smaller z-score. The grey area marks the two-tailed 1.65 deviance cut-off for abnormal score from the control group mean ($z \approx 0$), where only about 10% of the normal population is expected to score below/above it. Overall, APD children performance in both test versions was noticeably poorer, with higher median z-scores than compared with the TD children.

The next paragraphs will cover the examination and analysis of the individuals and group differences separately for each test version.

Table 3.4: Switching task descriptives for standardised residuals (z-scores) calculated for data measured with the ASL and CCRM speech material.

	APD					TD				
	N	median	sd	min	max	N	median	sd	min	max
ASL										
Quiet-NoAlt	20	1.81	1.39	-0.76	3.74	23	0.00	1.96	-1.69	6.05
Quiet-Alt	20	0.29	0.87	-0.79	2.12	23	-0.13	1.46	-1.72	5.27
AMSSN-Alt	20	1.79	1.45	-0.82	4.50	23	0.10	2.35	-2.18	9.04
MDR_F-Alt	20	0.99	1.75	-1.31	5.37	23	-0.13	1.11	-1.44	2.91
ENG_F-Alt	20	0.90	1.53	-2.96	3.09	23	0.12	1.55	-4.44	1.75
CCRM										
Quiet-NoAlt	20	0.36	1.75	-1.73	5.70	23	0.38	1.57	-1.92	5.72
Quiet-Alt	20	0.47	1.23	-1.66	3.58	23	-0.08	1.19	-1.68	3.44
AMSSN-Alt	20	1.62	2.03	-1.39	7.95	23	-0.28	1.09	-1.38	2.50
MDR_F-Alt	20	0.86	1.40	-1.12	4.06	23	0.28	1.11	-1.87	2.77
ENG_F-Alt	20	1.05	1.22	-0.80	3.25	23	0.26	1.14	-1.80	2.99
CCRM_F-Alt	20	1.11	0.89	-1.59	2.24	23	0.24	0.98	-1.76	1.47

ASL corpus Surprisingly, APD children were most susceptible to the non-switched condition where the target sentences were presented in quiet (Quiet-NoAlt) and the switched condition where the target sentences were presented with a non-speech distractor (AMSSN), with a median score laying above the norms range (see grey area). In addition, APD children performance was noticeably poorer for speech distractor spoken in an unfamiliar language (MDR_F), with performance just within the norms range. Approximately half of the APD children exhibited abnormal scores for Quiet-NoAlt (10/20, 50%) and AMSSN (11/20, 55%) conditions. Paradoxically, the APD children showed a greater difficulty in repeating the target ASL sentences in quiet without switching (Quiet-NoAlt) than with switching (Quiet-Alt). Another interesting observation is that the APD children did not benefit from release of masking for a speech distractor spoken in an unfamiliar language (MDR_F) as opposed to a familiar speech spoken in English (ENG_F). This sits well with our previous findings with adults where adults showed no benefit for MDR_F speech masker (see chapter ???). Moreover, not only that the APD children did not get any release of masking, they even showed a further decrement in performance, with a high proportion of APD children exhibiting abnormal score (9/20 = 45.0%), while only

few APD children were abnormal for ENG_F condition (4/20 = 55.0%). TD children performance on the other hand was relatively similar in both conditions, with only few children outside the norm (MDR_F: 3/23 = 13.0%; ENG_F: 1/19 = 4.3%).

CCRM corpus Figure 3.7 B reveals a similar trend as seen for the ASL sentences, however with a more modest differences between the two groups. AMSSN yielded the largest separation, where 45% (9/20) of the APD children had abnormal scores and with a median score of 1.62, which is close to the +1.65 upper deviance cut-off. In comparison, only 13% of the TD children (3/23) had abnormal performance for AMSSN condition. As seen in the ASL corpus, Quiet-NoAlt showed relatively larger separation between the groups where performance of 35% (7/20) of the APD children was abnormal, as opposed to circa 9% (2/23) in the TD group. As for the speech distractors, 30% (6/20), 25% (5/20) and 20% (4/20) of the APD children had abnormal score for ENG_F,MDR_F and CCRM_F conditions, respectively. The percentage of abnormal scores in the TD group were relatively low with 13% (3/20) for ENG_F, circa 4% (1/23) for MDR_F, 0% for CCRM_F.

nparLD() full 2x2x5 model

A three-way 2 x 2 x 5 factorial design with repeated measures was used to test the main effects of Group, Condition, and Material as well as their interaction on performance in the task (with z-scores as dependent variable). Note that the CCRM test condition with CCRM-type sentences as distractor (CCRM_F) was not was not included in the model since there was no comparable condition for the ASL speech material. Inspection of the standardised residuals z-scores revealed that the assumption of normal distribution (Shapiro Wilk test) was rejected for data measured with both speech material. Furthermore, homogeneity of the variance in the APD group was rejected for the ASL corpus (Levene's test; $F(4,95) = 2.71$, $p < 0.05$). Thus, due to the small sample size and the incomplete fulfilment of parametric statistical methods assumptions, a non-parametric approach was adopted. This was tested with a rank-based ANOVA-type statistic test (ATS) using the *nparLD()* function (nparLD package, REF), which is a robust method

for small datasets with outliers or skewed scores (see Jos Feys, 2016 for a good introduction for robust nonparametric techniques). The *nparLD ATS* test results is given in Table 3.3. No significant three-way or two-way interactions were found (Group x Condition x Material, Group x Material, Material x Condition, and Group x Condition; all *p's > 0.05*), while there was a highly significant main effect of Group (*p < 0.0001'*) and a strong main effect of Condition (*p < 0.001'*). Other non-parametric tests procedures gave similar results.

Discussion or here?

The lack of significant interaction (Group x Condition or Group x Condition x Material), is somewhat surprising and and do not reflect some of the differences seen in Figure ?? between the two groups for certain conditions or the overall difference in performance between the speech materials. These disagreement mat suggest that the model is under-powered to test that.

additional model for CCRM_F? An additional 2 x 6 model was included to test the effect of CCRM-type distractor on performance in the CCRM speech material, with Group and Condition as predictors and z-scores as dependent variable using *nparLD ATS* test for between x within design (f1.ld.f1). The ATS test found a strong significant difference between the groups (Statistic = 10.980, df = 1.000, p < 0.0), while no significant main effect was found for Condition (Statistic = 0.819, df = 4.618, p = 0.527) nor for Group x Condition interaction (Statistic = 1.215, df = 4.618, p = 0.301).

ROC curves table?

Switching effect (derived measures)

3.3.4 Spatialised speech-in-noise (LiSNS-UK)

SRTs by age

The distribution of the listeners SRTs by age, split by group, and their corresponding linear regression lines are shown in Figure 3.10 A. Due to time constrains and the large test battery, each participant was presented only once with each condition –

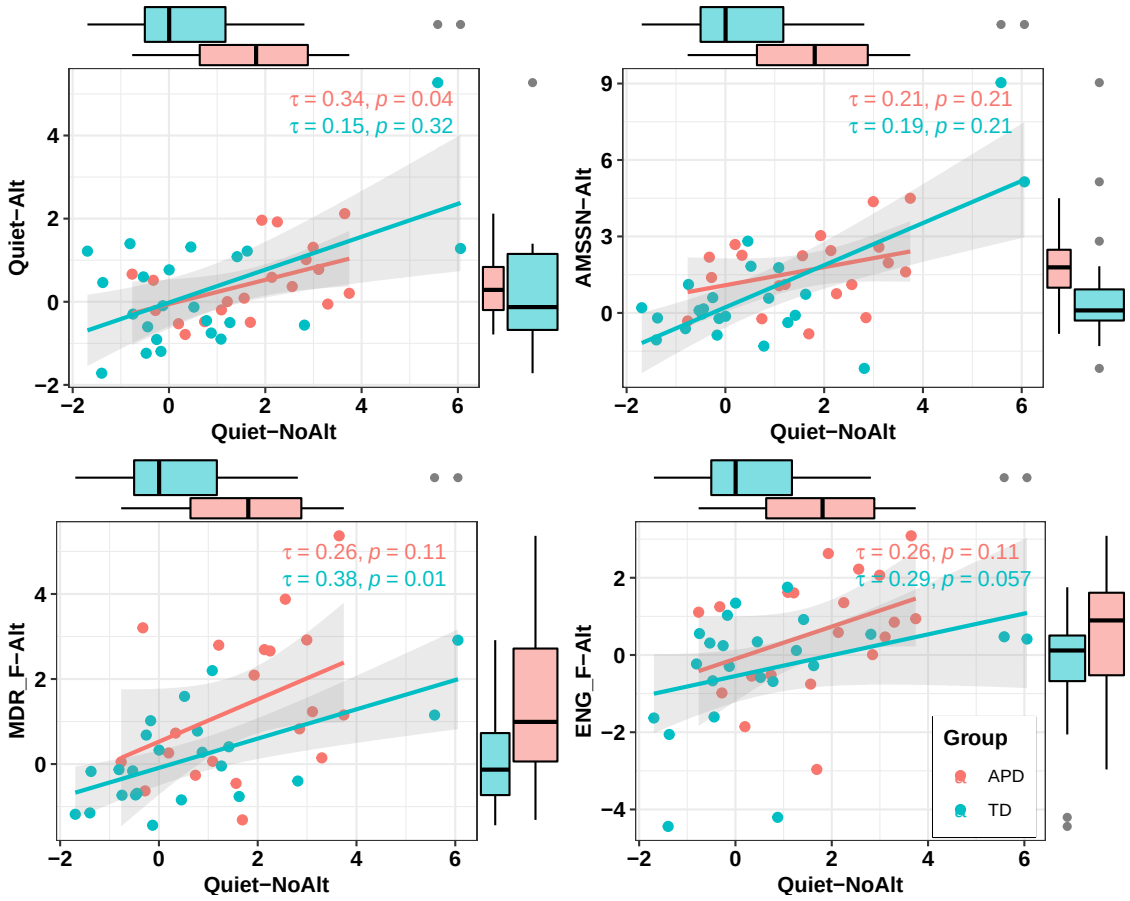

Figure 3.8: Add caption here.

Table 3.5: LiSNS standard residuals (z-scores) descriptives by group.

	TD					APD				
	N	median	sd	min	max	N	median	sd	min	max
SSN	23	0.16	0.98	-1.68	1.84	19	0.94	1.14	-1.48	3.07
S0N0	23	0.06	1.28	-1.81	3.28	19	1.22	1.31	-1.18	3.52
S0N90	23	-0.18	0.98	-1.55	1.85	19	0.44	1.11	-1.17	3.26
SRM	23	0.25	0.68	-0.71	1.61	19	0.61	1.37	-2.46	3.08

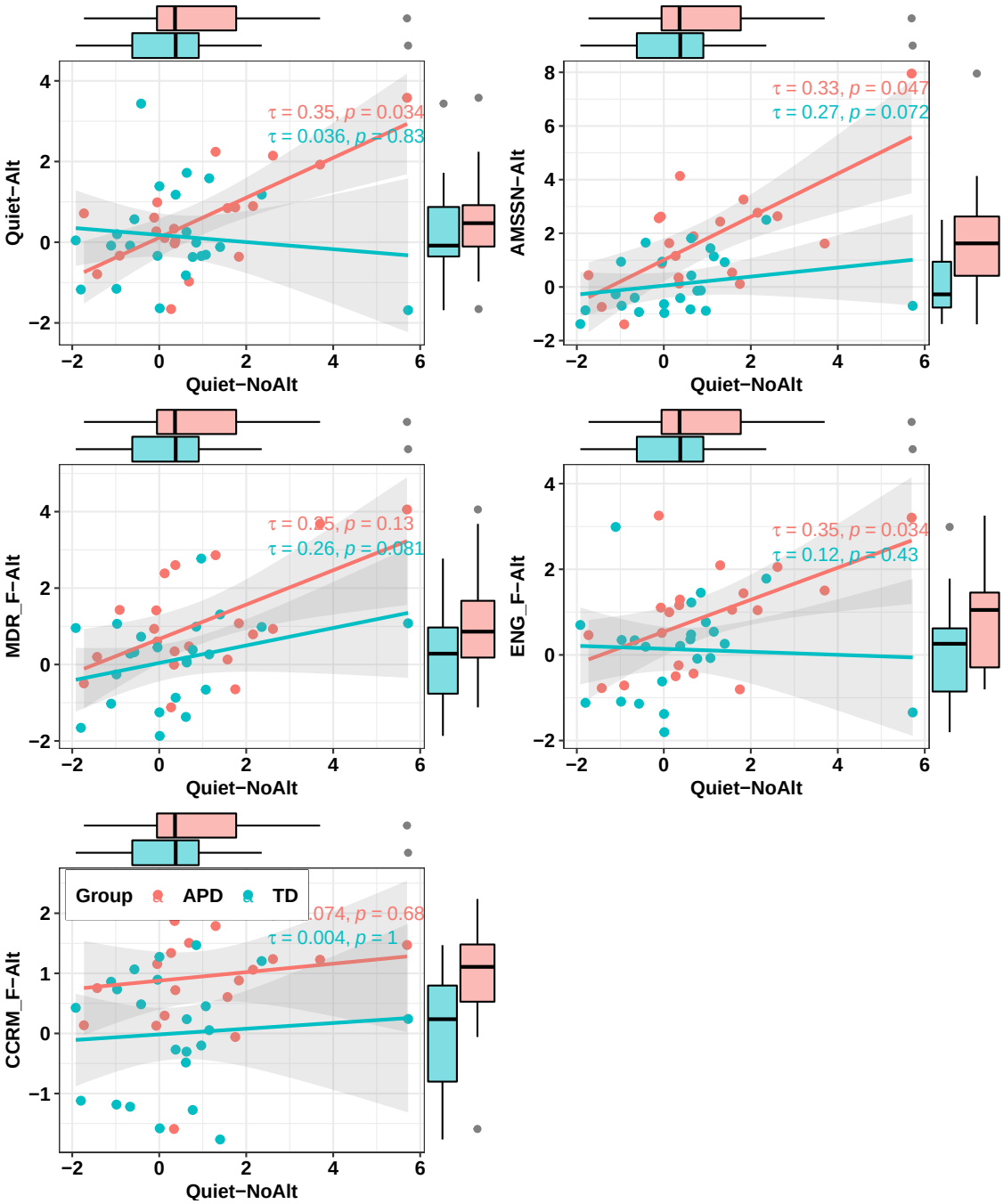

Figure 3.9: Add caption here.

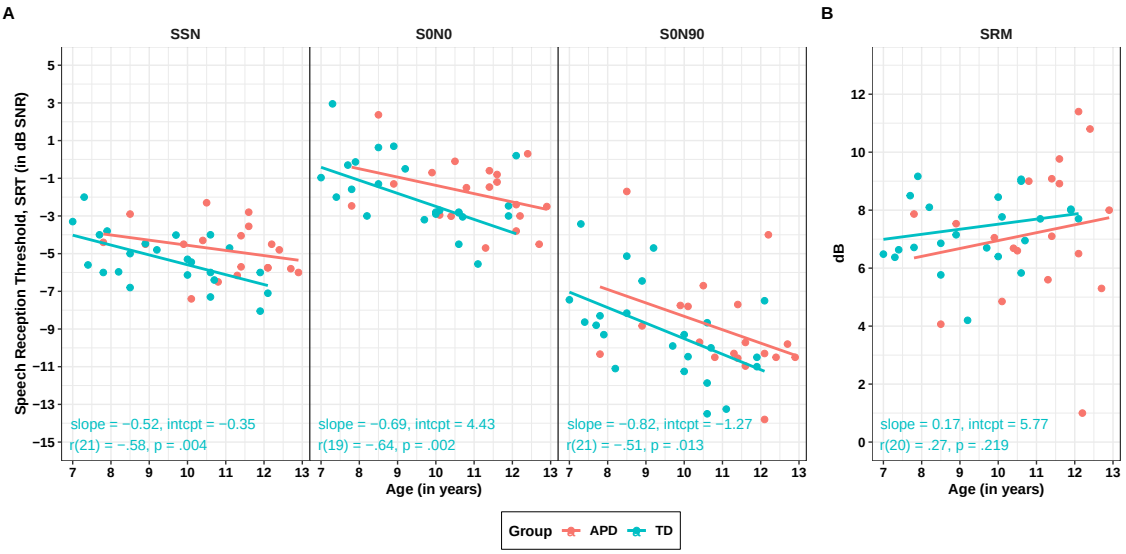

Figure 3.10: Scatterplot and linear regression lines for the LiSN-S UK and SPIN SRTs (A) and the derived measure SRM (B) as a function of the listeners age. Red indicates data from the APD group and green indicates data from the TD control group. The grey area represents the regression line 95 % confidence interval.

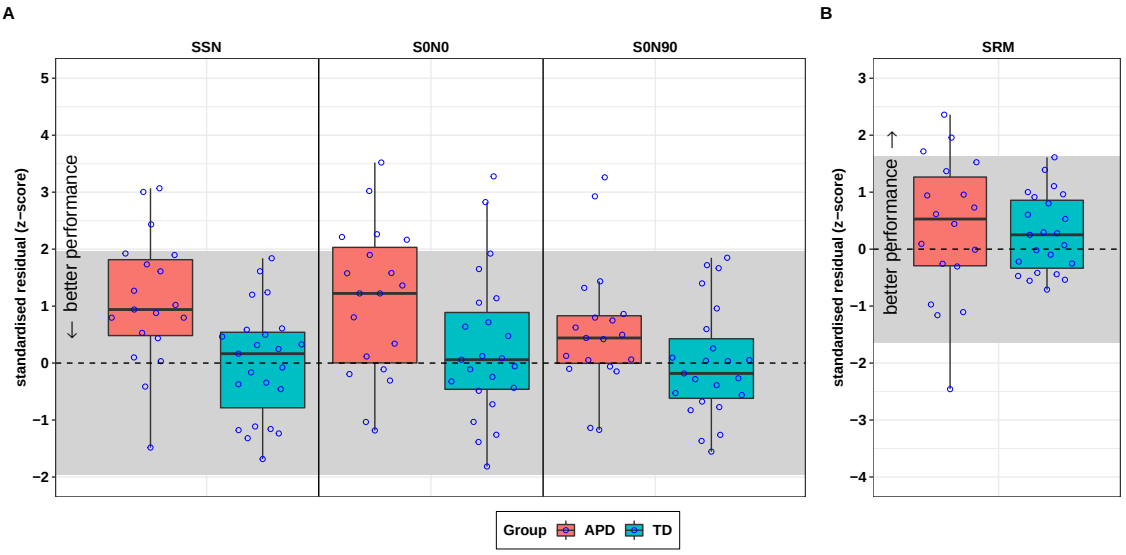

Figure 3.11: Add caption here.

spatially collocated (S0N0) and the spatially separated (S0N90). Participants were also presented with a non-spatialised condition, where the same speech material was presented in a speech-shaped-noise (SSN), resulting in a total of three SRTs obtained for each listener. Additionally, the listeners SRM, calculated by taking the individual's difference between scores in the two spatialised conditions (SRM = S0N0 - S0N90) is shown in Figure 3.10 B.

As previously reported by other LiSN-S test versions (e.g., Cameron et al., ,2007; REFs), the figures suggests a developmental trend in performance, with an overall improvement in performance with age. S0N90 showed the strongest age effect, with near to 1 dB improvement in performance per 1 year increase (TD slope = -0.84). The regression lines slope for S0N0 and SSN conditions where shallower, with roughly half a dB improvement in performance per 1 year increase (with TD slope of -0.51 and -0.47, respectively). Difference in performance with age for the SRM derived measure was negligible, with a predicted improvement in SRM of circa 1 dB between the age of 7 and 13 years.

Pearson correlation

Pearson's (ρ) correlation test confirmed the trend seen in the scatterplots, where all three test condition (SSN, S0N0 and S0N90) showed a significant moderate correlation between SRTs and age for the trimmed TD group, whereas no significant correlation was found for the SRM derived measure.

LMEM

A three-way 2 x 2 factorial design with repeated measures was used to test the main effects of Group (APD/TD), Condition (SON0/S0N90) and Age with SRTs as dependent variable. Interaction terms were included as well as a random intercept for subects. Note that only the spatialised test conditions were included. Since the parametric model assumptions (normal distribution and homogeneity of the variance) were met, a parametric approach was selected based on a linear mixed-effects regression model (LMEM'), using the **lmer()** function (lme4 package, REF). The model that was found to give the best fit and main effects are giving in Table 3.6. Model selection was based on a backward model selection procedure

Table 3.6: Add caption here.

SRT ~ Condition + Group + Age + (1 | Subjects)			
Main effects	Df	χ^2	p
Condition	1	140.373	***<0.001***
Group	1	2.756	***0.097***
Age	1	11.688	***<0.001***

* significance level = 0.05

using a likelihood ratio test (χ^2). The final model did not include interaction terms and thus indicating that the two groups behaved in a similar way across the two test conditions. The model revealed a highly significant main effect of Age and Condition as well as a strong main effect of Group.

Age-independent z-scores

Boxplots of the listeners age-corrected standardised residuals z-scores (blue circles) collapsed across the different test conditions are shown in Figure 3.11, separately for the APD group (red) and TD group (green). The z-scores were calculated in the exact same way as for the ST task. Again, the dashed lines indicate the TD group mean (~ 0), and the grey area indicate the lower and upper limit of normal population scores (TD mean $\pm$ 1.96). Descriptive statistics collapsed by group and test conditions are given in Table 3.5. Overall, when compared with the control group, the APD children exhibited poorer performance (i.e., higher z-score) across all three test conditions as well as for the derived SRM measure.

SSN and S0N0 test conditions yielded the largest separation between the groups, however the spread in scores was relatively large and the percentage of abnormal performance in the APD group was small, with circa 16% (3/19) and 26% (5/19), respectively. Whereas, only about 10% of the APD children exhibited an abnormal score for S0N90 test condition and for the SRM measure.

LMEM model

Differences between the groups for the spatialised conditions S0N0 and S0N90 was tested for a 2 x 2 factorial design LMEM model was used with z-score as

Table 3.7: Add caption here.

z ~ Condition + Group + (1 | Subjects)			
Main effects	Df	χ^2	p
Condition	1	5.134	***0.023***
Group	1	4.518	***0.034***

* significance level = 0.05

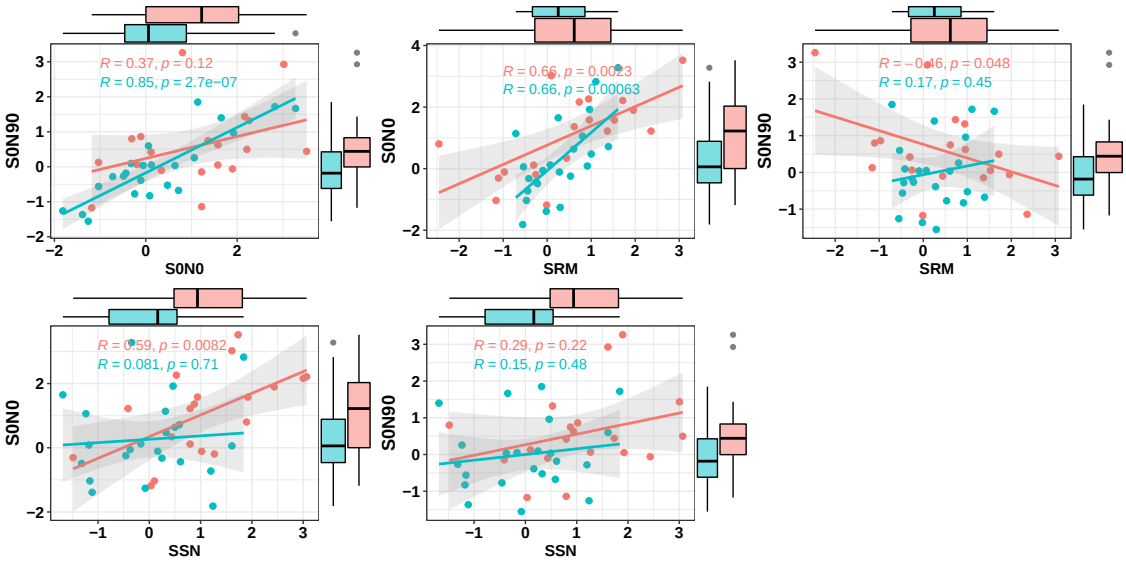

Figure 3.12: Add caption here.

a dependent variable and random intercept for subjects. Model assumptions for normal distribution and homogeneity of variance was verified. The model that was found to give the best fit did not include a two-way interaction term between the main effects Group and Condition, thus suggesting again that the two groups behaved in a similar way in the two test conditions (see Table 3.7. Comparison of the full model with a simplified model without each of the predictors revealed a significant effect of both Condition and Group.

Correlation between conditions (z-score)

comparison with clinical data?

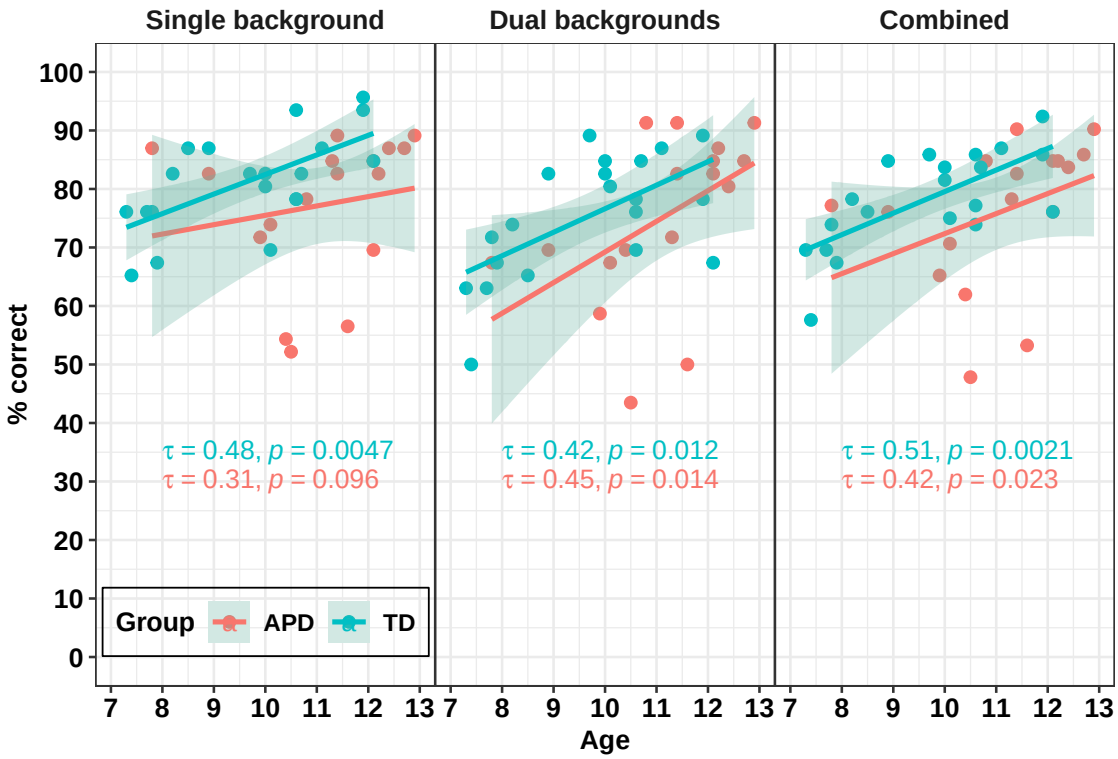

Figure 3.13: Scatterplot and linear regression lines for the ENVASA task %-correct as a function of age for single background, dual background and combined measure. Red indicates data from the APD group and green indicates data from the TD control group. The grey area represents the regression line 95 % confidence interval.

3.3.5 The Environmental Auditory Scene Analysis task (ENVASA)

Initial inspection of the individuals performance was performed to ensure that the task instructions were followed and well understood. Performance for the reference condition (single incongruent background at a high SNR), which is expected to least impact performance, was compared with a cut-off criterion of 56 %, calculated as 2 standard deviations from the TD group mean (84 % ± 14 %). Individuals with performance bellow the cut-off criterion were excluded from the analysis. One TD group aged 7 years old who scored 45 % was thus excluded. Of the remaining data 20 belonged to the TD group and 17 to the APD group.

Due to the small number of observations in each condition and for clarity, %-correct was calculated for three measures: i. a *single background*, ii. a *dual backgrounds*, iii. and a *combined* measure where scores in both conditions were

summed together[2]. Visual inspection of the listeners scores as a function of age (see Figure 3.13) showed a developmental trend in performance. This was suppurated by Kendall rank correlation test for non-parametric data with small sample-size, where performance significantly improved with increased age across the groups in all three measures ($p < .05$), excluding the APD group for single background ($p = 0.096$; see the complete test results in Figure 3.13). This is in agreement with Krishnan et al. (2013) study where they found a strong developmental effect across normal-hearing typically developing children in a similar age range to those measured in the present study.

Thus, for further analysis, age was controlled for using the same multiple-case approach method described in ????. Age-independent score was predicted using a linear regression model which was performed separately for each measure. The model was estimated from the TD data only with %-correct as the dependent variable and age as a predictor. Standard residuals were calculated for each listener, resulting in age-independent residuals that are comparable to z-scores for data with normal distribution, with a mean and SD of approximately 0 and 1, respectively. Furthermore, to reduce possible bias, outliers in the TD group were identified and removed from the model, while they were included back again for the calculation of the standardised residuals. The outliers criteria was set to TD scores more than 1.65 SD below the TD group mean. Boxplots of the age-independent z-scores for the three measures are shown in Figure 3.14. The red dashed line ($z = -1.64$) indicate the cut-off value for normal score, where only about 5 % of the normal population would be expected to score below it; the green dashed line ($z = +1.64$) indicate scores above which only 5 % of the normal population is expected to exceed.

Group difference was analysed using Wilcoxon rank-sum test for independent two samples test which is a t-test equivalent for non-parametric data [N = 999999 permutations and with bonferroni p-value adjustment; 'coin::wilcox_test()', REF]. Effect size (r) was analysed using 'rstatix::wilcox_effsize' (REF). Descriptive scores

[2]See Appendix ?? for a more detailed summary of the listeners scores split into the different test condition.

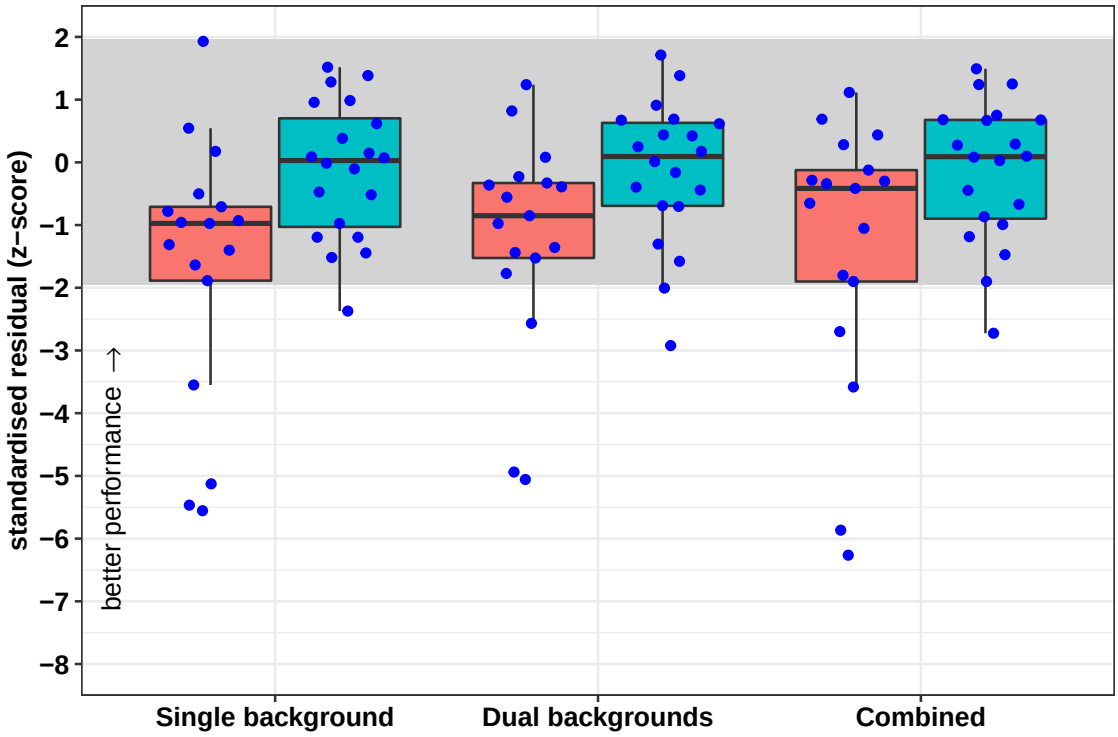

Figure 3.14: ENVASA standardised residuals for the %-correct in single background, dual background and combined measure. The dashed red line (z = -1.64) indicate the cut-off value for normal score, where only about 5 % of the normal population would be expected to score below it; the dashed green line (z = +1.64) indicate scores above which only 5 % of the normal population is expected to exceed.

Table 3.8: ENVASA standard residuals (z-scores) descriptive statistics and group difference analysis by test measures.

	TD					APD					Wilcoxon rank-sum test			
background	N	median	SD	min	max	N	median	SD	min	max	CI	p	effect size (r)	magnitude
single	20	0.03	1.08	-2.37	1.52	17	-0.97	2.11	-5.56	1.93	-2.27 - -0.34	*0.02*	*0.39*	moderate
dual	20	0.09	1.15	-2.92	1.71	17	-0.85	1.71	-5.06	1.24	-1.75 - -0.09	*0.04*	*0.34*	moderate
combined	20	0.09	1.13	-2.73	1.49	17	-0.42	2.15	-6.26	1.11	-1.80 - 0.10	*0.09*	*0.28*	small

by group collapsed by the three test measures as well as p statistics and effect size are given in Table 3.8. Performance of children in the APD group was significantly poorer than of the children in the TD group for both single and combined measure ($p < 0.05$, moderate effect size), whereas no significant difference in performance between the groups was found for the dual backgrounds measure ($p = 0.118$).

- Correlation with other measures

For appendix:

- similar figure to Kirshnan's figure with % correct & scores by age & z-scores

3.3.6 CELF-RS

```
##     item group1 vars  n  mean   sd median trimmed  mad min max range skew
## X11    1    APD    1 21  9.48 2.68      9    9.24 1.48   6  15     9 0.76
## X12    2     TD    1 23 13.61 3.14     13   13.58 2.97   7  19    12 0.07
##     kurtosis   se
## X11    -0.45 0.58
## X12    -0.77 0.66
##
##  Shapiro-Wilk normality test
##
## data:  residuals(w1)
## W = 0.95096, p-value = 0.05976
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  0.8731 0.3554
##       42
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  0.8731 0.3554
##       42
```

Boxplots of children's age-corrected scaled scores split into groups for the CELF-5 UK Recalling Sentences subtest are given in Figure 3.15. The grey area indicate the upper and lower norms limit among the normal population (± 1 SD).

On average, performance was within the norms range in both the APD group (median = 9) and the TD group, albeit laying within the upper norms limit (median = 13). Thus, although the majority of the APD children expressive language skills were within the norms, the figure reveals a clear difference in

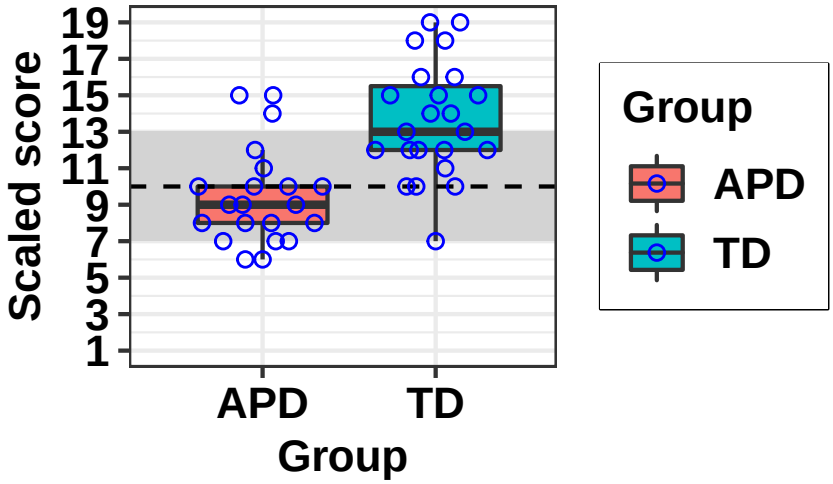

Figure 3.15: Boxplots for CELF-5 UK Recall Sentences subtest scaled scores by groups. The dashed line represents the norms mean and the grey area indicate the upper and lower limit of an average performance in the normal population (± 1 SD).

performance between the group, where the TD children expressive language skills are noticeably better than for children in the APD group.

Almost half of the TD children obtained a scaled score above the average and none exhibited abnormal scores. On the other hand, only three APD children performed above the average and performance of two children was considered to be abnormal (scaled score = 6).

An independent-samples t-test with bootstrapping (n=9999) was computed using **boot.t.test()** function (MKinfer package, REF) to compare the listeners scaled scores in the two groups (parametric data assumptions were met). There was a highly significant difference in scaled scores between the APD (M = 9.5, SD = 2.7) and TD group (M = 13.6, SD = 3.1) [t(41.81) = -4.71, p = < 0.001].

3.3.7 Questionnaires

CCC-2

data for on listener (TD13) was flagged as inconsistent and thus was removed.

What is the definition of abnormal score? 2 abnormal subcategories? Abnormal score == GCC < 55 GCC<55 + a negative SIDC is indicative of ASD. 8 APD children were flagged as possible ASDs.

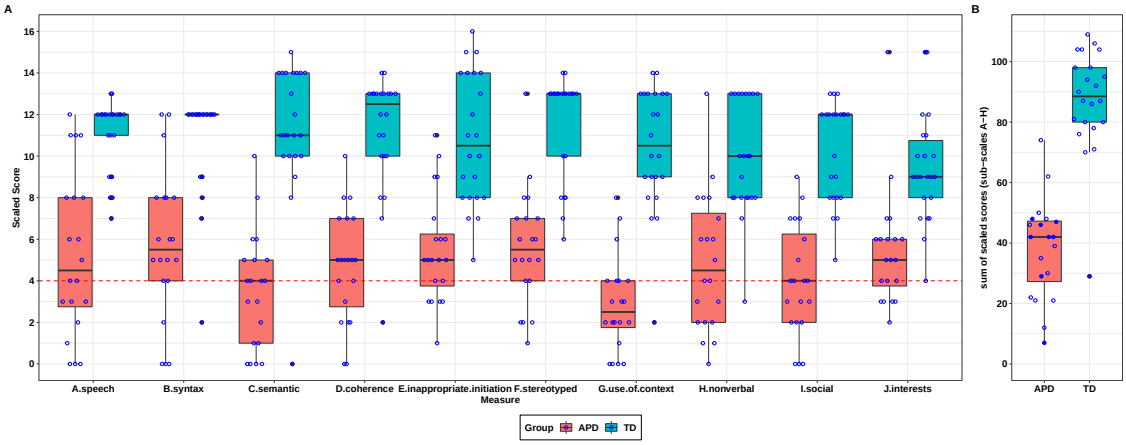

Figure 3.16: Add caption here.

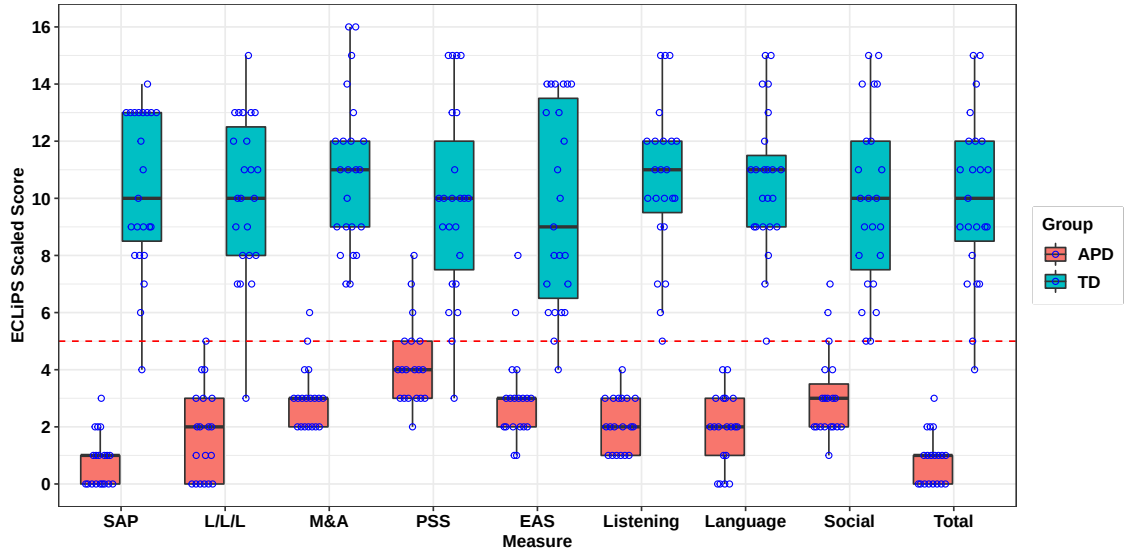

Figure 3.17: Add caption here.

3 out of 8 were diagnosed with ASD, additional 2 were undergoing an ASD assessment at the time of testing.

Again, despite our best effort to recruit APD children without

ECLIPS

3.4 Discussion

3.5 Conclusion

Alles Gescheite ist schon gedacht worden.
Man muss nur versuchen, es noch einmal zu denken.

All intelligent thoughts have already been thought;
what is necessary is only to try to think them again.

— Johann Wolfgang von Goethe (von Goethe, 1829)

General discussion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

Summary of main findings

Conclusion

Appendices

A

The First Appendix

This first appendix includes an R chunk that was hidden in the document (using `echo = FALSE`) to help with readibility:

In 02-rmd-basics-code.Rmd

```
library(tidyverse)

knitr::include_graphics("figures/chunk-parts.png")
```

And here's another one from the same chapter, i.e. Chapter ??:

B

The Second Appendix, for Fun

Works Cited

Algazi, V. R., Duda, R. O., Thompson, D. M., & Avendano, C. (2001). The CIPIC HRTF database. *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*. https://doi.org/10.1109/aspaa.2001.969552

Barry, J. G., & Moore, D. R. (2014). *Evaluation of Children's Listening and Processing Skills (ECLiPS)* (tech. rep.). MRC-T. London, United Kingdom.

Bench, J., Kowal, Å., & Bamford, J. (1979). The Bkb (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *British Journal of Audiology*, *13*(3), 108–112. https://doi.org/10.3109/03005367909078884 doi: 10.3109/03005367909078884

Cameron, S., & Dillon, H. (2007). Development of the Listening in Spatialized Noise-Sentences Test (LISN-S). *Ear and Hearing*, *28*(2), 196–211. https://doi.org/10.1097/AUD.0b013e318031267f

Cameron, S., Glyde, H., & Dillon, H. (2011). Listening in Spatialized Noise—Sentences Test (LiSN-S): Normative and Retest Reliability Data for Adolescents and Adults up to 60 Years of Age. *Journal of the American Academy of Audiology*, *22*(10), 697–709. https://doi.org/10.3766/jaaa.22.10.7

D. V. M., B. (2003). *The Children's Communication Checklist, Version 2 (CCC-2)* (tech. rep.). The Psycho- logical Corporation. London, United Kingdom.

H Wiig, E., Semel, E., & Secord, W. (2017). *Clinical Evaluation of Language Fundamentals - Fifth Edition UK (CELF-5UK)* (tech. rep.). PsychCorp, Pearson Clinical Assessment.

Krishnan, S., Leech, R., Aydelott, J., & Dick, F. (2013). School-age children's environmental object identification in natural auditory scenes: Effects of masking and contextual congruence. *Hearing Research*, *300*, 46–55. https://doi.org/10.1016/j.heares.2013.03.003

Leech, R., Gygi, B., Aydelott, J., & Dick, F. (2009). Informational factors in identifying environmental sounds in natural auditory scenes. *The Journal of the Acoustical Society of America*, *126*(6), 3147–3155. https://doi.org/10.1121/1.3238160

MacLeod, A., & Summerfield, Q. (1990). A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology*, *24*(October), 29–43. https://doi.org/10.3109/03005369009077840

Norbury, C. F., & Bishop, D. V. M. (2005). Children ' s Communication Checklist - 2 : a validation study. *Publie dans Revue Tranel*, *42*, 53–63.

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain*, *126*(4), 841–865. https://doi.org/10.1093/brain/awg076

von Goethe, J. W. (1829). *Wilhelm Meisters Wanderjahre oder die Entsagenden*. Cotta.

Wierstorf, H., Geier, M., Raake, A., & Spors, S. (2011). A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances. *AES130*.