Exploring the Neuroanatomical and Behavioural Correlates of Foreign Language Aptitude

Dissertation

zur Erlangung des Doktorgrades der Philosophie
an der Karl-Franzens-Universität Graz

eingereicht von Mag. phil. Sabrina Turker BA

am Institut für Anglistik

Erstbegutachterin: A.o. Univ.-Prof. Dr. Annemarie Peltzer-Karpf
Zweitbegutachterin: Assoz.-Prof. Dr. Susanne Maria Reiterer

2019
Acknowledgements

Firstly, I would like to express my sincere gratitude to my two official supervisors, Dr. Annemarie Peltzer-Karpf and Dr. Susanne Reiterer, and my two equally important unofficial supervisors and collaboration partners, Dr. Annemarie Seither-Preisler and Dr. Peter Schneider. These four outstanding researchers have supported and inspired me with their motivation and dedication over the past years while I completed my PhD studies. My dearest thanks goes to Dr. Reiterer – without even knowing me, she agreed to help me design a suitable project and provided me with data from one of her own studies and introduced me to Dr. Seither-Preisler and Dr. Schneider. Without her, I would never have had the opportunity to engage in the neuroscientific research world at all and I will remain thankful for that for the rest of my life. Thank you, Susanne!

Furthermore, I am truly indebted to Peter Schneider, his family, and his remarkable research group ‘Musik und Gehirn’ (Music and Brain) at the Department for Neuroradiology of the University Hospital Heidelberg, Germany. They have become like a second family to me and I will certainly keep the numerous hilarious and life-enriching moments of the past years as wonderful memories. Therefore, a big, big thank you for their valuable help goes to Bettina Zeidler, Kamila Jarosch, Jan Benner, Dorte Engelmann, Vita Schneider and Alexander Ludwig. Of course, I also owe thanks to the participants of my studies, in particular those of the longitudinal study, and all professional colleagues involved in the whole publishing, editing and reviewing processes.

Moreover, I am deeply indebted to the University of Graz, my home institution, that provided me with several scholarships for doctoral students to partly finance my studies and my research stays in Germany. The next main contributor to my PhD project is the Austrian Academy of Sciences who granted me the most valuable scholarships of all, a DOC-team fellowship. In this context, I would also like to thank Sabine Sommer-Lolei and Markus Christiner for asking me to become part of their DOC-team, and Univ.-Prof. Dr. Dressler for supporting the three of us. I would never have expected to finally receive a long-lasting scholarship that would really allow me to focus on my PhD thesis for more than two years. I would also like to thank Richard Parncutt and the Centre for Systematic Musicology for letting me become part of the team in the past two years.
Finally yet importantly, it would have been impossible to finish this PhD degree without the help of my simply wonderful, loving boyfriend Klaus and my caring family, including my lovely grandparents. Even in difficult times, they motivated and supported me personally and financially. Without them, I would not be the person I am now, and I love them with all my heart. This is as much their achievement as it is mine and I cannot say in words what they mean to me.

Graz, June 2019

(Recipient of a DOC-team scholarship of the Austrian Academy of Sciences 2017-2019)
Abstract

The aim of the present thesis was twofold. The first aim was to find out whether auditory cortex morphology is a potential neuroanatomical marker of foreign language aptitude. The second aim was to investigate the relationships between language aptitude, musicality and working memory in both adults and children. In the younger population, the relationships between intelligence, socio-economic status, arithmetic skills and language aptitude were further investigated.

Behavioural and neuroanatomical analyses were performed in two studies with different age groups, namely children and teenagers on the one hand, and adults on the other hand. On the neural level, both studies revealed a strong relationship between language aptitude and right-hemispheric auditory cortex morphology. Adults and children with high language aptitude scores possessed more complete posterior duplications in their right auditory cortex. Conversely, subjects with low aptitude more frequently possessed single gyri in their right auditory cortex. On the behavioural level, significant correlations between working memory and language aptitude were found, while a weak link between musical experience and language aptitude could only be found in the adults. In the child population, scores on the arithmetic competence test and school performance correlated with language aptitude, which was further found to be independent of general intelligence and socio-economic status.

To conclude, the findings suggest an intricate link between right auditory cortex morphology and language aptitude, so that a higher number of gyri and higher grey matter volumes in right auditory cortex seem to present neuroanatomical markers of high language aptitude. Additionally, language aptitude was not only significantly associated with working memory capacity, but it was linked to musicality (in adults) and arithmetic skills, as well as to school achievement, in the younger population.
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Auditory cortex</td>
</tr>
<tr>
<td>ADHD</td>
<td>Attention deficit hyperactivity disorder</td>
</tr>
<tr>
<td>AG</td>
<td>Angular gyrus</td>
</tr>
<tr>
<td>AMMA</td>
<td>Advanced Measures of Music Audiation</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>CANAL- F</td>
<td>Cognitive Ability for Novelty in Acquisition of Language - Foreign</td>
</tr>
<tr>
<td>CPD</td>
<td>Complete posterior duplication</td>
</tr>
<tr>
<td>CSD</td>
<td>Common stem duplication</td>
</tr>
<tr>
<td>CT</td>
<td>Computer tomography</td>
</tr>
<tr>
<td>CV(C)</td>
<td>Consonant-vowel(-consonant)</td>
</tr>
<tr>
<td>DLAB</td>
<td>Defense Language Aptitude Battery</td>
</tr>
<tr>
<td>DLPFC</td>
<td>Dorsolateral prefrontal cortex</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography</td>
</tr>
<tr>
<td>FDR</td>
<td>False discovery rate</td>
</tr>
<tr>
<td>FLL</td>
<td>Foreign language learning</td>
</tr>
<tr>
<td>fMRI</td>
<td>Functional magnetic resonance imaging</td>
</tr>
<tr>
<td>HG</td>
<td>Heschl’s gyrus</td>
</tr>
<tr>
<td>Hi-LAB</td>
<td>High-level Language Aptitude Battery</td>
</tr>
<tr>
<td>IFG</td>
<td>Inferior frontal gyrus</td>
</tr>
<tr>
<td>IMMA</td>
<td>Intermediate Measures of Music Audiation</td>
</tr>
<tr>
<td>IPL</td>
<td>Inferior parietal lobe</td>
</tr>
<tr>
<td>GM</td>
<td>Grey matter</td>
</tr>
<tr>
<td>L1</td>
<td>First/Native language</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>L2</td>
<td>Second language</td>
</tr>
<tr>
<td>LTM</td>
<td>Long-term memory</td>
</tr>
<tr>
<td>MD</td>
<td>Multiple duplications</td>
</tr>
<tr>
<td>MEG</td>
<td>Magnetencephalography</td>
</tr>
<tr>
<td>MLAT</td>
<td>Modern Language Aptitude Test</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>MTG</td>
<td>Middle temporal gyrus</td>
</tr>
<tr>
<td>PAC</td>
<td>Primary auditory cortex</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PET</td>
<td>Positron emission tomography</td>
</tr>
<tr>
<td>PLAB</td>
<td>Pimsleur’s Language Aptitude Battery</td>
</tr>
<tr>
<td>SLA</td>
<td>Second language acquisition</td>
</tr>
<tr>
<td>SMG</td>
<td>Supramarginal gyrus</td>
</tr>
<tr>
<td>STG</td>
<td>Superior temporal gyrus</td>
</tr>
<tr>
<td>STM</td>
<td>Short-term memory</td>
</tr>
<tr>
<td>STS</td>
<td>Superior temporal sulcus</td>
</tr>
<tr>
<td>TDCS</td>
<td>Transcranial direct current stimulation</td>
</tr>
<tr>
<td>TMS</td>
<td>Transcranial magnetic stimulation</td>
</tr>
<tr>
<td>Tukey HSD</td>
<td>Tukey honest significant difference</td>
</tr>
<tr>
<td>PT</td>
<td>Planum temporale</td>
</tr>
<tr>
<td>VBM</td>
<td>Voxel based morphometry</td>
</tr>
<tr>
<td>WM</td>
<td>Working memory</td>
</tr>
</tbody>
</table>
# Table of Contents

1. **INTRODUCTION** .................................................................................................................................................. 1

   1.1. **The Main Aims of the Present Thesis** ................................................................................................. 3

   1.2. **The Structure of the Present Thesis** .................................................................................................... 6

2. **The Multifaceted and Complex Construct of Language Aptitude** ................................................................. 9

   2.1. **What is ‘Aptitude’?** ................................................................................................................................ 10

       2.1.1. **A Neurocognitive Model of Aptitude and Learning** ................................................................. 14

   2.2. **A Brief History of Language Aptitude** ................................................................................................. 16

   2.3. **Defining the Concept of Language Aptitude** ....................................................................................... 18

   2.4. **Components of Language Aptitude** ...................................................................................................... 21

       2.4.1. **Working Memory and Language Aptitude** ............................................................................... 27

       2.4.2. **Aptitude vs. Aptitudes** ........................................................................................................... 28

   2.5. **Language Aptitude Tests** ................................................................................................................ 29

       2.5.1. **The Modern Language Aptitude Test (MLAT)** ........................................................................... 30

       2.5.2. **The LLAMA Language Aptitude Battery** .............................................................................. 33

       2.5.3. **The Hi-LAB (High-Level Language Aptitude Battery)** .......................................................... 38

       2.5.4. **Other Tests Designed to Measure Language Aptitude** ......................................................... 38

       2.5.4.1. **The Hindi Speech Imitation Test** ....................................................................................... 40

       2.5.4.2. **The MULT/AP** ................................................................................................................... 41

3. **Language Aptitude and the Brain** ............................................................................................................... 43

   3.1. **An Introduction to Methods of the Cognitive Neurosciences** ........................................................... 44
3.1.1. Structural Neuroimaging Methods .................................................................45
3.1.2. Functional Neuroimaging Methods ..............................................................46

3.2. FOREIGN LANGUAGE LEARNING FROM A NEUROSCIENTIFIC PERSPECTIVE ....47

3.3. THE HUMAN BRAIN .......................................................................................49
3.3.1. A Terminological Introduction ......................................................................49
3.3.2. The Development of the Human Brain .........................................................51
3.3.3. Neuroplasticity .............................................................................................52
3.3.4. The Core Language Regions in the Human Brain .........................................53

3.4. HESCHL’S GYRUS AND THE AUDITORY CORTEX ........................................57
3.4.1. Research on Heschl’s Gyrus and the Human Auditory Cortex .......................61
3.4.1.1. Functions of Right and Left Heschl’s Gyrus .................................................62
3.4.1.2. Specific Structural Properties of Heschl’s Gyrus ..........................................65

4. THE RELATIONSHIPS BETWEEN WORKING MEMORY, MUSICALITY AND
LANGUAGE APTITUDE .........................................................................................69

4.1. THE ROLE OF WORKING MEMORY FOR LANGUAGE APTITUDE ..................69
4.1.1. Now and Then: The Development of Working Memory Models....................70
4.1.1.1. Baddeley and Hitch’s Working Memory Models ...........................................71
4.1.1.2. A Terminological Clarification: Short-Term Memory vs. Working Memory ..........76
4.1.2. The Phonological-Executive Model of Working Memory for Language Aptitude ......78
4.1.3. Working Memory and the Brain .....................................................................79
4.1.4. Working Memory as an Important Predictor for Language Aptitude .............82

4.2. THE RELATIONSHIP BETWEEN MUSIC AND LANGUAGE APTITUDE ........86
4.2.1. Finding a Definition for Musicality / Musical Ability ......................................87
4.2.2. Can Musical Skills or Musical Training Enhance Foreign Language Learning? ....89
4.2.2.1. Music and Speech Perception .....................................................................90
4.2.2.2. Music and Speech Production ......................................................................92
4.2.3. Neural Changes Due to Musical Training and Experience ............................94
5. STUDY 1: RIGHT AUDITORY CORTEX MORPHOLOGY IS LINKED TO LANGUAGE APTITUDE, SPEECH IMITATION AND MUSICALITY ................................. 97

5.1. METHODS ............................................................................................................. 97

5.1.1. Subjects .............................................................................................................. 97

5.1.2. Language Aptitude Testing .................................................................................. 98
  5.1.2.1. An English Pronunciation Assessment ......................................................... 98
  5.1.2.2. The Hindi Speech Imitation Task ................................................................. 99
  5.1.2.3. The Modern Language Aptitude Test (MLAT) .............................................. 100

5.1.3. Musicality Assessment ....................................................................................... 100

5.1.4. Working Memory Capacity .............................................................................. 101


5.2. RESULTS ............................................................................................................ 106

5.2.1. Behavioural Results ......................................................................................... 106
  5.2.1.1. Descriptive Statistics .................................................................................... 106
  5.2.1.2. Results of the Correlational Analyses ......................................................... 108
  5.2.1.3. Results of the Principal Component Analysis ............................................. 110
  5.2.1.4. Comparisons of High and Low Aptitude Learners ..................................... 114

5.2.2. Neuroanatomical Results .................................................................................. 118

5.3. DISCUSSION ..................................................................................................... 126

5.3.1. Auditory Cortex Morphology as a Potential Marker for Language Aptitude and Musicality ........................................................................... 126

5.3.2. The Concept of Language Aptitude .................................................................. 131

5.3.3. Musicality and its Link to Language Aptitude ................................................ 134

5.3.4. Working Memory Capacity as a Predictor of Language Aptitude ................ 136

6. STUDY 2: RIGHT-HEMISPHERIC AUDITORY CORTEX MORPHOLOGY PREDICTS FOREIGN LANGUAGE APTITUDE IN CHILDREN AND TEENAGERS .................. 139

6.1. METHODS .......................................................................................................... 140
7.1.1. Summary of the Findings of Study 1 and 2 ................................................................. 176
7.1.2. The Significance of the Auditory Cortex for Language Aptitude ............................... 178
7.1.3. Functions of Right and Left Auditory Cortex ............................................................ 181
7.1.4. Individual Variation of Heschl’s Gyrus and the Significance of Gyrification Patterns .... 184
7.1.5. Duplications – Primary or Secondary Auditory Areas? ................................................ 188
7.1.6. Right Auditory Cortex and Musicality ........................................................................ 191
7.1.7. Auditory Cortex Morphology as a Stable Neuroanatomical Trait .............................. 193
7.2. The Concept of Language Aptitude Revisited ............................................................. 195
7.3. Behavioural Correlates of Language Aptitude ........................................................... 199
    7.3.1. The Relationship between Music and Language Aptitude ...................................... 200
        7.3.1.1. Replication Crisis and Publication Bias ............................................................... 205
    7.3.2. Working Memory as the Most Important Correlate of Language Aptitude .......... 207
        7.3.2.1. A New Concept of Language Aptitude? .............................................................. 211
    7.3.3. School Performance, Arithmetic Skills and Self- and Parent-Reported Aptitude .... 213
7.4. Implications of the Present Studies for Education and Science ............................. 216

8. CONCLUSION ..................................................................................................................... 220

9. REFERENCES ..................................................................................................................... 222

LIST OF FIGURES ................................................................................................................. 269

LIST OF TABLES ..................................................................................................................... 274
1. Introduction

The simple fact that some individuals appear to be much better at doing something, e.g., singing, playing the piano or learning a foreign language, has always intrigued people. Adding to the dilemma, the majority of these so-called ‘more gifted’ people put fairly little effort into acquiring a certain skill and need very little time to achieve a high proficiency or mastery level. In comparison to age-matched peers, their achievement seems outstanding given the ease of their progress and the little effort required. Regardless of whom you ask, almost everybody will agree that some people appear to be more gifted in certain domains for no obvious reason. Although these striking differences in giftedness or talent undoubtedly exist, few elaborate hypotheses have been brought forward to scientifically explain their origin. In the past decades, however, the notions of talent and aptitude have gained considerable momentum and the topic was addressed from different viewpoints. Since then, researchers have tried to provide a scientific, theoretical foundation of the vague concepts of aptitude and talent going beyond the simple and unsatisfying facts as stated above, e.g., quick learning without much effort.

Human beings possess traits, such as eye colour or size, that are inherited, that is biologically determined from birth on. We cannot influence how tall we grow or whether we have blue or brown eyes. The acquisition of specific skills, like learning a foreign language, on the other hand, cannot be easily attributed to genetics or the environment only. Doubtlessly, a variety of factors contribute to the overall achievement and proficiency of skills, i.e. both nature and nurture are necessary for their development. Environmental factors (e.g., educational background, social status), personality factors (e.g., intra-/extraversion, motivation), and other cognitive, biological factors (e.g., working memory, possible developmental impairments) largely affect an individual’s success in foreign language learning. Still, these internal and external factors fail to explain the full range of diversity encountered. The genetic component of giftedness, and thus also language aptitude, is undeniable and it is highly likely that language aptitude is at least partly biologically determined.

Language aptitude is a core factor influencing foreign language acquisition and due to its centrality in the present research project, it will be dealt with extensively in this thesis. From
a behavioural point of view, this thesis aims to explore the relationship between working memory, musicality and language aptitude in two different age groups, namely children and teenagers on the one hand, and adults on the other hand. A large body of research has addressed these cognitive abilities but hardly any studies have combined them into one study and have addressed their mutual interdependence not only from a behavioural point of view, but also from a neuroscientific one.

Thanks to the technological advances and inventions of the past decades, it has become possible to investigate the anatomy and functioning of the human brain in living individuals. First only allowed for clinical use, these methods have quickly become available for a larger audience including universities and research institutions all over the world. Nowadays, numerous disciplines, like medicine, psychology or linguistics, benefit largely from the now available neuroimaging methods for research purposes as well as for clinical assessment. Additionally, new disciplines, such as neurolinguistics or cognitive neuroscience, have established themselves during the so-called neuroscientific revolution and they attempt to integrate new research methods into older disciplines such as linguistics or psychology. Research on foreign language learning from a neuroscientific perspective has become increasingly popular in the past 25 years. Yet, few studies to date have addressed the neural underpinnings of a possibly innate foreign language learning potential, which is surely dependent on both the structure and functional connectivity of the brain and specific regions therein. Exploring the brains of individuals with a certain giftedness or an outstanding ability when it comes to learning a foreign language certainly is a unique opportunity to study the neurocognitive mechanisms and structures underlying this exceptional ability.

To sum up, in the present thesis I aim to shed further light on the interplay between musical abilities, language aptitude and working memory in adults and children, while further extending the behavioural research to school performance and arithmetic abilities in the younger population. On the neural level, I want to investigate a possible link between the morphology of auditory cortex, an area essential for the auditory processing of speech and language, and an individual’s potential to acquire foreign languages.
1.1. The Main Aims of the Present Thesis

There are two major reasons why I decided to invest several years of intense research, training and learning into this particular topic, not knowing whether I would end up with positive, publishable results at all. Both reasons can be explained by my personal and academic links to education and language learning.

Language aptitude is a hot topic in educational discourse and while its presence cannot be denied in foreign language classrooms, we know so little about its underlying characteristics and properties. Regarding education and general learning, the previously held belief that children need to be assessed continuously (e.g., obtain grades in all classes with continuous assessment) and need to perform at a certain level at a specific stage has been loosened. At the same time, individual variation has become a popular term in educational research. The aforementioned belief that proficiency or achievement are of central importance in classrooms has been more and more replaced by the view that we need to consider individual strengths and weaknesses, and help children make the most out of the potentials they possess. This was accompanied by the suggestion that each child holds individual talents and gifts and needs to be supported on an individual level to be able to fully develop their skills. While this shift is notably a step in the right direction, it has understandably overwhelmed schools and in particular the teaching staff. Having worked as a teacher myself, I know too well how extremely challenging it is to adapt to the individual needs of each child in classes of up to 30 children. All these changes in the way we deal with learning potential and our beliefs that everyone possesses a potential that needs to be fostered and supported, has led me to become interested in an individual’s (possibly fully predisposed) potential to acquire foreign languages. Being a fast and quite successful language learner myself without having to study much, I have always wanted to know why some people struggle with tasks others find so easy to master. To put it differently, I have always wanted to know why certain individuals are just so strikingly better at acquiring skills than others and whether their gift was biologically determined or acquired during early childhood or adolescence. Linking behavioural traits to language aptitude is only half of the full story and only by combining behavioural research with neuroscientific methods, we can find out what is going on in the brains of individuals with certain talents or aptitudes and what distinguishes them from those without these particular gifts.
The second reason why I pursued a PhD in neurolinguistics is a little more complex. Humans, whether researchers or not, are always looking for the reason why things are not the way we think they should be, or the way things used to be – even if we know that there might not be one single explanation accounting for the individual variation or the changes observed. I am fairly convinced that my need to find the underlying properties of foreign language aptitude arises at least partly from the general human need to get to the bottom of everything, to the nitty gritty so to say. It is only by finding these underlying properties that we might understand the complex aetiology behind everything. Of course, language aptitude is a particularly interesting construct for me since I am both a foreign language teacher and foreign language learner and I have made that clear in the previous paragraph already. You may call it scientific endeavour because I knew from the very beginning of my university studies that I wanted to gain in-depth knowledge far beyond what was offered in the degrees I pursued. Being a researcher is so much more than teaching at university and passing on knowledge, it is about discovery, exploration and your desire to expand not only your horizons but increase your knowledge day by day. Therefore, this PhD thesis is also a result of my motivation to further investigate and explore the world and especially those concepts and abilities that are most interesting for me personally.

The overall aim of my PhD thesis was to expand the current knowledge on the neurobiology of language aptitude by investigating whether auditory cortex morphology can be delineated as a neuroanatomical marker or correlate for high language learning potential, i.e. above-average language aptitude. Research in this domain has been scarce and few studies have addressed the neuroanatomy behind language aptitude compared to a considerably larger number of studies looking into the functional differences. Research has shown that the auditory cortex is responsible for the processing and partly also the decoding of speech and I assumed that it should thus be of central importance for developing outstanding linguistic ability as well. Bearing this in mind, I decided to use an innovative semi-automated manual segmentation method developed by PD Dr. P. Schneider, one of my collaboration partners and supervisors, to investigate small differences in gyral patterns in human auditory cortex to find potential neural markers for language aptitude in the human brain. As a second major aim, I was keen on confirming (or disproving) the intricate relationships between language aptitude, working memory and musicality on the behavioural level. Numerous studies have already confirmed that there is a connection between the aforementioned and overall language learning, but few
have actually addressed their relationship with foreign language aptitude. Another interesting aspect in the younger population of Study 2 was the exploration of the link between language aptitude, academic achievement and arithmetic competence, as well as self-reported language aptitude.

Two studies were conducted to address the aforementioned aspects and relations among concepts related to language aptitude. Study 1 was based on data collected by Susanne M. Reiterer and colleagues at the University of Tübingen in 2009. They undertook an extensive research project investigating musicality, working memory capacity, pronunciation and speech imitation aptitude (subsumed as language aptitude or talent) in combination with fMRI (functional magnetic resonance imaging). Altogether, about 130 German-speaking, monolingual subjects took part in their study (40 were selected for the present analysis), most of them being in tertiary education and all between 20 and 40 years of age at the time of the measurement (results of the overall study were published in a book edited by G. Dogil and S. Reiterer, 2009). For this thesis, data from a subset of participants was re-analysed and the morphology of auditory cortex was investigated and set in relation to the results of the behavioural analyses in this group of participants.

The second study was not a separate study per se, but data was collected as part of a longitudinal study at the University Hospital Heidelberg, Germany. The data presented in this thesis was collected between June 2016 and January 2018. Altogether, I assisted in the longitudinal study, and collected behavioural and brain data from all participants I included in the analyses. For Study 2, healthy children and teenagers between 10 and 16 years of age, all German-speaking and monolingually raised, were included. In the child study, I also included motivational and personal aspects as assessed by reports of the children themselves and their parents. In both studies, the same variables were assessed, namely verbal working memory, speech imitation skills (Hindi test), more general language aptitude and musicality, as well as musical experience. The neuroanatomical analyses were all conducted with the brain analysis software BrainVoyager QX (Maastricht, Netherlands).
Research was conducted to verify the following research hypotheses:

**Research Hypothesis 1:**

Given the left-lateralization of language processing and the significance of right auditory cortex for musical processing, left-hemispheric auditory cortex morphology (i.e. individual variation in the structure called Heschl’s gyrus) is expected to be a potential neuroanatomical trait linked to language aptitude. More specifically, more complete gyri in left auditory cortex should correlate with higher scores in the language aptitude testing batteries in children and adults.

**Research Hypothesis 2:**

While working memory is expected to show the strongest link to foreign language aptitude given its importance for overall learning processes, also musicality and musical experience will show striking correlations with language aptitude due to the positive transfer between these abilities.

**Research Hypothesis 3:**

In children and teenagers between 10 and 16 years of age, there should be a strong association between academic achievement, self-reported aptitude, arithmetic competence and performance on the language aptitude battery.

### 1.2. The Structure of the Present Thesis

Overall, this thesis will delve into the exciting world of language aptitude research and explore the main concepts necessary for an understanding of the present research projects. In total, it can be divided into three core parts, each dealing with several essential aspects worth mentioning, namely theoretical knowledge and the two research projects.

The first few chapters will serve as a theoretical introduction and discussion of previous research in the field. Since language aptitude has gained considerable interest mostly in foreign language learning (FLL)/second language acquisition (SLA) research, readers of other fields
may not be familiar with the concept itself. In a similar fashion, most scientists coming from a linguistic or educational background are not expected to have the profound neuroscientific foundations necessary for an understanding of the studies presented at the core of this thesis. Therefore, language aptitude shall be looked at from numerous perspectives, among them pedagogic and psycholinguistic, as well as neuroscientific ones.

In Chapter 2, the constructs of giftedness, language aptitude and talent shall be introduced and discussed thoroughly. For this, the importance of language aptitude for the acquisition of foreign languages will be highlighted and successful foreign language acquisition shall be discussed. Terminological issues will be tackled and the development of the concept of language aptitude from the birth of language aptitude research up to now will be paid sufficient attention. Also, the most frequently used language aptitude testing batteries shall be presented and compared to give the reader a good overview of what is currently on the market. However, the current lack of novel and comprehensive testing batteries considering the most recent theoretical developments of the concept and the inaccessibility of many quite famous tests will be of equal interest.

Chapter 3 is a journey into the cognitive neurosciences providing the reader with a basic understanding of the human brain including the areas essential for language processing with a focus on auditory processing. The most important aspects discussed in this chapter are the function and role of the primary auditory cortex not only for language processing, but also for language aptitude. Most specifically, the often-cited differences between the right and left hemisphere will be explained and discussed thoroughly in this section.

Chapter 4 focuses on working memory and musicality, especially considering them as possible cognitive indicators for exceptional language ability. It has been suggested that working memory is the strongest predictor for language aptitude and it is hence advisable to clarify the construct of working memory and explain how it can be measured in an individual. In addition, a brief overview of current research on working memory and to what extent these findings are of importance for language aptitude researchers shall be given. Regarding musicality, the reader will be provided with an introduction to the world of music and a thorough discussion to the inextricable link between language and music. Current research dealing with both music and language ability will be presented and shared neural networks and their involvement in both processing mechanisms shall be discussed comprehensively.
After the three theoretical chapters, the two research projects will be presented and discussed in detail in Chapters 5 and 6, respectively. First, the study with adults will be introduced including sections on methods, subjects, behavioural and neuroanatomic results and a short discussion. Likewise, the same details will be given for the second study. An interpretation of the results will only be marginally dealt with in these chapters since this is the aim of the following chapter.

Finally, the results of the two studies will be summarized and compared with each other in Chapter 7. Furthermore, they will be set in relation to the theoretical approaches presented in the previous chapters in an extensive discussion section. It will also be of great interest to see in how far the results achieved through these projects may have far-reaching consequences for language learning in various contexts and at various stages. Moreover, the theoretical constructs dealt with in previous chapters will be updated according to the results of the study.
2. The Multifaceted and Complex Construct of Language Aptitude

Identifying gifted children is an important but rather complex process that has become a hotly debated topic in the past years – not only in research, but even more so in society and politics. Children who display gifts in specific areas shall be identified and their skills need to be supported and fostered – so far, so easy. The truth is, however, that numerous gifts or aptitudes exist, and they are challenging to grasp and measure. In other words, we have difficulty defining what a gift in a domain is and we often lack instruments for identification. This is partly because the theoretical foundations and definitions of language aptitude or music aptitude, just to name two examples, are still not commonly agreed on and attempts are still made to build satisfying theoretical constructs. Additionally, we lack tools and specialists to test every child in every domain. How can we encourage and promote outstanding abilities, like foreign language aptitude, if we do not know who is gifted and how we can support these individuals?

As (Kerr, 2009) summarizes, the pedagogical implications the identification and support of gifted individuals bring along are diverse and it remains a highly controversial issue how to successfully encourage and promote gifts or talents. Only very few countries have well-funded, extensive programs or policies that do support the identification and teaching of children with special aptitudes. I refrain from going into detail about these political and societal issues since this is far beyond the scope of this thesis, but I want to emphasize that research on language aptitude is of high significance for more than just the research community. It is essential, not only for further defining the concept and exploring the underlying neural basis, but also for providing more useful tools and measures of language aptitude and discussing their educational relevance. This thesis focuses on individuals who have a high or particularly low language aptitude, i.e. a high or low potential for acquiring foreign languages.

This chapter will serve as a basic and profound introduction into the fascinating field of language aptitude research. First, it is necessary to distinguish between language aptitude or aptitude more generally and the plethora of related concepts often used interchangeably, e.g., ability, talent, gift, achievement. Then an introduction to the history of language aptitude and
its underlying theoretical assumptions will be provided. Thereafter, various models on the components of language aptitude and their validity shall be discussed in enough detail. Also, a short description of the most frequently applied language aptitude testing batteries and a critique of the unavailability of testing batteries mostly coming from the United States shall be included in this chapter (for a detailed review see Ameringer, Green, Leisser, & Turker, 2018).

2.1. What is ‘Aptitude’?

Numerous terms are currently in use to designate an individual’s high potential for learning a foreign language. Therefore, before discussing the concept of language aptitude, it is essential to explore and distinguish between a couple of very similar constructs, among them aptitude, giftedness, ability and talent. The terms seem quite different from one another at first sight but when trying to define them in simple terms, their relatedness becomes obvious. While the three terms ‘talent’, ‘aptitude’ and ‘ability’ have mostly been used interchangeably in research and in the literature, some researchers have proposed a clear differentiation between them.

Francoys Gagné (Gagné, 1999, 2004, 2010) is a French-Canadian aptitude researcher and also one of the most important researchers in the field of intelligence and giftedness who has significantly contributed to our understanding of these terms. In his works, he has criticised that researchers who conduct research on giftedness either leave out a differentiation between various related terms at all, e.g., they use giftedness and talent as synonyms, or they stick to one deliberate term in their studies. Since Gagné’s model of giftedness is of special interest for an understanding of language aptitude and has received widespread attention, it will be dealt with in more detail here.¹

In his paper “My convictions about the nature of abilities, gifts, and talents” (Gagné, 1999) he explained his thinking on all aspects related to the nature of abilities, gifts, and talents, particularly focusing on the development from gifts into talents and the processes accompanying this transformation. He clarified that it was necessary to differentiate between

¹ For attempts of other researchers, please refer to the summary provided in Gagné (2004).
so-called natural abilities (present at birth already) and what they develop into, termed achievement (sometimes referred to as ultimate attainment or proficiency as well). In his later published and very famous “Differentiated Model of Giftedness and Talent” (DMGT; Gagné, 2000), he replaced the terms natural abilities by gifts or aptitudes and systematically developed skills by talents. The following two definitions are proposed in his DMGT (Gagné, 2000, p.1):

GIFTEDNESS designates the possession and use of untrained or spontaneously expressed superior natural abilities (called aptitudes or gifts), in at least one ability domain, to a degree that places an individual at least among the top 10% of his or her age peers.

TALENT designates the superior mastery of systematically developed abilities (or skills) and knowledge in at least one field of human activity to a degree that places an individual within at least the upper 10% of age peers who are or have been active in that field or fields.

Gagné’s model differentiates between four mental and two physical aptitude domains. The mental aptitude domains are intellectual, creative, social and perceptual aptitude, and the physical aptitude domain comprise muscular and motor control aptitude (see left-hand side in Figure 1). He claims that these gifts are partially dependent on a genetic predisposition and are needed in every task a child must master during development. According to his model, speaking a foreign language and learning to read belong into the category of intellectual gifts. As stated in his definition of giftedness, these gifts or high aptitudes are directly observable in children since other factors, such as environmental influences or the development of systematic thinking and learning strategies, have not had a large impact at a young age yet. In accordance with J. B. Carroll, one of the fathers of language aptitude research, he believes that greater natural abilities will logically result in easier or faster learning processes.

According to his definition, gifts (used interchangeably with aptitudes) and talents differ in so far that talents are the final set of abilities that have developed out of aptitudes or gifts. Aptitudes, in this sense, would then be the raw gifts standing in sharp contrasts to the systematically developed, well-trained skills (i.e. talents). While natural abilities or gifts appear in two domains (mental and physical), talents are categorized into a larger variety of fields, such as academics, technics, sports and athletics and the like (see right-hand side in Figure 1). In Gagné’s model, natural abilities are partly influenced by chance (marked as ‘c’ in Figure 1) and have to undergo developmental processes to become competencies in specific fields.
During development, so-called influencing factors come into play. These are called catalysts and they include both environmental and intrapersonal factors that affect the development of gifts into talents.

**Figure 1** The development of natural abilities, termed ‘gifts’, to competencies, so-called ‘talents’, through environmental influences (catalysts) and developmental processes (taken from Gagné, 2005).

His model does not only highlight the developmental process taking place when gifts become talents, but it also clearly demonstrates that the developmental process is dependent on several variables, in this case the catalysts. Catalysts, except for chance, only come into play when natural abilities become competencies. This highlights that gifts are the genetic predisposition you are born with and environmental and intrapersonal factors influence this predisposition after birth. I will not go into detail about the various components and variables at this point, as they will be mentioned also later in the analysis of language aptitude. Let us
keep in mind that the more general classification of gifts and talents for specific domains also applies to language aptitude. According to Gagné’s differentiation, it is advisable to strictly distinguish between language talent and language aptitude, the first being a developed skill and the latter equalling an innate or biologically determined giftedness for acquiring foreign languages.

J. B. Carroll (Carroll, 1989, 1990), the forerunner of psycholinguistic research and first well-known language aptitude researcher, distinguished between the terms ability, aptitude and achievement. Whereas Gagné did not differentiate between the terms ability and aptitude, Carroll considered ability a measure of performance or a potential for performance, which could be stable over a longer period of time, but could change with time as well. This lack of stability would set it in contrast to aptitude and logically to achievement or talent. Dörnyei (2005), however, pointed out that ability and aptitude are still mostly used interchangeably by researchers nowadays and are thought to designate the same core concept, namely that of an a priori potential for the acquisition of a certain skill. Therefore, it seems to make sense to consider these two as synonyms that only have to be differentiated from talent/achievement. Comparing work by Carroll and Gagné, we can see that both differentiated clearly between aptitude and talent, but Carroll further suggested a differentiation between ability and aptitude.

Since I agree with Dörnyei and other researchers who have been using ability and aptitude interchangeably, I consider foreign (foreign) language learning ability and (foreign) language aptitude to refer to the same concept and they will be used synonymously from now on. Another point worth mentioning with regard to terminology is the adjective foreign as set in brackets before. Language aptitude does not necessarily impact first language acquisition but comes into play when foreign languages are acquired (Reiterer, 2019), meaning that the adjective ‘foreign’ in combination with language aptitude is redundant. Therefore, the term ‘language aptitude’ will be used throughout this thesis, although foreign is often added to point out the fact that only foreign language learning (FLL) is affected.

In summary, in the present thesis the term language aptitude will always refer to the untrained ability an individual is born with as described in Gagné’s DMGT (Gagné, 2000, 2004). Of course, various factors (catalysts in Gagné’s work) do have a great impact on the development of these aptitudes, among them environmental circumstances. Nonetheless, the concept at the core of this thesis is an individual’s natural ability for learning a foreign language.
and not the competence they have achieved (also often termed proficiency, talent or achievement).

2.1.1. A Neurocognitive Model of Aptitude and Learning

Most models explaining aptitude or giftedness only consider the behavioural level but fail to include the neural level. This is a major shortcoming as the brain is the seat of all our abilities. Just knowing that aptitude is linked to other abilities and partly genetically predisposed gives very little information as to which aspects might be genetically determined, how they manifest in individuals and how they might interact with environmental influences.

Recently, A. Seither-Preisler and P. Schneider proposed a model that captures the interaction of aptitude with neural and behavioural factors on various levels. Although it was primarily developed to explain musical aptitude and developments therein, it is suitable for the acquisition of a variety of skills, among them language. In their “Neurocognitive model for musical aptitude and competence development” (Schneider & Seither-Preisler, 2015; Seither-Preisler, Parncutt, & Schneider, 2014; see Figure 2), they describe how a biologically determined aptitude profile can lead to a certain competence profile, as also claimed by Gagné’s model. As Seither-Preisler and Schneider showed in their longitudinal study investigating musicality in children, those children with elevated musicality scores at the beginning (prior to testing and intervention) were most likely to take the opportunity to learn an instrument. This finding led them to develop their model in which the biological pathway is based on anatomical predispositions leading to higher intrinsic motivation. This higher intrinsic motivation, in turn, acts as modifying variables in the interplay between experiential, interventional variables of both training and neuro-genetic traits. In their view, anatomical properties, like grey matter volumes or the shapes of specific brain regions, interact with neural efficiency. As shown by their own research, grey matter volume of right Heschl’s gyrus (HG) was linked to a child’s motivation to learn and practice an instrument, which conversely intensified the neural efficiency of the auditory cortex as measured by magnetencephalography (MEG) in their study.

In their model, the right path describes the role of biological factors, e.g., genes, hormones or transmitters, which promote maturational plasticity and thus enable the developmental of cognitive and perceptual skills. As shown through the left path, advantageous
predispositions increase motivation, consequently leading to training-induced neural plasticity and enhanced neural efficiency. Pedagogic interventions are also part of the left path and may contribute to the manifestation of individual competence profiles. The images provided in their profile on the right-hand side are from their own studies, showing HG as a 3D mesh (above pre-existing anatomical factors) and responses to auditory stimuli (above neural efficiency).

**Figure 2** A neurocognitive model of competence development described by Schneider & Seither-Preisler (2015; English version used at conferences and presentations provided by Seither-Preisler in personal communication). The model attempts to capture the interaction between dispositional factors (i.e. a potential profile), natural maturation (right side) and training-induced plasticity (left side). Additionally, pedagogic intervention including socio-economic factors come into play and can positively influence motivation and practicing behaviour.

To my knowledge, their model is the first to try to relate the development of aptitude or potential to achievement while considering training-induced plasticity, maturational plasticity, and the importance of genetically determined structural variation in the human brain. Using
language aptitude as an example in this model, it appears that structural variation in certain brain regions could lead to maturational plasticity and the development of high skills. This could then lead to heightened intrinsic motivation to learn foreign languages. As such, language aptitude could not only lead to higher neural efficiency through structural variation in auditory cortex, for instance, but to intrinsic motivation, which could lead to the often-reported learning-induced plasticity in other brain regions. To date, few studies have been conducted looking into these specific factors and developments, which is why more research is needed to confirm the application of Seither-Preisler and Schneider’s neurocognitive model of learning in different contexts.

2.2. A Brief History of Language Aptitude

Language aptitude has gained momentum in the past years although the beginnings of aptitude research, also with regard to language, already date back to the first half of the 20th century. It all started with J.B. Carroll, an American psycholinguist who suggested that language aptitude was a natural ability and defined it by the rate of acquisition at which an unknown language was learned. In the past two decades, however, our conceptions of language aptitude have changed quite dramatically and research on the concept of language aptitude has redeemed interest in various disciplines.

At the beginning of language aptitude research it was believed that individuals who had an aptitude for FLL would learn very fast and make very quick progress in comparison to age-matched peers (Carroll, 1962, 1973, 1989). The founder of the discipline of language aptitude research, John B. Carroll, first published this definition. J.B. Carroll was a pioneer in this domain and his works and tests are still widely cited in the world of language aptitude research. While European scholars were discussing concepts such as language aptitude long before scholars on other continents, the first steps of language aptitude research were all taken in the United States. Logically, also almost all testing batteries were developed and published there.

Universities and colleges in the USA first showed an interest in the language skills of their pupils and students in the 1920s, a time often referred to as the birth of objective testing. It was not before 1960, however, that the US government began to invest in this trend for both political and social reasons. Ironically, very little attention was paid to the concept or construct
of language aptitude per se at that time. The government’s main aim behind exploring whether an individual was equipped for successfully learning a foreign language was to facilitate selection processes for the military. In other words, Americans were primarily interested in assessing linguistic competence on an academic level for practical purposes and this interest led to the birth of the first testing batteries in the 1950s and 1960s. While the American government was keen on assessing who was a good fit for the military, the underlying reasons for this exceptional ability and ways to foster it seemed to be of lesser interest (Spolsky, 1995).

Carroll, who coined the American psychological and linguistic world from 1950 onwards, was also one of the founders of the discipline termed ‘psycholinguistics’. Besides being the first graduate student of B.F. Skinner, and later a Harvard professor, he was also the first to organize seminars during which the psychology of language was suggested to be a separate, interdisciplinary field (for a summary, see Ameringer et al., 2018). Together with his colleague Stanley Sapon, he developed the well-known and still used Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959). Despite the decades that have passed, the MLAT is still widely administered today and was used as basis for almost all testing batteries developed afterwards (Carroll & Sapon, 1959; Spolsky, 1995). Wen & Skehan (2011) summarize Carroll’s four essential and quite innovative contributions to the field of language aptitude as follows: (1) He defined language aptitude as speed of learning; (2) He formulated a model of school learning which took into account different variables and their interaction; (3) He invented the MLAT, and (4) “his conceptualization of foreign language aptitude as containing multiple-components […] proved to be more enduring and interesting even than the MLAT test battery itself” (p. 18).

In the second half of the 20th century, or in the post-Carroll times, few tests were developed, and theoretical discussions took quite a long time to gain interest and be discussed in FLL/SLA research. According to Wen & Skehan (2011), all contributions to research in the years after Carroll can be placed in three simple categories, namely the development of test batteries similar to the MLAT, the investigation of aptitude-treatment interactions and research

---

2 For an excellent and complete overview of theoretical accounts including information on the development of the concept of language aptitude and different testing batteries, their advantages and disadvantages, please refer to Wen, Biedroń and Skehan (2017).
on the components Carroll had introduced. Some examples for studies between 1960 and 2000 would be the invention of the PLAB (Pimsleur Language Aptitude Battery, Pimsleur, 1966) or the CANAL-F (Grigorenko, Sternberg, & Ehrman, 2000). However, as Wen & Skehan (2011) clarify, language aptitude was seen as a stumbling block by publishers of language textbooks and teaching materials since it was impossible to adapt any book or material to the individual needs of each learner. This dilemma led to a marginalized interest in language aptitude in the second half of the 20th century. It was not before the beginning of the 21st century that the discipline suddenly started to attract attention again, also due to the rising interest in individual differences and the wish to help and support children in different educational contexts according to their very specialized needs, gifts or talents. Therefore, some updates have occurred in the past two decades, specifically regarding theoretical foundations of the concept of language aptitude and its subcomponents. Let us take a closer look at definitions of language aptitude and our current knowledge on what language aptitude consists of and how it can be measured.

2.3. Defining the Concept of Language Aptitude

It is hardly surprising that finding a uniform definition for language aptitude is just as challenging as finding one simple, compact definition of other rather vague concepts, e.g., intelligence. Given the multifaceted and complex nature of language aptitude per se, scholars had to accept the vagueness and fuzziness of the construct at some point and go with the minimal assumptions accepted by the scientific community. For assessing and investigating language aptitude, however, it is necessary to theorize, clarify and unify the construct of aptitude (S. Li, 2015). Defining and uniting the various existing concepts of language aptitude has become a hotly debated topic in the past 20 years, particularly in FLL research (Biedroń, 2015; Biedroń & Pawlak, 2016; Dörnyei & Skehan, 2003; S. Li, 2015; Wen, Biedroń, & Skehan, 2017; Wen & Skehan, 2011).

One of the first definitions of language aptitude proposed by J. B. Carroll himself defined aptitude as “an individual’s initial state of readiness and capacity for learning a foreign language, and probable facility in doing so [given the presence of motivation and opportunity]” (Carroll, 1981, p. 86). In his original description, he emphasized the need of motivation and opportunity and facility, but not speed, when learning a foreign language. Later, he strongly
emphasized that speed of learning was a core characteristic of foreign language aptitude, which was heavily criticised by other scholars.

Some fixed sets of characteristics of language aptitude have been brought forward in the past years. They shall be presented briefly and discussed more extensively afterwards to allow for a deeper understanding of the core characteristics of the concept. Most researchers working in the field of language aptitude and FLL agree upon these five sets of characteristics, although more research might be needed to fully disprove or approve these characteristics (for more details, see Kiss & Nikolov, 2005; Rysiewicz, 2009):

1. Language aptitude is independent of affective factors (e.g., anxiety, attitude, motivation) and more general cognitive capacity.
2. It is furthermore independent of intelligence and academic ability, even if overlap with these domains is possible.
3. Since it is neither dependent on prior learning experience, nor modifiable through training, it can be considered a relatively stable trait over longer periods of time.
4. It is a composite of various, mostly independent cognitive abilities and not a single, unitary capacity.
5. Language aptitude is a better predictor of L2 learning success and achievement compared to other individual differences.

From experience, one might argue that FLL success is highly dependent on good motivation and the right attitude (Characteristic 1). The existence of language aptitude, however, does not exclude an important role of motivation, attitude or anxiety during language learning. Rather, language aptitude can be considered one of the main reasons for high interest and motivation. Thus, an individual’s language aptitude is not necessarily dependent on personal (internal or external) motivation, but it might well be the reason why someone appears to be more motivated to learn foreign languages. In other words, one does not have to be motivated to do something in order to be good at it. Neither does general cognitive ability determine an individual’s success in learning a foreign language.

Regarding the relationship between language aptitude, intelligence and overall academic abilities (Characteristic 2), there is evidence to believe that intelligence and language
aptitude overlap in some respect, but they are not equal. Studies have at least shown that intelligence is not as good a predictor of FLL success as language aptitude (see also results of Study 2). To put it differently, one does not necessarily need to have an IQ above 130 (generally considered the score indicating outstanding intelligence) to be a good or even exceptional foreign language learner. Conversely, a very high IQ will not automatically mean that this person is outstanding at learning foreign languages. However, many intelligence tests depend on tasks that require linguistic abilities, explaining the overlap between IQ and language aptitude found in several studies (Sternberg, 2002).

Characteristic 3 is one of the most highly debated issues in language aptitude research. Sáfár & Kormos (2008) dedicated a whole article to three problems associated with language aptitude, one of them being the so-claimed stability of the concept. Regarding the stability – dynamics debate, they claimed that the assumption that language aptitude was a stable ability not susceptible to training and previous language learning experience has been under heavy critique by other scholars (McLaughlin, 1990). For instance, is widely accepted that working memory (WM) makes up a considerable part of language aptitude. Studies have shown, however, that trainings can improve WM scores and that WM capacity generally increases with age, reaching full capacity in early adulthood. Moreover, language experience and meta-linguistic awareness (Jessner, 2014; Singleton, 2014, 2017) are also thought to impact subsequent language learning processes. Consequently, language aptitude scores could be higher in older adults because they have learned several languages. Such an advantage caused by meta-linguistic awareness and language learning experience seems not to be the case for every aspect related to language aptitude – for instance, vocabulary learning is thought of as something that can be trained but does not get easier with the number of languages you learn, neither is this the case with phonetic memory. Grammatical analytic abilities, on the other hand, are likely to improve the more languages you learn and the more you learn about grammar, tense and the like. Robinson (2002) argued that aptitudes might be of different nature and importance at early or advanced learning stages. He further claimed that tests as the MLAT might be good at predicting initial progress, while they might fail to capture ability in advanced learning stages. In sum, it is easiest to argue that language aptitude is a ‘rather’ stable trait that is not ‘easily’ alterable through training or experience.

Characteristic 4 has been broadly discussed in the community in the past years, mostly since there is a lack of consensus of whether we can still speak of one single language aptitude,
given that it consists of various, potentially even unrelated subskills. Sáfár & Kormos (2008) state that this is a highly controversial topic because language aptitude still lacks an appropriate definition that is actually supported by language learning theories. They further criticize that language aptitude is still measured as a uniform concept in many studies despite a lack of agreement on what the subcomponents are and how they are best measured. These aspects related to the concept have not been clarified to date. It has even been proposed by some scholars that different language aptitudes (yes, the plural form) interact with each other and should be considered separate entities. There is no clear answer yet to this issue yet but it seems that language aptitude should not be treated as one single ability or entity that an individual possesses but related skills may vary substantially even in one single person.

To conclude, language aptitude is a construct that consists of several subcomponents, is rather stable and not easily modifiable through experience or training, while it is at the same time largely independent of general cognitive ability and IQ. Most likely leading to heightened motivation, it is also a stronger predictor for FLL success than other environmental variables. Unfortunately, recent studies focusing on language aptitude have until now only partly succeeded in showing whether aptitude is a purely innate capacity, to what extent it contributes to language proficiency and in how far it depends on other extra-linguistic factors, such as motivation, personality, or attitude. The findings presented in subsequent paragraphs might thus not be fully satisfying to a reader who demands clear answers and clear-cut categories. Still, they certainly provide the reader with a better picture of the research that has been conducted in the past years and help shed further light on the concept of language aptitude.

2.4. Components of Language Aptitude

The components and cognitive abilities underlying language aptitude have been under constant revision in the past decades and not all proposed adjustments and changes will be discussed in detail here. In the following paragraphs, I would like to focus on three very interesting approaches that have received widespread interest in the FLL community. These are Carroll’s four subcomponents of foreign language aptitude (in other words, the first model of foreign language aptitude; Carroll, 1962), Skehan’s ‘Processing Stages’ (Skehan, 1998, 2002) and Robinson’s ‘Aptitude Complex Hypothesis’ (Robinson, 2002). These three models
use very different approaches and deserve a separate discussion due to the validity of their underlying hypotheses.

As Carroll has already received substantial attention, his model shall be dealt with first. The four major components of language aptitude claimed by Carroll (1962, 1981) are presented in the following table (Table 1; provided by Dörnyei & Skehan, 2003).

<table>
<thead>
<tr>
<th>Aptitude Component</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonetic Coding Ability</td>
<td>Capacity to code unfamiliar sound so that it can be retained</td>
</tr>
<tr>
<td>Grammatical Sensitivity</td>
<td>Capacity to identify the functions that words fulfil in sentences</td>
</tr>
<tr>
<td>Inductive Language Learning Ability</td>
<td>Capacity to extrapolate from a given corpus to create new sentences</td>
</tr>
<tr>
<td>Associative Memory</td>
<td>Capacity to form associative links in memory</td>
</tr>
</tbody>
</table>

Table 1 Carroll's components of language aptitude (taken from Dörnyei & Skehan, 2003), including phonetic coding ability, grammatical sensitivity, inductive language learning ability and associative memory and descriptions of the four components.

Carroll suggested that individuals with a high potential for learning foreign languages must be capable of coding and retaining speech input, of identifying the functions of elements in a certain language and be able to apply newly learnt rules to create new language output, i.e. infer grammatical and semantic rules. Furthermore, he thought that an individual with high language aptitude must have an outstanding memory ability to keep all the novel items in both short- and long-term memory. These four components of language aptitude, namely phonetic coding ability, grammatical sensitivity, inductive language learning ability and associative memory, are basically still accepted nowadays. Regardless the validity of the components developed by Carroll, however, a major drawback of his model is that he did not refer to any underlying model for language learning to support his assumptions. Therefore, the model
clearly lacks a theoretical construct and it does not provide information on the relations between the four and their involvement at different stages and in different contexts of FLL.

Peter Skehan, one of the most important language aptitude researchers of the past twenty years, used the components proposed by Carroll and set them in a language-learning context (Skehan, 1998, 2002, 2016). His model is based on the three processing mechanisms underlying SLA, the developmental stages and the involved aptitude components including those in Carroll’s model. He aimed at combining aptitude theory with second language acquisition and thus effectively linked aptitude to various SLA developmental stages and their underlying cognitive processes. Not all aptitude constructs that he mentions (right side) are necessarily linked to language only (e.g., attentional control, WM), but most of them are. One major difference to the earlier components proposed by Carroll is that Carroll’s second and third category (inductive language learning ability and grammatical sensitivity) are therein subsumed under one more general term, namely ‘language analysis ability’ (Biedroń, 2015; Wen, Biedroń, & Skehan, 2017). Memory ability appears in the form of WM and retrieval memory only in this model (see Table 2). According to Skehan’s model (also extensively discussed in Biedroń & Pawlak, 2016; Hummel, 2009 and Wen et al., 2017), language aptitude would thus potentially comprise (1) phonetic coding ability (related to input processing), (2) language analysis ability (related to central processing) and (3) retrieval memory (related to language output). In addition, WM, attentional control and chunking are essential mechanisms for language aptitude according to his model.

Skehan used Carroll’s ideas and combined them with an underlying theory, which makes his approach extremely promising and successful. By looking at his processing stages, one can understand what happens during second or overall foreign language learning and in how far the different components come into play at the various stages. Yet, his model does not provide information as to the importance of all these steps in different contexts and at different stages of learning, giving the impression that the steps are the same in each context and at each stage.
Table 2 Skehan’s model of language aptitude focuses on SLA stages and their corresponding L2 cognitive processes. The three stages of SLA are language input, central processing and creating language output. The cognitive processes underlying these stages are segmentation, noticing, pattern recognition, complexification, handling feedback, error avoidance, automatization, creating a repertoire and lexicalization (taken from Skehan, 2016; see Wen, 2016).

The last model is Robinson’s so-named ‘Aptitude Complex Hypothesis’ (Robinson, 2001, 2002). It is very different from Carroll’s and Skehan’ approach given that it is based on pedagogic concepts and is adapted from Snow (Snow, 1994; Snow & Farr, 1987). In his work, Robinson argues that an aptitude for acquiring a foreign language always needs to be examined in relation to the varying processing demands of the given learning conditions and tasks. Consequently, an aptitude for learning under these processing conditions requires the interaction of different complexes of cognitive abilities. Claiming that there are various

<table>
<thead>
<tr>
<th>SLA Stages</th>
<th>L2 Cognitive Processes</th>
<th>Aptitude Constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Input</td>
<td>Input Processing (Segmentation)</td>
<td>Attentional Control</td>
</tr>
<tr>
<td>Central Processing</td>
<td></td>
<td>Working Memory</td>
</tr>
<tr>
<td>Language Output</td>
<td>Noticing</td>
<td>Phonetic Coding Ability</td>
</tr>
<tr>
<td></td>
<td>Pattern Recognition</td>
<td>Working Memory</td>
</tr>
<tr>
<td></td>
<td>Complexification</td>
<td>Language Analysis Ability</td>
</tr>
<tr>
<td></td>
<td>Handling Feedback</td>
<td>Language Analysis Ability</td>
</tr>
<tr>
<td></td>
<td>Error Avoidance</td>
<td>Working Memory</td>
</tr>
<tr>
<td></td>
<td>Automatization</td>
<td>Retrieval Memory</td>
</tr>
<tr>
<td></td>
<td>Creating a Repertoire</td>
<td>Retrieval Memory</td>
</tr>
<tr>
<td></td>
<td>Lexicalization</td>
<td>Chunking</td>
</tr>
</tbody>
</table>

The last model is Robinson’s so-named ‘Aptitude Complex Hypothesis’ (Robinson, 2001, 2002). It is very different from Carroll’s and Skehan’ approach given that it is based on pedagogic concepts and is adapted from Snow (Snow, 1994; Snow & Farr, 1987). In his work, Robinson argues that an aptitude for acquiring a foreign language always needs to be examined in relation to the varying processing demands of the given learning conditions and tasks. Consequently, an aptitude for learning under these processing conditions requires the interaction of different complexes of cognitive abilities. Claiming that there are various
aptitudes for learning a foreign language, he doubts that any test, like the MLAT, would ever be sensitive enough to capture the ongoing interaction of all these cognitive abilities.

According to Robinson’s definition, “foreign language learning is the ability to successfully adapt to and profit from instructed, or naturalistic exposure to L2” (Robinson, 2012, p. 57). In his model, he treats aptitude(s) as a set (or complex, as defined in the title of the model) of abilities that allow a foreign language learner to benefit from instructional interventions (pair-work, grammar rule instruction etc.). In his view, classroom learning conditions and pedagogic intervention need to be considered in aptitude models and aptitude tests need to involve actual learning with feedback. In contrast to Carroll and Skehan, whom he did not disagree with but whose ideas he implemented into his own ideas, he takes a rather pedagogic approach and emphasizes that implicit, incidental and explicit learning processes need to be clearly distinguished in any aptitude model or test (see Figure 3).

The most inner circle of his model consists of 10 core abilities (abbreviations as used in the figure provided3), namely processing speed (PS), pattern recognition (PR), phonological working memory capacity (PWMC), phonological working memory speed (PWMS), semantic priming (SP), lexical inferencing (IN), text working memory capacity (TWMC), text working memory speed (TWMS), grammatical sensitivity (GS) and rote memory (RM). Robinson states that these cognitive abilities contribute to higher order aptitude factors: noticing the gap (NTG), memory for contingent speech (MCS), deep semantic processing (DSP), memory for contingent text (MCT) and metalinguistic rule rehearsal (MRR). Without going too much into detail about the model here, abilities and aptitude complexes are the basic for task aptitude (e.g., open tasks, reasoning, single tasks, planning time) and pragmatic/interactional abilities or traits. The ‘Aptitude Complex Hypothesis’ is only one of the four interlocking hypotheses defining a framework which allows closer examination concerning individual differences in FLL. Generally, he bases aptitude on four complexes and distinguishes between primary (directly measurable) and secondary abilities. In his model (see details in Figure 3), abilities are considered hierarchical and related to instructional aspects. Finally, Robinson also pays

3 The abbreviations used in Robinson (2002) are only presented together with his model at this point to facilitate an understanding of the figure provided. The terms and abbreviations will not be used in the upcoming chapters, neither do they appear in the list of abbreviations.
tribute to pragmatics and pragmatic development highlighting that communicative ability in language aptitude has been poorly explored to date (Robinson, 2001).

**Figure 3** Aptitude complexes as described by Robinson (2002). The most inner circle consists of 10 core abilities that contribute to higher order aptitude factors (aptitude complexes; figure taken from Robinson, 2002). The hierarchical order goes from abilities to aptitude profiles, task aptitudes and finally pragmatic/interaction abilities and traits.

Although Robinson is certainly right when claiming that different learning contexts and stages must be considered, language aptitude is often understood as a being dependent on implicit learning without proper instruction, that is in a natural setting without instruction. When subjects complete language aptitude tests, they do not get proper instruction or support and must work out the rules themselves. Doubtlessly, this lacks a relation to usual foreign
language learning in classrooms and highlights the need to be able to translate findings from FLL research into classrooms and other educational settings. As mentioned by Robinson (2001), other researchers have also started questioning whether these distinctive components of language aptitude might be important at different stages of language acquisition and possibly even in different contexts of language learning (Abrahamsson & Hyltenstam, 2008; Artieda & Muñoz, 2016). According to aptitude experts, Peter Skehan’s ‘Processing Stages’, as well as Robinson’s ‘Aptitude Complex Hypothesis’ are considered the most important promising models on the market, so to say (Biedroń & Pawlak, 2016).

Finally, a discussion of the core components proposed by researchers should include a brief note on a rather recent model developed by Grigorenko and colleagues (refer to Grigorenko et al., 2000). Their model of triadic human intelligence led to a new interpretation of foreign language aptitude and a new test for measuring it. Unfortunately, these researchers have not pursued this theory and the test, the CANAL-F, afterwards and the test is not available upon request. Therefore, interested readers are referred to Wen et al. (2017) who give a concise account of the latest developments thereof.

To summarize, various scholars have addressed the underlying components and cognitive abilities of language aptitude from different angles. Although Robinson’ approach is particularly interesting and certainly worthy of discussion, it is not really applicable if a definition of language aptitude rather emphasizes the implicit learning mechanisms of individuals in a non-instructed setting without feedback. Therefore, whenever I refer to the components of language aptitude in later chapters, I am explicitly referring to Skehan’s three major processing stages, namely input processing, central processing and output processing and the abilities associated with them. The three components belonging to these three stages are phonetic memory, language analysis ability and memory. All of them are in some form assessed by the tests discussed beforehand and also in the two studies I conducted, I assessed these three major components of language aptitude through the MLAT (Carroll & Sapon, 1959) and the LLAMA language aptitude battery (Meara, 2005).

2.4.1. Working Memory and Language Aptitude

Due to its importance for overall FLL processes, WM seems to be a vital component of any form of aptitude. In the past 20 years, the proposal of “Working Memory as Language
Aptitude” has gained popularity and has been re-examined by numerous researchers (see Wen, 2019). In particular the cognitive (neuro)sciences have added empirical evidence and have helped overcome remaining issues and looming caveats concerning the implementation of WM as a major component of language aptitude. Studies have proven that the higher the WM capacity of an individual, the better the results in various domains of SLA/FLL (Singleton, 2017; Wen, 2016). Skehan’s model of processing stages shows at how many levels WM (also recognizable in the component ‘memory retrieval’) is essential and the importance of investigating WM is hence obvious. However, WM is only briefly mentioned at this point as it has been named several times in language aptitude models and deserves a discussion of its own. Therefore, it will be dealt with comprehensively in Chapter 4, the chapter in which cognitive abilities and their influence on foreign language learning shall be more closely investigated.

2.4.2. Aptitude vs. Aptitudes

The concept of language aptitude has undergone major changes in the past few years and the term ‘aptitude’ has been more and more frequently replaced by the plural form, ‘aptitudes’. Instead of speaking of components of ‘language aptitude’, for instance, it has been suggested to speak of different ‘language aptitudes’ that do not necessarily go hand in hand or interact. The idea of referring to gifts in the plural form instead of talking about one specific attitude that is possessed (or not) by an individual is not entirely new (see Gardner & MacIntyre, 1992). In recent articles, Suzuki & DeKeyser (2017), for example, talk about different cognitive aptitudes related to either explicit or implicit learning abilities. Robinson (2007) refers to aptitude(s) in his work to underline the necessity of further exploring which complexes of aptitudes are involved in second language learning aptitude. Bain, McCallum, Bell, Cochran, & Sawyer (2010), on the other hand, write about academic aptitudes among students.

To my knowledge, there is no theoretical framework or explanation that justifies or explains the advantage of using the plural form instead of the single form. In most works, as I see it, the plural is merely used to indicate that there are various forms of aptitude related to a domain, e.g., cognitive aptitudes, academic aptitudes, language aptitudes, but these works do not clarify the implications brought along by the introduction of these terms. Surely, I do not doubt that numerous related skills make up language aptitude and they do not necessarily
interact or overlap. Still, referring to all these subskills as language aptitudes makes them seem like very independent constructs, which is not completely justified if no theoretical model has been proposed regarding the nature of and relation between the various aptitudes proposed to be measured. Primarily, I want to say that I am aware that there has been a trend of referring to the various forms of language aptitude in the plural form. For the present thesis, which is not a theoretical but an empirical work, re-defining the concept of language aptitude and shedding light on the possibility of existence of different aptitudes is not a major concern, and much more research would be needed to develop a satisfying underlying theoretical framework. Moreover, in many recent publications from well-known aptitude researchers, language aptitude is still referred to in the singular (e.g., S. Li, 2018; Long & Granena, 2018; Winke, 2018). Therefore, the decision was made that language aptitude should be used as an umbrella term in the present thesis, with various different abilities contributing to overall language aptitude, even if they refer to very distinct skills.

2.5. Language Aptitude Tests

The tests designed in the early years of language aptitude research mainly assessed whether individuals were fast or slow learners of foreign languages, which is not fully in accordance with the views held today by aptitude researchers. The major issue at that time was that they were not based on any theoretical framework (intelligence tests containing linguistic elements are left out at this point). In the past decades, however, the scientific community has witnessed renewed enthusiasm for exploring the concept of language aptitude and therefore, many formerly neglected issues have been addressed (e.g., the relationship of aptitude to high levels of attainment or the influence of incidental exposure to a language). In summary, from 1960 to 2000 the focus was put on the speed of acquisition, whereas new conceptualizations have focused on the variation, so-called individual differences that can be found between different learners, and the reasons for these differences (Dörnyei & Skehan, 2003; Robinson, 2001; Spolsky, 1995).

Numerous tests have been designed to assess language aptitude but very few deserve a discussion of their own. The two most important language aptitude tests of the past decades are the MLAT and the LLAMA language aptitude battery (2.5.1 and 2.5.2.), which are actually closely linked since the LLAMA is based on the MLAT. The two will be discussed in much
detail since they have been predominantly used in language aptitude studies of the past years and these are also the tests applied in the two studies I conducted. Another famous but very novel test is the Hi-LAB, which will be briefly discussed as well. Especially with regard to the studies presented later on, the two other language aptitude tests that I used shall be presented briefly. They are the Hindi speech imitation test and the MULTAP, a speech perception test.

2.5.1. **The Modern Language Aptitude Test (MLAT)**

The best-known and most widely used test of the past decades is the MLAT (Carroll & Sapon, 1959), which was developed by John B. Carroll and his colleague Stanley Sapon. Carroll’s major interest was FLL and his model, the so-called ‘Carroll Model’, demonstrated that individuals with low aptitude needed more time to achieve certain criteria for learning in comparison to individuals with high aptitude. His model, which was based on the audiolingual teaching method, distinguished between five classes of variables accounting for variations in school achievement of children, one of them being language aptitude. Therefore, the MLAT consists of five different sections (see Table 3) that measure its four underlying components (see previous subchapters) in a hybrid manner, i.e. only all subtests together measure language aptitude.

Together with his colleague Sapon, Carroll chose a purely empirical approach and selected forty different cognitive and psychological tests to test their significance for FLL success. The tests that matched the language proficiency scores were then chosen to make up the MLAT. Such an approach was frequently applied in intelligence testing at that time and has been under critique for quite logical reasons (e.g., an obvious lack of an underlying theory of the cognitive concepts and abilities that were measured). Consequently, it might be highly problematic that this approach, from a theoretical perspective, only measures what the test parts measure and does not have much to do with the concept of language aptitude itself. This is one of the major weaknesses of Carroll’s model of language aptitude (Dörnyei, 2005; Sáfár & Kormos, 2008).
<table>
<thead>
<tr>
<th>Name</th>
<th>What is being tested?</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLAT I</td>
<td>Rote memory and auditory alertness – names of numbers in an artificial language are acquired.</td>
</tr>
<tr>
<td>MLAT II</td>
<td>Sound-symbol association ability – different sounds (in combinations) are presented and have to be associated with written forms.</td>
</tr>
<tr>
<td>MLAT III</td>
<td>Sound-symbol association ability and vocabulary knowledge – correct synonyms of disguised words must be selected (multiple choice).</td>
</tr>
<tr>
<td>MLAT IV</td>
<td>Grammatical sensitivity – components of sentences must be identified (grammatical function) and related to elements in other words.</td>
</tr>
<tr>
<td>MLAT V</td>
<td>Associative rote memory – as many Kurdish words as possible must be memorized (presented with English translations).</td>
</tr>
</tbody>
</table>

Table 3 The five subtests of the MLAT (Carroll and Sapon, 1957), their names, and what is being tested in each part of the test. For their relations to aptitude components, see Table 4.

Robinson describes the purpose of the MLAT, among others, as follows:

Prediction of rate, and feasibility of administration [, i.e. paper-and-pencil methods that could be done in one sitting and led to definite scores.,] with the goal of selection into programs (and subsequent assignment to languages of different levels of difficulty), were therefore the main purpose, constraint, and objective which the Modern Language Aptitude Test, the Defense Language Aptitude Battery, and also the VORD were funded and developed to meet (Robinson, 2005, p. 48).

What we know for sure is that the MLAT measures some components involved in language aptitude (see Table 4) but it also fails at fully measuring concepts that seem to be of importance as well (e.g., WM, inductive learning). Inductive learning, for instance, is only weakly measured in MLAT I and not in any other test, which basically means that the MLAT fails to measure implicit learning mechanisms underlying the acquisition of grammatical
analytical skills. This is surprising given that grammatical analytic ability was one of the components of language aptitude as suggested by Carroll himself.

<table>
<thead>
<tr>
<th>Component</th>
<th>Well</th>
<th>Weakly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonetic Coding Ability</td>
<td>MLAT II</td>
<td>MLAT III</td>
</tr>
<tr>
<td>Grammatical Sensitivity</td>
<td>MLAT IV</td>
<td>/</td>
</tr>
<tr>
<td>Rote Learning</td>
<td>MLAT V</td>
<td>MLAT I</td>
</tr>
<tr>
<td>Inductive Learning</td>
<td>/</td>
<td>MLAT I</td>
</tr>
</tbody>
</table>

Table 4 The four components of language aptitude as defined by J.B. Carroll and what each part of the MLAT measures well or weakly (taken from Rysiewicz, 2008).

It might seem logical to assume that testing large populations with different language aptitude batteries is also in the mind of the inventors and can only contribute to finding an adequate measure of language aptitude. That does not seem to be the case, however. Obtaining a version of a famous language aptitude battery is like finding the needle in the haystack. The US government likes to keep its tests as secret as possible making it hard for language aptitude researchers in other parts of the world to test their validity. Those researchers seeking to obtain a version of the MLAT, which is claimed to be available to a broader audience and open for research, will have difficulties ever getting their hands on it. As can be read online, “Unfortunately, due to the sensitive nature of the test, we only sell the MLAT to government agencies, missionary groups, and licensed clinical psychologists. We do NOT sell the test to individual researchers, teachers, or students” (Stansfield, 2019). Still, most papers on language aptitude cite the MLAT and state that it is a commercially available test (Yalçın & Spada, 2016). French, Hungarian and Italian versions of the MLAT exist, as well as one for younger individuals (8-11 years), and a new Japanese Language Aptitude test based on the MLAT has been developed (Robinson, 2001). Still, it is hard to understand why one would keep a successful test secret in an age when open access and data share are at the forefront of scientific research.
2.5.2. The LLAMA Language Aptitude Battery

The LLAMA test was developed by Paul Meara at Swansea University in 2005 and is a computer-based testing battery for language aptitude. It is a revised version of the Swansea LAT (language aptitude test) and, similarly to the MLAT, it is a combination of exploratory tests designed to assess foreign language aptitude (Granena, 2013). As the test is based on the MLAT, it appears quite similar, but consists of four parts only. The four parts, their corresponding names and information on what they test are provided in Table 5.

As Granena points out, the LLAMA test does not suffer from any limitation or restriction, e.g., difficult to obtain, only in pencil- and-paper format or only for military purposes. Although it is not extensively standardized as the MLAT, it has been used by numerous research groups and is available for free online.

<table>
<thead>
<tr>
<th>Name</th>
<th>What is being tested?</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLAMA B</td>
<td>Vocabulary Learning</td>
</tr>
<tr>
<td></td>
<td>Vocabulary learning of an unknown language in 2 minutes</td>
</tr>
<tr>
<td>LLAMA D</td>
<td>Phonetic Memory</td>
</tr>
<tr>
<td></td>
<td>Recognition and memory of heard unknown words</td>
</tr>
<tr>
<td>LLAMA E</td>
<td>Sound-Symbol Correspondences</td>
</tr>
<tr>
<td></td>
<td>Ability to understand and remember sound-symbol combinations</td>
</tr>
<tr>
<td>LLAMA F</td>
<td>Grammatical Inferencing</td>
</tr>
<tr>
<td></td>
<td>Understanding and application of grammar/syntax and semantics of an unknown language</td>
</tr>
</tbody>
</table>

Table 5 The different subtests of the LLAMA language aptitude battery, LLAMA B, D, E and F, aiming to measure the construct of language aptitude (Meara, 2005).

As Granena describes and can be read in the LLAMA manual (Meara, 2005), the LLAMA is based on the MLAT and is language-independent (based on material from an indigenous Central American language), i.e. it can be used for speakers of any native language.
unlike the MLAT. This is certainly a major advantage over other tests such as the MLAT, which is not even available in German, for instance. Language independence is a feature of great importance that has numerous advantages, e.g., facilitation of test administration, no influences from the native language or other languages spoken by participants (Granena, 2013). Since the LLAMA test has been used in the child population of Study 2, it seems worthwhile to discuss the tasks in detail at this point.

LLAMA B requires subjects to learn vocabulary by linking a written, alphabetic word form of an unknown language (mostly words consisting of two syllables) to a picture (see Figure 4).

**Figure 4** LLAMA B is a vocabulary learning task. Words are given in the middle (in written form) and each little figure denotes a different piece of vocabulary. In two minutes, participants have to remember as many words as possible by clicking on them and remembering the association between the new word and the symbol.
In LLAMA B, subjects have two minutes (120 seconds as indicated at the top of Figure 4) to study as many of the items (altogether 20) as possible and are then tested in a session where they have to link a written form to the correct picture. The written words mostly consist of multiple consonant-vowel syllables and can be clicked as often as wanted. In the testing phase, the written word is always given, and participants have to choose the correct image. No auditory input is provided during this task and the testing phase does not have any time limitation, as does the learning phase.

LLAMA D, testing phonetic memory, assesses in how far unknown words presented auditorily remain in a subjects’ memory without the possibility of applying explicit memory or learning strategies. During this task, participants hear a number of unknown words. Each word is only repeated once and the words are presented with only short breaks in between. This makes it impossible to remember the word forms and rehearse them during the listening phase. Thereafter, participants are always presented with novel words or one of the heard words and have to decide whether they have just perceived a novel or an already heard word. Several of the already known words appear numerous times during the testing phase to exclude success rates only due to chance.

In LLAMA E (Figure 5), subjects are presented with written combinations of symbols and letters (e.g., ‘0i’). They have two minutes to click on these written combinations and perceive their phonological form. To obtain the auditory speech input, they can click on the combination of sounds and letters (e.g., ‘0i’ corresponds to /ta/, while ‘3é’ corresponds to /ni/ – these are just examples and not the correct solutions for the test). During the learning phase, subjects may press each written symbol combination as often as they want to perceive the auditory form in these two minutes. In the testing phase, participants always hear a combination of two pairs of symbols and letters and are given two possibilities of what the written forms of this combination might look like. To give a final example, a participant will hear the syllables /nita/ and has to decide whether the corresponding correct written form is ‘3é0i’ or ‘0é3i’.
Figure 5 LLAMA E tests the acquisition of sound-symbol correspondence. Each combination of a letter and a digit (e.g., 0ì, 3ì) is pronounced differently in the given, new language and subjects have to understand how the system works and learn the associations between two sounds and two symbols.

LLAMA F (Figure 6) is probably the most complex task and participants are therefore allowed to take notes during the task. In LLAMA F, subjects see sentences of an unknown language combined with a picture. By reading the sentences and seeing the pictures, participants have to understand the meanings of words in that language, as well as the syntactic structure and the morphological and syntactical rules. After 5 minutes of learning, they enter the testing phase. In the testing phase, participants see pictures (as those beforehand) and have to choose between two sentences that these pictures correspond to. In other words, they have to choose the correct sentence (syntactically and semantically). After the first half of the testing phase, the task becomes more difficult as completely new images are shown. Subjects can only manage to choose the correct ones by applying the rules that they have learnt during the learning phase. In the example of Figure 6, ‘eked-ilad-za’ stands for ‘one green square’. ‘Za’ is the word or affix for singular and comes in two variants, as ‘sa’ or ‘za’, depending on whether
it is preceded by a voiced or voiceless plosive (voiced in this case requiring ‘za’). ‘Eked’ stands for ‘square’ and always precedes the colour adjective, in this case ‘ilad’ for green.

Figure 6 LLAMA F tests language analysis ability. Each grey square presents a sentence and a picture, the sentence describing in the new language what is present and happening in the picture.

It is easy to explain why the LLAMA has become so popular in the past decades: it is easy to administer, suitable for children, language-independent, available for free and gives you scores for each part without having to calculate them yourself (Granena, 2013; Rogers, 2017). It is very popular among children since they enjoy the tasks (I report this from a study with 220 children, who all undertook the LLAMA language battery during our behavioural testing). In each part of the LLAMA, subjects receive a % performance score between 0 and 100%, 100% being the highest score. During each of the subtests, a correct answer gives a number of points while an incorrect answer leads to the deduction of already scored points. Subjects receive direct feedback during the testing phase as correct and incorrect answers are always followed by specific auditory signals.
2.5.3. The Hi-LAB (High-Level Language Aptitude Battery)

The Hi-LAB is one of the newest and most complex language aptitude testing batteries on the market. It was developed by researchers at the Center for Advanced Study of Language at the University of Maryland (Doughty, 2010, 2019). Doughty and colleagues aimed to incorporate not only explicit learning, as measured by the MLAT, for instance, but also implicit learning mechanisms and all components of the WM system separately (e.g., inhibition, updating, switching etc.). They designed the Hi-LAB as a complementary test battery to the MLAT, motivated by research on aptitude complexes (Robinson, 2001) and L2 processing stages (Skehan, 2016). They hoped to be able to predict and explain high levels of L2 proficiency, specifically in post-critical period learners, compared to other language aptitude tests focussing on the success in early stages of language learning (Doughty, 2013).

The Hi-LAB has been described as an assessment-oriented test that attempts to provide more diagnostic and remedial purposes in combination with the MLAT and was declared as a richer, theoretically grounded conceptualization of language aptitude (Dörnyei, 2014; Doughty, 2013; Wen & Skehan, 2011). Unfortunately, the Hi-LAB is not available for researchers and only very few details on the tests have been provided.

To my knowledge, one study to date explored links in cognitive aptitude assessment and L2 speaking proficiency using and comparing the LLAMA and the Hi-Lab (Granena, 2018). In one of the only studies conducted with the Hi-Lab, the authors reported that neither executive WM, nor auditory processing speed or auditory testing could distinguish successful and very successful learners, which is not conclusive in terms of validation for this test (Linck et al., 2013; Wen et al., 2017).

2.5.4. Other Tests Designed to Measure Language Aptitude

Without any doubt, the LLAMA and the MLAT are the most famous and most widely used language aptitude batteries. In earlier years, various tests very similar to the MLAT were designed, but with less success. One of them is the PLAB. The PLAB, so-called Pimsleur’s Language Aptitude Battery (Pimsleur, 1966), is very similar to the MLAT and was designed to meet the same purposes. The main difference to the MLAT at that time was that it could be administered to a younger population and that it included a tiny section on motivation for FLL.
The MLAT, the PLAB, the VORD (one of the forerunners of the MLAT) and the DLAB (Defense Language Aptitude Battery; Peterson & Al-Haik, 1976) are all very similar and measure practically the same. I would have loved to show examples of all these tests but they are not available for researchers, not even upon request. The DLAB and the VORD are protected tests that are only administered to United States government personnel (Robinson, 2002). As already mentioned, all these tests were invented for one purpose, namely to differentiate between students who had very good proficiency and those with very low proficiency. Interestingly also, the other tests were all based on the MLAT and relied on the fact that what Carroll had described as measuring language aptitude was correct.

Some years after the birth of all these quite similar language aptitude tests, Grigorenko, Sternberg and Ehrmann (Grigorenko et al., 2000) presented a new theory of foreign language aptitude and a testing battery measuring their theoretical construct, the CANAL-F (Cognitive Ability for Novelty in Acquisition of Language – Foreign; briefly mentioned in previous subchapters). According to the model of the three authors, there are four aspects essential for language aptitude, namely (1) knowledge acquisition processes, (2) levels of processing, (3) modes of input, and (4) memory processes (Rysiewicz, 2009). These abilities and factors operate through different processes. What is striking about the CANAL-F is that it is dynamic given that the individual learns a new (artificial) language during the test. The language is called Ursulu and resembles natural languages but does not correspond to any existing language. As interesting as the CANAL-F sounds, requesting a copy of it for use in research will have you sigh in despair just as with the others. In his pencil-and-paper test cannot be obtained because there is no online version and no paper-and-pencil versions are available anymore (this was the official response received by collaboration partners). Apart from that, it was also designed for use in government contexts (Robinson, 2002).

To sum up, the MLAT and the LLAMA language aptitude battery are very similar testing batteries, but the LLAMA has the advantage that it is language-independent, available for free and suitable for children. Almost all other language aptitude tests (VORD, PLAB, DLAB, MLAT) were only designed for the US government and versions cannot be obtained upon request. It is thus hardly surprising that language aptitude research has seen practically no progress between 1960 and 2000 given the obvious lack of willingness to share knowledge and experience. Researchers at the University of Vienna, however, have worked on developing their own tests for assessing different aspects of language aptitude. These two tests are the
Hindi speech imitation test and the Mult/AP. Both will be discussed due to their relevance of the studies conducted for this thesis.

2.5.4.1. The Hindi Speech Imitation Test

The Hindi speech imitation test was developed by Susanne Reiterer, now affiliated with the University of Vienna, and is only described in publications of studies it has been used in (Reiterer, 2019; Reiterer et al., 2011; Turker, Reiterer, Seither-Preisler, & Schneider, 2017; Turker, Reiterer, Schneider, & Preisler, 2018). The test was developed to assess an individual’s speech imitation ability, not language aptitude per se, which is why it was not included in previous chapters. In other words, the major aim behind the development of this speech imitation task was to assess if and to what extent subjects could imitate with native-like proficiency without receiving proper instruction and information on the speech material and without ever having come across that particular language. Therefore, participants are neither before nor after repetition aware of the semantic content of the words or sentences they are asked to repeat. Prof. Reiterer termed the ability assessed by the Hindi test ‘pronunciation aptitude’. In the present subchapter, only the rationale behind the test and the original procedure as followed by Reiterer shall be discussed. Specifics about chosen stimuli and number of raters for the present thesis will be presented in the corresponding methods sections.

The whole test uses four simple Hindi words and four Hindi sentences of different length and phonetic complexity (syllable length: 7/7/9/11). The speech material that was used for the test was recorded in a sound-proof room as spoken by a male native speaker of Hindi. The task requires subjects to repeat immediately (direct imitation) after binaural presentation for three times. A pilot study was conducted shortly after test development and showed that performance was very low for average individuals after only a single exposure to the stimuli. Therefore, a repetition after listening to each stimulus three times was chosen to ensure that everybody would be capable of producing all stimuli, in particular sentences, or at least parts of them (Dogil & Reiterer, 2009).

The assessment of the quality of the Hindi speech imitation of the aforementioned stimuli is performed with an online (i.e. blind) rating of native speakers. In the first study by Reiterer and colleagues using the Hindi stimuli, MP3 files were inserted into the internet evaluation database that was created only for this speech imitation test. To verify if acoustic
quality was sufficient, the native raters were asked to provide feedback not only on the online rating procedure, but also on the acoustic quality of the speech material provided. Feedback was very positive and no loss in quality was reported. Raters were naïve regarding linguistic or phonetic background of the speech imitators and were only instructed to transmit their global intuitive impression of whether the sample they were listening to could have been spoken by a Hindi native speaker or not. To ensure the quality of the evaluation procedure, recordings of native Hindi speakers who had imitated the speech samples were also randomly inserted into the database. For the intuitive rating scale (Jilka, 2009), a rating bar was used. It ranged from 0 to 10 (least to most representation of ‘native-speaker-likeness’) and had to be clicked after listening to the sample. For the online ratings, a balanced number of female and male raters was always used. In their study of 2009 (Jilka, 2009), 30 native speakers performed the rating online, 15 females and 15 males.

2.5.4.2. The MULT/AP

The MULT/AP (Christiner, 2019), short for multilingual aptitude test, is a newly developed speech and language perception test developed by Susanne Reiterer and Markus Christiner from the University of Vienna, Austria. The test was designed to assess basic phonetic aptitude for perceiving unfamiliar language input, that is, an individual’s potential to discriminate language differences without specific instruction and/or training. The MULT/AP tests the perception of strings of language fragments of different syllable length, which are then compared to either a single phrase or various phrases. Depending on the difficulty of the task and the age of participants, the lengths of the strings can vary. Children, for instance, are tested for strings of two different constituents to eight constituents, which are always compared to a single phrase. Adults or adolescents listen to strings of up to twelve constituents, which are then compared to one to three phrases in which ordering needs to be considered.

The MULT/AP builds upon earlier studies (Christiner & Reiterer, 2013, 2015) and overall, a large variety of languages is included, such as Chinese, Tagalog, Persian, Hindi, Japanese or Russian. For Reiterer and her colleague it was important that different rhythmic components are considered in the test, which is why stress-timed, syllable-timed and mora-timed languages, as well as tone languages, were included. Apart from assessing phonetic
aptitude, a questionnaire that collects further information about other factors, e.g., age, gender, cognitive abilities, language experience, was also implemented.

All languages subjects are familiar with are excluded from the testing afterwards. Currently, six forms of the MULT/AP exist (three long and three short versions) and are used to assess phonetic aptitude in different age groups (e.g., pre-school children, primary school children, schoolchildren etc.). Although the test has already been administered to a large number of people, details on test development and results of the validation have not been published yet but will be the main focus of Markus Christiner’s thesis (Christiner, 2019; thesis currently in preparation).
3. Language Aptitude and the Brain

Everything we perceive (feel, hear, see) or do is processed and stored in our brain. Even the simplest activities or tasks such as listening to a bird sing, thinking about what to buy in a supermarket or pointing at an object engage various brain regions and require the interaction between numerous others. This interaction is based on neural transmission, and thus on the existence of neurons, axons and synapses connecting these neurons. The brain has long been a mystery and despite decades of intense research, we are still far from fully understanding the mechanisms and processes underlying the human brain. In other words, although technology has made outstanding progress and thousands of researchers have dedicated their careers to further unravelling the processes behind our brain, we are not yet capable of answering from our status of knowledge why the most complex organ in our body works exactly the way it does.

Since the advent of cognitive neurosciences thanks to the technological advances and inventions of the past decades (e.g., the birth of computer tomography, positron-emission tomography or magnetic resonance imaging), it has become possible to investigate the functioning of the human brain in living individuals. Earlier, conclusions on the brain could only be drawn from post-mortem studies. These technological advancements have triggered off a so-called cognitive neuroscientific revolution. Today, numerous disciplines such as medicine, psychology, speech-language pathology or linguistics, benefit largely from the now available neuroimaging methods for research purposes, as well as for clinical assessment. Additionally, new disciplines have come to existence due to this revolution and the possibilities of investigation it has brought along. Research fields such as neuropsychology, cognitive neuroscience or neurolinguistics have successfully established themselves and have attracted great interest in the past decades. These new disciplines attempt to integrate new research methods in the aforementioned disciplines by linking specific forms of behaviour to the functioning and the neurobiology of our brain or specific regions therein. It is important to mention that these two approaches, i.e. behavioural testing and neuroscientific investigations, only together provide a complete framework and this is also the reason why both deserve sufficient attention when discussing topics like language aptitude.
This chapter provides the reader with a more general introduction as to why it is essential to explore the neural underpinnings of language learning in the human brain. Most importantly, the reader will be provided with some basic information on the human brain and our current possibilities of further exploring it. As I applied one of the most commonly used neuroimaging methods, namely magnetic resonance imaging (MRI), I will briefly explain the benefits of and possibilities of investigation with MRI and similar methods. Together with an introduction to language-related brain regions, the concept of neuroplasticity, which is often named together with FLL, shall be discussed. This is integral since the studies that were conducted for this PhD thesis do not measure neuroplasticity and thus changes in the brain due to experience or training. The studies focus on biologically determined structures in the auditory cortex (henceforth AC) that are most likely independent of experience or training. Moreover, a considerable part of this chapter deals with the particularities of the human AC and how functional or structural variation therein might be important for music or language processing, and logically language aptitude. More specifically, HG, comprising the primary AC, and studies investigating differences in HG morphology and functioning shall be dealt with in sufficient detail.

3.1. An Introduction to Methods of the Cognitive Neurosciences

Despite the technological advances and more than half a century of intense research, our brain is still considered the greatest mystery of all times. Earlier, if researchers wanted to know what specific areas of the brain looked like, e.g., to investigate brain lesions or individual variation, they had to cut open the heads of dead people and explore the anatomy of their brains post-mortem. For decades, this was the only possibility of gaining information about the structure of the brain. Since the invention of computerized tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET), i.e. since the advent of modern neuroimaging research, in vivo brain imaging has replaced post-mortem investigations. By providing images of the brain in a couple of minutes without severe negative consequences for the individual, these methods have caused the large hype of the cognitive neurosciences since the second half of the 20th century.

As already indicated, there are two major approaches to an investigation of the human brain, namely by doing structural or functional analysis. Investigating structural brain scans
such as those produced by MRI or CT give us the opportunity to closely examine the structure of our brain and detect abnormalities or individual variation. This can be very useful for clinical assessment or diagnosis, for instance. More recently, anatomical variation has also become an interesting topic for research in different fields, such as learning disabilities and FLL. Before functional neuroimaging was widely applicable, most information about the normally functioning brain was inferred from abnormal processing after a lesion, for instance. This was particularly the case for knowledge on speech and language processing, which was primarily inferred from studying aphasic patients and the lesions or brain damage they had suffered. Nowadays, both lesion studies and functional neuroimaging in typical and atypical populations are used to shed further light on the functioning of our brain, especially with regard to language (Wise & Price, 2006).

Before delving into these exciting topics, a brief introduction on the conduction and analysis of structural and functional neuroimaging methods shall be given. Specific attention will be paid on MRI as it is also the method of choice in the two studies that will be presented later on.

3.1.1. Structural Neuroimaging Methods

X-ray technology was the first method used for imaging the structure of a living human brain and it was also the only available technique until the 1970s. Given the high radiation exposure and the poor two-dimensional view, the invention of CT was a great advance. CT combines x-rays with computer technology and allows to gather intensity information of different positions around the patient’s head. Thus, a three-dimensional view can be created using matrices. It further allows for the visualization of internal structures such as grey and white matter and the differentiation of ventricles (Purves et al., 2001). CT was in fact the first non-invasive radiological method that was capable of generating tomographic images of each part of the human body with superimposition of adjacent structures (National Research Council & Institute of Medicine, 1996).

The importance of CT in clinical imaging has even exceeded the most optimistic hopes of researchers, especially for diagnostic and therapeutic measures (Liguori et al., 2015). Although CT scans were a great advancement in the fields of medicine and psychology and triggered off a new era in brain imaging, the later developed MRI, which uses a strong magnetic
field, has become an even more valuable imaging tool. Especially for research purposes, MRI is the method of choice over CT. In contrast to CT scans, MRI is non-invasive and harmless – two obvious advantages it has over CT. Another advantage of MRI over CT is that views of the brain can be obtained from any angle and many different contrast mechanisms can be applied. “By adjusting the pulse parameters […] the same scanner can generate images in which grey and white matter are invisible but the brain vasculature stands out in sharp detail” (Purves et al., 2001).

MRI has seen a remarkable growth as a tool for studying brain function and brain anatomy and pathology in the past years. The magnetic field at the basis of MRI is created by a scanner and expressed in units of Tesla (the most frequently used MRI scanners worldwide using 1.5 or 4.0 Tesla scanners, although 7.0 and 11.0 Tesla scanners are becoming available now). The scanner uses changing magnetic gradients and oscillating electromagnetic fields (pulse sequences) to create images. Different tissue properties may be detected with different pulse sequences. MRI scanners are tuned to the frequency of hydrogen nuclei in the human body. These atomic nuclei absorb electromagnetic fields and they then emit electromagnetic energy (Huettel, Song, & McCarthy, 2004). All that subjects need to do in an MRI is to lie down on a table and their heads are placed at the very centre of the scanner to receive structural images.

3.1.2. Functional Neuroimaging Methods

Functional brain imaging allows for a closer investigation of what exactly the brain is doing during a certain task and which areas are involved in this process. The most frequently applied functional neuroimaging methods are based on a simple principle. In fact, the brain needs a large amount of our energy resources, namely around 20% of circulating glucose. Active nerve cells consume more glucose and oxygen and the local blood flow thus increases in active brain areas. Neuroimaging methods such as PET and fMRI are capable of detecting and mapping these changes in blood flow in our brain. In other words, through the use of these methods, we can see differences between more active and less active brain regions during different tasks, such as reading words or listening to a song (Purves et al., 2001).

Most studies that have revealed interesting insights into speech and language processing have used PET, fMRI, EEG (electroencephalography) or MEG (magnetoencephalography)
The obvious advantages of the latter two are that they can record electrical activity directly and that they have a very high temporal resolution. Whereas almost every psychology department at universities around the world owns their own EEG (cheap but cannot be used for an in-depth view into the brain), MEGs are a financial burden and are seldom found. They provide much higher spatial and temporal resolution but cost even more than most MRIs. Unlike MEG and EEG, fMRI and PET have a much greater spatial resolution and given that they depend on a change in local blood flow, they are in particular of interest for functional anatomy and physiological responsiveness (Wise & Price, 2006).

As most linguists and psychologists have access to fMRI and it can be applied an unlimited number of times in contrast to PET, this method has been most frequently used in the past although the technique is extremely noisy and makes linguistic investigations of continuous speech and speech production quite challenging. PET has two major advantages in this respect, one being that the method is silent and the second that it is equally sensitive to signal throughout the brain as fMRI. The latter, however, allows event-related measurements (i.e. the response of the brain to a specific stimulus), which is impossible using PET scans, and, as already mentioned, an unlimited number of scans can be acquired using fMRI as no radiation is required for this method compared to PET (National Research Council & Institute of Medicine, 1996; Wise & Price, 2006).

3.2. Foreign Language Learning from a Neuroscientific Perspective

Since brain imaging has become widely available for research at the end of the 20th and the beginning of the 21st century, investigating typical and atypical language processing in the human brain has become a wide and popular field in academic research. Whereas processes and mechanisms underlying speech and language, especially FLL, have generally been one of the hot topics in neuroscientific research, language aptitude has received comparably little attention until recently. There is no doubt that the cognitive neurosciences offer a great opportunity to gain better insights into what is going on in the brains of very gifted individuals. Biedroń, who has spent the last years providing us with a clearer and more comprehensive view
of the most recent developments in language aptitude research, highlights the necessity of taking research in the cognitive neurosciences into account in FLL research:

[...] at the moment there is no academic discussion about individual differences, especially cognitive factors, without neuroscientific research. Applied linguists and language educators cannot fail to include these breakthroughs from neuroscience into FL aptitude research [as] neurolinguistics has become the most informative and ground-breaking source of knowledge about SLA. (Biedroń, 2015, p. 15)

As she makes quite clear, there is no reason why the dazzling results achieved through neuroscientific research should be neglected in FLL-related disciplines, including language aptitude. A major strength of the cognitive neurosciences lies in the possibility of combining the results gained through neuroimaging with behavioural research and trying to interpret findings to update constructs and theories. We should therefore see it as an excellent additional resource that certainly helps to gain a much more advanced understanding of the underlying properties and mechanisms of people’s behaviour.

Regarding language, the mental faculty that allows us to communicate with one another, various attempts were made to find the often-described core language region(s) in the human brain. At the beginning of neuroscientific research many decades ago, it was believed that parts of the brain each possess one specific function or are responsible for one specific behaviour and thus, damage to this area would result in the loss of the ability or trait. This approach was similar to thoughts in phrenology (introduced by advocated by Joseph Gall, 1796, and Johann Spurzheim; Critchley, 1965), a pseudo medicine of the early 19th century, often referred to as the first popular mapping of brain function. Phrenologists claimed that bumps and indentations in individuals’ skulls provided information about the magnitude of character traits and behaviours supported by the underlying brain area. Being unable to measure cortical volume directly, they just assumed that increases in brain size in certain areas could be translated into measurable bumps on the skull, e.g., thieves would have a flattening of the skull above the specific area claimed to support ‘honesty’ (Huettel et al., 2004). The initial novelty of these functional maps wore off and scientists started to dispute their validity. It soon became clear that every behaviour, skill or ability required the interaction of numerous areas and could not be attributed to one specific location only. This holds true for language as well, meaning that there is not one single region responsible for the storage and processing of language. As a
matter of fact, when an individual hears or speaks a language, for instance, a wide network of different areas spanning over both hemispheres is activated, which has been proved by lesion studies and functional neuroimaging (Price, 2012).

But how is it possible to measure a concept like language aptitude in our brain? This is a difficult question and it needs to be emphasized at this point that research involving brain structure or function is an outstanding opportunity to better understand how our brain works and why it is built the way it is. One has to be cautious, however, since most neuroimaging studies are correlational and can only allow us to see that one ability is related to a specific part of the brain or activity of this very region. In other words, we cannot explain the reason and directedness of these relations but only interpret our results according to our hypotheses. This is easier with a related example. Let us imagine that the results of our studies show that skilled readers use a less distributed network during reading compared to less skilled readers. By looking at brain activations only we cannot determine whether it is the well-developed skill that is associated with a certain pattern of activation or the pattern that is at the basis of the development of excellent reading skills. Likewise, through the application of neuroimaging to linguistically gifted or less gifted subjects, we cannot determine whether functional activation differences in highly gifted and poorly gifted individuals during a number of tasks are a result or the reason for aptitude. This means that we always have to be careful about the conclusions we draw and make sure not to take results as causal evidence. Nevertheless, there is no doubt that we can gain considerable information about a certain skill by looking at it from a neuroscientific perspective and there are several ways in which an outstanding ability like language aptitude can be investigated in more depth.

3.3. The Human Brain

3.3.1. A Terminological Introduction

Before an analysis of the most important language-related regions in the brain is provided, it is essential to introduce the reader to basic neuroscientific and brain-related vocabulary that will be used frequently in the subsequent chapters (Table 6).
### (Cerebral) Cortex
A collection of neurons that forms a thin sheet of about 2-5 mm thickness at the brain’s surface. In MRI scans it usually appears as a dark gray strip as the outer boundary of the brain beneath the skull and the meninges.

### Grey Matter (GM)
A generic term for a collection of neuronal cell bodies in the brain. When a freshly dissected brain is cut open, neurons appear grey. Also, in MRI scans, the layers of neurons appear grey in contrast to the connections between neurons (i.e. axons).

### White Matter
A generic term for a collection of axons (long connection fibers between neurons) in the central nervous system. They are wrapped in myelin, which is responsible for the electrochemical transmission of signals from one region to the other. Due to the white appearance of myelin, axons appear white and are referred to as white matter.

### Sulcus
A type of convolution in the brain. Sulci are the grooves in the surface of the cerebrum.

### Gyrus
The second type of convolution in the brain. Gyri are the bumps in the cerebrum.

---

**Table 6** Important neurological terms with their corresponding definitions necessary for an understanding of neuroscientific studies (Bear, Connors, & Paradiso, 1996, pp. 23-25 & 163–175, 2015).

Why do we need to know and be able to differentiate between all these terms? When we look at the human brain, we immediately realize that it is not an even surface but is characterized by many sulci and gyri. Sulci and gyri vary considerably in individuals, even to an extent that it could be argued that the brain is like a snowflake – each brain is unique. Still, the main structures and functions of the human brain are basically the same in every healthy individual. Therefore, whenever I talk about structural variation in the brain, I refer to sulci and gyri with different names, depending on their exact location in the brain.
In an MRI scan, contrasts in matter are shown in black-and-white contrasts. There are parts of the brain that appear whitish on a scan, in contrast to layers appearing grey and the black background. The so-called grey matter (henceforth GM) is a thin layer of the brain that is convoluted and contains the 100 trillion nerve cells estimated to be found in every single brain. GM consists of different types of nerve cells and this thin outer layer of the brain is called ‘cortex’. Whereas GM consists of neurons, white matter refers to the subcortical layer of axons (fibres connecting neurons) and glia cells (cells responsible for the stability of the brain functioning as a glue). When analysing the morphology of certain brain regions, I always speak of the surface volume of GM or its thickness (then termed cortical thickness). In different parts of the brain, this layer may vary in thickness and consists of layers comprising different types of nerve cells. This is the subject of investigation of cytoarchitectonic studies, in which the microstructure of the cerebral cortex is the focus. Any research looking at the connection between different parts of the human brain requires a look at structural connections of white matter, on the other hand (Bear et al., 2015; Friederici & Wartenburger, 2010).

### 3.3.2. The Development of the Human Brain

Mammalian brain development is an exciting field, which has seen significant progress in the last decades. In a human being, brain development begins in the third gestational week and is a series of dynamic and adaptive processes with the aim of promoting the emergence and differentiation of new neural functions and structures (Stiles & Jernigan, 2010). The process takes place during the embryonic period throughout adolescence. Actually, the basic structures of the brain and the central nervous system are developed during the embryonic period. The production of neurons, for instance, starts 42 days after conception and no neurons are produced after mid-gestation. Until the infant is born, only rudimentary connections between neurons are established and only major fibre pathways are complete. The brain is far from fully developed and established at the time of the birth of an infant. During the pre-school period, the brain increases four-fold in size and already reaches around 90% of adult volume at age 6. Changes in the structure of the brain, e.g., grey and white matter changes, occur until late adolescence paralleling changes in functional organization (ibid, 2010).

During the developmental stages, numerous processes must take place. Different types of neurons have to be built and they have to connect with other neurons in order to pass on
information, i.e. to build a network. Every single neuron can connect with more than 1 000 neurons in other parts of the brain and it is estimated that an adult brain possesses no less than 60 trillion connections. Why and how does the brain shape itself in the form of sulci and gyri? “The enfolding of the mature brain is thought to be an adaptation to the dramatic growth in the size of the brain during the course of evolution […] allow[ing] large brains to fit in comparatively small cranial vaults that had to remain small to accommodate the birth process” (ibid, 2010, p. 329).

Most basic steps, such as the production of neurons, the formation of the brain and the development of the structure of the brain and basic connections between neurons take place in the foetus. Other processes, in contrast, require a long time after birth and continue throughout childhood. These processes are for instance synaptic exuberance and pruning. Precise connections between neurons need to be formed and as no or not sufficient environmental feedback (the brain is said to be ‘experience expectant’) can be acquired in the first weeks, months and years of life, there is an exuberance of connections formed. The number of synapses reaches its maximum at nearly twice the number of connections found in the adult brain, before it slowly declines to normal adult levels during childhood and adolescence. This decline is called as ‘pruning’ and only the most important and efficient connections are kept by the brain of the infant. Concerning the structure of the brain, the gross brain structure only undergoes subtle changes but processes like myelinisation or pruning take place a long time after the birth of a child. What needs to be emphasized is that both genetic components and the environment play an essential role and only an ongoing interplay between the two can lead to successful neural development (ibid, 2010).

Although most steps of brain development take place in the infant brain, the adult brain, though being considered fully developed, can also undergo minor changes. These minor changes can be caused by learning processes or be the result of a lesion or brain damage, and are subsumed under the term ‘neuroplasticity’.

3.3.3. Neuroplasticity

The human brain has an outstanding ability to change or reconfigure both functionally and anatomically in response to cognitive demands, environmental stimulation, lesions or behavioural experience. FLL has been linked to neuroplastic changes in numerous studies,
dealt with in several reviews (e.g., Li, Legault, & Litcofsky, 2014). In their review, Li et al. investigate the processes and mechanisms driving experience-dependent anatomical and functional changes. Similarly, the recovery of functional neural networks and the thereby caused compensation of a cognitive skill after damage to a certain brain region can also only take place due to the plasticity of the human brain (C. K. Thompson, 2000). This property of the brain is termed ‘neuroplasticity’.

When an adult acquires a new skill, e.g., they learn a foreign language, cortical changes take place on different levels. These changes are mostly of synaptic nature, i.e. connections between neurons and on a more general level areas are developed or strengthened, and irrelevant connections may be lost, such as it is the case in pruning. Neuroplasticity is most often associated with changes in grey matter or the development of synapses in the adult brain due to intense learning phases in the acquisition of a skill (e.g., Draganski et al., 2004; Münte, Altenmüller, & Jäncke, 2002).

While there is no doubt that functional and structural changes take place consistently throughout the life span, some studies have reported that in longitudinal comparisons, macrostructures, such as morphology, gyrification and shape of AC, for instance, are expected to remain highly stable over almost a decade (Seither-Preisler et al., 2014). Therefore, AC morphology as measured and investigated in the present thesis is not related to changes due to the neuroplastic human brain. Rather, the structural variation encountered in subjects is biologically determined and cannot be easily altered through training or experience, which is why none of the gross structural variation described in the upcoming chapters should be confounded with other brain mechanisms related to neuroplasticity.

3.3.4. The Core Language Regions in the Human Brain

The core language areas are those regions in the human brain that contribute centrally to language processing and the damage of which results in impairments that hinder the production (i.e. speaking or writing) or comprehension (i.e. listening or reading) of language. In other words, if one of the core language regions in the cerebral cortex is injured or damaged, a number of linguistic impairments will occur (Petrides, 2014). Numerous areas of the human brain contribute to the processing of language at a certain point of time and depending on the
specific task performed, but there are some regions that are of greater importance and more frequently involved than others.

No discussion on language areas of the brain would be complete without a short introduction to the works of Paul Broca and Carl Wernicke who were pioneers at their time regarding the investigation of language impairments. Drawing their knowledge from post-mortem studies and brain lesions, they were the first to come up with hypotheses explaining where language was situated in the brain. Paul Broca became famous for first describing the posterior two thirds of the inferior frontal gyrus (IFG; later termed ‘Broca’s area’) as the seat of the ability to speak (Broca, 1865). Shortly after, a German colleague named Carl Wernicke reported brain lesions in two patients who could not understand spoken language despite fluent articulation (Wernicke, 1874). He assumed that the impaired area in this case, namely the superior temporal gyrus (STG), must be crucial for language comprehension (see summary in Tremblay & Dick, 2016). This area then became known as ‘Wernicke’s area’, the seat of language perception and comprehension. These two areas, Broca’s and Wernicke’s area, were then combined in the first classical model of speech language, known as the ‘Broca-Wernicke-Lichtheim-Geschwind model’ (see also Hickok & Poeppel, 2004) or simply ‘Wernicke-Lichtheim model’ (Graves, 1997). In the past years, however, this fixation on Broca’s and Wernicke’s areas and the corresponding model has been criticized, nicely displayed in the title of Tremblay and Dick’s paper “Broca and Wernicke are dead, or moving past the classic model of language neurobiology” (2016). In their paper, they argue that brain research has made substantial progress in the categorization of specific brain areas and there is clearly a considerable lack of precise anatomical definitions of the two areas. Therefore, they suggest replacing the terms with more precise anatomical definitions.

Let us get further into the details of language processing. What happens when we perceive speech? The posterior STG is activated (also known as primary auditory cortex, PAC, and Wernicke’s area). As speech also contains numerous nonverbal cues, activation will relate to more than the phonetic cues only. This will allow mapping of long-term representations of word meaning (semantic system) and syntax, as well as motor areas for speech production. It has been shown in nonhuman primates that there seems to be a hierarchy of processing involving the PAC and parabelt auditory areas. Projections then involve the medial temporal lobe (memory), the prefrontal cortex via the superior longitudinal fasciculus, external capsule and the uncinated fasciculus. Anatomically, there is no agreement with regard to the anatomical
route leading from speech perception to the semantic system, i.e. the part of the brain where meaning is stored. Some studies have shown that supramodal access to meaning is provided by the temporo-parieto-occipital junction/AG and the inferior lateral temporal cortex. It has also been proved that the anterior part of the temporal lobe is important for mapping language onto meaning. This has been confirmed by other studies in which activation in this area was reduced given a speech impairment, lesion in the temporal lobe or a distortion of the speech signal (Wise & Price, 2006).

What happens when we produce speech, on the other hand? Interestingly, speech and nonspeech sounds lead to an equally strong response in the planum temporale (PT), an area connecting the auditory areas (STG) to the IPL (SMG). It is the key element of the temporo-parietal junction. It has therefore been proposed that the PT, known to play a role in music perception, has a number of functions unrelated to speech processing. Sequences of speech are known nowadays to be held in STM and heard speech can then be mapped onto articulatory movements and gestures. Reciprocal projections involving the arcuate fasciculus, i.e. the connection between superior temporal areas and inferior frontal areas, seem to explain the pathway from the posterior auditory regions to the prefrontal/premotor regions. Just perceiving speech has also been proved to activate articulatory muscles which are responsible for speech output (Wise & Price, 2006).

Leaving aside this issue, almost all core language regions are situated around the so-called Sylvian or Rolandic fissure (see Figure 7; the Sylvian fissure is located between the major blue area, the IFG, and the green area, the AC) and are described as belonging to the peri-sylvian region. These core areas include regions in different lobes of the brain and in both hemispheres. In the frontal lobe, core areas are the IFG, earlier known as Broca’s area, and the pre- and postcentral gyri, which are responsible for the physical aspects behind speech production, e.g., orofacial muscles. In the parietal lobe, the supramarginal (SMG) and angular gyri (AG) are essential areas for speech comprehension and meaning. Last but not less importantly, auditory areas are situated below the Sylvian fissure and in close proximity to the SMG. They comprise the superior and middle temporal gyri (STG and MTG) and are vital regions when it comes to processing any kind of auditory input such as music or speech. These core language areas have not only been confirmed by lesion studies, i.e. studies with patients having suffered from damage to one of them and then showing a deficit in language production.
and/or comprehension. Evidence also comes from stimulation studies and most recently fMRI, EEG and MEG studies (Petrides, 2014).

In their model explaining the cortical organization of speech processing, the researchers G. Hickok and D. Poeppel argue that a dual-stream model (see Figure 7) can best explain the functional anatomy of language (Hickok & Poeppel, 2000, 2004).

**Figure 7** Anatomical areas involved in the dual stream model proposed by Hickok and Poeppel (figure adapted from Hickok, 2012). The left hemisphere is presented on the left, the right hemisphere on the right. Blue: dorsal stream (articulatory network and sensorimotor interface), pink: ventral stream (combinatorial network and lexical interface), green: spectrotemporal (auditory) analysis, yellow: phonological network. *pIFG* posterior inferior frontal gyrus, *PM* premotor, *a/pMTG* anterior/posterior middle temporal gyrus, *a/pITS* anterior/posterior inferior temporal sulcus.

At the earliest stage of speech processing, namely spectro-temporal analysis, ACs (bilaterally) are required. After this initial analysis, the system diverges into two broad streams, namely a ventral and a dorsal stream. The dorsal pathway is responsible for mapping sensory
or phonological information on articulatory motor representations. For this process, an interaction between the sensory motor areas (parietal-temporal) and frontal and motor areas (IFG, premotor cortex, anterior insula) is necessary. The ventral stream, on the other hand, maps sensory or phonological representations onto lexical conceptual representations, connecting posterior and anterior temporal areas. While the ventral stream is largely bilaterally organized despite slight computational differences between the hemispheres, the dorsal pathway is dominantly left-hemispheric. The main idea behind this model is that the acoustic speech network needs to interact with both conceptual and articulatory-motor systems, preferably at the same time. This model with two concurrent parallel anatomical pathways for language processing has been confirmed in studies with healthy and language-impaired subjects (e.g., Kümmerer et al., 2013; Saur et al., 2008).

No account of the core language regions is complete with a brief comment on the major connection between the aforementioned vital language regions, the arcuate fasciculus. The arcuate fasciculus is a fibre tract (white matter connections) that originates in the temporal lobe and curves around the Sylvian fissure, finally projecting to the frontal lobe and thus connecting the earlier believed key regions of language in the brain, Broca’s and Wernicke’s area. However, the arcuate fasciculus seems to be strongly left-lateralized, with very few individuals possessing an STG pathway in the right hemisphere (Glasser & Rilling, 2008). These authors also found that in those subjects who possessed an STG pathway on the right side, the pathway did not overlap with phonological activations that were found more anteriorly in mid-STG/STG compared to the left side. They suggested that unlike the left hemisphere, the right auditory association cortex (secondary auditory areas) in STG, vital for phonological processing, were not directly connected to frontal areas.

### 3.4. Heschl’s Gyrus and the Auditory Cortex

The primary focus of the present thesis is on the AC, which is why HG and the PAC are dealt with in more detail in this chapter. Referring to the aforementioned model, STG and the AC situated therein are the starting points of the analysis of speech and primarily responsible for spectro-temporal analysis, as well as for phonological processing. A detailed account on the variation found in AC and more specifically in HG so far requires a profound understanding of the various functions this brain region fulfils and its neuroanatomy.
The AC is the location of auditory processing in the human brain. Research on auditory processing in the mammal, including more detailed descriptions and analyses of their anatomical features, started in the second half of the 19th century. It was first revealed that responses to auditory stimuli could be detected in monkeys and chimpanzees in the temporal lobe, more specifically in the superior temporal plane (i.e. in STG). The primary auditory region contains a representation of the basilar membrane that can be found in the cochlea in the inner ear. Concerning the gross anatomy of AC, it can be said that it occupies a great part of the STG, as well as a lower bank of the lateral sulcus in the temporal lobe (Hackett et al., 2001). Auditory areas are usually classified into (1) primary (belt) and (2) secondary (parabelt) areas that perform different tasks (Hickok & Poeppel, 2004; see Figure 8 for exact locations; primary areas are marked in grey, secondary areas in blue and multimodal hubs required for further phonological analyses in pink).

Figure 8 Lateral view of the left hemisphere of the superior temporal lobe. Colours used in the figure correspond to rough cytoarchitectonic parcellations into primary (grey), secondary fields (blue) and the multimodal integration cortex (pink; taken from Hickok & Poeppel, 2000). AG angular gyrus, MTG middle temporal gyrus, STG superior temporal gyrus, pSTP posterior superior temporal plane, aSTP anterior superior temporal plane, H Heschl’s gyrus, I Insula, FO frontal operculum, PCG precentral gyrus, PO parietal operculum
The temporal lobe (in the above illustration starting next to and beneath the insula in blue and going back to the end of the pink region, comprising all regions to the most anterior and most inferior parts of the brain), which is home to the AC, consists of three gyri, namely the superior (STG), the middle (MTG) and the inferior temporal gyri (the latter is not shown in Figure 8; Petrides, 2014). This superolateral region is seen as a centre for auditory and multisensory comprehension of linguistic input. “Deep within the lower bank of the lateral fissure, approximately at the level of the postcentral gyrus, lies the transversely oriented Heschl’s gyrus [...] where the primary and secondary auditory cortical areas are located” (Petrides, 2014, p. 12). Figure 9 shows HG and anterior STG in the right hemisphere (red) and the left hemisphere (blue) from a sagittal (a), an axial (b) and a rotated frontal, 3D constructed view (c).

**Figure 9** AC/HG from different viewpoints in the author’s brain. (a) Sagittal view of left HG, (b) axial view of right and left STG including HG, (c) 3D reconstructed AC and STG of both hemispheres; right hemisphere marked in blue, left hemisphere marked in red.
Architectonic studies so far have not provided evidence as to whether primary auditory regions reside on the same specific area despite large structural differences in HG morphology – including single gyri, gyri with partial or complete duplications and multiple gyri occurring in the AC of an individual. Early studies (Penhune, Zatorre, MacDonald, & Evans, 1996; Rademacher, Caviness, Steinmetz, & Galaburda, 1993) suggested that PAC occupies always the first division of HG duplications. Da Costa et al., 2011, on the other hand, found that anatomical variants of HG are part of a continuum rather than very distinct subtypes and argued that PAC spans both divisions of duplications of HG. This was confirmed by Zoellner et al. (2019), who found that the exact position of PAC on HG showed high intra-individual variability, often PAC was also located on HG duplications, reflecting earlier work (e.g., Rademacher et al., 1993, 2001).

Functionally, AC areas are connected to frontal, central and inferior parietal areas through the arcuate fasciculus as already mentioned. Rauschecker & Scott (2009) proposed a dual auditory processing model, in which they provided a complete framework of auditory analysis and the interaction of involved regions (Figure 10).

![Figure 10 Model of dual auditory processing and connections of auditory areas to the sensory system. The model was designed to close the loop between speech perception and production (red: postero-dorsal stream, green: antero-ventral stream; taken from Rauschecker and Scott, 2010). PMC premotor cortex, CS central sulcus, IPL inferior parietal lobe, IFG inferior frontal gyrus, STS superior temporal sulcus](image)
According to their model, core and parabelt areas of AC are in contact with either frontal regions, like IFG, or parietal regions like IPL. Communication with the premotor areas, on the other hand, occurs secondary and through the other regions. The terminology that they report is in analogy to the model of language processing suggested by Hickok and Poeppel (Hickok, 2012; Hickok & Poeppel, 2000, 2004) – the interaction between AC and IPL is named the dorsal stream, while the connection between AC and frontal areas is displayed as the ventral stream.

To conclude, AC or auditory regions are rather umbrella terms including primary and secondary auditory areas. The first is the location of primary spectro-temporal analysis and includes HG or the PAC, also termed the core auditory region. The secondary auditory areas are also referred to as para-belt regions, comprising mainly anterior and posterior portions of STG. Whereas the first is situated in the medial part of HG, the second is estimated to start in its lateral part, spreading to the STG. HG is considered a distinct morphological entity (Petrides, 2014).

3.4.1. Research on Heschl’s Gyrus and the Human Auditory Cortex

Already at the beginning of neuroscientific research via post-mortem investigations, intra- and inter-individual asymmetries in gyrification and size were found in AC of subjects. The first one to point out this fact was R. L. Heschl during the anatomical investigation of 1087 post-mortem brains (Heschl, 1878). After Heschl’s insights, further post-mortem studies confirmed the morphological variation he had found. In other words, it became clear that sulci and gyri in AC did not only vary substantially between individuals, but also in the same individual there were major structural differences in right and left AC. It was only decades later that investigations using MRI could confirm the results found in the post-mortem analyses (Auerbach, 1906; J. Benner et al., 2017; Brodmann, 1909; Geschwind & Levistky, 1968; Penhune et al., 1996; Rademacher et al., 2001; Serrallach et al., 2016; v. Economo & Horn, 1930).

Differences in individuals were extensive, hard to categorize and only very recent research has explicitly dealt with the morphological variation found in HG in different populations. Among the populations that have received widespread interest with regard to morphological variation of HG are musicians, schizophrenics and dyslexics. The findings of
these studies will be dealt with more explicitly in the following paragraphs. For now, it suffices to mention that different populations show different patterns of the size and number of gyri in their AC.

3.4.1.1. Functions of Right and Left Heschl’s Gyrus

In the past years researchers have hotly debated the topic of language lateralization, agreeing that AC in the left was mainly responsible for the processing of linguistic information, whereas AC in the right was supposed to be primarily involved in music-related auditory processing. These assumptions, however, have been under critique, especially by studies finding bilateral activations for both language and music processing in AC.

One of the first studies pointing towards a specialization of left auditory areas for speech perception was conducted by Liégeois-Chauvel, de Graaf, Laguitton, & Chauvel (1999). They argued that in order to perceive speech, the human brain needed cortical mechanisms capable of analysing and encoding changes in frequency (spectral) in acoustic signals. These subtle differences in the timing of acoustic elements were thought to constitute the basis for language perception more generally. In their study, they studied the responses of right and left AC to simple consonant-vowel (CV) syllables only differing in voice onset time. As they only found the left hemisphere to be involved in this process, they argued that left HG reflected sequential processing of the different components of syllables, not only in speech sounds but also generally with non-verbal sounds. However, in their study, they tested only 17 subjects and all of them were epileptic, which clearly limits the overall validity of their results for healthy populations.

In their review, Zatorre, Belin, & Penhune (2002) argued that speech and musical sounds take advantage of a modulation of acoustic parameters and thus differently exploit acoustic cues. They suggested that while speech depends on rapidly changing broadband sounds, tonal patterns that are essential for music require processing of smaller and more precise changes in frequency. Therefore, they concluded that temporal resolution should be more efficiently processed in left auditory cortical areas (due to the left-lateralization of language) and spectral resolution should show an advantage of processing in right auditory cortical areas (Zatorre & Belin, 2001). The same research group further suggested that there were functional hierarchies, meaning that the core areas bilaterally (i.e. in left and right AC)
could be involved in the early stages of auditory processing, while higher word-recognition mechanisms is likely to be associated with anterior and ventral left auditory regions (Zatorre et al., 2002).

In another study, the same research group (Hyde, Peretz, & Zatorre, 2008) investigated passive listening to pure-tone melodic-like sequences with varied pitch in individuals. Their main aim was to assess the relative contribution of left and right auditory regions in this process. The results of their study showed that the right PT (at the intersection of secondary AC to temporo-parietal areas) was linearly responsive to increasing pitch distance. The left side, on the other hand, was constant and reflected only large pitch changes. They concluded that there must be a relative hemispheric specialization as fine-grained differences in pitch were more reliant on right AC. Zatorre and colleagues became known as advocates for a model in which left auditory fields are considered more sensitive to temporal structure, whereas right auditory fields are considered more closely linked to spectral structure (see also McGettigan & Scott, 2012).

Focusing on language, Poeppel, Guillemin, Thompson, Fritz, & Braun (2004) used different auditory tasks in a PET study to explore involvement of left and right auditory areas in lexical decision, categorical perception and frequency modulation direction discrimination. They reported that frequency-modulated tones, categorical perception and non-word/word judgments led to bilateral activation of AC, even if the lexical decision tasks led to stronger responses in the left and frequency-modulated tones to stronger responses in the right. They argued that right-hemispheric, non-primary AC might confer a slight rightward advantage for the processing of syllable-sized acoustic units, supported by findings from other studies (Giraud et al., 2007; Luo & Poeppel, 2012). In a position paper (Poeppel, 2003), the author brought forward a theory informed by the patterns of oscillations termed the ‘Asymmetric Sampling in Time Hypothesis’. It proposed that the left temporal lobe preferentially sampled information over short time windows (e.g., duration of phonemes; ~20-40ms), compared to longer time windows (length of syllables and intonation profiles; ~150-250ms) in the right hemisphere. Suggesting a time-based perspective, the model proposed that an incoming speech signal had a bilaterally symmetric neural representation at an early representational level, which was elaborated asymmetrically depending on the length of the temporal integration window (McGettigan & Scott, 2012; Poeppel, 2003).
In another early work on the lateralization of AC functions, Tervaniemi & Hugdahl (2003) summarized findings on the structural and functional basis of human brain lateralization in the auditory modality. They criticized that there was a lack of correspondence between structural and functional measures of asymmetry (e.g., no right-left asymmetry anatomically but functionally). They stated that the PT was larger in the left hemisphere and related to speech perception and could thus be seen as the epicentre of a mosaic of left-hemispheric language areas. Given that the left-larger-than-right asymmetry was earlier confirmed in great apes, they hypothesized that it could be related to sound-based language development. They further argued that many studies had shown that PT was activated by speech and non-speech sounds (often slight asymmetries in the superior temporal sulcus were found as well) possibly due to early auditory processing, which took place in the PT. Later language-related processing then would occur in an extended, multimodal area in STS and MTG. On the behavioural level, they claimed a substantial right-ear advantage, which has been mostly interpreted as a behavioural measure for left-temporal lobe processing superiority for phonological stimuli. Contralateral projections were stronger and more preponderant and could thus block the processing of ipsilateral projections. Schneider et al. (2009) reported that ipsilateral projections were stronger involved in PAC than contralateral projections. Possibly, this is different in secondary auditory areas, where contralateral projections are stronger.

McGettigan & Scott (2012) further reviewed cortical asymmetries in speech perception, focussing on differences in auditory processing between the two hemispheres. They argued that there was no simple left-right dichotomy of acoustic processing as often assumed. They also mentioned that it should be considered that there might be large discrepancies in languages concerning how many separate phonemes are used, how much languages rely on rhythm, melody and pitch. Therefore, different languages could be more dependent on one or the other hemisphere and core or para-belt area due to their specific nature. Moreover, there are even languages that communicate without phonemes at all (e.g., through humming or whistling; Brandt, Slevc, & Gebrian, 2012). McGettigan and Scott argue that the right temporal lobe could be strongly selective for specific aspects of the speech signal, whereas the left temporal lobe could be indifferent to such differences (please read the full review for more details).

To summarize, while the view that language is primarily associated with the left hemisphere and music with the right has been upheld for a long time, auditory processing does
not follow this easy left-right dichotomy. Both the right and left AC are essentially involved in the processing of different acoustic features related to language and music.

### 3.4.1.2. Specific Structural Properties of Heschl’s Gyrus

While differences in the function of right and left AC have been extensively discussed in the literature, little information has been provided so far as to the relevance and occurrence of structural variation and potential relationships between structural variation and behaviour. Those studies focusing specifically on the anatomical properties of right or left AC primarily dealt with disordered populations (e.g., patients with tinnitus, schizophrenia, learning disabilities, attention deficit hyperactivity disorder), or individuals with high musical abilities (e.g., professional musicians). Only a small amount of studies, in contrast, looked into the occurrence of specific HG morphotypes in very large populations with the aim of providing further insights into the neurobiology of AC.

The most common form of HG is a single gyrus, in which a gyrus is bordered by two sulci, the first transverse sulcus and Heschl’s sulcus. However, HG also appears in the form of common stem duplications (CSD), meaning that HG is partially divided along its length by a sulcus, termed sulcus intermedius. This sulcus can be short or long and of varying depth. It is vital to distinguish a CSD from a complete posterior duplication (CPD). A CPD, in contrast to a CSD, does not possess a sulcus intermedius but presents as two complete gyri along the medial two thirds of its mediolateral extent. In the case of a CSD, the sulcus does not extend down to the medial base of HG and leaves two divisions connected by a common stem, explaining the origin of the term (see also Da Costa et al., 2011). Apart from the three morphotypes, many studies have confirmed the occurrence of multiple gyri, often also termed multiplications or multiple duplications (sometimes abbreviated as MD). Regarding hemispheric differences of AC, the right hemisphere has been consistently reported to be more variable in comparison to the left. Marie et al., (2015) analysed an impressive sample of 430 individuals and showed that duplications of HG (that is more than one single HG) occurred in a bit more than a third of individuals in the left hemisphere but in half of all participants in the right hemisphere. A concrete schematic representation of the four HG morphotypes, including multiple gyri as a separate type (MD), was provided in Benner et al. (2017; see Figure 11).
Figure 11 Schematic representations of the four main HG morphotypes: a) single gyrus, b) CSD, c) CPD and d) MD (adapted from Benner et al., 2017). *aHG* anterior HG, *PP* planum polare, *FTS* first transverse sulcus, *SI* sulcus intermedius, *HS1* first Heschl’s sulcus, *HS2* second Heschl’s sulcus, *cHS* first complete Heschl’s sulcus

As already mentioned, most research on the neuroanatomy of AC has been performed in atypical populations. Schneider et al. (2009) found a relationship between HG volume and chronic tinnitus, finding smaller medial HG volumes ipsilateral to the affected ear in tinnitus patients. They further found that there was also a relation to tinnitus-related hearing loss. In another study, the same research group discovered that left HG GM volumes correlated with a preference in fundamental pitch listening (Schneider, Sluming, Roberts, Bleeck, & Rupp, 2005). Spectral pitch listeners, on the other hand, had a dominant right lateral HG. Furthermore, they found that the absolute size of the neural HG substrate was linked to musical aptitude (pitch perception) as well (Schneider et al., 2002, 2005).

Investigating longitudinal changes in GM in children with autism, Prigge et al. (2013) found an increase in GM volume in HG with age, mostly due to global hemispheric GM effects. Another finding of their study was that GM volume of right HG significantly differed between typically developing and autistic children. In schizophrenic patients, right and left HG volumes were reduced, in contrast to both controls and bipolar patients (Hirayasu et al., 2000). Additionally, HG was found to be functionally involved in auditory hallucinations in schizophrenic patients, with a trend for a higher number of duplicated HGs in those patients.
with hallucinations and higher HG volume in the right hemisphere in the same group of patients (Hubl et al., 2010).

In another study by Schneider and colleagues (Seither-Preisler et al., 2014), the authors reported that anatomical characteristics of HG were correlated with reading and spelling skills in a large sample of children. Children with AD(H)D, for instance, showed reduced volumes of HG and enhanced volume of PT, possibly explaining auditory processing problems associated with attentional and literacy problems. In the same study, children with musical training showed considerably larger HGs, especially in the right hemisphere. In the same longitudinal study, right-hemispheric AC morphology could also be linked to dyslexia and attention deficits (Serrallach et al., 2016).

Despite the large number of studies looking into HG variation in atypical populations, several studies have also linked HG morphology to high musical ability or musical training (e.g., Schneider et al., 2002). Schneider and colleagues, for instance, reported that musicians presented with larger GM volumes in the right hemisphere. Apart from these structural findings, Schneider et al. further found a strong relation between the neurophysiological response to auditory stimuli (individual peaks as measured through MEG) and the anatomical parameters. Moreover, GM volume correlated with music aptitude in non-musicians and amateurs in their study. In Benner et al. (2017; the same research group but a different sample) musicianship was also found to correlate with right-hemispheric AC morphology. Looking into AC morphology in 41 musicians, they distinguished between four morphotypes (the three aforementioned and MD) and found that the vast majority of musicians (90%) possessed HG multiplications, either in one hemisphere (39%) or even in both (51%). They concluded that HG multiplications occurred more often in professional musicians as compared to the general population and comprised a functional unit within HG.

To summarize, research on AC functioning has extensively discussed the roles of right and left AC, suggesting that the right hemisphere is responsible for finer resolution in the frequency domain and less important for temporal resolution. Left AC, in contrast, has been reported to be sensitive for the processing of fast acoustic events (e.g., speech sounds) and thus possesses very high temporal resolution. Concerning structural variation in left and right HG, several disorders could be linked to differences or anomalies in AC morphology. This was the case for tinnitus patients and patients with schizophrenia, but also for children and teenagers
with AD(H)D and dyslexia. While studies on these disordered populations usually showed a decrease in grey matter in HG, the occurrence of multiplications of HG were significantly linked to musicianship in other studies.
4. The Relationships between Working Memory, Musicality and Language Aptitude

There is substantial evidence that musicality and WM capacity play a vital role in the acquisition of foreign languages. While the first is expected to positively influence FLL through enhanced perception of speech-related cues and consequently facilitates speech perception, the second is often suggested to be a prerequisite for overall learning processes that positively influences the acquisition of any foreign language. In the following paragraphs, both concepts, musicality (musical ability) and WM, will be discussed comprehensively and special emphasis will be put on their possible links to foreign language aptitude. Since the present thesis explores behavioural links, as well as neuroanatomical ones, also the present chapter will approach the two aforementioned concepts from various angles, including a neuroscientific approach.

4.1. The Role of Working Memory for Language Aptitude

Even if we do not realize it most of the time, we remember elements like digits or letters, just to give two examples, frequently in our daily life. Whenever we need to keep a new pin code or password in mind, we save that information in our temporary storage system – our so-called ‘working memory’/WM. We need WM in all tasks, from language comprehension (e.g., combining earlier sentences with later spoken sentences and making meaning of the whole conversation), problem solving (e.g., in arithmetic processes) or planning (useful in daily lives). Unfortunately, this storage system is limited, which means that we can only store a limited amount of information, and a couple of hours or days later we will be very likely to have forgotten the code or password if it was complex and if we did not consciously attempt to keep it in mind – or write it down. The truth is that without rehearsal, elements of either written or spoken nature do not enter long-term memory (henceforth LTM) and are lost since WM can only store them for a limited amount of time.

One could write an entire thesis on the role of WM for FLL and an introduction to this topic at this point can only serve as a glimpse into this highly intriguing topic (for a great
overview, see Wen, 2016). Due to its centrality for concepts such as intelligence or overall cognitive capacity, research on WM has blossomed in the last two decades in a variety of disciplines, two of them being FLL and SLA. Given the strong link between WM and language learning reported in a plethora of studies, aptitude researchers like Zhisheng Wen and Peter Skehan have even argued that what makes certain individuals so good at learning foreign languages is mainly their high WM capacity, promoting it as the essential predictor for FLL capacity in individuals. This proposal of the mid-1990s has been termed “Working Memory as Language Aptitude” (Wen, 2019) and was positively received by the research community. Other researchers have even gone as far as to state that WM can be seen as a pure measure of children’s (a priori) learning potential (Alloway & Alloway, 2010), highlighting its significance for all learning processes, not only for language learning.

Before delving into this discussion and exploring the interconnectedness between WM and FLL, it is necessary to throw a closer look at what it is, what it consists of and in how far these components of WM could play a role in the acquisition of other language-related skills. Additionally, there will be a separate section on the confusion regarding the terms short-term memory (henceforth STM) and WM, mostly used interchangeably by researchers. Although the neurological underpinnings of WM will also be discussed marginally in the present chapter, this link will not be explored in the studies of this thesis.

4.1.1. Now and Then: The Development of Working Memory Models

According to Wen (Wen, 2016), W. James’ (1890) demarcation of primary and secondary memory (reported in Baddeley, 2007) was the first and quite early trace of current WM models. James distinguished between primary and secondary memory, the first being equivalent to today’s concept of STM the second being our LTM. In his view, primary memory was limited, and knowledge was transferred via rehearsal to secondary memory that presumably had unlimited storage capacity. In a similar vein, D. O. Hebb (1949) proposed that every individual possesses a STM and a LTM and this was supported by studies showing that information is lost if it is not rehearsed and thus enters LTM (Cooper, 2005). It is believed that STM is the result of the temporary activation of neural connections between cells, while LTM is most likely the result of permanent synaptic changes (Andrade, 2014).
In fact, the question whether there are two separate systems, one for temporary and one for long-term storage, was highly debated until the 1960s when many new models entered the market. Nowadays we know from lesion studies, for instance, that LTM might be fully intact, while STM is severely impaired (for examples, see Baddeley, 2010). For instance, there are patients who fail to keep any new incoming information although they are still able to access memories from childhood or the time before their accident. This is most likely possible due to the different regions in the brain associated with WM/STM and LTM.

An early quantification of the capacity limit of WM was proposed by Miller, (1956), who claimed that seven was the magical number and young adults could not keep more than seven ‘chunks’ (letters, digits, words, other units) in memory at a time. Later it turned out that the length and phonological complexity of these so-called chunks largely influences storage capacity. As an example, there is a lower span for longer or phonologically more complex words, while one can keep a larger number of short or morphologically simple words in mind (Service, 1998). Cowan (Chen & Cowan, 2009; Cowan, 2001) finally proposed that WM has a capacity of around three chunks of meaningful items in young adults (fewer in children and older individuals), confirmed by results of his own studies. Supporting Cowan’s suggestion, Gilchrist, Cowan, & Naveh-Benjamin (2008) reported that young adults were only capable of recalling three to four long verbal chunks, like idioms or short sentences. In Chen & Cowan (2009) they used word lists and word pairs and showed that individuals remembered only three units, which led to their suggestion that the central WM faculty was limited to three to five chunks for adults. According to Wen (2019), both proposals remain the best-known competitors: Miller’s ‘7 plus or minus 2’ and Cowan’s ‘4 plus or minus 1’. Wen (2019) argues that “most WM theorists would agree to accept that the capacity of WM can be set at between four chunks of unrelated items (Cowan’s view) and seven units of information (Miller’s view)” (p. 190). It should be mentioned, however, that WM capacity develops with age from childhood to adulthood and slightly decreases with age (Myerson, Emery, White, & Hale, 2003). It is estimated to peak at around 30 years of age (Alloway & Alloway, 2013).

4.1.1.1. Baddeley and Hitch’s Working Memory Models

Although some researchers like Hebb had already claimed beforehand that there must be a system storing information temporarily in the human brain, Alan Baddeley and colleagues were the first who used the term ‘WM’ to refer to this temporary storage system (still mostly
used interchangeably with STM) more than 30 years ago. Baddeley & Hitch (1974) were the first who proposed a multi-component model of WM replacing this idea of a unitary system of memory that is capable of storing every form of input (e.g., visual, spatial). This was their so-called classical model of WM (generally referred to as the three-component model; see Figure 12). It consisted of three major components, namely a central executive (in the center of Figure 12) that was assisted by two subsidiary storage systems, a phonological loop (right-hand side) and a visuospatial sketchpad (left-hand side). The phonological loop was responsible for storing sound and language, and the visuospatial sketchpad was responsible for storing visual and spatial input. The function of the central executive was that of a control organ that directed attention to what was being stored at what time, and it was thus the central element of the three-component WM model.

![Figure 12](image)

**Figure 12** The three-component WM model proposed by Baddeley and Hitch (proposed in 1974; taken from Baddeley, 2003a), consisting of the visuospatial sketchpad, the central executive and the phonological loop interacting with each other while temporarily maintaining any incoming information.

According to Baddeley and Hitch’s first model, WM was a limited capacity system that served for temporary maintenance and storage of incoming information and differentiated between auditory and visual input. They further emphasized its role in reasoning, language comprehension and overall learning (see also Baddeley, 2003b).

To understand how the different components of WM interact, it seems plausible to consider an everyday situation as an example. If you walk through the street, you will encounter a variety of different objects, colours and shapes in the form of people walking, cars passing, house doors on each side and the like. However, our brain is not made for taking in all that information and storing it – especially since there is limited capacity, only what is relevant or
important to us deserves ‘space’ in our brain, so to say. What you remember afterwards might be what some of the people looked like that you passed but you would probably have difficulty remembering which cars you saw, the colours of all houses in the street, whether there were a lot of bikes next to the street and the like, unless you were effectively directed toward that information for some specific reason. For instance, you would surely remember how many scooters passed you if you were thinking about buying a new scooter. Likewise, a woman who wants to get pregnant is often said to literally report hundreds of pregnant women or women with children on the street. The fact is, we only see, hear and remember information that catches our attention because it is relevant to us at a specific time – and for being relevant, we need to direct our attention to that element, whether this happens voluntarily or not is not important.

As already mentioned, the phonological loop is the element of the WM model that is responsible for the storage and maintenance of auditory and thus language input, which makes it the most important component to look at in terms of foreign language aptitude. Baddeley (2003a,b) states that the phonological loop consists of a phonological store and an articulatory rehearsal mechanism. Both are considered equally important for storing incoming linguistic information. The so-called articulatory rehearsal mechanism is analogous to subvocal speech, meaning that it is equal to hearing yourself repeating a sequence in your head. Through re-articulation and retrieval, memory traces can be refreshed, and elements are longer stored in the phonological loop, which is useful if we want to remember something. Imagine you had to remember a 10-letter password that did not convey any meaning, but consisted of arbitrary letters (e.g., consider the difference between the passwords ‘mypassword’ and ‘bWmQR4tl6L’). The only way to remember the second password without writing it down is by using the articulatory rehearsal mechanism and by explicitly repeating the letters over and over again in your mind.

The problem with the phonological loop and language is that only a certain number of items may be stored since a point is then reached at which the first item will have faded before it can be rehearsed and thus refreshed. Baddeley (2003a, 2017) holds the view that articulation and rehearsal mechanisms are strongly linked and studies have shown that the longer the words (one vs. five syllables in their study), the fewer elements an individual can remember. In other words, the longer the elements you are trying to keep in your WM, the less storage space you have and your immediate memory span declines. He argues that longer words will lead to
slower and more complex rehearsal processes and thereby less information can be stored (see 4.1.1).

The three-component model of WM was welcomed and widely acknowledged in the research community because it explained well how information was temporarily stored in our brain. However, when taking a closer look at the model, it becomes clear that it suffers from some major shortcomings that should be addressed by a model of WM. One of these major shortcomings is that it failed to answer questions as to how STM/WM and LTM were connected (problem 1). This connection is vital given that all our knowledge is stored in LTM and the model clearly lacks an association between these two key memory storage systems. It is particularly surprising that this link was not included despite the earlier published work by Atkinson & Shiffrin (1968), who presented a model of information processing according to which information from all modalities (auditory, visual, spatial) passes sensory registers before entering STM. The authors claimed that STM was not only vital for encoding information, but also for feeding it into and out of LTM, suggesting an intimate link between the two (see also Wen, 2016). Moreover, the model by Baddeley and Hitch failed to explain how the human brain managed to maintain serial order in WM (problem 2) and some researchers further criticized that Baddeley and Hitch did not explain the biological function of the phonological loop (problem 3).

To overcome the first problem, a new model which also included LTM had to be developed. For the second critique point, please refer to Baddeley (2003a) since a discussion of that topic would go far beyond the scope of this thesis. With regard to the third problem, Baddeley and colleagues argued that the biological function of the phonological loop must have been to facilitate language learning. By claiming that phonological loop capacity was a reliable predictor of an individual’s ability to learn a second language, he clearly argued for an evolutional component to this part of WM. After years of intense research on WM and the aforementioned issues raised concerning their first model, Baddeley and Hitch came up with changes to their original model. The novel model of WM (Baddeley & Hitch, 2000) included an additional component that was introduced to explain the link between STM and LTM to eliminate the first problem of the old model. This storage entity connecting WM with LTM was termed the ‘episodic buffer’ (see Figure 13).
Figure 13 The multi-component model of WM (Baddeley and Hitch, 2000; taken from Baddeley, 2003a): the central executive controls the phonological loop, the episodic buffer and the visuospatial sketchpad, which are then linked to LTM, more specifically to visual semantics, episodic LTM and language.

The novel model of WM still relied on the central executive as the main controlling organ with direct access to the visuospatial sketchpad, the episodic buffer and the phonological loop. All three components, however, were also linked to LTM but to different parts therein. The phonological loop was shown to possess direct access to language storage, while the visuospatial sketchpad directly accessed visual semantics and the episodic buffer served mainly as a storage unit connecting the central executive and the episodic LTM storage system. The new model explained the criticized lack of connection between STM and LTM and gave a fuller account of human memory as compared to the previous model. It should also be mentioned that this model additionally distinguished between fluid and crystallized systems, two terms originally formulated in the framework of intelligence theory. ‘Fluid’ generally refers to constantly changing and updating systems and is therefore linked to biological parameters (e.g., processing speed, capacity), whereas ‘crystallized’ describes our core
memory that cannot be altered immediately, equalling our already acquired knowledge (Cattell, 1963). When considering this new model, it is interesting that no direct connection between the phonological loop and the visuospatial sketchpad in this visualization. This is surprising since reading, for instance, is most likely to rely both on phonological and visual information processing in a very short amount of time. Therefore, an interaction between the two in periods shorter than milliseconds – even if through the episodic buffer - seems very likely.

In the present chapter, only two models were dealt with comprehensively. It shall be noted that different camps proposed different WM models but not as successfully as Baddeley and colleagues⁴. Especially Cowen, one of the most famous and influential WM researchers of the past years, has suggested models of his own. However, as Baddeley put it “[i]n practice, Cowan and I tend to agree with each other on most aspects of our respective theories, despite using very different theoretical metaphors, his derived from initial focus on attention, my own influenced by studies of short-term verbal memory” (Baddeley, 2010, R140).

4.1.1.2. A Terminological Clarification: Short-Term Memory vs. Working Memory

A complete account of the concept of WM should include information on related terms that are often used interchangeably in studies. Therefore, it is vital to consider potential differences in terminology between (phonological) WM and (phonological) STM.

Cowan (2008) assessed the strength of evidence for three types of memory in human beings, namely LTM, STM and WM. Regarding the difference between WM and STM, he suggested that WM was a temporary information storage system that was closely related to STM. Thus, he argued that the two were not completely distinct from one another and considered STM to be a substantial part of WM. He stated, however, that other researchers (e.g. Engle, 2002) saw WM as the attention-related aspect of STM.

According to Unsworth & Engle (2007), the easiest distinction between WM and STM had been made with reference to simple and complex memory tasks, the simple being associated with phonological STM (Kolb & Whishaw, 2008) and the second with WM. In this

⁴ Please refer to the extensive works of Wen (2012, 2016) and the book edited by Wen et al. (2019) for more detailed accounts on the various WM models that have been brought forward to date.
manner, the first only involved the repetition of input (e.g., repeating non-words), whereas the second also involved the manipulation of input (e.g., repeating a number of digits in the reversed order). Hence, the capacity to simply repeat elements was considered to be a measure of STM, while any task with higher processing demands involving the manipulation of stored speech input was attributed to WM (see also Gathercole, 1999; Hale et al., 2011). This was supported by findings from Kail and Hall (2001). In contrast to the easy verbal memory tasks, the complex tasks were claimed to assess functions of the central-executive, mostly responsible for the regulation of attention (Sáfár & Kormos, 2008). Results of studies, on the other hand, have disagreed with the distinction between STM and WM, finding that both simple and complex span tasks measure the same processes (e.g., Unsworth & Engle, 2007).

In Baddeley’s model, which is also the theoretical framework I am adapting in this thesis, STM is not mentioned explicitly as it is seen as an essential subpart of WM. The first comprises the two slave systems, the phonological loop and the visuospatial sketchpad and the latter refers to the whole multi-componential model. Although the research literature is still blurred by the ambiguous use of these two constructs, many researchers acknowledge the coexistence of both, some seeing WM as a replacement of STM, others seeing WM as a theoretical conception of STM (Aben, Stapert, & Blokland, 2012). In their review, the authors conclude that there are strong arguments in favour of a large or complete overlap of STM and WM, also in line with recent neurobiological findings. They also criticize that differences in results between complex and simple span tasks might be merely due to higher demands and consequently additional processing demands.

To summarize, STM and WM are still often used interchangeably and are most likely two overlapping constructs that heavily influence one another. I agree with Baddeley’s proposal that STM is part of WM but will use the term WM throughout this thesis to refer to our temporary storage system. Most recently, however, WM has been suggested to require a phonological and an executive component that need to be distinguished when assessing and talking about WM capacity.
4.1.2. The Phonological-Executive Model of Working Memory for Language Aptitude

Few researchers have attempted to integrate most recent developments of WM into FLL situations and contexts. Wen (2016, 2019), was one of the few, and designed his own phonological-executive WM model to combine language learning processes with WM and help shed further light on the significance of WM for language aptitude. Wen points out that it is vital not to speak of WM as a unitary concept and to apply different tests that actually measure quite distinct parts and components thereof (Figure 14; his ideas are based on the aforementioned distinction also between STM and WM).

![Figure 14 Wen’s phonological-executive WM model explaining the relation between phonological and executive WM with regard to language and tasks associated with the two (taken from Wen, 2019).](image)

For verbal WM, he proposes the distinction between phonological WM and executive WM and suggests to test and report results accordingly in studies. In his model, the architecture consists of five levels integrating LTM (Level I – including declarative and procedural memory), WM components (phonological and executive WM respectively, as well as related
visuospatial WM; Level II) and specific WM functions associated with SLA domains and skills (e.g., rehearsal, updating; Level III). The last level, Level IV, comprises tasks assessing the domains and specific skills located on Level III (for details, see Figure 14).

In the two studies presented in this thesis, three tests will be used to measure WM, namely digit span forward, digit span backward and non-word span. Since digit span forward and non-word span are clearly simple memory span tasks and only require phonological short-term store and articulatory rehearsal, they are considered to measure phonological WM. Digit span backward, on the other hand, requires the manipulation of stored input and is therefore not identical to the other tasks with respect to processing demand and complexity. Still, it is not equal in complexity to the tasks mentioned in Figure 14, such as reading span or n-back span). Given the slightly more complex nature of digit span backward but the lack of functions of executive WM involved, I argue that digit span backward is at the intersection of phonological WM and executive WM.

To summarize, whenever only the term WM appears from now on, I am speaking of phonological WM. According to the works of Wen, it refers to a temporary storage system requiring both a phonological short-term store and an articulatory rehearsal mechanism. The later presented studies do not provide information on the role of executive WM, nor on visuospatial or LTM in any form.

4.1.3. Working Memory and the Brain

In the past years, neuroscientific research on WM has become very popular, and specifically the development of WM in infants during childhood and into adolescence has been a matter of discussion. Although the neural correlates of WM will not be addressed in the studies presented later on, it is still interesting to see the overlap of brain regions involved in WM on the one hand, and language comprehension and production on the other hand.

With respect to the exact interaction between regions involved in WM and the mechanisms behind memory storage, no consensus has been achieved so far. A dominant theory of WM is the so-called ‘Persistent Activity Hypothesis’, which basically states that prefrontal areas are responsible for both encoding and maintenance of WM items through sustained elevated activity (Fiebig & Lansner, 2017). In other words, during WM processes, the prefrontal cortex shows higher permanent engagement. However, there exist alternative
models of WM maintenance based on synaptic plasticity. They suggest that short-term non-associated synaptic facilitation is at the base of WM. In their recent work, Fiebig and Lansner take up an old idea and argue that all WM-related processes, including encoding, maintenance and reactivation, manifest in discrete oscillatory bursts rather than persistent activity. Following this argumentation, they suggest that another potential explanation is a fast-expressing form of Hebbian synaptic plasticity. To summarize, there is no consensus on the neurofunctional properties underlying our ability to temporarily encode and store input, but various theories have been brought forward to date.

Regarding the neuroanatomy of WM in the cerebral cortex, most studies providing us with insights into the exact location of areas important for WM in the brain were lesion studies from patients who had lost their ability to store new incoming information. Baddeley (2003b; see Figure 15) summarized information from lesion studies by Vallar & Papagno (1995) and Smith & Jonides (1997) in a review on the neurobiology of WM. The two lesion studies by the aforementioned authors reported that the central executive was mainly processed in the dorsolateral prefrontal cortex (DLPFC) bilaterally, while processing of phonological information relied heavily on the left SMG/IPL only. The articulatory rehearsal mechanism, an essential part of phonological WM, involved the motor cortex and the IFG close to the Sylvian fissure in the left hemisphere, as well as the right cerebellum. Last, the inner scribe for spatial rehearsal depended on processing of the right hemisphere only and involved the superior parietal lobule, the IFG and the premotor cortex. A major shortcoming of lesion studies, however, is that it is impossible to know whether a lesion in a specific area might have influenced the functioning of other surrounding areas or the underlying functional connectivity between areas in the near vicinity. Therefore, it is advisable to be cautious when interpreting the consequences of lesions to specific brain regions.
Figure 15 According to Baddeley, the regions most involved in WM are the right cerebellum, the right dorso-parietal cortex, the inferior frontal cortex bilaterally, the fusiform gyrus in the left hemisphere and the left SMG (taken from Baddeley, 2003a). AR articulatory rehearsal, CE central executive, PS phonological storage, VC visual cache/storage, IS inner scribe (spatial rehearsal).

Apart from lesion studies, studies with PET and fMRI investigating WM capacity have gained increasing popularity in the past decades. Smith & Jonides (1997) explored the neural architecture of WM dysfunction through PET applying spatial and verbal memory conditions. To sum up their findings, the mental blackboard we depend on for performing computations for higher-level cognitive processes needed an array of brain regions to work in concert for correct functioning. Christophel, Klink, Spitzer, Roelfsema, & Haynes (2017) state that there is an ongoing debate on where contents are generally stored, but current views heavily emphasize the role of prefrontal, parietal and sensory cortices in WM. In the past years, non-invasive brain stimulation methods have become increasingly popular and researchers have tried to improve WM skills through non-invasive brain stimulation of the DLPFC to test its involvement in the aforementioned functions. A meta-analysis conducted by Brunoni & Vanderhasselt (2014) included 13 studies applying non-invasive brain stimulation, that is either transcranial direct current stimulation (tDCS) or transcranial magnetic stimulation (TMS), on that very region to increase WM performance measured by an n-back task. They found that tDCS stimulation only led to faster answers, thus an improvement in reaction times, but not in performance. Studies applying TMS, however, succeeded in improving all WM measures in
individuals whose DLPFC was stimulated, suggesting that the dorsolateral prefrontal cortex is a vital area for WM in the human brain.

Since WM is closely related to academic achievement, reading skills and mathematical competence, developmental scientists have become intrigued by exploring how WM develops in children. With a wide range of methods applied and numerous small-scale studies reporting different results, consensus is far from reachable now. Still, I would like to mention some interesting studies that have shed more light into that poorly understood area so far. Bathelt, Gathercole, Johnson, & Astle (2018) assessed WM in children and teenagers between 5 and 16 years and found that greater involvement of the corpus callosum and posterior temporal white matter could be associated with the central executive of WM. In the older children, the central executive seemed to be linked to the occipito-temporal cortex. Ullman, Almeida, & Klingberg (2014) on the other hand, found that structural and functional MRI data could explain variance in WM capacity in children, even 2 years after measurement. Current WM correlated with fronto-parietal engagement, whereas future capacity could be explained by basal ganglia and thalamus function and structure.

To summarize, the WM network is quite vast, as is the case with language more generally, and there are numerous overlapping areas. Apart from lesion studies, it is quite challenging to test which specific regions are necessary for which integral part of WM (e.g., the phonological loop, the subvocal rehearsal mechanism). The problem with lesion studies is that lesions are often vast, many skills can be impaired and the network may be interrupted as connectivity is disrupted while the region may not be implicated in that specific task at all. This is true in particular due to the large overlap with other skills and the difficulty of assessing only WM. Furthermore, even if fMRI studies have found several specific regions that are involved in short-term storage of information, we know that the brain is far too complex as to rely on such a modulatory basis.

4.1.4. Working Memory as an Important Predictor for Language Aptitude

Following the definition and history of WM, its different mechanisms and its underlying neurofunctional and neuroanatomical bases, I will now explore previous research on the relationship between WM and FLL, as well as language aptitude. In the past three decades, research on WM and its significance for language learning has blossomed, and various
learning disorders (e.g., reading impairments) have been linked to deficits in WM, supporting the link between linguistic skills and WM.

Baddeley and colleagues (Baddeley & Hitch, 1974; Baddeley, Gathercole, & Papagno, 1998; Baddeley & Hitch, 2000; Papagno, Valentine, & Baddeley, 1991) were some of the first to suggest a strong link between parts of the WM system and novel word learning by introducing the phonological loop as the word learning device per se. When they first investigated this component in more depth, they claimed that it was critical for the learning of foreign languages since language learning heavily depended on vocabulary learning. This hypothetical link has been broadly discussed and supported afterwards by the former (Baddeley, 2017) and was taken forward by language aptitude researchers around the world (Ellis, 1996; Kormos & Sáfár, 2008; Miyake & Friedman, 1998; Wen et al., 2017). Following Baddeley’s ideas, the phonological loop was thought to be capable of explaining the observed inter-individual variety of language learning outcomes in different contexts. There has been evidence, however, that verbal STM (a sub-part of WM) and verbal WM might be relevant for different aspects of language learning, e.g., the first for vocabulary learning and the second for grammatical aspects (Verhagen & Leseman, 2016). In other words, in how far the different components of WM interact when it comes to biologically determined language aptitude, is still a matter of debate.

Zhisheng Wen and Peter Skehan (Wen & Skehan, 2011) are two well-known and highly appreciated FLL researchers who have focused on investigating the concept of WM and its relationship to language aptitude in their past works. In their paper of 2011, they first argued overtly for an incorporation of WM in the construct of language aptitude, which has received widespread support from researchers in similar fields. WM is a quite young concept that comes from cognitive psychology. According to the authors, WM should be included in language aptitude theories for three reasons, namely (1) WM capacity differs between learners of languages, (2) WM is a concept that can be measured reliably and validly and (3) WM certainly plays a role in SLA/FLL, which could make it an essential part of language aptitude as well. The first argument has been supported by a plethora of studies, meaning that foreign language learning success, both in the receptive and productive domain, and WM capacity are correlated. The second argument, despite claiming that WM is a unitary concept, is also partly true. Researchers use different tests for the assessment of WM, even though we cannot be completely sure to what extent such tests measure the involvement of the various components
thereof (refer to Baddeley, 2000). As we will see, the third argument provided by Wen and Skehan is strongly supported by the behavioural data we present in the upcoming chapters. Linck, Osthus, Koeth, & Bunting (2014) performed a very large meta-analysis confirming the robust, positive relationship between WM and L2 comprehension and production. However, he also concluded that work still remained to be done, specifically to advance theoretical models and further fully understand the contributions of WM to L2 outcomes.

Given the importance of WM for language learning, it seems worthwhile to discuss the findings of studies concerning their interaction. As Sáfár & Kormos (2008) explain, the component of Hitch and Baddeley’s model that has received most interest in FLL research is the phonological loop because it is easy to measure and highly relevant for speech. Most studies have used a digit span or non-word repetition task in which a subject must recall numbers or non-words of various lengths. Studies investigating language ability and its relation to WM have confirmed the role of the latter on a variety of linguistic abilities concluding that it leads to faster and more successful first and second language learning (Kormos & Sáfár, 2008; Linck et al., 2014; Miyake & Friedman, 1998). With respect to first language acquisition (L1 acquisition), Gathercole and Baddeley (1989) first reported that vocabulary size in L1 could be predicted by WM. Masoura and Gathercole (1999), who reported a link between high phonological STM (WM) and vocabulary knowledge in 4-year-old children’s native language, confirmed these findings. Another study looking into the relationship between WM/STM and L1 learning found that STM for words (word repetition) could predict the mean length of utterances as used by children between 1 to 3 years of age (Blake, Austin, Cannon, Lisus and Vaugahn, 1994). Similarly, phonological WM was found to be associated with longer, grammatically more complex sentence productions in children aged 3 in a study by Adams and Gathercole (1995).

Concerning FLL, the British camp (Baddeley and Gathercole) has shown keen interest in the role of the phonological loop for L2 vocabulary learning, compared to the North American camp (Daneman, Engle and Miyake) that has focused on differences in the central executive (see discussions in Sáfár and Kormos, 2008, and Wen and Skehan, 2011). Baddeley, Gathercole and colleagues (e.g., Papagno, Valentine and Baddeley, 1991; Papagno et al., 1995; Masoura and Gathercole, 1999) concluded that the phonological loop is the language learning device per se (e.g., see Baddeley, Gathercole and Papagno, 1998) since it is responsible for the acquisition of new phonological forms, i.e. essential for acquiring new vocabulary. Majerus,
Poncelet, van der Linden and Weekes (2008) confirmed that lexical learning is not only associated with STM in monolinguals (e.g., as suggested by Baddeley et al., 1988) but also in bilingual learners in their own study. Martin and Ellis (2012), for instance, also reported that phonological STM and WM (simple vs. complex span tasks) were not only strongly associated with L2 vocabulary, but also with the acquisition of new grammatical forms. Going beyond vocabulary learning as well, Leeser (2007) reported that reading comprehension was predicted by performance on WM tests. Similar results were found by O’Brien et al. (2006), who concluded that the role of phonological STM was much greater in less proficient learners. In other words, those with less proficiency relied much more heavily on their WM. They suggested that WM must be of greater importance in early stages of learning and less proficient learners. While we still cannot predict the exact amount of variance of learning behaviour explained by WM to date, numerous studies have confirmed that WM not only affects L1 acquisition, but it also positively influences SLA (Juffs and Harrington, 2011).

Regarding the involvement of WM in foreign language aptitude, some researchers have suggested that high WM is mainly responsible for high language aptitude and thus by far the strongest predictor thereof (Miyake & Friedman, 1998; Sawyer & Ranta, 2001; Wen et al., 2017; Wen & Skehan, 2011). Given that studies have proved that learners with higher WM capacity significantly outperformed age-matched peers when learning another foreign language (Biedroń, 2015; van den Noort, Bosch, & Hugdahl, 2006), it seems logical to assume that high language aptitude depends on high WM. However, there is major disagreement on the specific WM components involved, how they can be reliably tested, and whether and how they really relate to the known components of language aptitude (Jacquemot & Scott, 2006). Additionally, other studies have questioned this substantial impact WM is often reported to have on language aptitude as they could not confirm this strong link between the two in their own studies (Winke, 2013). Moreover, there has been an ongoing debate as to whether an inclusion of WM in a concept of language aptitude would mean that language aptitude is alterable through training and experience, as has been suggested to be the case with WM. However, there is still disagreement on the meaningfulness and efficiency of WM trainings and thus also partly to the dynamic nature of WM (refer to the meta-analysis by Melby-Lervåg & Hulme, 2013).

To conclude, language learning is a learning process that is highly reliant on temporary storage of speech material and subsequent retrieval (speech production). Doubtlessly, an
individual has to be capable of maintaining information in WM while doing a number of cognitive tasks simultaneously in order to successfully master the acquisition of a (foreign) language. Therefore, it might not be fully adequate to speak of WM capacity as the only relevant predictor for foreign language learning success. Rather we should consider it a prerequisite for overall learning since no learning takes place without WM (Alloway & Alloway, 2010). Still, the extent to which language learning is dependent on different components of the WM system (e.g., central executive, phonological loop) and whether all WM tasks are suitable for measuring these components remains controversial. In addition, one aspect that has been poorly explored to date is also how WM is involved during different contexts of learning (e.g., implicit, explicit, with or without instruction) and at different stages of learning (e.g., at the very beginning vs. achieving high proficiency).

4.2. The Relationship between Music and Language Aptitude

Language and music are two partly distinct and at the same time equivalent auditory phenomena. Language is a means of communication that is not only necessary for the expression of emotion and thought but also for sharing knowledge and intentions. The same applies at least to a certain extent to music. There has been a long lasting and hot debate between philosophers and scientists as to the question whether language and music have a common or separate origin (for a good summary, see Besson & Schön, 2001). This debate has been re-examined in the past decades with the mutual interdependence of both being agreed on by researchers of different domains. The question that has been at the core of these debates was whether music and language evolved from one another (e.g., language developing as a means of communication out of music or vice versa) and are therefore of the same origin and how the two influence one another.

Language and music share various similarities and the most interesting point regarding definitions is that definitions for language often apply for music as well and vice versa – supporting an evolutionary perspective linking the two. Music and language are both structural systems, i.e. they consist of sequential events that unfold in time and they follow a hierarchical organization. Furthermore, both phenomena are conveyed by sounds and appear to be specific to humans. The last point, however, has been shown to be only partly true since animals like birds, for example, also communicate through singing, even if their singing is not equal to the
complexity of linguistic structures in speech and language. Therefore, the long debate about what makes language so typically human has most often focused on the aspect that language is a human universal, that is, that no other species has produced a means of communication as complex and astonishing as human language. Moreover, also functional similarities exist between the two auditory phenomena. Both rely on intentionality (that is, they are produced for a certain purpose) and require a theory of mind (i.e. an understanding of the relationship between mental states and actions; Wellman, 2018). Regardless of the culture or locality, every child is able to learn to speak and sing and to intuitively comprehend music. Language is also based on various musical qualities, such as prosody or vocal timbre, while language appears very often in combination with music (e.g., singing). Another functional aspect relating the two is that memory is further involved in both systems. These are just some of the similarities that are striking even when considering the differences (Besson & Schön, 2001; Jackendoff, 2009; Jäncke, 2012).

From what we know now, music does not only influence FLL on the behavioural level, but it also changes the way our brain responds to sound and thus it changes the way our brain responds to speech and language. At this point, I would like to refer the interested reader to a recently published paper written by Turker, Sommer-Lolei and Christiner (2019) for the Austrian Journal für Begabtenförderung on the shared similarities of language learning and music and the pedagogical implications we can draw from the findings of studies to date. Before looking at the interaction between musical abilities and FLL/SLA, it seems worthwhile to specify the terms ‘musicality’ and ‘musical ability’ since they are often used interchangeably in the literature and have also been used as synonyms in the present chapter so far.

4.2.1. Finding a Definition for Musicality / Musical Ability

When speaking about the intricate relationship between music and language, one should clearly distinguish between two very different concepts, namely musicality/ musical ability (mostly used synonymously) and musical training/experience. The first refers to one’s skills at mastering certain music-related tasks, like pitch or rhythm perception, and the second refers to gained experience through singing or playing an instrument (also known as musical expertise or musicianship). An individual may have practically no musical training and expertise but still have high musical ability in a pitch discrimination task, for example. Similarly, an individual
might score weakly on a rhythm discrimination task despite years of formal musical training, which is rather the exception than the rule, however. Overall, the terms designating an individual’s potential for acquiring musical skills are even slightly more confusing and less differentiable compared to the terms for language aptitude.

Levitin (2012) states that different abilities, such as reading or remembering music, listening to attributes of musical performance or even playing a musical instrument seem to be intuitively included in our subjective profile of musical abilities. When we think of someone who has good musical abilities, we usually imagine this person as being good at the aforementioned. But some people seem to be outstanding when it comes to improvising, while others perform marvellously in a choir or an orchestra. In dictionaries, one finds very different definitions of the term ‘musicality’. It is sometimes described as “a knowledge of music and the ability to be sensitive to it” (Macmillan Online Dictionary), or “a sensitivity to, knowledge of or talent for music” (Merriam-Webster). If we look at scholar approaches to the topic, we find even more eloquent definitions that are in accordance with the ideas presented before. Hoving, ten Cate, Peretz, & Trehub (2015) describe musicality as “a natural, spontaneously developing trait based on and constrained by biology and cognition” (p.12; supported by Ravignanni, Thompson, & Filippi, 2018). Gembris (2013), an expert in music research, defines music aptitude as a potential that develops through practice and is influenced by the environment. It enables us to experience music emotionally, understand it intellectually and be able to create it by singing, playing an instrument, composing or improvising.

Gordon (1989, 1999, 2001) dedicated most of his life to investigating musical aptitude, an individual’s predisposition to ‘learn’ music. He suggested that “music aptitude is a measure of a student’s potential to learn music [although] everyone has some level of music aptitude, is musical, and can learn to listen to and perform music with some degree of success” (Gordon, 2001, p. 4). According to him, 1 or 2 individuals in 100 have an exceptionally high music aptitude (2% as suggested in Gembris, 2013). He suggests that after the age of nine, before which levels of musicality change with the quality of the musical environment, the level of music aptitude remains stable and is practically unalterable (the stages before age nine and at age nine are termed developmental and stabilized music aptitude; Gordon, 1989). He considered audiation, i.e. the inner capacity to imagine and perceive fundamental aspects such as rhythm, pitch, timbre, loudness, the basis of music aptitude and compared audiation to thought with regard to language. Gordon identified 32 different music aptitudes, two of the
stabilized music aptitudes being music creativity and music improvisation. Cutietta (1991) states that Gordon has endeavoured to define music aptitude as a many-faceted ability consisting of at least 20 dimensions of stabilized music aptitude. The most dominant of these are rhythm and tonal aptitude; although the author states that it is still not possible to date to determine if Gordon was right.

As can be seen above, a definition of musical aptitude often clearly includes an innate component, whereas musicality is mostly rather defined as a skill, be it innate or acquired after an infant’s birth. Gordon (see above) considered nature and nurture as equally important and insisted that quality exposure was necessary for music abilities to develop in a child with an innate potential. The genetic link, however, was practically unexplored at the time of Gordon’s research. With this respect, Oikkonen & Järvelä (2014) highlight that our genes are not only responsible for the development of our inner ear, but also for our auditory pathways and neurocognitive processes, all features underlying musical abilities. They state that it is of high interest that more than 10,000 hours of intense training can make any individual a professional musician (an idea developed by Ericsson, Krampe, & Tesch-Römer, 1993). Still, this does not necessarily mean that the person will be a talent comparable to Mozart or Beethoven. According to them, musical talent or aptitude is an innate ability to understand and perceive musical features, e.g., rhythm, pitch, timbre and tone duration. It depends on and is influenced by environmental circumstances during and after childhood, such as it is the case with language. According to these accounts, musicality and musical aptitude would be similar, one of the two, namely aptitude, highlighting the biological predisposition towards musicality, musicality itself designating the present skills. In the present thesis, I will refer to the assessed skills as musicality or musical ability, contrasted with musical training or experience. The extent to which these musical abilities are a priori dispositions and could thus reflect music aptitude will not be discussed in this thesis.

4.2.2. Can Musical Skills or Musical Training Enhance Foreign Language Learning?

We are facing times of fast and unprecedented global change and even more so, we are facing the challenges these significant transformations bring along. The world has become diverse and globalized and our ability to communicate and collaborate with speakers of
different linguistic backgrounds depends on our ability to speak foreign languages. Due to this rising importance of FLL in the past decades, researchers have been particularly keen on finding out how we can learn foreign languages easier and faster. One of the major foci of this research has been the positive transfer of musical skills on the acquisition of other languages. As has been discussed beforehand, language and music are highly intertwined, and it thus seems logical that higher ability or training in one of the two could positively affect the acquisition of the other.

4.2.2.1. Music and Speech Perception

Higher musicality and substantial musical training have been consistently found to have an impact on foreign language perception. It is believed that especially the processing and perception of non-native speech sounds benefits from enhanced auditory perception and discrimination ability, such as enhanced pitch or rhythm discrimination ability. This enhanced pitch or rhythm perception is often a result of musical aptitude or musical experience. Chobert & Besson (2013) highlight in their review that most research has been performed on the effect of musical expertise on pitch and duration perception, showing that it confers linguistically relevant advantages.

Speakers show considerable individual differences for distinguishing subtle phonetic contrasts in non-native speech sounds (e.g., Chandrasekaran, Sampath, & Wong, 2010). As Kempe, Bublitz, & Brooks (2015) claim, these differences might arise “from genetically mediated differences in the neuroanatomy of subcortical and cortical pathways that result in superior auditory processing, which may benefit the processing of acoustic features” (p. 350) for both language and music. They argue that musical expertise could exert a direct causal influence on speech-sound discrimination because of more precise processing of acoustical features and higher attentional control. In their own study, they assessed musical ability (AMMA), musical training and had English speakers discriminate Norwegian tonal contrasts and vowels (that is, tones that differed in temporal, pitch, and spectral characteristics). They controlled for language-learning experience and intelligence and found that musical aptitude and musical experience were linked to frequency-related processing in subjects, but not temporal processing. Surprisingly, however, musical aptitude and non-native speech sound processing were not directly linked, only through the mediating effect of enhanced spectral processing.
Focusing on pitch perception, Schön, Magne, & Besson (2004) reported that extensive musical training (non-musicians vs. professional musicians) led to a facilitation of pitch contour processing in a foreign language. In their study, musicians were more successful in detecting manipulations of pitch contour in both language and music, pointing towards an advantage in the auditory processing of language through musical training. These findings highlight musical training but do not exclude higher musicality in the subjects of this study. Besson, Schön, Moreno, Santos, & Magne (2007) also focused on pitch perception and concluded that musical expertise can positively improve pitch processing in music and speech, which demonstrates a transfer of training between music and language. In a similar study, Marques, Moreno, Castro, & Besson (2007) tested French people’s pitch perception of sentences in Portuguese and found a clear advantage of professional musicians over non-musicians.

Moreover, the learning of tone languages (e.g., Mandarin) has often been linked to better pitch ability in learners for obvious reasons (Bowles, Chang, & Karuzis, 2016; Saito, Sun, & Tierney, 2018). Bowles, Chang, & Karuzis, (2016), for instance, found that pitch ability was a better predictor for second language aptitude in a tone language than musicality and general cognitive ability. Furthermore, Saito, Sun, & Tierney (2018) reported that suprasegmental attainment in Chinese learners of English was associated with phonemic coding and rhythmic imagery in their study. However, the strongest predictor of suprasegmental attainment was the quality and quantity of L2 learning. Delogu and colleagues tested Italian speakers who performed perceptive tasks in Mandarin. In children and adults, melodic abilities and musical training were associated with enhanced discrimination of lexical tone. However, they did not find any differences concerning the discrimination of segmental variations (e.g., consonant or vowel changes; Delogu, Lampis, & Belardinelli, 2010; Delogu, Lampis, & Olivetti Belardinelli, 2006). Altogether, it can be said that musicality most likely leads to enhanced processing of foreign language lexical tone variation.

Focusing on rhythm processing, Bhatara, Yeung, & Nazzi (2015) tested French speakers’ melody and rhythm perception skills, finding that there was a positive correlation between FLL experience (in years) and rhythm perception. They suggested that both musical training and language learning were linked to rhythm perception in their French speakers, but not to their perception of melody. Nardo & Reiterer (2009) reported a strong link between musicality, pronunciation talent and grammatical sensitivity, one component of language...
aptitude as measured by the MLAT. In their study, the rhythm subscore of the AMMA test showed the highest correlations with language measures. Fonseca-Mora, Toscano-Fuentes, & Wermke (2011) investigated the role of rhythm in FLL. They hypothesized that the acquisition of foreign languages could benefit from structural and motivational properties of music in songs, in particular due to the refinement of auditory perception and the acquisition of metacognitive knowledge. Their proclaimed positive effects of musical training on foreign language learning, however, could not be confirmed in Spanish children, when they tried to improve foreign language acquisition through musical intervention (Fonseca-Mora, Jara-Jiménez, & Gómez-Domínguez, 2015). For a more extensive review that includes the neurophysiological level as well, please refer to Chobert & Besson (2013).

4.2.2.2. Music and Speech Production

Apart from the link between musical skills and language perception, numerous studies have further confirmed that musicality has an impact on foreign language production, especially second language pronunciation skills (Mireille Besson et al., 2007; Dogil & Reiterer, 2009; Fonseca-Mora et al., 2011; Ludke, Ferreira, & Overy, 2014; Schön et al., 2004). Christiner & Reiterer (2013) compared foreign language imitation skills of non-musicians/non-singers, professional singers and professional instrumentalists. The results showed that both instrumentalists and singers outperformed people without musical training or experience. Ludke (2010) also focused on singing and FLL but from a different perspective. To be more specific, they trained subjects to imitate Hungarian (an unfamiliar language to all subjects) using different strategies, namely rhythmical speaking, normal speaking and singing. They found that singing could facilitate short-term paired associate phrase learning in an unfamiliar language. They further claimed that their study suggested that a ‘listen-and-sing’ learning strategy could facilitate verbatim memory for spoken foreign language material.

Milovanov and colleagues in Finland performed numerous studies on music and speech production in their research group (Milovanov, Huotilainen, Välimäki, Esquef, & Tervaniemi, 2008; Milovanov, Pietilä, Tervaniemi, & Esquef, 2010; Milovanov & Tervaniemi, 2011). They reported a highly successful interplay between musicality and L2 learning (English) in Finnish native speakers. They found enhanced pronunciation skills in different populations of Finnish native speakers who had musical training or a relatively high musicality. They further reported that higher musical aptitude went not only hand in hand with better second language
pronunciation, but also with more accurate chord discrimination ability and more prominent sound-change-evoked brain activation to musical stimuli. Moreover, their results suggested that regular music practice led to a modulation of the brain’s linguistic organization and might have altered hemispheric functioning, thus being advantageous for pronunciation skills. In a similar study, Vangehuchten, Verhoeven, & Thys (2015) found a significant relationship between English pronunciation skills and musical skills in Spanish native speakers. Dolman & Spring (2014), on the other hand, revealed that excellent skills in specific musical abilities of Japanese learners (e.g., discrimination of pitch, rhythm and loudness) highly correlated with better pronunciation in the second language of these individuals (English). In accordance with previous studies, Slevc & Miyake (2006) reported a consistent relationship between phonological aspects of FLL and musical aptitude. In addition to the already known phonological advantage (e.g., pronunciation), they also tested syntactic and semantic skills. Musical aptitude, however, could not be linked to syntactic or semantic skills in the same population. Shabani & Torkeh (2014) conducted a study to examine the link between musical intelligence (assessed by a musical intelligence questionnaire of Howard Gardner’s multiple intelligence questionnaires) and language proficiency in English, finding a particularly strong link ($r = .829^{**}$). A major shortcoming, however, was that Gardner’s concept of intelligence is very broad, meaning that other traits like creativity, social behaviour and musicality are tested and included as well to a certain extent (for a critique on Gardner’s intelligence concept, see Neubauer & Stern, 2013).

Based on these results, one could argue that musical abilities, regardless your first language, might positively influence the learning of any foreign language. Most studies exploring the link between language skills and music have focused on the learning of non-native sounds and pronunciation skills in a foreign language. Although the syntactic relationships between language and music are a topic of discussion (Slevc, 2012), there has been practically no research on the influence of musical training or expertise on syntax and grammar skills in a foreign language. When throwing a closer look at the aforementioned studies, one may wonder how musical training or expertise were assessed. It may be seen as a limitation that most studies did not use musical training or apply musical learning strategies but used the number of instruments and the years of musical training as measure of musical expertise only (Zeromskaitė, 2014).
4.2.3. **Neural Changes Due to Musical Training and Experience**

Given the common features of music and language, the suggestion of similar underlying brain functions and hence a common network in the human brain seems logical. Music training and its possible influences on other skills, e.g., language, have been intensely investigated given the often-cited effects of music on a variety of cognitive functions. In other words, music training does not only seem to have a positive effect on language processing and the development of linguistic skills, but also on other cognitive functions, such as reading and attention (Besson & Schön, 2001; Seither-Preisler et al., 2014; Serrallach et al., 2016). These benefits due to musical training or experience are often referred to as ‘positive transfer’.

It all started with the so-called ‘Mozart effect’ (Jenkins, 2001; Rauscher & Shaw, 1998) according to which just listening for about ten minutes to music by Mozart led to better spatial reasoning skills and a higher IQ in normal subjects. Although this effect was only temporary (10-15 minutes), it gave rise to a controversial discussion. Rauscher and Shaw stressed that the effect they had found was only limited to spatial temporal reasoning and effects on general intelligence must have been due to inappropriate test procedures (Steele, Bass, & Crook, 1999). It is now known that enjoyment arousal explains the positive impact of music on other skills but still, this discovery has been the starting point of further investigations as to whether music training leads to robust skill transfer and changes in brain functioning (Moreno & Bidelman, 2014; Oikkonen & Järvelä, 2014).

As musical aptitude has a long history of research, the brains of musicians have been a matter of interest to neuropsychologists since the beginning of the neurocognitive era. Musicians appear to exhibit more concentrated brain waves in several auditory areas of the brain, starting in the brainstem (Kraus & Nicol, 2018). Investigations of auditory processing in the brains of musicians have shown a number of differences between musicians and non-musicians. Non-auditory functions, e.g. auditory-motor loop, spatial cognition or multisensory networks, are more often involved in the processing of sound in musicians, which leads to a more differentiated and diverse perception of music (Schneider & Seither-Preisler, 2015). The regions most activated by music apart from the AC are the hypothalamus, the hippocampus and the nucleus accumbens (Oikkonen & Järvelä, 2014). It is important to distinguish between temporary effects, like higher cerebral blood flow in specific regions, and long-term plasticity effects induced by musical training. Indeed, musicians’ brains have been used as a model of
neuroplasticity in the past years, with studies finding that musical training leads to functional and structural changes in the auditory system (for details, see Kraus & Chandrasekaran, 2010). Such differences are for example larger GM volumes in motor, auditory and visuo-spatial regions, or altered neurophysiological responses when listening to speech or music.

Why and how does music change our brain? Music requires motor and visual components and thus leads to the recruiting of an array of brain networks that involve visual, auditory, motor, as well as memory-related processes. Benefits due to musical training are seen at subcortical and cortical levels of the auditory system and more so even in regions outside of the lemniscal hearing pathway, such as in the frontal lobes (Kempe et al., 2015; Kraus & Chandrasekaran, 2010). Musical training has indeed been shown to change the human brain and has long-lasting biological effects. Let us consider the brainstem just to give an example. Initial stages of auditory sensory processing are already improved by training. The brainstem can be considered an essential part of the auditory pathway as it serves as processor for signals prior to cerebral cortex processing. Music training over a longer period of time leads to an enhancement of the magnitude with which the brainstem responds to features such as musical pitch. Furthermore, neural activity is temporally more precise in musicians (Moreno & Bidelman, 2014; Wong, Skoe, Russo, Dees, & Kraus, 2007).

In simple terms, musicians have not only well-trained ears, as is often stated, but also well-trained brainstems and possibly well-trained auditory cortices in the cerebral cortex. Kraus & Nicol (2018) clarify that the auditory brainstem comprises numerous structures specialized for certain sound-processing functions and all of these are involved in the encoding of speech and music. Moreno & Bidelman (2014) state that a considerable amount of research has proved that music training has positive, long-lasting biological benefits on auditory functioning and may also lead to neuroanatomical differences such as morphological changes in the precentral gyrus, the motor brain areas or AC/HG, for instance. It has been shown that auditory training changes the processing of sound stimuli in the brain (e.g. Seither-Preisler et al., 2014; Serrallach et al., 2016) and therefore specific life experiences may cause specific

---

5 For a current and complete overview, see Kraus and Nicol (2018), who devoted a whole book chapter on the mechanisms underlying sound processing and the relationship between auditory processing of music and its relation to language.
functional changes. As a matter of fact, musicians’ perceptual, language and high-level cognitive processing (e.g. WM, verbal intelligence) are enhanced (e.g., Moreno et al., 2011; Roden, Kreutz, & Bongard, 2012). In other words, these benefits caused by music extend far beyond simple enhancements regarding perceptual abilities and have been shown to positively impact non-auditory functioning involving WM and intelligence, i.e. high-order aspects of cognition (Schellenberg, 2011). However, it is still debated whether enhancements in specific skills like language learning or performance on intelligence tests are due to enhanced executive functioning or WM of those with musical training. As an example, see Schellenberg (2004, 2011), disagreeing with his mediation effect, and Degé, Kubicek, & Schwarzer (2011) agreeing that executive functions mediate the link between musical training and intelligence.

Usually, there is a correlation between the years of musical training and the enhancements and there seems to be a neuroplastic change or modification created by the amount of musical exposure. For instance, Corrigall & Trainor (2011) found that length of musical training was significantly linked to reading comprehension in children, while Moreno et al. (2009) found higher linguistic abilities in musically trained 8-year-old children. Forgeard, Winner, Norton, & Schlaug (2008) confirmed that duration of training from childhood on could predict auditory discrimination abilities, fine motor skills, vocabulary and reasoning skills in adults. It is expected that the earlier children start musical training (see White-Schwoch, Carr, Anderson, Strait, & Kraus, 2013), the better the transfer effects and the higher the functional and anatomical changes in the brains of individuals (Penhune, 2011).

In sum, music has been one of the hot topics regarding plasticity and it has been confirmed by numerous studies that musical training and experience leads to robust changes in the human brain, mostly in the auditory areas, that are also essential for other cognitive skills. In particular, music has not only been linked to the perception of speech (e.g., non-native speech sounds, prosody, tonal pitch contours in tone languages), but it has also been shown to positively influence FLL skills. Of course, correlations cannot reveal causal relationships between these skills. While it might be the case that high musicality and considerable musical training is positively linked to better foreign language pronunciation and perception, we cannot say much about the concrete interaction between the assessed skills and the directedness of transfer.
5. Study 1: Right Auditory Cortex Morphology is linked to Language Aptitude, Speech Imitation and Musicality

The major aim of the first study was to investigate a potential relationship between AC morphology and language aptitude in adults. Language aptitude was assessed by a language aptitude test (MLAT), an English pronunciation task and the Hindi speech imitation test. Another objective was to explore the relationships between musical experience, musicality, WM capacity and language aptitude in the very same population.

This study is an analysis of data that was collected by Assoc.-Prof. Dr. Susanne Reiterer and colleagues in an extensive study at the University of Tübingen in 2008. The overall results were published in a book edited by G. Dogil and S. Reiterer (Dogil & Reiterer, 2009). For the present thesis, behavioural and neuroanatomical data from a subgroup of participants of that study were analysed. The results of the present study have already been published in Turker et al. (2017) and Turker et al. (2018) and will be dealt with in more detail in the present chapter.

5.1. Methods

5.1.1. Subjects

All subjects ($N = 30; 13$ male/$17$ female)\(^6\) were healthy German native speakers who had been raised monolingually, i.e. with one native language only in the first years of life. All subjects were between 20 and 40 years ($M_{\text{age}} = 26.8 \pm 5.0$ y) and had started learning their second language, English, at $10 \pm 1$ years of age, which is the usual case in the German school system. The subjects were either students or had even achieved higher positions at an institution of higher education around Tübingen (e.g., doctoral students, post-doctoral researchers, post-doctoral researchers, post-doctoral researchers).

---

\(^6\) The sample for this study consisted of a subgroup of participants of the aforementioned study at the University of Tübingen, Germany. This subgroup consisted of the most gifted and the least gifted individuals of the overall sample of 79 subjects.
professors). Participants were excluded from the present analysis if they showed any medical or neurological impairment, or (had) suffered from speech or language impairments. All subjects were paid for participation and provided written informed consent before participating in the experiment. The data were analysed anonymously, and the study was approved by the ‘Ethikkomission des Universitätsklinikums Tübingen’ in 2008.

To test differences on assessed scales on a group-statistical level, subjects were further classified according to two scores, namely the Hindi speech imitation test and the MLAT. After distinguishing between high and low speech imitators (i.e. according to the Hindi score; high referring to above average performance, low to below average performance), the statistical tests were used to compare the performances of the high speech imitators ($N = 14$, $M_{age} = 27.9 \pm 5.5$, 3 females) with the low speech imitators ($N = 16$, $M_{age} = 25.8 \pm 4.3$, 10 females). First, high and low aptitude learners were distinguished according to the results on the MLAT. The groups were divided into an above-average and a below-average group and statistical tests were used to compare the high aptitude ($N = 14$, $M_{age} = 27.3 \pm 5.8$, 4 females) and the low aptitude language learners ($N = 15$, $M_{age} = 26.1 \pm 4.2$, 9 females).

5.1.2. Language Aptitude Testing

The participants underwent an extensive battery of language learning tests, including an English pronunciation assessment, a Hindi speech imitation task and three parts of the well-known MLAT (Carroll & Sapon, 1959). All three tests were considered to measure different aspects and facets of language aptitude. While the first was a measure of proficiency or achievement and expected to be the result of aptitude, the second assessed perceptive skills and their translation to productive skills. Finally, the MLAT was designed to measure different components of language aptitude and the three tests of the MLAT that were administered assessed vocabulary learning, phonetic coding ability and language analysis ability.

5.1.2.1. An English Pronunciation Assessment

All subjects had completed their studies (at least a bachelor’s degree) and had achieved a high level of proficiency (B2) in English, at least according to the criteria as defined in the school leaving examinations in Germany. Therefore, all participants were expected to possess about the same level of English. In the English pronunciation assessment, all participants were
asked to read “The North Wind and the Sun”, a text widely known in the English community. The task was to read the text in their best English, paying specific attention to their pronunciation. The test is a typical test for pronunciation assessment as it contains very frequent words of the English language but includes vocabulary with a wide range of challenging sound combinations, especially for non-native speakers of English. Participants only had one attempt and were told that specific attention would be paid to their pronunciation, both with respect to how single sounds were pronounced and their overall speech melody and intonation. This would give a precise account of their overall pronunciation and thus be an excellent indicator for their English pronunciation skills.

Subjects were recorded on a professional speech recording equipment in a sound-proof basement room of the former phonetics laboratory of the Institute of Natural Language Processing of the University of Stuttgart, Germany. To assess subjects’ pronunciation skills, 30 native British English speakers were recruited for a rating. Each rater was provided with enough details to be able to rate subjects’ English pronunciation performance. The rating took place online and raters were paid hourly for their work. The raters gave each participant a score between 0 and 10 according to the correctness and quality of the pronunciation based on a global intuitive impression rating procedure (the rating procedure is detailed in Jilka, 2009 and Reiterer et al., 2011). The inter-rater reliability was very high because of the unusually high amount of raters. Thus, the corresponding scores can be considered as highly reliable.

5.1.2.2. The Hindi Speech Imitation Task

The Hindi speech imitation task is thought of as a measure of phonetic memory, being highly reliant on WM as it requires the correct storage and retrieval of newly acquired information. To assess their speech imitation skills, subjects were asked to repeat four words and four sentences (consisting all of more than two syllables) in Hindi, a language unknown to them. They heard each speech input three times and were then asked to imitate once what they had just heard. During their imitation, speakers were recorded on a professional speech recording equipment in a soundproof basement room of the former phonetics laboratory of the Institute of Natural Language Processing of the University of Stuttgart, Germany. Similar to the English assessment, thirty native Hindi speakers were recruited and paid for their assessment of subjects’ speech imitation skills. They were provided with the speech material online and gave ratings from 0 to 10 on an intuitive Likert-scale-like bar for the quality and
‘nativelikeness’ of the speech material they were provided with. For the rating, sound files of Hindi native speakers were added to the overall rating sample of German speakers (without knowledge of the raters) to additionally verify the validity of the whole rating procedure. For the rating, an inter-rater reliability was calculated, and the corresponding scores can thus be considered as highly reliable.

5.1.2.3. **The Modern Language Aptitude Test (MLAT)**

As it is still considered a good measure of foreign language aptitude, the MLAT was administered to all participants. The main aim behind choosing parts of this test was to assess different components of language aptitude. The three subtests chosen provide measures of phonetic coding ability (Part III, ‘Spelling Cues’), grammatical sensitivity (Part IV, ‘Words in Sentences’) and associative memory (Part V, ‘Paired Associates’).

In Part III, letter strings and non-words, which are phonologically very similar to specific English words, were presented and test takers were provided with five English nouns and had to tick the one that was semantically very close to the letter strings and non-words. In Part IV, test takers were provided with full sentences with various elements that performed different functions in the sentences (e.g., adjectives, adverbs). Looking at a new sentence, test takers had to specify the function of an element by indicating which other word it corresponded to in grammatical function. Part V was a Kurdish vocabulary learning task in which individuals had to learn as many Kurdish words as possible with their English translation in two minutes. The testing phase of MLAT V consisted of a multiple-choice task.

As explained in 2.4., these are three core components of language aptitude that also appear in Skehan's (1998, 2002) most recent model of foreign language aptitude and processing. The MLAT is a paper-pencil test that was administered to one participant at a time without specific instruction. MLAT III yielded a maximum of 50 points, MLAT IV of 45 and MLAT V a maximum of 24 points. For the later statistical analyses, the raw scores were always used, as well as a total sum score, combining scores of all three subtests.

5.1.3. **Musicality Assessment**

In order to assess aptitude in the musical domain, the AMMA test (Advanced Measures of Music Audiation; Gordon, 1989) was administered. It is a well-established tool for assessing
musical aptitude and consists of two parts assessing rhythmic and tonal perception. The subjects were asked to complete both tasks, (1) the rhythm discrimination task and (2) the pitch discrimination task, with a maximum of 40 points achievable per subpart. For the following analyses, the general raw scores and raw scores of the two subparts were used.

In addition to the musicality assessment, a questionnaire was used to find out the number and type of instruments that the subjects had learnt in the course of their life. An instrument had to be played for at least six months to count as a musical experience. In other words, any musical experience shorter than half a year was not considered as relevant for the present analyses. Singing was also included as musical experience if it was done in a choir or professionally.

5.1.4. Working Memory Capacity

To assess phonological WM, three different verbal WM tasks were administered. In the three verbal WM tests, subjects repeated digits, both forward and backward, and non-words. Whereas digit span forward and non-word span are considered basic phonological STM tasks (i.e. classical phonological WM tasks), digit span backward is not entirely a phonological STM task as it also requires the manipulation of input. In other words, while the first two only require storage and retrieval of verbal input, the third requires operations on the stored input, namely saying the heard digits in reversed order and not in order as heard. For reasons as specified in 4.1.2, I consider all three administered span tasks as phonological WM tasks despite the slightly higher cognitive demand in digit span backward compared to digit span forward and non-word span. Therefore, the three tests that were applied in the study were all expected to provide a good measure of phonological WM.

The digit stimuli for digit span forward and backward were taken from the ‘Kurztest für allgemeine Intelligenz (KAI)’, a German test for general intelligence (Lehrl, Gallwitz, Blaha, & Fischer, 1992). The digit span forward task required participants to overtly repeat a number of digits provided as auditory input. The test started with the repetition of three random digits (e.g., 2-6-1) and ended with a maximum of nine digits (e.g., 9-8-6-1-4-3-1-7). The digit span backward test, on the other hand, required subjects to repeat heard digits backwards, i.e. in the reversed order as heard (e.g., 2-6-4 had to be repeated overtly as 4-6-2). Length in the backward task ranged from two to eight digits. For both tasks, each participant had two chances.
to correctly repeat the same number of digits. Then another digit was added to increase complexity. If a participant failed to repeat the first attempt correctly, they were given another chance. Failure of the second attempt, on the other hand, indicated the end of the test.

The non-word repetition task worked similarly to the digit span tasks. During the test, subjects needed to repeat German non-words that were created from a syllable database developed according to German phonotactic rules (e.g. “knol”, “pflax”, “bamp”; always CVC syllables) at the Institute of Natural Language Processing, University of Stuttgart, Germany (described in more detail in the master thesis of U. Benner, 2005). Specific attention was paid to the correct pronunciation of each sound of each non-word.

For each of the three tasks, one correct repetition yielded one point and a total of 14 points (raw score) could be achieved in each of the separate tasks. The maximum score for all three WM tasks together was thus 42 points.

5.1.5. Neuroanatomical Analysis: Semi-Automated Manual Segmentation

For this study, high-resolution T1-weighted structural magnetic MRI (Siemens, Magnetom SonataVision; 1.5 Tesla; software version: syngo MR 2004A, 176 DICOM slices, sagittal orientation, slice thickness 1 mm) was performed at the University of Tübingen, Germany. The procedure applied to obtain 3D reconstructions of AC, PT and the overall STG, was developed by Peter Schneider at the Department for Neuroradiology at the University Hospital Heidelberg, Germany. This innovative method can best be described as a semi-automated manual segmentation of the specific region of interest in the human brain. The method has been applied in a number of studies by Peter Schneider and his research group (J. Benner et al., 2017; Schneider et al., 2009; Seither-Preisler et al., 2014; Serrallach et al., 2016; Turker et al., 2017; Turker, Reiterer, et al., 2018) and has proven to provide very detailed and precise depictions of human AC. The software of choice for this manual segmentation method was Brain Voyager Software QX 2.8 (Brain Innovation B.V., Maastricht, Netherlands).

In contrast to usually done structural analyses of the brain, the method takes into account individual variation as each brain and even each hemisphere is segmented. In other neuroanatomical analyses, like voxel-based morphometry (VBM), standardized atlases are used to calculate GM or white matter volumes in specific brain regions. While these approaches are less time-consuming and easier to calculate, they fail to consider small details in specific
brain regions that might be of great importance. To be more specific, while the method developed by Schneider and colleagues successfully takes individual variation in AC into account, other methods for analysing gyrification and GM volumes do not.

Various pre-processing steps were applied to make the obtained structural scans comparable on a variety of levels. These included the adjustment of brain images in both contrast and brightness, followed by a correction for inhomogeneity. All scans were rotated in direction of the antero-posterior commissural line before images were normalized in stereotactic space (Talairach & Tournoux, 1988). The most time-consuming and most exciting part of this innovative method was the segmentation itself (for details of the steps, see Figure 16). During this process, the superior temporal plane that includes HG, the anterior STG and the PT was manually marked in each slice of the brain containing this region before segmentation (see Figure 16, b). This was done on sagittal brain slices along the lateral fissure using the standard definition of the landmarks of human AC. This whole process is best described as a semi-manual slice-by-slice segmentation (for details, see Schneider et al., 2005; Wengenroth, Blatow, Bendszus, & Schneider, 2010). After successful segmentation (i.e. all regions outside of the region of interest were eliminated; see Figure 16, c), all slices containing the target region were reconstructed in a 3D Mesh reconstruction. In the final steps, the 3D mesh was slightly smoothed for better visualization (see the results in Figure 16, d).

The ACs of all participants were 3D-reconstructed, and two researchers did the reconstructions independently to guarantee objectivity. The exact method as presented in this chapter can be read up in other publications by the research group as mentioned above. The method per se has not been published separately but can be applied by following the exact procedure as detailed in these papers or this thesis. A detailed account in the form of a protocol on the pre-processing and segmentation steps can also be obtained by contacting the author or P. Schneider, head of the research group ‘Musik und Gehirn’ at the Department for Neuroradiology at the University Hospital Heidelberg, Germany.
Figure 16 Detailed steps of the semi-automated manual segmentation in BrainVoyager QX. First, the whole brain data is pre-processed (a), and then STG including AC is manually segmented in each slice of the brain (b). After eliminating all areas outside of STG (c), AC is reconstructed as a 3D mesh displaying the whole STG including PT and HG (d).
Three types of HGs were distinguished for the present analysis. These three types were: (1) single gyrus, (2) common stem duplication (CSD) and (3) complete posterior duplication (CPD; see Figure 17). As already mentioned, the first type refers to one single gyrus that is followed by a complete sulcus. The second type, CSD, designates one single gyrus that is divided by a sulcus that does not reach the medial end of HG but proceeds towards the lateral end of HG and thus STG. Therefore, the two parts of HG in this case are connected at the medial end but not at the lateral end. The third form, CPD, designates one HG followed by a complete duplication. This duplication is also called adjacent posterior duplication since it is posterior to HG and there is no space in between HG and the duplication. These three types have been successfully distinguished between and related to behavioural outcomes in other studies (J. Benner et al., 2017; Seither-Preisler et al., 2014).

Figure 17 The three types of HGs distinguished in the present study: single gyrus, common stem duplication (CSD), complete posterior duplication (CPD; from left to right). aSTG anterior superior temporal gyrus, HG Heschl's gyrus, PT planum temporale

In the group of participants of this adult study, only two individuals possessed more than two complete gyri and would therefore have fallen into a category of ‘multiple gyri’, as distinguished in the second study that will be presented later on (see Chapter 6). Given this low
rate, I decided not to distinguish between multiple gyri and CPDs in this study and counted the two instances of multiple gyri as CPDs.

5.2. Results

The findings of the behavioural analyses show that the assessed core abilities (language aptitude, musicality and WM capacity) not only correlated positively with one another, but several assessed scales were linked to the differentiation between high and low speech imitators and high and low language learners (MLAT score). The results of the neuroanatomical analyses, on the other hand, showed a strong link between the occurrence of CPDs in the right hemisphere, i.e. in right AC, and high speech imitation and high language aptitude.

First, the behavioural results will be discussed in detail. Especially the relationship between WM capacity, rhythm and pitch perception and language aptitude will be emphasized therein. After a brief summary of the descriptive results, findings of the correlational analyses, a principal component analysis (PCA) and a comparison between high and low aptitude subjects (either according to the Hindi or the MLAT score as indicated) will be presented.

The behavioural analysis will be complemented by a detailed account of the neuroanatomical results. Relating AC morphology to language aptitude, musicality and speech imitation aptitude was done through three one-way ANOVAs (analyses of variance) and Chi-Square tests, commonly used to evaluate tests of independence when bivariate tables, or a crosstabulation are used. ANOVAs provide the opportunity to look into the relationship between several variables of interest (e.g., musicality, language aptitude, speech imitation) and a number of fixed factors, like AC morphology (i.e. single gyrus, CSD and CPD). All statistical analyses were performed with SPSS 25 and visualizations were computed with R or Matlab.

5.2.1. Behavioural Results

5.2.1.1. Descriptive Statistics

A table of the descriptive results is provided (Table 7). Performance in the English pronunciation task ($M = 6.40, SD = 1.72$) was much higher than in the Hindi speech imitation task ($M = 4.81, SD = 1.64$). The number of instruments subjects played ranged from zero to
three ($M = 1.23, SD = 0.97$). The number of foreign languages that participants had learnt ranged from one to nine ($M = 2.59, SD = 1.72$). Results on the AMMA pitch discrimination task (AMMA tonal; $M = 28.72, SD = 5.68$) were similar to the results of the AMMA rhythm discrimination task ($M = 31.10, SD = 4.61$; AMMA total: 42 to 79, $M = 59.80, SD = 10.05$).

<table>
<thead>
<tr>
<th>Variable</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>English score</td>
<td>6.40</td>
<td>1.72</td>
</tr>
<tr>
<td>Hindi score</td>
<td>4.81</td>
<td>1.64</td>
</tr>
<tr>
<td>$n^\circ$ of instruments</td>
<td>1.23</td>
<td>0.97</td>
</tr>
<tr>
<td>$n^\circ$ of languages</td>
<td>2.59</td>
<td>1.72</td>
</tr>
<tr>
<td>AMMA tonal</td>
<td>28.72</td>
<td>5.68</td>
</tr>
<tr>
<td>AMMA rhythm</td>
<td>31.10</td>
<td>4.61</td>
</tr>
<tr>
<td>AMMA total</td>
<td>59.80</td>
<td>10.05</td>
</tr>
<tr>
<td>Digit span forward</td>
<td>9.59</td>
<td>1.88</td>
</tr>
<tr>
<td>Digit span backward</td>
<td>8.76</td>
<td>2.13</td>
</tr>
<tr>
<td>Non-word span</td>
<td>7.55</td>
<td>1.74</td>
</tr>
<tr>
<td>MLAT III</td>
<td>36.69</td>
<td>8.62</td>
</tr>
<tr>
<td>MLAT IV</td>
<td>29.28</td>
<td>5.58</td>
</tr>
<tr>
<td>MLAT V</td>
<td>17.31</td>
<td>5.09</td>
</tr>
<tr>
<td>MLAT total</td>
<td>83.41</td>
<td>14.23</td>
</tr>
</tbody>
</table>

Table 7 Mean scores and standard deviations for all assessed scales for all participants. Scores given with two decimal places. AMMA Advanced Measures of Music Audiation, MLAT Modern Language Aptitude Test, $n^\circ$ number

Concerning scores on the WM scales, digit span forward ($M = 9.59, SD = 1.88$) and digit span backward ($M = 8.76, SD = 2.13$) gave similar results. The lowest WM scores were found for non-word span ($M = 7.55, SD = 1.74$). MLAT total scores ranged from 49 to 109.
points (raw score; \( M = 83.41, SD = 14.23 \)), displaying a considerable gap between ‘highly gifted’ and ‘poorly gifted’ language learners. The MLAT total score summarized the results of all three sub-parts, MLAT III (\( M = 36.69, SD = 8.62 \)), MLAT IV (\( M = 29.28, SD = 5.58 \)) and MLAT V (\( M = 17.31, SD = 5.09 \)). In the language aptitude battery, the best performance was found in MLAT III, measuring phonetic coding ability, while the least successful performance was found in MLAT V, the vocabulary learning task.

There were slight differences in performance between males and females in two tasks. \( t \)-tests for independent samples were performed to test the statistical significance of these differences between the two groups. In digit span forward (males: \( M = 10.4 \pm 1.6 \); females: \( M = 8.6 \pm 1.8 \); \( t(28) = -2.8, p = .009 \)) and MLAT V (males: \( M = 19.4 \pm 4.1 \); females: \( M = 14.7 \pm 5.1 \); \( t(28) = -2.8, p = .01 \)), males performed on average significantly better than females, explaining why more males were in the high language aptitude group.

5.2.1.2. Results of the Correlational Analyses

A correlational analysis allows to uncover potential relations between the assessed scores, e.g., the links between scores on language aptitude tests, WM and musicality. Given that multiple variables are usually used in a correlational analysis and multiple comparisons are thus performed, alpha error accumulation is an issue. In other words, doing multiple comparisons may significantly increase the risk of chance findings at a critical \( p \)-value of .05. To control for that, the method of false discovery rate (FDR-correction; Benjamini & Hochberg, 1995) was chosen. In their work of 1995, the authors suggest 10-20% as an acceptable proportion of false discoveries. Therefore, I decided to choose 20% as an acceptable proportion of false discoveries to control for alpha error accumulation in order not to risk losing true positives.

Results of the correlational analyses are provided in Table 8 with both results before and after correction being marked. Results that remained significant after correction for multiple comparisons are shown in bold and with three asterisks, while results that were only significant at an uncorrected level are marked with a single asterisk. Only correlation coefficients are provided in the table and significance values are discussed in the upcoming paragraphs.
<table>
<thead>
<tr>
<th></th>
<th>n° of instr.</th>
<th>n° of lang.</th>
<th>Hindi score</th>
<th>English score</th>
<th>AMMA tonal</th>
<th>AMMA rhythm</th>
<th>MLAT V</th>
<th>MLAT IV</th>
<th>MLAT III</th>
<th>nonword span</th>
<th>digit span b</th>
<th>digit span f</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>n° of instr.</td>
<td>1</td>
<td>.159</td>
<td>.394***</td>
<td>.310</td>
<td>.455***</td>
<td>.407***</td>
<td>-.021</td>
<td>.506***</td>
<td>.182</td>
<td>.350</td>
<td>.282</td>
<td>.145</td>
<td>.169</td>
</tr>
<tr>
<td>n° of lang.</td>
<td>1</td>
<td>.003</td>
<td>-.066</td>
<td>.218</td>
<td>.213</td>
<td>.048</td>
<td>.217</td>
<td>.131</td>
<td>.103</td>
<td>.273</td>
<td>.056</td>
<td>.591***</td>
<td></td>
</tr>
<tr>
<td>Hindi score</td>
<td>1</td>
<td>.390***</td>
<td>.284</td>
<td>.278*</td>
<td>.356</td>
<td>.384*</td>
<td>.290</td>
<td>.483***</td>
<td>.369*</td>
<td>.447***</td>
<td>.248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English score</td>
<td>1</td>
<td>268</td>
<td>.393***</td>
<td>.333</td>
<td>.557***</td>
<td>.756***</td>
<td>.163</td>
<td>.251</td>
<td>.195</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMMA tonal</td>
<td>1</td>
<td></td>
<td>.911***</td>
<td>-.023</td>
<td>.333</td>
<td>.243</td>
<td>.312</td>
<td>.260</td>
<td>-.101</td>
<td>.168</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMMA rhythm</td>
<td>1</td>
<td></td>
<td></td>
<td>.367</td>
<td>.365</td>
<td>.313</td>
<td>.224</td>
<td>.034</td>
<td>.042</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLAT V</td>
<td>1</td>
<td></td>
<td></td>
<td>.278</td>
<td>.299</td>
<td>.012</td>
<td>.165</td>
<td>.212</td>
<td>-.028</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLAT IV</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>.590***</td>
<td>.237</td>
<td>.246</td>
<td>.062</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLAT III</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.012</td>
<td>.268</td>
<td>.320</td>
<td>.022</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonword span</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.335</td>
<td>.323</td>
<td>.534***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>digit span b</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.259</td>
<td>.313</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>digit span f</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.145</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 8** Pearson correlation coefficients (rounded to three decimal places) for all assessed scales. *digit span b* digit span backward, *digit span f* digit span forward; *instr.* instruments, *lang.* languages. *** significant at p<.05 after FDR correction (Benjamini and Hochberg, 1985); * significant at p<.05 before correction.
Only AMMA rhythm and AMMA tonal \((r = .911, p < .001)\) showed a very strong correlation with each other, which indicated a strong relationship far beyond chance. Likewise, two parts of the MLAT correlated significantly with each other, namely MLAT IV and MLAT III \((r = .590, p = .001)\). Significant correlations could further be found between the English pronunciation score and parts of the MLAT: MLAT III and English \((r = .756, p = .002)\) and MLAT IV and English \((r = .557, p = .002)\). MLAT IV correlated with the number of instruments learnt by a subject \((r = .506, p = .005)\). The age of participants (ranging from 20 to 40 years) highly correlated with the number of foreign languages acquired \((r = .501, p = .005)\) and with non-word span \((r = .534, p = .003)\). After correction for multiple comparisons, non-word span and the Hindi score \((r = .482, p = .008)\), as well as digit span forward and the Hindi score \((r = .447, p = .015)\) still correlated significantly. The same was true for results on Hindi and English tasks \((r = .390, p = .033)\), and scores on the Hindi test and the number of instruments \((r = .394, p = .032)\). Last, AMMA tonal correlated significantly with the number of instruments \((r = .455, p = .013)\), whereas AMMA rhythm correlated both with the number of instruments \((r = .407, p = .028)\) and the English pronunciation score \((r = .393, p = .035; \text{all significant correlations after correction are marked in bold and with} ***\).)

5.2.1.3. Results of the Principal Component Analysis

A principal component analysis (henceforth always PCA) is a very valuable tool to gain insights into the most influential variables contributing to performance on different test scales. This is particularly useful if a large variety of tests is included and they are expected to load onto very different cognitive abilities. Another goal behind conducting a PCA is also to verify whether the multiple variables included in the overall analyses really measure similar, more general constructs. My hypothesis was that music-related tests and language-related tests would weigh heavily on two largely independent components. If this were true, the PCA would result in at least two different components with the many scales loading heavily on the one or the other, e.g., AMMA rhythm results loading heavily on a music-related component, compared to LLAMA D loading heavily on a language-related component.

A pre-analysis of the data proved that the requirements for the application of the PCA were fulfilled. These requirements were the following: (1) the determinant as an indicator of multicollinearity should be below \(p = .05\) \((p = .002\) in the present analysis); (2) the Kaiser-Meyer-Olkin criterion as a measure for the suitability of the sample should be above \(p = .50\) \(p \).
= .65 in the present study); (3) the Bartlett-test for sphericity should be significant at \( p < .05 \) (\( p < .001 \) in the present study). For the PCA, the same variables as shown in the correlation matrix of Table 8 were included (altogether 12 variables; results are displayed in Table 9).

<table>
<thead>
<tr>
<th>Rotated component matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Components</strong></td>
</tr>
<tr>
<td>(1) Musicality</td>
</tr>
<tr>
<td>(2) Language Aptitude</td>
</tr>
<tr>
<td>(3) Working Memory</td>
</tr>
<tr>
<td><strong>Variables</strong></td>
</tr>
<tr>
<td>AMMA tonal</td>
</tr>
<tr>
<td>AMMA rhythm</td>
</tr>
<tr>
<td>( n° ) instruments</td>
</tr>
<tr>
<td>( n° ) languages</td>
</tr>
<tr>
<td>MLAT III</td>
</tr>
<tr>
<td>MLAT IV</td>
</tr>
<tr>
<td>MLAT V</td>
</tr>
<tr>
<td>English score</td>
</tr>
<tr>
<td>Hindi score</td>
</tr>
<tr>
<td>Non-word span</td>
</tr>
<tr>
<td>Digit span forward</td>
</tr>
<tr>
<td>Digit span backward</td>
</tr>
</tbody>
</table>

**Table 9** Rotated component matrix displaying three core components measured by all scales in the PCA: Component 1 (Musicality), Component 2 (Language Aptitude) and Component 3 (Working Memory). No loadings below .3 are shown.

For the PCA, varimax rotation with Kaiser-normalization was used, which yielded a solution with three factors with eigenvalues clearly above 1. The rotation converged in 6 iterations. The variance explained by the model was 61.2%, which confirms the
appropriateness of the method on the data. Table 9 displays the rotated component matrix with the coefficients of each scale on the three identified components (as entitled). For the sake of clarity, coefficients below .3, which signify very small and hence negligible loadings, are not shown in Table 9. In contrast, coefficients above .5 signify particularly relevant contributions.

According to the criteria for strong and relevant loadings (loadings above .5), the first of the three components comprised scales related to music, on a general level, and musical ability on a more specific level. The name chosen for this component was ‘Musicality’. Variables that loaded heavily onto this component were AMMA tonal and rhythm perception scores and the number of played instruments by participants. The second component included concepts related to language skills. Therefore, I chose the appropriate name ‘Language Aptitude’ for this component. It included the English pronunciation score and MLAT parts III, IV and V. The third component ‘Working Memory’ which was extracted by the PCA included measures related to WM capacity (heavy loadings by digit span forward, digit span backward, non-word span and the Hindi score).

The number of instruments played by subjects also loaded on WM capacity, while the number of languages spoken was associated with the component musicality. MLAT IV, measuring grammatical sensitivity, did not only load onto the second component (‘Language Aptitude’), but also on the first component (‘Musicality’). In a similar fashion, non-word span also moderately loaded on this component. The Hindi score, a scale most strongly related to WM, also moderately loaded on the component ‘Language Aptitude’.

The results of the PCA were not only a good indicator to show which scales loaded on a specific cognitive component, but they were also ideally suited to provide closer insights on possible neuroanatomical correlations. Although the main aim of the present study was to find out in how far specific HG types correlated with language aptitude, it was also particularly interesting to relate the three components identified by the PCA with the results of the neuroanatomical analyses as well. By such, the different scales that contributed to the overall component (e.g., language aptitude) and their significance thereon could also be included. Therefore, in a next step, I turned the participants’ individual factor scores on the aforementioned components into separate variables (positive/negative: above/below average; \( M = 0, SD = 1 \)). Then I compared the three types of right-hemispheric HG morphology (i.e.
single, CSD and CPD) with these individual factor scores. The results of these comparisons are graphically illustrated in Figure 18 (published in Turker et al., 2017).

![Graph showing factor scores for the three components of the PCA analysis (musicality, language talent and WM) compared in subjects with a single gyrus, CSDs and CPDs. Factor scores range from -1 to +1.](image)

**Figure 18** Factor scores for the three components of the PCA analysis (musicality, language talent and WM) compared in subjects with a single gyrus, CSDs and CPDs. Factor scores range from -1 to +1.

*t*-tests were conducted to test differences between the groups and revealed that for all three components of the PCA, performance of subjects with CPDs was highest. With regard to musicality, subjects with a CPD were significantly better ($M = 0.48$, $SD = 0.97$) than subjects with CSDs ($M = -0.4$, $SD = 0.76$; $t_{(20)} = -2.4$, $p = .027$). In the language aptitude testing, subjects with CPDs ($M = 0.48$, $SD = 0.80$) were significantly superior to subjects with CSDs ($M = 0.15$, $SD = 0.75$; $t_{(16)} = -2.7$, $p = .015$). Moreover, they were also superior to those with a single gyrus ($M = -0.99$, $SD = 1.0$; $t_{(16)} = -3.4$, $p = .004$). Looking at the component WM, subjects with CPDs ($M = 0.65$, $SD = 0.91$) performed significantly better than subjects with CSDs ($M = -3.2$, $SD = \ldots$)
0.86; \( t_{(20)} = -2.6, p = .019 \) and subjects with a single gyrus \( (M = -0.52, SD = 0.88; t_{(16)} = -2.7, p = .016) \).

5.2.1.4. Comparisons of High and Low Aptitude Learners

In a next step, I wanted to find out which of all administered tests would show substantial differences between those subjects with high and low scores on (1) the Hindi speech imitation task, and (2) the MLAT (total raw score). \( t \)-tests for independent samples were performed for that purpose in both cases. First, differences according to the speech imitation score will be discussed (high speech imitators: \( N = 14 \), low speech imitators: \( N = 16 \), see information in section 5.1.1. and Table 10). Only significant results and trends are reported. Results including differences in mean scores, including \( t \)- and \( p \)-values are provided, as well as the corresponding effect sizes.

<table>
<thead>
<tr>
<th>Instruments</th>
<th>High imitators</th>
<th>Low imitators</th>
<th>( t )-value</th>
<th>( p )-value</th>
<th>( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instruments</strong></td>
<td>1.6 ± 0.9</td>
<td>0.9 ± 0.9</td>
<td>-2.317</td>
<td>.028</td>
<td>.362</td>
</tr>
<tr>
<td><strong>English score</strong></td>
<td>7.2 ± 1.7</td>
<td>5.7 ± 2.1</td>
<td>-2.095</td>
<td>.045</td>
<td>.365</td>
</tr>
<tr>
<td><strong>Digit span forward</strong></td>
<td>10.5 ± 1.7</td>
<td>8.8 ± 1.7</td>
<td>-2.725</td>
<td>.011</td>
<td>.447</td>
</tr>
<tr>
<td><strong>Non-word span</strong></td>
<td>8.4 ± 1.8</td>
<td>6.9 ± 1.4</td>
<td>-2.532</td>
<td>.017</td>
<td>.422</td>
</tr>
<tr>
<td><strong>MLAT total</strong></td>
<td>89.6 ± 9.0</td>
<td>78.4 ± 15.9</td>
<td>-2.267</td>
<td>.032</td>
<td>.400</td>
</tr>
<tr>
<td><strong>MLAT IV</strong></td>
<td>31.5 ± 5.4</td>
<td>27.5 ± 5.2</td>
<td>-2.001</td>
<td>.056</td>
<td>.353</td>
</tr>
<tr>
<td><strong>AMMA rhythm</strong></td>
<td>32.8 ± 3.9</td>
<td>29.8 ± 4.8</td>
<td>-1.826</td>
<td>.073</td>
<td>.324</td>
</tr>
</tbody>
</table>

Table 10 Results of students’ \( t \)-tests. Differences in mean scores \((M \pm SD)\) between subjects in the high and low speech imitators (mean scores and standard deviations are rounded to one decimal place). \( t \)-values, \( p \)-values and correlation coefficient \( r \) (effect size; three decimal places) are provided.
Several assessed scales were linked to the differentiation between high and low speech imitators. These were the number of instruments ($t_{(28)} = -2.32$, $p = .028$), the English pronunciation score ($t_{(28)} = -2.1$, $p = .045$), results on digit span forward ($t_{(27)} = -2.73$, $p = .011$), the non-word span score ($t_{(27)} = -2.5$, $p = .017$) and the MLAT score ($t_{(27)} = -2.27$, $p = .032$). In sum, more gifted speech imitators had higher WM (digit span forward and non-word span) and better scores on the MLAT and the English pronunciation. Moreover, they played more instruments. Results are graphically displayed in Figures 19 and 20 respectively. They show differences in performance on WM scales and the MLAT of individuals with high and low speech imitation skills. Although differences in the results on all three phonological WM scales are visually detectable (see Figure 19), only the scores on non-word span proved to be significantly different in the low and high Hindi speech imitators.

![Box plot showing differences in WM capacity](image)

**Figure 19** Differences in WM capacity in adults with high or low speech imitation ability. Scores range from 0 to 14 in all three tests.
With reference to the results in the MLAT language battery, only the overall score showed a statistically significant difference between the two groups, although there is a clear visual trend observable in Figure 20 for the subtests as well.

**Figure 20** Comparison of high and low speech imitators concerning their results on the various language aptitude scores: MLAT III, IV and V.

In addition, I chose a linear multiple regression analysis (method: step-wise forward) to specify which other scales were the best predictors for the criterion variable Hindi speech imitation score. The predictors included were AMMA tonal, AMMA rhythm, English proficiency, MLAT III, MLAT IV, MLAT V, digit span forward, digit span backward, non-word span, number of instruments, and number of learned languages. The model yielded a corrected $R^2$-value of 0.375, which corresponds to an explained variance of 37.5%. Beta-values, indicating the relative importance of contributing variables and summing up to a maximum of 1, were .44 for non-word span, .31 for the number of instruments played, and .25
for part V of the MLAT. In sum, the regression revealed that the three most important predictors for the Hindi speech imitation score were performance on non-word span, the number of instruments played by an individual and results of MLAT V, a measure of grammatical sensitivity. Overall, these three scales explained 37.5% of variation of the Hindi score, which pointed to a high explanatory value of the considered variables for phonetic coding ability.

To detect potential differences in the variety of tests based on aptitude according to the MLAT total raw score (high aptitude group: \(N = 14\); low aptitude group: \(N = 15\)), \(t\)-tests were performed for each assessed scale. Only significant results and statistical trends are displayed in Table 11, where differences in mean scores, including \(t\)-values, \(p\)-values and the corresponding effect sizes (\(r\)) are provided.

<table>
<thead>
<tr>
<th></th>
<th>High aptitude</th>
<th>Low aptitude</th>
<th>(t)-value</th>
<th>(p)-value</th>
<th>(r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruments</td>
<td>1.57 ± 0.94</td>
<td>0.87 ± 0.92</td>
<td>-2.048</td>
<td>.050</td>
<td>.352</td>
</tr>
<tr>
<td><strong>Hindi score</strong></td>
<td>5.64 ± 1.47</td>
<td>3.89 ± 1.23</td>
<td>-3.566</td>
<td>.001</td>
<td>.542</td>
</tr>
<tr>
<td><strong>English score</strong></td>
<td>7.64 ± 1.67</td>
<td>5.13 ± 1.54</td>
<td>-4.230</td>
<td>&lt;.001</td>
<td>.615</td>
</tr>
<tr>
<td><strong>Digit span forward</strong></td>
<td>10.21 ± 1.53</td>
<td>9.0 ± 2.04</td>
<td>-1.806</td>
<td>.082</td>
<td>.318</td>
</tr>
</tbody>
</table>

**Table 11** Results of students’ \(t\)-tests. Differences in mean scores (\(M \pm SD\)) between subjects in the high and low aptitude groups according to the Hindi speech imitation score (mean scores and standard deviations are rounded to one decimal place). \(t\)-values, \(p\)-values and correlation coefficient \(r\) (effect size; three decimal places) are provided.

The results show that those with high language aptitude scores on the MLAT (overall raw score) also played more instruments, performed better in the Hindi speech imitation testing, the English pronunciation assessment and there was a tendency that they had higher WM as measured by digit span forward. Especially the link with the English score was highly significant, meaning that English pronunciation results and results on the MLAT were tightly linked.
5.2.2. Neuroanatomical Results

The major aim of the neuroanatomical analyses was to investigate whether AC morphology could be delineated as a marker for language aptitude. In other words, I was particularly interested in finding out whether a specific form of shape or gyrification of HG could be related to one of the assessed scales for language aptitude. In order to verify this potential link, the 3D reconstructed HGs of each participant in both hemispheres had to be set in relation to performance in language aptitude tests. The same was done for the results on the musicality tests to confirm previously suggested links between right-hemispheric AC morphology and musical ability (J. Benner et al., 2017).

In a first step, all HGs were categorized by description (i.e. CPDs and CSDs were defined and HGs were placed in the corresponding categories) like done in previous studies (J. Benner et al., 2017; Schneider et al., 2005; Seither-Preisler et al., 2014). Two researchers did this categorization independently (myself and an experienced colleague) and if we disagreed on the appropriate category, the case was discussed in much detail and the most suitable categorization was then chosen (see Table 12).

<table>
<thead>
<tr>
<th>Types of HG</th>
<th>Right AC</th>
<th>Left AC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>number (%)</td>
<td>high/low</td>
</tr>
<tr>
<td>Single</td>
<td>8 (26%)</td>
<td>1/7</td>
</tr>
<tr>
<td>CSD</td>
<td>11 (37%)</td>
<td>3/8</td>
</tr>
<tr>
<td>CPD</td>
<td>11 (37%)</td>
<td>10/1</td>
</tr>
<tr>
<td>Total</td>
<td>30 (100%)</td>
<td>1/7</td>
</tr>
</tbody>
</table>

Table 12 Overall distribution of HG types in right and left hemisphere (RH/LH), as compared to distribution in high and low aptitude groups according to the Hindi score.

Regardless of group status, a single gyrus was the most frequently occurring type of HG in the left hemisphere (60%), while it was the least frequently occurring type in the right hemisphere (26%). The other two types, CSDs and CPDs, were evenly distributed in the right
hemisphere. In the left hemisphere, on the other hand, CPDs occurred three times as often as CSDs. The proportion of the types of HG found in the high- and low aptitude groups, according to the Hindi speech imitation score, is also displayed in the second column of Table 12. Regarding the distribution in the hemispheres according to speech imitation group status, CPDs were the most common form found in the right hemisphere of individuals in the high speech imitation group (10 out of 14 in the high aptitude group, compared to 1 out of 16 in the low aptitude group). In the low speech imitation group, on the other hand, CSDs and single gyri in the right hemisphere were almost identically distributed (7 and 8 out of 16). In the left hemisphere, the most frequently occurring HG type for the high and low speech imitation groups were single gyri (9 occurrences in both aptitude groups).

To look into group-related differences in AC morphology between high and low speech imitators, the 3D reconstructed ACs of all subjects were combined to calculate an average AC surface map. This was once done for the overall language aptitude score, and once for the Hindi task (a visualization for the latter is provided in Figure 21).

![Averaged AC surfaces](image)

**Figure 21** Averaged AC surfaces in the left and right hemispheres in individuals of the two aptitude groups (high Hindi speech imitation score on the left, low Hindi speech imitation score on the right). Left hemisphere (L) in blue, right hemisphere (R) in red. AC auditory cortex
Apparently, high and low speech imitators most frequently possessed single gyri with marginal lateral duplications in their left hemisphere. Therefore, no significant difference was suspected with this regard, although statistical analyses were nonetheless performed. In the right hemisphere, on the other hand, it was clearly visible that those with a high Hindi speech imitator score possessed more complete gyri (that is, CPDs). This visualization, however, can only help to observe trends but it cannot confirm whether a difference is also statistically significant. Therefore, it was only computed as an example for the differentiation between high and low speech imitators. In fact, almost all subjects who were in the high speech imitation aptitude, were also in the high overall language aptitude group (according to the MLAT score), i.e. there was large overlap between the different aptitude groups. This was also suggested by the group comparisons that were presented beforehand and showed that those in the high speech imitation group also had significantly higher scores on the MLAT and vice versa. Thus, an averaged surface map of the high vs. low overall language aptitude groups would have looked strikingly similar.

Since averaged surfaces only allow to see patterns, it is meaningful to look at the whole variation of AC morphology in both groups of high or low aptitude speech imitators and each learner individually. For this purpose, Figure 22 summarizes all forms of HGs found in the subjects of the present study. While 14 individuals were in the high speech imitation aptitude group, the low speech imitation aptitude group contained 16 subjects. All types of HG found in the left and right hemisphere respectively are also indicated in the figure itself (s, *, +). The figure contains exactly those 3D mesh reconstructions that were also used for the categorization into the three types of HG as specified earlier.

In order to verify the significance of the previous observations that CPDs appear more often in the high speech imitation and high language aptitude subjects, I performed two one-way ANOVAs, for each hemisphere separately. The ANOVAs were applied to detect score-related links between the three following morphological HG characteristics in the right and left hemisphere: (1) single gyrus, (2) CSD, and (3) CPD. In a one-way ANOVA, variances in the group means within a sample are compared. There is only one independent variable or factor organised in categorical groups in a one-way ANOVA, which is the type of HG (single, CSD and CPD) in this case. The null hypothesis was that there is no difference in the scores on the Hindi speech imitation test between the three types of HGs for either the right or left hemisphere.
Figure 22 AC morphology of all individuals classified as either high or low aptitude individuals according to the Hindi speech imitation score. High aptitude subjects are in the top row, compared to low aptitude individuals in the bottom row. Each specific type is also given for each HG. Left hemisphere in blue, right hemisphere in red. s single gyrus, *CSD common stem duplication, +CPD complete posterior duplication
The first one-way ANOVA found a significant effect for HG morphology in right AC and the Hindi speech imitation score \((F_{(2,27)}= 9.2, p < .001, \eta^2 = .41; \text{see Figure 23})\). According to the results of the Tukey HSD post-hoc tests, subjects with CPDs in their right AC achieved significantly higher scores in the Hindi speech imitation test \((M = 6.1 \pm 1.2)\) than subjects with a single gyrus \((M = 3.9 \pm 1.4; p = .002)\), and subjects with CSDs \((M = 4.1 \pm 1.4; p = .004)\). However, there was no significant difference in the Hindi speech imitation skills between individuals with a single gyrus and CSDs. The results of the analyses for the left hemisphere will not be discussed since none of the ANOVAs yielded any significant results with regard to left-hemispheric AC morphology.

![Figure 23](image)

**Figure 23** Mean total Hindi score (between 0-10) for all three types of AC morphology, namely single gyrus, CSD, and CPD in right AC (A) and left AC (B) of all subjects. CPDs are clearly associated with higher mean scores in the Hindi speech imitation test, as supported by results of the statistical analyses.
Moreover, Chi-Square-tests were performed on the frequency distributions of these neuroanatomical characteristics in the right hemisphere. CPDs were the most frequently occurring type of HG in the high speech imitation group (71%), whereas those subjects in the low speech imitation group had CPDs most rarely (6%; $\chi^2(2) = 14.1, p < .001$). Evidently, CDPs in right AC were associated with particularly high speech imitation scores, whereas variation in the left hemisphere could not be linked to any characteristic form of HG.

To further investigate the potential link between AC morphology and the overall language aptitude score (MLAT total score), another two one-way ANOVAs were performed. As in the first ANOVA, there was only one independent variable, namely HG morphology in the form of the three types distinguished in this study (single, CSD, CPD). The ANOVA for right AC found an association between right AC and the overall language aptitude results as measured by the MLAT (total raw score; $F_{(2,26)} = 12.579, p < .001, \eta^2 = .492$; see Figure 24).

**Figure 24** The relationship between HG types (single, CSD, CPD) and the overall MLAT raw scores (ranging from 0 to 120). In right AC (A), higher scores in the MLAT were linked to the possession of CSDs and CPDs. No relationship was found in the left hemisphere (B).
To verify which HG types were distinctly related to MLAT scores, post-hoc tests were performed. According to the results of the Tukey HSD post-hoc tests, subjects with CPDs achieved higher scores in the MLAT ($M = 92.55 \pm 3.17$) than subjects with single gyri ($M = 67.14 \pm 3.98$; $p = .005$) and subjects with CSDs ($M = 84.64 \pm 3.17$; $p < .001$). The ANOVA performed for left-hemispheric AC morphology did not yield any statistically significant results.

Previous studies by P. Schneider and his research group ‘Music and the Brain’ found that high musicality or musicianship was related to structural variation in right-hemispheric AC (J. Benner et al., 2017; Seither-Preisler et al., 2014). In their studies, a higher number of gyri (multiple gyri) were linked to musicianship or high musical ability. In the present study, I primarily aimed at finding a similar relationship with various measures of language aptitude. However, many subjects of the present study had considerable musical experience and training and played several instruments. Although they were not professional musicians, their musical ability could as well have been the result of right AC morphology as suggested by the aforementioned studies. Therefore, another two one-way ANOVAs were performed to detect potential relationships between HG morphological characteristics and results on the musicality test (AMMA) for the left and right hemisphere respectively.

The results of the one-way ANOVA for the right hemisphere showed a very similar pattern, speaking for a right-hemispheric involvement of AC for musicality as well. More precisely, the mean of the total raw AMMA score in the right hemisphere for individuals with single gyri was $55.7 \pm 3.5$, compared to $56.5 \pm 2.8$ for CSDs, and $65.8 \pm 2.8$ for CPDs. The ANOVA confirmed that those subjects with CPDs achieved slightly higher scores than subjects with a single gyrus and CSDs ($F(2,26) = 3.8$, $p = .036$, $\eta^2 = .23$; see Figure 25). In contrast, statistical analyses failed to detect a difference between single gyri and CSDs for musical ability in the right hemisphere, while analyses of the left hemisphere did not yield any statistically significant results.
Figure 25 Comparison of the mean total AMMA scores (ranging from 0 to 80) for individuals displaying the three distinguished HG characteristics, namely single gyrus, CSD and CPD. Evidently, the mean score for CPDs was significantly higher than that for the other two types.

The findings of the ANOVAs were in accordance with the results of the PCA, which found that all three components were associated with CPDs, but not with the other two types of HGs. Thus, the findings of the statistical analyses clearly point towards an involvement of the right hemisphere for both language aptitude and musicality. It seems that two complete gyri, i.e. CPDs, are advantageous for the acquisition of foreign languages and high musical ability, which will be discussed in more detail in the following discussion and in Chapter 7 of the present thesis.
5.3. Discussion

The neuroanatomical results of the present study suggest that AC morphology is closely related to both musical ability and language aptitude, while there seems to be a strong behavioural link between language aptitude and WM, and a weak link between language aptitude and musical ability. These findings deserve sufficient attention, in particular with respect to the possible involvement of AC for the acquisition of foreign languages and an innate potential for doing so. In order to facilitate the overall discussion and comparison of results of both studies, also regarding the theoretical underpinnings of both language aptitude and musicality, the results will be thoroughly discussed at this point. The later overall discussion of Study 1 and Study 2 will mainly serve to detect similarities and differences and summarize both single studies in a broader context (see Chapter 7).

5.3.1. Auditory Cortex Morphology as a Potential Marker for Language Aptitude and Musicality

According to the findings of the neuroanatomical analyses, right-hemispheric AC morphology seems to be critically involved in the manifestation of two core concepts investigated in this thesis, namely language aptitude and musicality. In the present study, I used a very different and novel approach by applying an innovative method with the aim of detecting subtle differences that were not considered in studies to date. The findings are only related to structural variation that is very likely to be biologically determined and suggest that the possession of two gyri instead of a single gyrus in right AC is associated with high speech imitation skills, high overall language aptitude (measured by tasks comprising vocabulary learning, phonetic coding ability and language analysis ability) and high musical ability.

The neural basis of foreign language aptitude has gained momentum in the past years and specifically the number of functional studies has increased drastically. Although studies looking at functional activation differences between learners of high and low aptitude in language-related domains help to further explore the underlying functional properties of FLL, only investigating structural variation in the human brain with regard to specific skills can shed light on the neurobiology underlying these abilities (Berken, Gracco, Chen, & Klein, 2016). Concerning language, studies so far have suggested that left AC is involved in learning novel
elements of a language (tonal pitch contrasts and phonetic differences; Golestani & Zatorre, 2009) and also in the production and perception of novel speech sounds (Golestani & Pallier, 2007; Golestani, Paus, & Zatorre, 2002). One of the main hypotheses of this study was that AC could be critically involved in outstanding language aptitude. Differences were expected mainly in the left hemisphere as has been demonstrated in numerous studies (Golestani, Price, & Scott, 2011; Golestani & Pallier, 2007; Golestani & Zatorre, 2009; Hu et al., 2013; Wong et al., 2007). However, the results of the present study suggest that individuals with high Hindi speech imitation scores and high overall language aptitude have more than one single HG in their right AC and right AC morphology might thus be a neuroanatomical marker of language aptitude.

In the current study, the occurrence of duplications in the left hemisphere was considerably lower compared to the rather balanced distribution in the right hemisphere. This has been suggested by other studies (e.g., Marie et al., 2015), although the reasons thereof remain unclear. The present study supports this claim that the left AC displays less variation than right AC, but it cannot answer the cause of this imbalance. Regardless of this imbalanced distribution of morphological characteristics of HG, it might well be the case that the left hemisphere, or at least left-hemispheric primary and secondary auditory areas, are not as important for the development of language aptitude as initially hypothesized. Even if the findings are somewhat surprising, an involvement of the right hemisphere for language aptitude is not an entirely new idea. Quite recently, a number of appealing studies by Kepinska and colleagues (Kepinska, de Rover, Caspers, & Schiller, 2017a, 2017b; Kepinska, Pereda, Caspers, & Schiller, 2017) on language analytic ability also highlighted the significance of the right hemisphere for language aptitude. In their studies, Kepinska et al. managed to link differences in functional connectivity, brain activation and brain structure in the right hemisphere to more skilled language learners. In their work, they categorized individuals according to the LLAMA language aptitude test (LLAMA F). They confirmed that the highly skilled grammatical learners displayed significant differences in functional connectivity, functional activation and GM volume in several right-hemispheric language-related areas. Moreover, Prat et al. (2016) reported that highly successful second language learners functionally relied more on their right hemisphere than less successful learners, supporting the vital role of the right hemisphere for a variety of linguistic skills.
The present study suggests that a larger number of HGs in an individual’s right AC is related to both linguistic and musical skills but the exact function of more gyri or duplications in AC is far from understood to date. Not only with regard to language aptitude, but also even on a more general level, the occurrence of CPDs in HG and their specific functional or structural role remain unclear. Looking at the brain more generally, research has not succeeded so far to delineate the functions of certain sulci or gyri, and also inter-individual variation associated to certain skills. It is common knowledge that HG hosts primary auditory areas, while HG duplications are thought to host among others associated language-related parabelt areas, which could be more important for structural and functional connectivity between AC and other brain areas. Thus, the connectivity between first HG and its duplications may have a hitherto unknown impact on auditory functioning, which could positively affect the development of linguistic and musical skills (see discussion in Chapter 7). Leaving aside the differences between the right and left hemisphere, there is convincing evidence for a specific structure-function relationship of HG duplications, which remain yet to be uncovered.

It is noteworthy that not only language aptitude, but also musical ability, showed a link to AC morphology in the right hemisphere. A considerably larger prevalence of HG duplications has been found not only in right AC of musicians (J. Benner et al., 2017), but also in left AC of linguistically more talented subjects (Golestani & Pallier, 2007; Golestani et al., 2002). The results of the present study confirmed the prevalence of HG duplications in subjects with high language aptitude and high musicality. The major difference to earlier studies, however, is that both were linked to right AC. The finding that speech imitation, language aptitude and musicality yielded parallel results regarding structural variation in right AC was very unexpected due to the lack of correlation between the three in the statistical analyses. In the analyses, there was no significant relationship between the Hindi speech imitation score and neither part of the AMMA musicality test. Only the Hindi test and one part of the MLAT correlated with each other. Evidently both linguistic and musical skills, even if not directly correlated with each other, appear to be closely linked to right HG. Moreover, when I compared participants’ individual factor scores on each of the three components for the three HG types, it was obvious that CPDs went hand in hand with higher scores in all three components of the PCA. In other words, for all three factors found by the analysis (musicality, working memory, and language aptitude), performance of subjects with CPDs was highest.
It is common knowledge that musical ability is closely linked to the right hemisphere and also recent research has highlighted the significance of right-hemispheric AC for the processing of musical cues (Kraus & Chandrasekaran, 2010). The findings of the present study thus support the view that the right hemisphere is not only functionally but also structurally essential for the perception and processing of music and music-related features. The Hindi speech imitation task is basically a WM task which requires excellent use of the articulators to produce speech and a functionally highly efficient AC to perceive subtle differences in the speech input. Is it possible that only phonetic coding ability is highly dependent on (1) musicality (rhythmic and melodic discrimination abilities) or (2) processing of right AC? If decoding of speech input were more dependent on auditory processing of music-relevant features in the right hemisphere, this would explain why only the right hemisphere was linked to speech imitation ability. It could also partly explain why the overall MLAT score could be linked to right AC as well. However, the correlation between phonetic coding ability (MLAT IV) and the Hindi task was only moderate and did not survive the correction for multiple comparisons.

A major challenge is to explain if and to what extent differences in AC are due to language aptitude or due to musical ability. A first explanation is that the present study confirms the importance of CPDs in the right hemisphere for the processing of music-related cues and because the Hindi speech imitation involves non-speech processing expertise, analyses revealed a similar relationship between the two. This could not explain the link with the overall language aptitude results, however. A second explanation is that I found a neuroanatomical marker for foreign language aptitude, explaining the significant link with both language aptitude measures, and this marker influences musical processing as well, leading to a high expertise in both domains. Alternatively, the results could suggest that the right hemisphere is more important at the initial stages of auditory processing, in particular in children. At such a young age, aptitudes can still be developed into talents and an advantage in auditory analyses could therefore prove to be advantageous for linguistic, as well as musical processing.

The neuroanatomical traits investigated in the present study are very likely to be biologically determined markers for language aptitude and musicality than to be caused by environmental influences and experience/training. A recent study found that the individual morphology of structures like HG was extremely stable from childhood into adulthood despite
high inter-individual variation (Seither-Preisler et al., 2014). In their longitudinal study (see Study 2), Schneider, Seither-Preisler and colleagues (see also Serrallach et al., 2016) followed children over years (nine years since the onset of the study) and looked at structural properties of AC every second year through MRI scans. They found that the ACs of these children (overall, 220 children participated in the study from the beginning on) changed as little as 0.001% over a period of 6 years. Most interestingly, however, these children were either musically trained or not and it was expected that the morphology of their AC would grow like a muscle due to musical training and experience. This was not the case and they therefore argued that the principal structures of our brain do not change during our lifetime. Rather, it appears that the specific structures of our brain (sulci and gyri) have a strong biological component, which may be determined genetically, prenatally or during very early post-natal periods. In other words, each person is already born with a certain number and a certain shape of gyri and sulci, and these shapes cannot be easily modified or altered through training or experience. Thus, individual differences in AC morphology in the present study are very unlikely to be the results of environmental influences or practicing behaviour.

The findings that AC morphology is a rather stable, neuroanatomical trait not easily alterable through training or experience do not necessarily mean that language aptitude is an innate, biologically determined ability. I am aware that the view of language aptitude has changed in the past decades and is more and more frequently referred to as a dynamic construct that may indeed undergo change over time. Still, these developments do no imply that language aptitude cannot be partly biologically determined, with stable neuroanatomical traits that are favourable for the development of language learning skills being present at birth already. From the present study, it is impossible to say whether language aptitude and AC morphology are innate and have not changed in the individuals I assessed, however.

To conclude, the right hemisphere might be more important for language acquisition and processing than initially assumed. However, more research will be needed to explore the involvement of the right hemisphere, in particular of the right HG, in different aspects of language aptitude. Also, given the variety of regions in the brain that are essential for language processing, it will be necessary to develop segmentation methods to structurally analyse other essential language processing areas, such as the IPL (Dogil & Reiterer, 2009; Golestani & Pallier, 2007; Golestani et al., 2011; Golestani & Zatorre, 2009; Hu et al., 2013). Overall, the findings of the present study do not contradict theories claiming that the left hemisphere is
more important than the right for language processing in healthy adults. Nevertheless, they highlight the need to consider specific areas in the right hemisphere to be predominantly involved in certain linguistic skills at certain stages or during specific tasks. The results further highlight that language aptitude and musicality are both linked to right-hemispheric AC morphology. If it is possible to determine neuroanatomical markers that remain highly stable from early infancy into adulthood, this challenges the belief that the capacity to acquire associated behavioural skills can be markedly altered throughout lifetime. Furthermore, if the structures of certain brain regions are strongly related to specific behavioural skills, a main aim of researchers should be to find out how they control the natural unfolding and development of these skills. Although it is clear that numerous external variables also affect the development of linguistic and musical skills, the findings firmly support the claim that there are strong innate and/or prenatally determined neurological factors that remain to be uncovered in the near future.

5.3.2. The Concept of Language Aptitude

Let us briefly recapitulate what the present findings can tell us about the overall concept of language aptitude and the assessed scales supposed to measure language aptitude (to different extents). The Hindi speech imitation test was introduced as a measure of language aptitude, although statistical analyses showed that it is more strongly linked to WM. It was expected that the Hindi score should primarily measure phonetic coding ability or phonetic memory, i.e. in how far subjects were capable of decoding speech input, memorizing it and transferring it into correct motor movement to produce the exact same sounds as previously heard. In the analyses, the Hindi speech imitation was primarily dealt with as a subskill of language aptitude, although it served as a good example that WM is a primary component of language aptitude. Overall, however, the measures used for language aptitude and the results of the PCA (revealing that both the English pronunciation score, the MLAT and the Hindi load onto one variable) strongly support the three tests’ validity for measuring different components of language aptitude.

While the Hindi task was expected to be primarily a measure of WM and auditory perception skills, the English pronunciation task is thought to give a good estimate of pronunciation proficiency in an already acquired language. High pronunciation proficiency
relies on a certain ability for phonetic coding and it thus seems very likely that the two actually go very well in hand. In this sense, the Hindi score could be seen as a precursor for the English score since high phonetic coding ability and phonetic memory should lead to an excellent pronunciation in any language acquired by an individual. Even after correction for multiple comparisons, a positive, moderate correlation could be found between the two ($r=0.395$, $p=.031$).

A major issue with regard to the choice of the language aptitude tests was that there were hardly any language aptitude tests suitable and available for adults and in German. Some of the adults of the study had spent considerable time in an English-speaking country or had studied English. Despite the fact that everyone who grows up in Germany is expected to speak good English (at least level after finishing school), these differences in skills might have largely affected performance on the MLAT, which had to be applied in English. Moreover, the relationship between the Hindi score and the English score was not as strong as hypothesized, indicating that speech imitation of unknown material and good pronunciation skills do not necessarily go hand in hand. In my view, language aptitude is a predominantly biologically driven capacity that unfolds over time only in interaction with the environment. Therefore, it seems more likely to assume that a high Hindi score should lead to a better English score, i.e. the better you are at decoding, retaining and reproducing speech material, the easier it should be to develop excellent pronunciation skills in a given language and not vice versa. This was not the case, however.

Clear significant positive correlations were found between the English pronunciation score and MLAT parts III and IV. The results of the correlational analyses in these two cases were quite robust, leading to two possible interpretations: (1) the English pronunciation score and the MLAT both rely on one specific aspect, namely overall English proficiency, or (2) high pronunciation proficiency is the result of particularly high language aptitude. The first interpretation highlights the aforementioned problem that the MLAT had to be administered in English, obviously giving subjects with excellent English skills an advantage, even in the absence of high language learning potential. If we assume, however, that excellent English skills are the result of high language aptitude, we can minimize the influence of English skills on the MLAT and suggest that phonetic ability (MLAT III) and grammatical sensitivity (MLAT IV) are essential for acquiring native-like proficiency in any language.
Interestingly, the correlational analyses failed to detect any relationship between English pronunciation proficiency and vocabulary learning. This could be an indicator that vocabulary learning is more closely associated to other factors that were not assessed in this study, such as learning strategies or motivation. Additionally, the proposal that verbal WM skills (attributed to the phonological loop) are essential for novel vocabulary learning (Atkins & Baddeley, 1998; A. Baddeley et al., 1998; Gathercole & Baddeley, 1989) could not be supported by the findings of the present study. Unfortunately, no German version of the MLAT exists to date, but in the past years language-independent tests, such as the LLAMA language aptitude battery (Meara, 2005), have gained popularity (Granena, 2013, 2018; Rogers, 2017) and are therefore probably a better choice for assessing language aptitude.

One result that deserves further discussion is that the number of languages spoken by each participant could not be linked to any other score, except for age. The first logical assumption would be that the more languages you have learnt, the better you should be at learning different aspects of a new language (as measured by the MLAT; relating to meta-linguistic awareness). This is an intuitive assertion that readers will probably agree with instantly. Likewise, the more languages you speak and the more sounds you have learned to pronounce, the better your English pronunciation and your Hindi speech imitation should be. Some of this could probably be attributed to heightened meta-linguistic awareness (Jessner, 1999, 2014), although pronunciation is less likely to be influenced by this factor. None of these assumptions, however, could be supported by the findings of the present study. Vice versa, I could also argue that those subjects with very high language aptitude (high MLAT scores and indirectly good English and speech imitation) tend to learn more languages due to the facility and ease with which they acquire a foreign language.

One major limitation of including previous language experience was that information on subjects’ proficiency, the learning context and experiences could not be included. In the present study, it was impossible to control in any form how well the participants had learnt the languages, how well they were able to speak them at the time of the assessment and in which context they had learnt them. In the current sample, the number of foreign languages spoken by a subject ranged from one to nine, which is quite outstanding. It would have been both necessary and at the same time impossible to include proficiency measures (e.g., grammar, vocabulary, pronunciation) in all foreign languages while also considering the learning context, duration and getting a detailed account of the overall learning processes. On a very positive
note, on the other hand, the fact that the number of languages did not interact with any language aptitude core strongly supports the initial claim that language aptitude is a rather innate and inflexible capacity that does not change substantially due to learning or practice at least. Both Eisenstein (1982) and A. S. Thompson, (2013) reported that previous language experience or language training had an impact on the results on language aptitude tests in their study, but the present study could not corroborate these findings. In the beginning of language aptitude research, the construct was seen as stable and unmodifiable through practice or training (Carroll, 1962). These assumptions were questioned in the past decades (Klein, 1995; Sáfár & Kormos, 2008; A. S. Thompson, 2013) but the findings might support the claim that aptitude may not really be as dynamic a construct as suggested by some researchers.

5.3.3. Musicality and its Link to Language Aptitude

The results of the PCA defined three highly influential factors for the musicality component, namely rhythm perception, pitch perception (both measured by the AMMA test) and the number of instruments played by each subject. Nonetheless, also a number of other factors loaded less strongly but still noteworthy onto the variable ‘Musicality’. The fact that non-word span could be related to musicality probably reflects the fact that musical processing depends on temporally structured, but semantically undefined information. This information has to be stored in WM and thus non-word span could be a good measure to capture this process. On the other hand, the relationship between MLAT IV, measuring grammatical sensitivity, and musical ability is less transparent and conclusive. Looking at it from a broader perspective, however, it becomes less surprising as the understanding of both language and music depends on internalized grammatical rules. Additionally, the number of languages loaded onto this component, although neither the AMMA, nor the number of instruments, correlated with the number of languages. It hence seems that the PCA factor ‘Musicality’ has a higher validity than the single subscales alone and the number of languages might be related to musicality on a more general level, not sufficiently captured by the single tests alone.

If we think about musicality generally, a relationship with the number of instruments played by an individual is quite obvious. The number of instruments has been linked to a variety of cognitive skills, language acquisition just being one of many (Milovanov & Tervaniemi, 2011; Nardo & Reiterer, 2009). Kraus & Chandrasekaran (2010) have accurately argued that
music is a resource leading to auditory fitness and thus positively influences the acquisition of skills in other domains. This phenomenon is also referred to as positive transfer. The results of the present study support this view by revealing a relationship between the musical domain and several language-relevant skills. More specifically, those subjects who scored very well on language aptitude tests also played more instruments and had better results on the musicality tests. These findings are in accordance with recent research and confirm the strong relationship between language and music (Mireille Besson et al., 2007; L. Lee & Lin, 2015; Milovanov et al., 2010; Schön et al., 2004).

The AMMA test is supposed to be a measure of an individual’s musical potential but has been widely used as a tool for measuring musical ability in various disciplines. The correlational analyses confirmed that rhythm and pitch perception, the subparts of the AMMA test, correlated highly and significantly with each other. This, in turn, supports the fact that subjects with a certain musical ability or so-called musicality, are very good in different musical domains, in this case rhythmic and melodic discrimination abilities. Apart from correlating with each other, both parts also correlated positively and significantly with the number of instruments an individual played. A functionally more efficient AC (Kraus & Chandrasekaran, 2010) could provide better and more advanced auditory discrimination abilities, lending again support to the hypothesis that musical training leads to enhanced auditory processing.

A very interesting finding was the strong correlation between one subpart of the AMMA, rhythm perception, and English pronunciation skills. Other studies have also linked rhythmic skills to the learning of specific foreign languages (Bhatara et al., 2015) and have emphasized that they are vital during first language acquisition (Nazzi, Jusczyk, & Johnson, 2000). It is likely that rhythmic discrimination and perception skills facilitate the acquisition of a language and due to considerable practice and experience over time, this could lead to improvements in pronunciation. We should not forget that the English score might as well be a good indicator for overall English skills and not only pronunciation. In addition, the text that had to be read for the English pronunciation rating (“The North Wind and the Sun”) is very lyric-like. It is a fable for children, to be more precise, and performance of reading this fable might have given an advantage to musically more gifted subjects. Overall, I favour the view that English skills are related to rhythm perception, not only because of musical training, but also because of the importance of rhythm for prosodic and phonological skills.
In sum, for future research it will be vital to spend more time investigating the concept of musical aptitude/musicality by using various measures with the aim of fully grasping the construct. More factors might need to be taken into account and even if the AMMA test is considered a standard measure for musicality, it could be very useful to additionally calculate an index of musical practice (as done in Seither-Preisler et al., 2014, Serrallach et al., 2016). Such an index of musical practice is a fine-grained measure for musical expertise and training, and it enables researchers to implement numerous aspects of musicality (such as different music-related skills and associated variables like amount of practice, singing interest etc.). Of course, it is only meaningful to undertake such a calculation with individuals with considerable musical practice and training over years, but it will surely be of high relevance when further investigating the relationship between language and music.

5.3.4. Working Memory Capacity as a Predictor of Language Aptitude

It is hardly surprising that speech imitation requires excellent WM and the claim that WM is essential for language aptitude is doubtlessly not far-fetched (Miyake & Friedman, 1998; Wen, 2016). (Coady & Evans, 2008; Conti-Ramsden, Botting, & Faragher, 2001). In the present study, there was a consistent, positive relationship between speech imitation skills and different measures of phonological WM capacity. The three assessed WM scales were digit span forward, digit span backward, and non-word span. Two of the three correlated positively and significantly with the Hindi speech imitation score, non-word span showing the highest correlation in this regard. Overall, however, non-word span could only be linked to two variables, namely age and the Hindi score. While WM surely improves from childhood to adolescence and reaches final capacity in adulthood, it seems very unlikely that WM improves substantially between 20 and 40 years of age. However, there have been studies showing that WM reaches its highest capacity at around 30 years of age (Alloway and Alloway, 2013). This might be an explanation for the link of non-word span with age, although it is surprising then that the other two WM scales did not correlate with age at all.

It further comes as a surprise that the correlational analyses failed to detect any relationships between the three WM scores, on the one hand, and WM and other variables, on the other hand. A major limitation must have been the very small sample size because the WM scales correlated with each other in the overall study by Dogil and Reiterer (2009).
Correlational analyses are interesting to detect relationships between assessed scales but with small sample sizes, it is impossible to achieve statistically significant results, in particular when you have to apply correction as well. Therefore, the lack of association is most certainly the result of a lack of statistical power. WM comprises different components that influence each other and are thus expected to share at least some form of common basis (Baddeley, 2017; Baddeley & Hitch, 2000). Results of the PCA also revealed a clearly defined component for WM capacity. The variables that loaded highest on this component were the Hindi speech imitation task, non-word span, digit span forward and digit span backward. Additionally, the number of instruments loaded on this component as well, supporting a possible interaction between musical expertise and high WM capacity. The PCA also pointed toward a stronger relationship between the three WM scales, and also with the Hindi score, than confirmed in the correlational analyses.

One major hypothesis was that phonetic coding ability (i.e. the ability to decode and retain speech input) was the one component of language aptitude that should be best measured through the Hindi speech imitation task. Non-word span, however, is a strikingly similar task. In spite of slight differences, both the non-word span task and the Hindi speech imitation task rely on speech material that consists of simple syllables consisting of a consonant and a vowel (CV syllables). During non-word span, the WM load increased gradually with one syllable being added at a time. Whereas this task got more complex, the Hindi speech imitation used two categories of similar elements only, namely words and sentences. The words were all of similar length and the large increase in complexity came abruptly when changing from imitating words to sentences. Regardless these differences, both heavily relied on WM capacity and I therefore propose that both tests are equally useful and trustworthy measures of WM on the one hand, and phonetic coding ability on the other hand. Possibly due to the small sample, no relationship between results on the MLAT and WM measures could be found.

Non-word span has been used as an indicator for specific language impairment (Coady & Evans, 2008; Conti-Ramsden, Botting, & Faragher, 2001) and dyslexia (Moll, Göbel, Gooch, Landerl, & Snowling, 2016). It would have been logical to assume that high or low language learning aptitude depend highly on non-word span, as it seems to be implicated in language difficulties specifically. This supports the proposal to further develop non-word tasks to ameliorate language aptitude testing batteries and integrate WM more explicitly in language aptitude testing as earlier suggested by Chan, Skehan, & Gong (2011).
Apart from non-word span, also digit span forward correlated positively and significantly with the Hindi speech imitation task. Since digit span forward and non-word span are both easy WM tasks that require storage but not manipulation of input, it is hardly surprising that they both correlate highly and significantly with the Hindi speech imitation. What is surprising, is that other studies have shown that high ability and success in FLL, in this case the second language was English, correlated with WM tasks of differing complexity (e.g., easy and complex span; Biedroń, 2015; Kormos & Sáfár, 2008; van den Noort et al., 2006). The English pronunciation score, however, was not correlated with any of the WM scales.

To summarize, the small sample size might have severely limited the statistical power of the present study, which would explain why the WM scores did not correlate with one another and with very few of the other variables. The link with the Hindi speech imitation score, on the other hand, was strong and favours the assumption that WM is related to language aptitude. Specifically, speech imitation was found to rely heavily on WM. It might be more or less significant at different stages and for different language-related skills, however.

As the first study had already revealed interesting insights into the potential role of right AC for language aptitude and speech imitation, I was keen on finding out whether the findings could be replicated in the younger population. The present study (Study 2) can therefore be seen as a replication study on the one hand, and an original study on the other hand. It is original in the sense that it is one of the first to closely investigate the relationship between WM, language aptitude, arithmetic competence, musicality and school achievement in a group of children and teenagers. Moreover, it is the first study that closely investigated AC morphology in a younger population, particularly with regard to musical abilities and language aptitude. In this study, not only the different gyrification patterns in HG were compared, but also differences in GM volumes in the ACs of the subjects. The younger population provided the unique opportunity to further explore whether the cognitive abilities at the core of this thesis would also relate to academic achievement and another important cognitive ability, namely arithmetic fluency. Arithmetic fluency was only touched upon briefly in the introduction, which is why it will be described in much detail in the methods section of this study.

The second study presented in this thesis included altogether 64 children (for behavioural analyses), 42 of which were then included in the neuroanatomic analyses (in the form of extreme groups with either particularly high or particularly low language aptitude). As mentioned in Chapter 1, the study was conducted at the University Hospital Heidelberg, Germany, and the children were taking part in a longitudinal study organized by P. Schneider and his research group there. The typically developing children who were included in the analysis were tested on selected weekends between 2016 and 2018. While assessment took place in Heidelberg, analyses were primarily conducted in Graz, only marginally in Heidelberg. In the following chapter, specific attention will be paid to giving precise information on all tests additionally used in the child population that were not part of the adult study (Chapter 5). Apart from that, more attention will be paid to behavioural outcomes that could not be included
in the adult group, such as school achievement, arithmetic fluency and self-reported aptitude, as well as parent-reported aptitude.

6.1. Methods

6.1.1. Subjects

All subjects were healthy, monolingually raised German native speakers between 10 and 16 years of age. All of them took part in a longitudinal study that aimed to explore the effects of musical training on overall development and related cognitive abilities at the Medical University Heidelberg, Germany. The whole study was conducted at the University Hospital Heidelberg and children were recruited from different parts of Germany, mostly Heidelberg and surroundings but also parts of Baden-Württemberg, Hessen and the city of Hamburg. All of them had started learning their second language, namely English, when they were 10 ± 1 years old, which is the typical case for children in Germany. They spoke between one (English) and three foreign languages (mostly French, Spanish or Latin), depending on the type of school they were attending. Criteria for exclusion were a history of neurological or psychiatric conditions, as well as learning disabilities (e.g., dyslexia or dyscalculia; assessed by clinicians and paediatricians) or developmental disorders, such as autism spectrum disorders or attention deficit hyperactivity disorder (ADHD). Additionally, children who had been raised bilingually, that is, with two native languages, were also excluded from the present study.

First, behavioural results are reported from the complete sample of subjects (N = 64; M_{age} = 14.4 ± 1.1 y; 32 females). Since one aim was to reveal potential behavioural and neuroanatomical differences on a group-statistical level, the participants were then divided into three equal groups according to the percentile ranks achieved in the LLAMA language aptitude battery (Meara, 2005). For these groups, the cut-off points were set at the 33rd and the 66th percentile, and the intermediate group was discarded from further group-statistical analyses. The final sample (N = 42; M_{age} = 14.5 ±1.3 y; 19 females) included the two extreme groups, namely a high aptitude group (N = 21; M_{age} = 14.6 ± 0.9 y; 9 females) and a low aptitude group (N = 21; M_{age} = 14.4 ± 1.4 y; 10 females).
6.1.2. Testing Procedure

All testing took place at the Department of Neuroradiology at the University Hospital Heidelberg, the research MRI lab therein and additional rooms situated next to the MRI lab. Subjects’ parents received monetary reward for their participation in the study including travel costs and accommodation for the night in between testing. Before participating in the experiment, all subjects and their parents provided written informed consent as approved by the Ethics Committee of the Medical University Heidelberg.

Testing took place on two days during fixed weekends. A maximum of six to ten children participated in both the behavioural testing session and the structural MRI session on two consecutive days. Each child had to complete the behavioural testing, consisting of several paper-and-pencil and online tests on musicality, language learning and the like, and the T1 scan of the MRI session. All tests were conducted in pseudorandomized order with enough breaks in between for subjects to rest. Children who did not manage to complete all tasks during these two days were asked to complete the MULT/AP online test at home and communication happened via e-mail after the completion of the test.

6.1.3. Questionnaires and Interviews

Parents were asked to fill out questionnaires about their child’s history of language learning including possible speech and language disorders in early childhood, speech and language therapy and bilingual upbringing in order to make sure none of the children met these criteria (e.g., language disorder, bilingual upbringing). If they did, they were excluded from the analyses. The questionnaire further asked parents to give a concrete estimation of their children’s language learning potential on a scale from 0 to 10 (0 equalling a lack of potential, 10 equalling an extremely high potential for learning foreign languages). Furthermore, the questionnaire included detailed accounts of subjects’ school grades and the languages they were acquiring at the time of the assessment. Although the same information was also asked from children, parents’ responses were checked to verify children’s reports and in case of a discrepancy, parents were asked for clarification at the end of the testing session.

Participants of the study, i.e. the children and teenagers, were interviewed in short sessions about their self-reported potential for the acquisition of foreign languages (using an
identical scale to that for the parents, ranging from 0 to 10). Children were further asked about their school grades in German, English and other potential foreign languages and whether they considered themselves good singers and how much they liked singing (both separately on a scale from 0 to 10 points). Moreover, they provided details of their musical experience. They provided information on how many instruments they were playing at the time of the assessment and how many they had learnt (for a duration longer than 6 months) in their lives. Special attention was paid that subjects and their parents could not interact during the interview or while their parents completed the questionnaire.

6.1.4. General Intelligence Testing

In order to make sure that none of the observed differences in language aptitude in the children were due to differences in intelligence, which can be a substantial confounding factor, children’s nonverbal IQ was tested with the revised version of the Culture Fair Intelligence Test (CFT20-R; Weiß, 2008). The four chosen subtests, namely ‘substitutions’, ‘classifications’, ‘matrices’, and ‘reasoning’, are considered as measures of general intelligence (general fluid ability) according to the classical intelligence model (Cattell, 1943). In the present study, age norms were used to compare results between subjects. According to the ICD-10 scheme (WHO, 2010; F70: IQ 50-69: mild intellectual disability) the cut-off criterion for exclusion from the study was an IQ below 70 points. As all subjects performed better than that, nobody had to be excluded. However, to verify that intelligence did not play a role, student’s $t$-tests for comparison between groups were performed (see 6.2).

6.1.5. The LLAMA Language Aptitude Battery

Children performed the LLAMA language aptitude battery (see 2.5.2) instead of the MLAT (see 2.5.1) used in the first study. The MLAT is only suitable for adults and only available in English and was therefore not an option for the child study. For the present study, the LLAMA was chosen as it has recently become very popular because it is easy to administer, language-independent and available for free online. Moreover, it is suitable for children of any age. Scores of the LLAMA lie between 0 and 100%, guidance for interpretation of results is suggested in the LLAMA manual. Since the LLAMA does not include a detailed instruction, specific instructions for each task that explained the training and testing phase of each of the
subparts of the LLAMA were formulated in written form and read to subjects. Subjects always received information on how much time they would need for the specific task. The LLAMA test was already described in previous chapters and the separate parts will only be briefly summarized at this point. The specific instructions were always read before beginning the testing of each subpart of the LLAMA.

In the ‘Vocabulary learning’ subtest (LLAMA B) children had two minutes to learn as many words associated with tiny figures as possible (only visual input). This subtest assessed their ability to quickly form links in memory. The ‘Phonetic memory’ subtest (LLAMA D) assessed the recognition of previously heard words in an unknown language. Participants were auditory presented with words of an unknown language one quickly after the other. After this auditory presentation, they were presented with one word at a time and had to decide for each stimulus if it was part of the sequence presented beforehand or not. In the ‘Sound-symbol correspondence’ subtest (LLAMA E), children had two minutes to learn associations between 27 simple combinations of digits and letters (e.g., 0ī) and CV syllables (e.g., that 0ī corresponds to the spoken syllable /pa/). In the testing phase, they were presented with auditory combinations of two pairs of digits and letters (e.g., /patu/) and had to find the correct written form of these (e.g., 0ī3ē). Finally, in the ‘Grammatical inferencing’ subtest (LLAMA F), children learnt the syntax and semantics of an unfamiliar language in five minutes by being provided with pictures and corresponding sentences. One picture always corresponded to one sentence that gave information about the syntax of the language and the meaning of elements in that sentence. After the learning phase, they saw the same pictures and completely new pictures and had to choose the grammatically correct sentence to describe the pictures. For the new pictures, they must have understood the semantic and syntactic rules of the language in order to be capable of choosing the correct sentence.

All LLAMA tests were administered starting with LLAMA B and proceeding as suggested to LLAMA F. Breaks in between tests were avoided, and children were not allowed to ask questions concerning the answers of the test. During LLAMA F, children were allowed to take notes on a sheet during the learning phase. They were allowed to have a look at their notes during the testing phase then.
6.1.6. **The Hindi Speech Imitation Task**

The Hindi speech imitation test (see 2.5.4.1) was the same that was used in the adult group with the exact same number of stimuli used. Although subjects were much younger than in Study 1, all words and sentences were included, even if children were expected to show considerably worse performance than adults. Subjects were instructed to listen carefully via headphones, and they heard each language input three times in a row before being asked to repeat once what they had just heard.

In a separate online rating, native speakers of Hindi \(N = 8\); 4 females) were paid to rate the speech samples of all children on a scale from 0 to 10 indicating how well the participants had managed to imitate the speech input. This time only 8 raters were hired since the number of raters in the adult group was unnecessarily high. During the rating procedure, all samples were randomized and the samples the children had heard were always provided together with their imitation attempts. Inter-rater reliability was calculated, and Hindi score was calculated from the mean ratings of each rater.

6.1.7. **Musicality Assessment**

As in the adult group, the AMMA test by Gordon (1989), which assesses the accuracy of tonal and rhythmic perception by comparing a given standard and a comparison melody that can be the same or slightly modified regarding pitch or rhythm, was used for the assessment of musicality. Since the conduction and analysis of Study 1, the AMMA has been externally validated by comparing performance to the Musical Aptitude Profile (MAP) developed by the same author (Hanson, 2019). Correlations of the different scales ranged from \(r = 0.7 – 0.8\). The AMMA had also been re-validated by the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen, Gingras, Musil, & Stewart, 2014) that assesses self-reported musical skills and behaviours on various dimensions in the general population.

6.1.8. **Working Memory Capacity**

Similar to the adult group, children had to perform three span tasks: digit span forward, digit span backward and non-word span. None of the tests were changed in any form because of the younger age of the participants. The scoring procedure remained the same for this set of
subjects. For each of the three tasks, one correct repetition yielded one point and altogether, 14 points in total could be achieved in each of the separate tasks (maximum for all three tasks: 42 points).

6.1.9. Arithmetic Competence

Mental arithmetic is a fundamental ability underlying all mathematical calculations an individual performs. While arithmetic competence had not been assessed in the adult group, I was keen on administering this test in the child population. A research group at the Department of Educational Neuroscience at the University of Graz (led by Prof. Dr. Roland Grabner) adapted the arithmetic fluency test as used in the present study. The arithmetic fluency test is loosely based on the French Kit test of arithmetic skills (developed by French, Ekstrom, & Price, 1963) and was used in the present form with adults in a study by Vogel et al. (2017). It included four types of calculations, namely additions, divisions, subtractions and multiplications divided into separate units in a booklet. Overall, subjects were asked to perform as many calculations as possible in a short period of time. The time frame was usually set to 50 seconds, whereas the introductory training session with divisions only lasted for 25 seconds. Children were given clear instructions at the beginning of the task since the test is very time-sensitive and children often feel under pressure when having to perform calculations in short periods. First, children were instructed to open the arithmetic fluency booklet, then wait for the start signal (given by the tester) and start with the calculation at the top of each column and proceed vertically. They were specifically told not skip any calculations in between, for instance if they considered them too difficult.

The practicing trial with divisions was followed by two pages of additions, two pages of subtractions and one page of multiplications (always 50 seconds per page). Additions and subtractions included either simple calculations (with numbers ranging from 1 to 9; first page) or more complex calculations (with numbers ranging from 1-20; second page). One correctly solved calculation yielded one single point and a total raw score for each type of calculation was noted. The maximum number of calculations per page varied with complexity, with more complex tasks (complex subtractions, for instance) consisting of fewer calculations overall. None of the children (and adult test takers in a pilot phase) came close to finishing all
calculations of a page in 50 seconds. An overall sum score was calculated that was then used for all statistical analyses, except for the PCA.

6.1.10. The MULT/AP

In addition to the LLAMA and the Hindi speech imitation task, the MULT/AP (see 2.5.4.2) was administered to test children’s ability to perceive differences between Hindi speech utterances. Since the Hindi speech imitation task is a pure measure of imitation (motor control and pronunciation), one aim was to determine whether subjects with high imitation scores would be better in the Hindi speech perception task as well. Looking into the relation between the two test scores could certainly help disentangle the relationship between production and perception. Therefore, the participants of the study also performed the MULT/AP.

The MULT/AP was administered online on a laptop to one child at a time during the whole testing session. If children failed to complete the test during the fixed weekend, they were given online access to complete the test at home on their own. This was no problem since the test is self-explanatory with explicit instructions online.


For the neuroanatomical analysis in the younger population, not only overall morphology and gyrification patterns, but also GM volumes of the corresponding HGs were calculated. The segmentation (see also 5.1.5 for a comparison) is therefore slightly different from that of the first study. For segmentation, Brain Voyager software QX 2.8 (Brain Innovation, B.V, Maastricht, NL) was used again.

A T1-weighted structural magnetic MRI (Siemens, TrioTim, 3 Tesla) was performed to investigate the anatomy of AC. A standardized individual approach of three-dimensional GM surface reconstruction of auditory subareas (HG, PT) was applied to account for individual morphology and gyrification patterns. All brain images were adjusted in contrast and brightness, precisely corrected for inhomogeneity and rotated in direction of the antero-posterior commissural line. The superior temporal plane, including HG, the aSTG and PT, was segmented into sagittal MRI slices along the Sylvian fissure. This was done by using the
standard definition of the landmarks of AC and approved additional criteria: the first complete Heschl's sulcus with a large mediolateral extent (> 97%) and pronounced depth was used as the posterior boundary in the case of single HGs and CSD, and the last complete Heschl’s sulcus in the case of CPDs or multiple HG duplications. The crescent-shaped first transverse sulcus was used as the anterior boundary of HG, thereby dividing AC into two parts, namely (1) an anterior stream including HG, HG duplications and aSTG and (2) a posterior stream including PT. HG was separated from aSTG by an anterior borderline with y = 0. The range of the included image grey values was calculated individually. A box was marked around left and right AC to generate intensity histograms of these areas. The ‘grey value inclusion range’, which was used for surface reconstruction and morphometry, was defined on the basis of two criteria: (1) the value of the GM peak multiplied by the factor 0.28, which characterizes an appropriate cut-off value to separate liquor from GM tissue, (2) the saddle point between GM and white matter peaks. The grey and white value voxels embedded in this inclusion range were marked and used for 3D reconstruction; for morphometry, only GM values were used.

In the previous study (results published in Turker et al., 2017, 2018; see Chapter 5) three types of HGs were distinguished, namely single gyri (single), CSDs and CPDs (including two cases of multiple gyri). In other studies, such as Benner et al. (2017), at least four types of gyri were distinguished, namely the three aforementioned, and multiple gyri (at least three complete HGs) were considered a separate category (see Figure 26). Due to the larger variability in HG types in the present study and the larger sample size (N = 42), this fourfold categorization was applied this time. Therefore, all HGs for each hemisphere were again separately and independently classified by two researchers and then placed in the appropriate category.

When distinguishing between CPDs and multiple gyri, it is important to check whether the duplication, i.e. the second gyrus, is followed by a complete sulcus from the lateral to the medial end. ‘Multiple gyri’ is the category used when a CPD is followed by another gyrus that might be connected to the first duplication or not. For clearly distinguishing between the types, it is vital to look at the 3D reconstruction from various angles and also look at the MRI scan to closer investigate the form of sulci and gyri before they are segmented.
Figure 26 Four types of HGs distinguished in Study 2: single gyrus (Single), CSD, CPD and multiple gyri (Multiple). aSTG anterior superior temporal gyrus, HG Heschl’s gyrus, PT planum temporale

6.2. Results

The focus of the first part of the results section will be on the results of the behavioural analyses. In particular, the relationship between WM capacity, musicality/musical experience and language aptitude will be stressed therein. However, also the results of the arithmetic fluency test, the intelligence test and perceptions of language aptitude were looked into. Following a brief summary of the correlational analyses, the results of the PCA and a comparison between high and low aptitude subjects (according to the LLAMA score as indicated) will be dealt with in detail.

In addition to the behavioural analysis, a detailed account of the neuroanatomical results will be provided. The relationship between AC morphology and the three core scales, namely language aptitude, musicality and speech imitation aptitude, was tested through three one-way ANOVAs conducted for each hemisphere. All statistical analyses were performed with SPSS 25 and visualizations were computed with R. Results are currently in press and will hence be published in the upcoming weeks (see Turker et al., 2019).
6.2.1. Behavioural Results

First, the descriptive results will be discussed as done in Study 1. Then links between all assessed scales in the large group of participants are investigated ($N = 64$). Subsequently, the results of a PCA that was performed to explore the main abilities assessed through the variety of tests will be presented. Finally, differences in abilities between the low and high aptitude group ($N = 21$ per group) are presented. All following statistical analyses were computed with SPSS 25 and to correct for multiple comparisons, FDR correction (also termed Benjamini-Hochberg procedure; Benjamini & Hochberg, 1995) was applied. The procedures are thus identical to those in Study 1.

To guarantee that potential differences on the scales of interest were not caused by differences in subjects’ intelligence and socio-economic status, results of the age-normalized IQ achieved in the CFT20-R (Weiß, 2008) and measures of socio-economic background (Schneider & Seither-Preisler, 2015; Seither-Preisler et al., 2014) were compared between the two aptitude groups. The latter were determined from a general questionnaire for parents, developed by the researchers of the longitudinal study. A PCA revealed three relevant social dimensions: (1) education environment (including the mother’s and father’s highest professional degree and the number of books at home); (2) parental engagement (including the amount of parent–child communication, the frequency of common participation in cultural events, and the parents’ personal interest in children’s activities); and (3) resources and leisure activities (including courses in sports, arts, etc., and children’s resources, such as their own room, personal computer, etc.). Parental income loaded as well on factors 1 and 3. The individual factor scores on each dimension were compared across groups. Children and teenagers in the two groups neither differed in IQ (low aptitude group: $M = 107 \pm 16.2$; high aptitude group: $M = 114.4 \pm 12.8$; $t_{(38)} = -1.57, p = .126$), nor in socio-economic status (education environment: $t_{(38)} = -1.96, p = .06$, parental engagement: $t_{(38)} = -0.29, p = .77$, resources and leisure activities: $t_{(38)} = 0.85, p = .40$). Therefore, none of the observed group-statistical differences in language aptitude and other variables assessing cognitive skills can be accounted for by differences in intelligence and socio-economic status.

6.2.1.1. Descriptive Statistics
Descriptive results are provided in Table 13, including mean scores and standard deviations for all assessed scales.

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade L1 (German)</td>
<td>2.29</td>
<td>0.71</td>
</tr>
<tr>
<td>Grade L2 (English)</td>
<td>2.47</td>
<td>0.94</td>
</tr>
<tr>
<td>Hindi score</td>
<td>3.55</td>
<td>1.33</td>
</tr>
<tr>
<td>Digit span forward</td>
<td>7.27</td>
<td>2.19</td>
</tr>
<tr>
<td>Digit span backward</td>
<td>6.81</td>
<td>2.04</td>
</tr>
<tr>
<td>Non-word span</td>
<td>4.85</td>
<td>1.60</td>
</tr>
<tr>
<td>LLAMA B</td>
<td>44.92</td>
<td>15.28</td>
</tr>
<tr>
<td>LLAMA D</td>
<td>35.48</td>
<td>16.55</td>
</tr>
<tr>
<td>LLAMA E</td>
<td>62.06</td>
<td>29.90</td>
</tr>
<tr>
<td>LLAMA F</td>
<td>40.00</td>
<td>26.21</td>
</tr>
<tr>
<td>AMMA rhythm</td>
<td>26.08</td>
<td>4.46</td>
</tr>
<tr>
<td>AMMA tonal</td>
<td>23.95</td>
<td>3.78</td>
</tr>
<tr>
<td>Arithmetic fluency</td>
<td>141.04</td>
<td>40.99</td>
</tr>
<tr>
<td>Self-reported aptitude</td>
<td>6.18</td>
<td>1.49</td>
</tr>
<tr>
<td>Parent-reported aptitude</td>
<td>6.31</td>
<td>1.95</td>
</tr>
</tbody>
</table>

Table 13 Results on all assessed scales, including school grades in German (first language) and English (second language). Mean scores and standard deviations provided rounded to two decimal places. L1 first language, L2 second language.

German grades ranged from 1 to 4 ($M = 2.29, SD = 0.71$), while English grades ranged from 1 to 5 ($M = 2.47, SD = 0.94$). In the German school system, 1 is the best achievable grade, while 5 indicates a failure of the class. Ratings for the Hindi score ranged from 0.69 to 6.72, indicating a very low score for the best speech imitator in this population ($M = 3.55, SD = 1.33$). When comparing the Hindi score in this younger population ($3.55 \pm 1.33$) with the mean Hindi
score achieved by adults in study 1 (4.81 ± 1.64) it becomes clear that there is a remarkable
difference in performance between younger and older subjects. Likewise, there are obvious
differences in WM performance in all three WM-related scales, as well as in AMMA tonal and
rhythm. In this study, the three WM scales ranged from 2-13, with digit span forward showing
the highest mean score (M = 7.27, SD =2.19). The part of the LLAMA that was most easily
mastered by children and teenagers was LLAMA E, the sound-symbol association task, with
scores ranging from 0 to 100 (M = 62.06, SD =29.90). The maximum scores achieved in the
other three language aptitude tests were 95 in LLAMA B, 70 in LLAMA D and 80 in LLAMA
F. In other words, apart from LLAMA E, none of the 64 children achieved a 100% score in the
three other language aptitude tests.

The musicality tests yielded results between 13 and 34 points (maximum achievable
score: 40) in AMMA rhythm (M = 26.08, SD = 4.46) and between 14 and 32 points in AMMA
tonal (maximum achievable score: 40; M = 23.95, SD =3.78). The arithmetic fluency test results
in children ranged from 44 to 255 points, with a mean of 141.03 and a standard deviation of
40.99. Last, self- and parent-reported language aptitude ranged from 2 to 9 and 3 to 10
respectively. Mean scores on these two scales were very similar (self-reported aptitude: M =
6.18, SD = 1.49; parent-reported aptitude (M = 6.31, SD =3.78).

6.2.1.2. Results of the Correlational Analyses

Following the descriptive results, a correlational analysis was performed for all assessed
scales (number of languages, self-reported aptitude, parent-reported aptitude, grades in German
and English, Hindi score, digit span forward, backward and non-word span, LLAMA B,
LLAMA D, LLAMA E, LLAMA F, arithmetic fluency) in all tested children (N = 64).
Spearman’s rank correlation coefficients (Spearman’s rho; the data was not normally
distributed, which is why Spearman’s rank correlation coefficients are reported instead of
Pearson’s correlation coefficients) and p-values of related scales are displayed in Table 14..
Since the MULT/AP did not correlate with any other score, it was not included in Table 14.
Table 14 Spearman correlation matrix for all assessed scales. Correlation coefficients (r) and significance values (p) are provided for each correlation, rounded to three decimal places.

---

<table>
<thead>
<tr>
<th>Languages</th>
<th>SR aptitude</th>
<th>PR aptitude</th>
<th>German Grade</th>
<th>English Grade</th>
<th>Hindi Score</th>
<th>Digit span forward</th>
<th>Digit span backward</th>
<th>Nonword span</th>
<th>LLAMA B</th>
<th>LLAMA E</th>
<th>LLAMA F</th>
<th>Arithmetic skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Languages</td>
<td>r = .374*</td>
<td>r = .379*</td>
<td>r = -.246</td>
<td>r = -.261*</td>
<td>r = .207</td>
<td>r = .294*</td>
<td>r = .078</td>
<td>r = .234</td>
<td>r = .088</td>
<td>r = .264*</td>
<td>r = .214*</td>
<td>r = .254*</td>
</tr>
<tr>
<td>SR aptitude</td>
<td>r = .787*</td>
<td>r = .407*</td>
<td>r = -.592*</td>
<td>r = -.082</td>
<td>r = .340*</td>
<td>r = -.273*</td>
<td>r = -.059</td>
<td>r = -.116</td>
<td>r = .265*</td>
<td>r = .214</td>
<td>r = .248*</td>
<td>r = .183</td>
</tr>
<tr>
<td>PR aptitude</td>
<td>r = -.364*</td>
<td>r = -.592*</td>
<td>r = -.338</td>
<td>r = .475*</td>
<td>r = -.248</td>
<td>r = -.176</td>
<td>r = -.164</td>
<td>r = -.225</td>
<td>r = .303*</td>
<td>r = .275</td>
<td>r = .223</td>
<td>r = .222</td>
</tr>
<tr>
<td>German</td>
<td>r = -.506*</td>
<td>r = -.250</td>
<td>r = -.230</td>
<td>r = -.169</td>
<td>r = -.101</td>
<td>r = -.240</td>
<td>r = -.171</td>
<td>r = -.198</td>
<td>r = -.195</td>
<td>r = .199</td>
<td>r = .199</td>
<td>r = .212</td>
</tr>
<tr>
<td>English</td>
<td>r = -.042</td>
<td>r = -.165</td>
<td>r = -.202*</td>
<td>r = -.049</td>
<td>r = -.179</td>
<td>r = -.235</td>
<td>r = -.224</td>
<td>r = -.170</td>
<td>r = .100</td>
<td>r = .042</td>
<td>r = .042</td>
<td>r = .042</td>
</tr>
<tr>
<td>Hindi score</td>
<td>r = -.234</td>
<td>r = -.601</td>
<td>r = .196</td>
<td>r = -.323*</td>
<td>r = .145</td>
<td>r = .201</td>
<td>r = .152</td>
<td>r = .113</td>
<td>r = .067</td>
<td>r = .998</td>
<td>r = .130</td>
<td>r = .009</td>
</tr>
<tr>
<td>Digit span</td>
<td>r = .524*</td>
<td>r = .519*</td>
<td>r = .261*</td>
<td>r = .379*</td>
<td>r = .405*</td>
<td>r = .296*</td>
<td>r = .373*</td>
<td>r = .373*</td>
<td>r = .373*</td>
<td>r = .373*</td>
<td>r = .373*</td>
<td>r = .373*</td>
</tr>
<tr>
<td>forward</td>
<td>p = .001</td>
<td>p = .001</td>
<td>p = .040</td>
<td>p = .002</td>
<td>p = .001</td>
<td>p = .020</td>
<td>p = .004</td>
<td>r = .458*</td>
<td>r = .458*</td>
<td>r = .458*</td>
<td>r = .458*</td>
<td>r = .458*</td>
</tr>
<tr>
<td>Digit span</td>
<td>r = .198</td>
<td>r = .283*</td>
<td>r = .315*</td>
<td>r = .379*</td>
<td>r = .166</td>
<td>r = .458*</td>
<td>r = .458*</td>
<td>r = .458*</td>
<td>r = .458*</td>
<td>r = .458*</td>
<td>r = .458*</td>
<td>r = .458*</td>
</tr>
<tr>
<td>backward</td>
<td>p = .123</td>
<td>p = .025</td>
<td>p = .014</td>
<td>p = .003</td>
<td>p = .200</td>
<td>r = .001</td>
<td>r = .001</td>
<td>r = .001</td>
<td>r = .001</td>
<td>r = .001</td>
<td>r = .001</td>
<td>r = .001</td>
</tr>
<tr>
<td>Nonword</td>
<td>r = .107</td>
<td>r = .243</td>
<td>r = .164</td>
<td>r = .257*</td>
<td>r = .006</td>
<td>r = .006</td>
<td>r = .006</td>
<td>r = .006</td>
<td>r = .006</td>
<td>r = .006</td>
<td>r = .006</td>
<td>r = .006</td>
</tr>
<tr>
<td>spin</td>
<td>p = .413</td>
<td>p = .059</td>
<td>p = .208</td>
<td>p = .002</td>
<td>p = .966</td>
<td>r = .300</td>
<td>r = .001</td>
<td>r = .001</td>
<td>r = .001</td>
<td>r = .001</td>
<td>r = .001</td>
<td>r = .001</td>
</tr>
<tr>
<td>LLAMA B</td>
<td>r = .268*</td>
<td>r = .318*</td>
<td>r = .050</td>
<td>r = .294*</td>
<td>r = .294*</td>
<td>r = .294*</td>
<td>r = .294*</td>
<td>r = .294*</td>
<td>r = .294*</td>
<td>r = .294*</td>
<td>r = .294*</td>
<td>r = .294*</td>
</tr>
<tr>
<td>LLAMA D</td>
<td>r = .276*</td>
<td>r = .080</td>
<td>r = .250</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
</tr>
<tr>
<td>1</td>
<td>p = .029</td>
<td>p = .532</td>
<td>r = .277*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
</tr>
<tr>
<td>LLAMA E</td>
<td>r = .209</td>
<td>r = .023</td>
<td>r = .115</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
<td>r = .058</td>
</tr>
<tr>
<td>LLAMA F</td>
<td>1</td>
<td>p = .115</td>
<td>r = .277*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
<td>r = .299*</td>
</tr>
</tbody>
</table>

*significant at p<.05 before correction; bold = significant at p<.05 after correction (Benjamini-Hochberg correction for multiple comparisons [Benjamini and Hochberg, 1985]); Languages = number of languages, SR = self-reported, PR = parent-reported
The number of foreign languages learnt by a child correlated positively with self-reported \( (r = .374, p = .002) \) and parent-reported aptitude \( (r = .379, p = .004) \), and with two measures of the language aptitude battery, namely phonetic memory \( (r = .364, p = .003) \) and sound-symbol association \( (r = .314, p = .012) \). Self-reported aptitude showed a striking correlation with parent-reported aptitude \( (r = .787, p < .001) \) and school grades in L1, German \( (r = -.407, p = .005) \) and L2, English \( (r = -.592, p < .001) \). Furthermore, self-reported aptitude correlated with digit span forward \( (r = .340, p = .006) \), while parent-reported aptitude also correlated with the English grade \( (r = -.592, p < .001; \) German grade \( p \)-values did not survive statistical correction) and digit span forward \( (r = .475, p < .001) \). Parent-reported aptitude was further linked to LLAMA F, measuring grammatical inferencing \( (r = .393, p = .003) \). Moreover, German and English grades correlated highly with each other \( (r = .596, p = .001) \).

The Hindi score, considered a measure of non-word span, was only linked to vocabulary learning \( (\text{LLAMA B}, r = .323, p = .009) \). Digit span forward, on the other hand, was not only strongly associated with digit span backward \( (r = .524, p < .001) \) and non-word span \( (r = .519, p < .001) \), but it correlated with all language aptitude scores, two correlations being strong enough to survive correction for multiple comparisons \( (\text{LLAMA D}: r = .379, p = .002; \) LLAMA E: \( r = .405, p = .001) \). Digit span backward was not related to non-word span, but to LLAMA E \( (r = .379, p = .003; \) correlations with LLAMA B and D did not survive correction). While all LLAMAS were related to another before correction, only the link between LLAMA B and E (vocabulary learning, sound-symbol association) remained significant \( (r = .318, p = .011) \). Arithmetic skills correlated highly with digit span forward \( (r = .373, p = .004) \), digit span backward \( (r = .458, p < .001) \) and vocabulary learning \( (\text{LLAMA B}, r = .394, p = .002) \). Correlations in the table are given both at an uncorrected level \( (p < .05 \text{ before correction}) \) and after FDR correction (Benjamini & Hochberg, 1985).

\[ 6.2.1.3. \quad \text{Results of the Principal Component Analysis} \]

To detect potential differences in comparison to the results of the PCA in Study 1, a PCA was performed for the scores of this younger population as well. The various subscores of the arithmetic fluency test, as well as all other assessed scales were included in the PCA. The results are presented in Table 15.
Table 15 The results of the varimax rotated component matrix for the PCA including all 17 variables resulted in three components. All r-values below .35 were discarded in the analysis (and thus do not appear in the table). AF arithmetic fluency

Given the particularly strong link between WM and language aptitude as revealed by the correlational analyses, a main aim was to reveal whether WM scales would construct an own entity (that is, a separate component largely or at least partly independent from language aptitude) or not. All criteria for the PCA were fulfilled (see description of criteria in 5.2.1.3.) and the PCA was performed for the different subscores of all tests: divisions, subtractions,
additions and multiplications belonging to the overall arithmetic competence score; digit span forward, backward and non-word span; LLAMA B-E and Hindi; singing ability, singing passion, AMMA rhythm, AMMA tonal and the number of instruments played by each subject.

Results of the PCA suggest that arithmetic fluency is a separate construct and all types of calculations, as well as WM (digit span backward) load heavily onto this component. Concerning the relationship between language aptitude and WM, the picture is more complex. All variables related to language learning and WM loaded heavily onto the same variable, which is why it has been termed ‘Language Aptitude/Working Memory’. The highest loadings were digit span forward, followed by non-word span, sound-symbol association ability, digit span backward and vocabulary learning. Furthermore, the Hindi speech imitation score and LLAMA D and LLAMA F loaded only onto this component as well. The third component was termed ‘Musical Ability’ because only aspects related to music and musical ability seemed to load onto this component. Despite the lack of correlation between music-related variables in the correlational analyses, it is noteworthy that singing ability, singing passion, the AMMA results and the number of instruments all loaded onto one concept.

6.2.1.4. Comparisons between Low and High Aptitude Learners

While the PCA and the correlational analyses aimed to detect relationships between assessed scales, a comparison between those subjects with poor scores on language aptitude and those with high scores thereon could help detect those variables most useful for a differentiation between the two. In other words, comparing mean scores between a high and low talent group (again, these are extreme terms – they are more suitably described as above average language learners and below average language learners) and finding out whether the observed differences were statistically significant, was a vital aspect for defining an outstanding language learner. In order to verify the potential behavioural and neuroanatomical differences on a group-statistical level, however, a score was needed to arrive at two clearly distinct groups that could be compared.

In the adult study, comparisons were made once with the Hindi speech imitation score and once with the MLAT overall score. In the present study, the Hindi score showed very little variation, meaning that picking two extreme groups would have been very challenging and differences between the two groups would have been marginal. Generally, performance in the
speech imitation was rather low, which is why the Hindi score was neglected in this case. Therefore, I decided to use the LLAMA score as a general raw score to differentiate between two groups of foreign language learners, which will later on be termed ‘highly gifted’ (or high aptitude group) and ‘poorly gifted’ (low aptitude group) subjects. To obtain these groups, the cut-off points were set at the 33rd and the 66th percentile, and the intermediate group was discarded from further group-statistical analyses. The final sample ($N = 42$; $M_{\text{age}} = 14.5 \pm 1.3$ y; 19 females) consisted of a high aptitude group ($N = 21$; $M_{\text{age}} = 14.6 \pm 0.9$ y; 9 females) and a low aptitude group ($N = 21$; $M_{\text{age}} = 14.4 \pm 1.4$ y; 10 females).

In Table 16, differences between high aptitude and low aptitude individuals are shown for several measured scales. Mean scores, standard deviations and results of the Mann-Whitney U Test are shown (Z-scores), together with the corresponding $p$-values. The Mann-Whitney U Test was chosen in this case since performance on many assessed scales was not normally distributed and a non-parametric test had to be chosen instead of the $t$-test. Effect sizes were calculated as well. The correlation coefficient $r$ as an effect size was calculated by dividing the $Z$ score by the square root of the total sample size ($N = 42$ in the present study) and values between 0 and 1 were obtained. $r$ values between .3 and .5 are medium effects (accounting for 9% of total variance) and effects above .5 are large effects (accounting for 25% of variance; see also Rosenthal, 1991, 1994).
Table 16 Differences in performance on measured test scales between high-aptitude and low-aptitude children (Mann Whitney U Test). Z- and p-values, as well as effect size (correlation coefficient r; Z-value divided by the square root of the total sample size; Field, 2013) are displayed.
Subjects in the high aptitude group had not only learnt more languages ($Z = -3.13, p = .002$), but also had higher self-reported ($Z = -3.19, p = .001$) and parent-reported aptitude ($Z = -3.07, p = .002$). Moreover, their grades in language subjects at school were significantly better, most evidently in English ($Z = -2.56, p = .010$), their second language, but also in German ($Z = -2.29, p = .022$), their first language. Children with high language aptitude scores also had significantly better results in the speech imitation ($Z = -2.77, p = .006$), two of the three WM tasks (digit span forward: $Z = -3.87, p < .001$; digit span backward: $Z = -3.27, p = .001$) and better arithmetic skills as assessed by the arithmetic fluency test ($Z = -2.64, p = .008$).

Results are graphically displayed in Figure 27, showing the comparison of high and low aptitude children on WM tasks and the Hindi test.

**Figure 27** A comparison of performance of the high and low aptitude groups on all three WM tasks and the Hindi speech imitation task. Raw scores on the three WM scales ranged from 0 to 14, while the Hindi speech imitation scores ranged from 0 to 10. All comparisons are significant except for performance in the non-word repetition.
It is striking that almost all assessed scales could be linked to the distinction between above average and below average language aptitude. In particular, the difference between the two groups in digit span forward and backward was highly significant, with a very large effect size as well. Also, school achievement in the form of mean scores in English and German showed striking discrepancies according to group status. To summarize, all measured scales could be linked to the distinction between above average and below average language aptitude, although there was no significant difference in intelligence or socio-economic status between the two groups. To put it differently, subjects in the high language aptitude group outperformed those in the low aptitude group in almost all administered tasks, ranging from arithmetic abilities to WM and the Hindi speech imitation task.

6.2.2. Neuroanatomical Results

Since differentiating between poorly and highly gifted language learners yielded significant results for the behavioural scales, the results of the neuroanatomical analyses were expected to be of similar nature. As explained in the methods section, four types of HG were distinguished in the child population, namely singly gyrus, CSD, CPD and multiple gyri. Following the procedure of Study 1, left and right HGs were manually segmented and morphologically classified according to the predefined types by two researchers independently. In this study, however, an additional neuroanatomical analysis was performed, namely GM volumes in left and right AC were compared between high and low aptitude subjects (for details on the procedure, please refer to section 6.1.11). First, results of the differences in shapes of HG, i.e. differences according to the predefined morphological types of HG, shall be discussed. Therefore, the frequency distributions of the four HG types were calculated. As indicated in Table 17, the total number of participants was 42 and results for each hemisphere are displayed with the percentage value included.

Similar to the results found in the first study, the left hemisphere showed slightly less variation in AC morphology compared to the right hemisphere. In the right hemisphere, the predominant HG type was multiple gyri (N = 17), followed by a single gyrus (N = 15), CPDs (N = 7) and CSDs (N = 3). The significance in difference of occurrence was confirmed by a chi-square test: $\chi^2_{df = 3, N = 42} = 12.5, p = .006$. In the left hemisphere the predominant HG type
was single gyrus ($N = 20$), followed by CPD ($N = 12$), multiple ($N = 6$) and CSD ($N = 4$); $\chi^2_{df = 3, N = 42} = 14.7, p = .002$.

<table>
<thead>
<tr>
<th>Types of HG</th>
<th>Left AC</th>
<th>Right AC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single</strong></td>
<td>20 (47.6%)</td>
<td>15 (35.7%)</td>
</tr>
<tr>
<td><strong>CSD</strong></td>
<td>4 (9.5%)</td>
<td>3 (7.1%)</td>
</tr>
<tr>
<td><strong>CPD</strong></td>
<td>12 (28.6%)</td>
<td>7 (16.66%)</td>
</tr>
<tr>
<td><strong>Multiple</strong></td>
<td>6 (14.3%)</td>
<td>17 (40.5%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>42 (100%)</td>
<td>42 (100%)</td>
</tr>
</tbody>
</table>

**Table 17** The distribution of HG types, namely single, CSD, CPD, multiple gyri, in both hemispheres with percentage values given as well.

For the categorization of HG types, all individually segmented HGs of both hemispheres were checked and classified by two researchers independently. If there was disagreement, the case was discussed and the most appropriate morphotype was chosen. As the number of subjects in the present study was considerably higher than in the first study, a complete illustration of all HGs is not provided. Especially as the number of cases increases, the more confusing and difficult it gets to see clear patterns and observe differences according to group-status. Therefore, only four exemplary high- and low aptitude children’s ACs are displayed in Figure 28. Although single illustrations provide readers with a more concrete experience of the neuroanatomical analyses and the combinations associated with/without language aptitude, they make it very difficult to see overall patterns associated with group-status. Consequently, in addition to the individual AC illustrations, averaged surfaces of AC for both hemispheres were calculated for each group and are graphically displayed on a standard AC in Figure 29.
**Figure 28** An exemplary illustration of four HGs in the high and low aptitude groups respectively. For details of the four designated types, see the legend. The right hemisphere is marked in blue, the left in red. *CSD* complete posterior duplication *CPD* complete posterior duplication
Figure 29 Averaged HG surfaces of subjects in the high (left panel; \(N = 21\)) and low aptitude group (right panel; \(N = 21\)). High language aptitude is associated with CPDs and multiple gyri in the right hemisphere, while low language aptitude results are associated with a single gyrus in the right hemisphere. Left hemisphere (L): blue; right hemisphere (R): red.

To investigate the relationship between the four right-hemispheric HG types and language aptitude, frequency distributions were compared for the high aptitude and low aptitude subjects as classified by performance on the language aptitude test (LLAMA). HG types were significantly differently distributed in the two groups (\(\chi^2 \, df = 3, N = 42 \approx 10.01, p = .012\)), with multiple gyri being an indicator for high aptitude. Single gyri, on the other hand, were an indicator for low aptitude (\(\chi^2 \, df = 3, N = 42 \approx 8.4, p = .009\); 57% of low aptitude children possessing a single gyrus in right AC compared to 14.3% of the children in the high aptitude group). The same tests applied to the left hemisphere yielded no statistically significant results. A comparison of HG types of right and left AC and their association with results of the LLAMA (ranging from 0 to 100%) are displayed in Figure 30.
For the present study, GM volumes were also calculated and right- and left-hemispheric GM volumes between the two aptitude groups were compared. Results of a Mann-Whitney-U-Test revealed that GM volumes in right AC differed significantly between the low and high aptitude group ($U_{(40)} = 78.00$, $Z = -3.59$, $p < .001$). In other words, GM volume in right AC was significantly higher in the high aptitude than in the low aptitude subjects. The same test applied to GM volume in left HG yielded no statistically significant results ($U_{(40)} = 149.00$, $Z = -1.8$, n.s.).

This finding linking GM volume in right AC to high language aptitude was further corroborated by correlational analyses, which revealed that right, but not left HG GM volumes were associated with a variety of language-relevant variables assessed in the current study. These included the number of spoken foreign languages ($r = .35$, $p = .022$), self-reported language aptitude ($r = .36$, $p = .019$), parent-reported language aptitude ($r = .38$, $p = .024$), school grade in German ($r = .46$, $p = .003$), Hindi speech imitation score ($r = .39$, $p = .011$),
and scores on LLAMA B \((r = .34, p = .028)\), LLAMA E \((r = .39, p = .010)\), LLAMA F \((r = .45, p = .003)\). For LLAMA D there was at least a corresponding non-significant trend \((r = .28, p = .069)\). Moreover, a correlation was found for the WM measure digit span forward \((r = .38, p = .015)\).

Leaving aside the distinction between high and low aptitude subjects, two linear regression analyses were performed to investigate to what extent the LLAMA overall score predicted left and right GM volumes respectively. The prediction was only significant for right HG \((R^2 = .214, p = .002)\), where 21.4% of variance in GM volume was explained by high language aptitude. The linear regression analysis for left AC yielded no statistically significant results \((R^2 = .044, p = .182)\). As can be seen in Figure 31, the higher the LLAMA overall score, the higher the GM volume in right AC.

**Figure 31** Results of the linear regression analyses showing the relationship between LLAMA overall performance mean scores (ranging from 0 to 100%) and GM volumes (mm³) in right (A) and left (B) AC for all tested individuals. 21.4% of variance in language aptitude are explained by right-hemispheric HG GM volume.
All the aforementioned statistical tests were also performed to test the relationship between right AC morphology and musical ability. None of the analyses yielded any significant results and they will hence be neglected in this section and the following discussion section.

To summarize, the results of the neuroanatomical analyses suggest an intricate relationship between right-hemispheric AC morphology and high language aptitude. The results of the linear regression reveal that the higher the GM volume in right AC, the significantly higher a subjects’ performance on the LLAMA test. Therefore, the findings of the neuroanatomical analyses do not only suggest that gyrification patterns of right AC are significantly associated with foreign language aptitude, but also GM volume shows a strong link to language aptitude scores. In sum, the more gyri individuals had in right AC and the higher their GM volumes in right AC were, the better they performed in the language aptitude testing. If they possessed a single gyrus in right AC, they received the lowest scores in the language aptitude testing.

6.3. Discussion

The results of the present study propose that individual differences in right AC morphology are significantly linked to language aptitude in children and teenagers. More specifically, more complete HGs in the right hemisphere (i.e. multiple gyri) and higher corresponding GM volumes could be delineated as neuroanatomical markers for language aptitude as measured by the LLAMA language aptitude battery. On the other hand, those subjects with lowest scores in the language aptitude testing possessed single gyri most frequently in their right AC. Therefore, while multiple gyri were an indicator for high aptitude, single gyri were an indicator for low aptitude. Furthermore, the intricate link between WM and language aptitude was successfully confirmed, supporting the claim that WM is strongly involved in FLL processes and thus also language aptitude. While also arithmetic fluency and academic achievement were strongly associated with language aptitude, the often-cited relationship between musical and language ability could not be corroborated by the findings of the present study.
6.3.1. **Right Auditory Cortex Morphology as a Marker for Language Aptitude**

In accordance with the results of Study 1 (see Chapter 5), the findings of the present study suggest that the morphology of right HG is a potential neuroanatomical marker for foreign language aptitude in children and teenagers aged 10 to 16 years. While the occurrence of multiple gyri in right AC was linked to high language aptitude, those children and teenagers who possessed single gyri in their right AC were most likely to be in the low language aptitude group. These results also highlight the importance of right-hemispheric auditory processing for language learning (see Kepinska, de Rover, et al., 2017b, 2017a; Kepinska, Pereda, et al., 2017) and call for more extensive research on the involvement of AC in foreign language learning.

Based on the present findings, I suggest that AC morphology, both in the form of GM volumes and morphological characteristics, is a stable trait that can be seen as a potential marker for foreign language aptitude. The extent to which inter-individual variation in AC morphology is a result of experience-dependent, intrauterine, or genetic influences or a combination thereof, has been discussed in both cross-sectional (Golestani et al., 2011; Herholz & Zatorre, 2012; Ressel et al., 2012) and longitudinal studies (Hyde et al., 2009; Penhune, 2011). Prenatal studies have shown that HG develops by week 31 of gestational age. In most cases, right HG develops 1 to 2 weeks earlier than the left (Chi, Dooling, & Gilles, 1977). The morphology and asymmetry of PT, which is located posterior to HG, becomes visible at gestational week 30 and appears to remain fairly stable across fetuses, newborns, children, and adults (Preis, Jäncke, Schmitz-Hillebrecht, & Steinmetz, 1999). Studies with monozygotic and dizygotic twins have demonstrated that morphometric differences of AC are predominantly attributable to genetic factors (Pol Hulshoff et al., 2006). Results of longitudinal studies with children and teenagers (Seither-Preisler et al., 2014; Serrallach et al., 2016) have shown a high inter-individual variability of HG morphology and GM volumes, but an almost perfect intra-individual stability over years, regardless of interim auditory and musical training. Even though clear learning-induced changes were established on the neurofunctional level over time, these were not reflected on the neuroanatomical level at all. Consequently, I argue that AC morphology and GM volume can be considered highly stable neuroanatomical characteristics that could represent potential markers of aptitude in different auditory-related cognitive domains.
The right hemisphere, and in particular right AC, is implicated in various music-related processes, while studies on language learning so far have found little evidence for involvement of right AC in language learning processes. In previous studies, it was found that right-hemispheric HG morphology and GM volumes were closely linked to musical aptitude (J. Benner et al., 2017; Seither-Preisler et al., 2014), which in turn predicted the motivation to practice a musical instrument. In other words, the larger GM volumes children had, the higher their intrinsic motivation to musically engage. These findings motivated the search for potential neuroanatomical markers for language learning, which are also highly reliant on auditory processing and integration. The claim that AC morphology might play a role in FLL has been rarely discussed, in particular since only few studies (Golestani et al., 2011; Golestani, Molko, Dehaene, LeBihan, & Pallier, 2007) have explicitly addressed a link between the two. This may be since left-hemispheric language dominance has been a well-established and accepted fact for the past decades, which might have discouraged extensive research in language-relevant right-hemispheric brain regions and functions.

The significance of right AC for FLL is also supported by findings on the involvement of right AC in first language acquisition. Perani et al. (2011) found a preponderance in right primary and secondary AC for speech input in infants, while Dehaene-Lambertz and colleagues (Dehaene-Lambertz, Hertz-Pannier, Dubois, & Dehaene, 2008; Dehaene-Lambertz et al., 2006) found a leftward asymmetry for speech-like stimuli from birth on. Homae, Watanabe, Nakano, Asakawa, & Taga (2006) emphasize that speech prosody is one of the most important sources of information for infants acquiring their native language. They provided evidence that speech processing in the human infant brain developed from the analysis of pitch information to comparing and integrating information in speech sounds with acquired prosodic structures. According to their neuroimaging findings in 10-month old infants, cortical activation in response to manipulated speech prosody was clearly right-lateralized in temporal and temporoparietal regions. This suggests that in early infancy right-hemispheric auditory functions are essential and predominant, until more refined left-hemispheric language comprehension skills step into the foreground. This is also consistent with findings of an earlier foetal and postnatal maturation of the right hemisphere in general, which enables global feature extraction before more complex analytic skills become predominant (for a review see Chiron et al., 1997). As right-hemispheric auditory processing is focused on the recognition of vocal timbre and prosody, it is likely that right-hemispheric brain areas play a crucial role in early native
language acquisition by infants. Moreover, they should be relevant for the ability to learn new languages in later life, which also requires the recognition and memorization of prosodic contours of the unfamiliar language before explicitly using and manipulating its elements. Therefore, an advantageous neuroanatomical morphology and gyrification of right AC and corresponding larger GM volumes could consequently be associated with high language aptitude.

Given the potential involvement of right AC morphology in foreign language aptitude, it seems worthwhile to look at the various functions of primary and secondary auditory areas in the processing of speech. While primary auditory regions are crucial for the analysis of simple sound features, secondary areas enable a larger integration and hence more complex auditory pattern recognition, relevant for the processing of music and speech (Leaver & Rauschecker, 2010). HG includes a posteromedial primary and an anterolateral secondary part and also shows a characteristic hemispheric specialization. Different studies have provided evidence that the left AC possesses a sensitivity for the processing of fast acoustic events (e.g., speech; Zatorre et al., 2002), while the right AC is specialized for finer resolution in the frequency domain (Hyde et al., 2008; Poeppel et al., 2004; Zatorre et al., 2002). Concerning HG, the left side seems to be more sensitive to the temporally conveyed fundamental pitch of complex tones, while the right side seems to be more important for the spectral discrimination of different timbres of musical and vocal sounds (Schneider et al., 2005). It has also been suggested that left and right AC have distinct temporal integration windows with the superior temporal sulcus being specialized in the analysis of short segments (25-50 ms) in the left and of long segments (200-300 ms) in the right hemisphere (Boemio, Fromm, Braun, & Poeppel, 2005; Poeppel, 2003). That predisposes right AC not only to analyse slow musical rhythms and melodic contours, but also to process the syllabic structure and prosodic modulations of spoken language. The latter also conveys important emotional information about the mood and intentions of the speaker. In addition, there is evidence that lateral right AC is specialized in spatial sound processing (Zatorre & Belin, 2001; Zatorre & Penhune, 2001). In previous studies, the research groups of Seither-Preisler and Schneider found clear evidence that an efficient integration of these left and right-hemispheric auditory functions promotes phonological awareness, literacy skills and attention (Seither-Preisler et al., 2014; Serrallach et al., 2016).
Each AC is a hub for numerous neural circuits to other brain areas that are involved in multimodal processing, motor activation and speech production (Friederici & Gierhan, 2013). Having more efficient and faster processing and integration of speech units (be it single sounds, syllables or whole words) could definitely prove advantageous for FLL/SLA. Higher GM volumes in certain parts of AC might provide better structural links to other language-relevant areas, e.g., the inferior parietal cortex, and prove advantageous for functional connectivity. To summarize, variations in the individual neuroanatomy of specific brain regions, such as HG or AC more generally, should not be neglected and could be used to further develop theories on language learning and aptitude.

To conclude, the neuroanatomical results of the present study confirm initial hypotheses that foreign language aptitude is significantly associated with right-hemispheric AC morphology and GM volumes. In the present study, multiple gyri and larger GM volumes of HG in the right hemisphere were highly correlated with high language aptitude. Single gyri, on the other hand, were linked to low language aptitude.

### 6.3.2. The Link between Language Aptitude, Working Memory and Arithmetic Competence

Researchers have argued that language aptitude and WM are so intricately linked that WM capacity equals outstanding language learning ability or at least deserves to be seen as a dominant subcomponent (Wen, 2016; Wen et al., 2017). The present study supports this hypothesis in so far that language aptitude scores and WM scores highly correlated with one another. This was also reflected in the results of the PCA, which indicated that WM and language aptitude loaded onto one single factor. Thus, WM appears to be a crucial constituent of language aptitude. Both language aptitude and WM also correlated significantly with the arithmetic score. Obviously, parts of the language aptitude tests and the arithmetic fluency test touch upon WM and analytic abilities, leading to a link between the scores. This is consistent with earlier results (DeStefano & LeFevre, 2004; Peng, Namkung, Barnes, & Sun, 2016) suggesting that WM is an essential, shared foundation of arithmetic and language skills.

It has further been suggested that the Hindi speech imitation test captures aspects of language aptitude and WM capacity (Reiterer et al., 2011). In the present study, children from the high language aptitude group had significantly higher Hindi speech imitation scores.
However, correlational analyses showed only one significant correlation with the subtest LLMA B (vocabulary learning), but not with the three other language aptitude scales. Furthermore, there was a lack of correlation between the Hindi and WM scores. This is particularly surprising, since one might assume that the Hindi test is a mere non-word span task, which comes in two complexities, namely single words and whole sentences. A possible reason for the observed lack of association is that repetition of real speech material can be more or less difficult depending on the native language. The non-word span stimuli, in contrast were only combinations of simple CVC syllables that follow the phonotactic rules in German. Consequently, imitating Hindi words and sentences might require skills that go beyond simple WM capacity and also require complex auditory pattern recognition and generalization of common language-specific acoustic features.

6.3.3. Behavioural Findings: School Grades, Self- and Parent-Reported Aptitude

Children who scored significantly better in the language aptitude testing also had better school grades in both English and German, were learning more languages at the time of testing and thought of themselves as better language learners. The positive relationship between the number of foreign languages being learnt and the language aptitude scores can be explained by two theories: (1) Children who learn languages quite effortlessly and fast know about their gift through the ease with which they learn English, for instance, and therefore they chose to learn more foreign languages at school; (2) Children who have learnt more languages could have profited from the language classes in so far that their meta-linguistic awareness was particularly higher and therefore they scored better in the language aptitude testing. Although no causal evidence can be gained through correlational analyses in the present study, both possibilities are actually good explanations for the observed correlations.

Certainly, school grades are not an optimal indicator for real potential, but one would assume that the ease with which foreign languages are learned (by remembering vocabulary, understanding grammatical complexity) should be at least reflected in those grades. It furthermore seems that both parents and children have a very good intuition regarding their child’s/ their own language learning potential and this intuition is well supported by their scores in the language aptitude batteries. This self-reported aptitude might impact school grades as
well, since more confidence could boost motivation and thereby influence behaviour at school in language classes.

6.4. Conclusion

The findings of the present study with children and teenagers suggest that (1) right AC morphology is crucial for high language aptitude, (2) language aptitude and arithmetic skills are strongly linked and (3) WM is an essential component of language aptitude. A higher number of HGs and corresponding higher GM volumes in right AC were associated with higher performance in the LLAMA language aptitude battery. Moreover, language learning ability showed a strong link with WM capacity, speech imitation skills and arithmetic abilities. Those children with high language aptitude also had better school grades in English, German, considered themselves more gifted for language learning and were estimated to be better language learners by their parents. The behavioural findings of this study suggest that language aptitude is associated with WM and arithmetic abilities and influences school performance and self-perception, as well as parents’ opinions of their children. In contrast to earlier studies, no association between musicality and language aptitude could be found. This suggests that the neuroanatomical link between right-hemispheric AC morphology and a subject’s individual aptitude profile diverges for musical and language skills.

6.5. Acknowledgements and Funding for the Present Study

The data presented in this chapter stems from the aforementioned longitudinal study at the University Hospital Heidelberg, organised by Dr. P. Schneider his research group “Music and Brain”. This longitudinal project was first supported by the German Federal Ministry of Education and Research (BMBF): Grants 01KJ0809/10 and 01KJ1204 (project: “AMseL: Audio- and Neuroplasticity of Musical Learning I + II”; this was part of the accompanying research of the cultural education program “An Instrument for Every Child”). It was further financially supported by the German Research Foundation (DFG) as part of the Heisenberg program (“Sound perception between outstanding musical abilities and auditory dysfunction: The neural basis of individual predisposition, maturation, and learning-induced plasticity in a lifespan perspective”) of Dr. Schneider. Furthermore, since 2017, I am recipient of a DOC-
team-fellowship of the Austrian Academy of Sciences and I received financial support through this DOC-team scholarship during the analysis and publication of the results (see Turker et al., 2019).
7. Discussion

Language aptitude is a multifaceted and complex concept that has been extensively researched in the past two decades but only few studies to date have succeeded in linking this outstanding ability to specific brain regions, functionally and anatomically. Despite the research community’s growing interest in the underlying neurobiology of language aptitude, we are still struggling to understand how it manifests structurally and functionally in an individual’s brain and the implications caused thereby. Moreover, we are still far from fully grasping the extent to which language aptitude is biologically predisposed or environmentally driven, and consequently can or cannot be altered through experience or training. On the behavioural level, it is still hotly debated whether and to what extent environmental factors (e.g., socio-economic background, motivation) may contribute to the development of language aptitude. Studies of the past years have primarily focused on disentangling the contribution of WM capacity, musical ability and/or musical training for language aptitude, while others have focused on the impact of educational settings and learning contexts on FLL. However, most of these studies have yielded largely inconsistent results and many have suffered from severe limitations (e.g., small sample sizes, different methodologies).

In the first few chapters of this thesis I aimed to give a very brief overview of the research that has been done on language aptitude in the past years. Since researchers have looked at such a variety of concepts related to language aptitude, it is practically impossible to keep track of all the studies that have been conducted in the past years. What complicates matters even more is that several terms have been used to designate language aptitude, including the terms language talent, proficiency, or learning ability. Other closely associated terms and concepts are language learning success or language achievement, both referring to a final state of performance and not an initial potential. Hence, the theoretical underpinnings of language aptitude have been extensively discussed recently, in particular to provide a valid, general theoretical framework relating language aptitude with FLL processes on behavioural and neural levels. J.B. Carroll proposed a quite successful model of language learning many years ago but his conceptions of language aptitude and the subcomponents he defined as underlying language aptitude have been under heavy critique in the past 20 years. Numerous researchers (e.g., Peter Skehan, Peter Robinson, and Zoltan Dörnyei) have elaborated and
advanced the concept of language aptitude in the past years by addressing previously neglected issues, such as the involvement of other cognitive abilities and the context of language learning. The newer models proposed by these researchers have certainly added very valuable insights into the construct of language aptitude and they are, in my view, complementary to each other, tackling the issue from different viewpoints. However, while I agree that language learning takes place in classrooms and aspects such as communication, instruction and feedback deserve their space in SLA research, this approach does not completely go hand in hand with the notion of language aptitude that I presented at the very beginning of this thesis. In my view, your potential to learn foreign languages should be independent of the instructions you receive, that is, it should be independent of teachers, classrooms and methodological factors like feedback or error correction. I see it more as an implicit learning process that takes place in the absence of a concrete learning context and should be similar for each individual, regardless their background, linguistic knowledge and learning situation. Nevertheless, I also agree with researchers who have pointed out that meta-linguistic awareness (Jessner, 1999, 2014), for instance, could be a dimension worth looking into as it might influence scores on language aptitude tests.

The present thesis aimed primarily at investigating potential neuroanatomical makers for high language aptitude in the human AC, while at the same time further exploring the interconnectedness between musicality, WM and language aptitude in two different populations. The two experimental studies at the core of this thesis were the first to confirm a potential relationship between the gross structural variation in human AC and an individual’s a priori disposition for acquiring foreign languages. Interestingly, the results of both studies found that only right-hemispheric AC morphology correlated with language aptitude. High language aptitude was associated with the possession of more than one complete gyrus (CPDs and multiple gyri, respectively) in both populations, while low language aptitude was associated with the occurrence of single gyri in right AC. Behaviourally, the earlier hypothesised connection between language aptitude and musicality/musical experience could only be partly confirmed in the first study with adults, where musical experience and musical ability showed weak links with different language aptitude measures. In the child study, in contrast, no relation between linguistic and musical abilities was found. WM capacity, on the other hand, was closely linked to language aptitude and was found to be the most important predictor thereof, supporting claims from numerous studies of the past decades. In the second
study with children and teenagers, language aptitude further correlated highly with arithmetic skills, school performance and self- and other-reported language aptitude. This provides evidence that a predisposition for learning languages might not only be advantageous for language-related learning processes but could further have a positive impact on the development of mathematical skills and academic achievement.

This thesis aimed to verify three research hypotheses, one dealing with potential neuroanatomical correlates of language aptitude, the other two with behavioural links between language aptitude and other skills. Neuroscientific and behavioural results will therefore be discussed in separate subchapters. While research hypothesis 1 will be dealt with in the first subchapter, the other two research questions will be summarized and dealt with simultaneously in the second part of this discussion dealing with potential behavioural correlates of language aptitude. Additional findings that are of interest with regard to the overall topic of this thesis will be added to the corresponding subchapters. In the two previous chapters (Chapters 5 and 6, respectively) the two studies at the core of this thesis were dealt with in some detail already. These chapters have hopefully provided comprehensible methodological approaches alongside with a thorough presentation of the results and detailed, but limited interpretations of the outcomes of the studies. Therefore, the discussions of Chapters 5 and 6 have already touched upon many topics and possible interpretations thereof. Still, I have refrained from drawing general conclusions and have not compared the results of the two studies with respect to the underlying research hypotheses since this is the general aim of this overall discussion section.

7.1. Auditory Cortex Anatomy as a Neuroanatomical Marker for Foreign Language Aptitude and Musicality

The first research hypothesis addressed the potential link between language aptitude and HG, an essential part of the AC that is known to host both primary and secondary auditory processing areas. While the functional properties of HG have received widespread interest in the cognitive neurosciences, few studies on language learning have actually dealt with variation in the structure of that specific region. The research hypothesis formulated before the onset of the two studies was the following:
Research Hypothesis 1:

Given the left-lateralization of language processing and the significance of right auditory cortex for musical processing, left-hemispheric auditory cortex morphology (i.e. individual variation in the structure called Heschl’s gyrus) is expected to be a potential neuroanatomical trait linked to language aptitude. More specifically, more complete gyri in left auditory cortex should correlate with higher scores in the language aptitude testing batteries in children and adults.

After a brief summary of the findings of the two studies, the potential role of the right hemisphere for language aptitude will be discussed, in particular concerning differences between right and left AC and how they might be vital for foreign language aptitude. While a short, separate discussion was already provided in Chapters 5 and 6 respectively, an in-depth discussion and interpretation of right-hemispheric involvement in language aptitude shall be provided. Subsequently, the topic of individual variation in AC and the potential role(s) of HG duplications for primary and/or secondary auditory processing shall be addressed. Last, a very sensitive topic, namely that of the innateness of gross structural variation in AC, shall be addressed. There is no reason to believe that the differences in morphology as observed in the present studies were induced by training or experience of any sort, be it of musical or linguistic nature. To put it differently, the children and adults who participated in the two studies were very likely born with the exact gross structures and GM volumes in AC (that is, the gyrification present at the time of measurement). This suggests that the correlates I found could be genetically predisposed markers for language aptitude.

7.1.1. Summary of the Findings of Study 1 and 2

The findings of the two separate studies, one with adults and the other with children and teenagers, confirmed that there seems to be a relation between right but not left AC morphology and language aptitude. While multiple gyri and CPDs were associated with high speech imitation ability (in adults) and high overall language aptitude (in children and adults), possessing a single gyrus, on the other hand, was associated with low language aptitude. Based on these findings, it seems possible, if not even likely, that right-hemispheric AC morphology is involved in the manifestation of foreign language aptitude.
Let us throw a closer look at the results again. Both conducted studies found that right-hemispheric AC morphology is a potential neuroanatomical marker of foreign language aptitude. More specifically, gyrification of AC was linked to either high or low language aptitude. In the first study with adults, only gross structural variation of AC was analysed. In other words, only differences in the major shapes and number of gyri was considered. In the group of children and teenagers, an additional analysis of GM volumes in AC was applied to add a further dimension to the results. In the adult group, more complete gyri (CPDs) of HG in the right hemisphere correlated highly and significantly with higher speech imitation skills in the Hindi task and with higher overall language aptitude as measured by the MLAT (measuring vocabulary learning, phonetic coding ability and language analysis ability). Additionally, right AC morphology also correlated with higher musical ability as measured by the AMMA test, assessing rhythm and pitch discrimination ability. In other words, those adults who possessed CPDs, that is those whose HG comprised at least two complete gyri in the AC of the right hemisphere, had the highest scores in the Hindi speech imitation task, the MLAT and the AMMA musicality test.

The second study was a partial replication study of the first. The major differences between the two studies was the age of participants, namely 20 to 40-year-old adults in the first study, and 10 to 16-year-old children and teenagers in the second study. Another difference was the test that was administered for assessing language aptitude. As the MLAT was not suitable for children and teenagers, the LLAMA language aptitude battery was chosen instead. Despite this difference in the choice of the test battery, results were expected to be similar. Hence, it was expected that children with multiple gyri or at least two complete gyri in their right AC would also present with higher speech imitation and higher language aptitude, as well as with higher musicality scores. The findings of the second study showed that those children with multiple gyri in their right AC were more likely to be in the high language aptitude group as measured by the LLAMA language aptitude battery, assessing phonetic coding ability, vocabulary learning, sound-symbol association learning and language analysis ability. Moreover, the higher the GM volumes in subjects’ right AC was, the higher their results on the language aptitude testing were. In this study, however, neither musical ability, nor speech imitation ability (identical tasks as used in the first study) could be linked to left- or right-hemispheric AC morphology.
To sum up, the results of the two studies indicated that right AC morphology was linked to language aptitude in so far that at least two complete gyri in a subject’s right AC were significantly associated with (1) high speech imitation ability in adults and (2) overall language aptitude in children and adults. Moreover, the findings further revealed that high musical ability, comprising rhythm and pitch perception, was further related to CPDs in right AC of adults. The findings concerning the link to musicality, however, could not be confirmed in the child group, where musicality failed to correlate with any other skill measured in the whole study.

7.1.2. The Significance of the Auditory Cortex for Language Aptitude

The AC was chosen for the present thesis for a variety of reasons: its importance as starting location of speech processing in the brain, the availability of already successfully applied manual segmentation methods and the often found strong relationship between language and music, with musical abilities having been linked to AC morphology in the right hemisphere (J. Benner et al., 2017; Schneider et al., 2005; Seither-Preisler et al., 2014). These reasons will be considered again briefly before going into detail about the significance of the findings of the two experimental studies.

First, any kind of speech input has to pass the AC before it is forwarded to other regions that are said to integrate linguistic meaning (e.g., the parietal lobe), including phonological, semantic and syntactic analyses. As mentioned briefly in the introduction to language aptitude and the brain, Hickok & Poeppel (2000, 2004) suggest that speech processing takes place along two routes, explaining why their model is also termed the dual route model. Acoustic speech signals, processed in AC regions, are linked to conceptual-semantic representations and at the same time, they are passed on to the motor speech system where they are reproduced with the vocal tract. The first part, namely the link between auditory processing and conceptual-semantic analyses is the so-called dorsal stream. The second part, the ventral stream, is responsible for the reproduction of speech signals in the motor cortex and inferior frontal regions. The two routes function independently and in a parallel fashion, explaining why we can effortlessly repeat a non-word, which has no semantic content for us, but might be unable to produce speech after a lesion while having intact speech comprehension (e.g., Wernicke’s aphasia; Dronkers & Baldo, 2010). The main point is that speech processing first requires
bilateral processing in auditory regions on dorsal STG and STS before diverging into these two broad streams. Therefore, the STG, hosting primary and secondary auditory areas, is the central point for speech perception and primarily involved in speech production as well. Logically, differences in auditory processing or AC anatomy (possibly the latter interfering with the first) could at some point prove to be advantageous or disadvantageous for overall speech comprehension and processing. The STG/AC is linked to frontal and parietal regions through the arcuate fasciculus, which makes the dual route processing possible. As has been mentioned earlier already, the arcuate fasciculus is a fibre tract linking the lateral temporal cortex with the frontal cortex, mostly projecting around the Sylvian fissure. It is assumed to be highly involved in human language as it is considerably smaller or even absent in non-human primates. The arcuate fasciculus and its two subsegments, one linking the STG with frontal regions, one linking it to middle temporal regions, are not equal in both hemispheres, with the left hemisphere usually having a stronger connection to the frontal regions (Glasser & Rilling, 2008). To date, hardly any studies have investigated the link between structural variation in AC to differences in functional activation, functional connectivity between STG and parietal/frontal regions or structural connectivity (differences in the arcuate fasciculus) between the very same regions.

The second reason why AC was chosen as the prime location of the present thesis was that no other brain region has received such widespread attention regarding its gross structural variation, making it an ideal location for studying the neuroanatomy of this specific region and its possible correlates. Schneider and his research group (J. Benner et al., 2017; Schneider et al., 2002; Seither-Preisler et al., 2014; Serrallach et al., 2016; Wengenroth et al., 2010; Zoellner et al., 2019) have extensively researched the gross morphology of AC, developing their own method for detailed structural analysis thereof. Their innovative semi-automated manual segmentation analysis considers individual patterns of HG and stands in sharp contrast to most other conventional structural analysis methods. Doubtlessly, there are various challenges to studying speech anatomy through structural imaging (see review by Marie & Golestani, 2017), explaining why structural research is generally still sparse, even more so on an individual level. The semi-automated segmentation method developed by Schneider and colleagues is an excellent choice that manages to take into account individual variation by manually segmenting the AC in each hemisphere. This provides the unique opportunity to study small differences in the structure of this very region.
Complementing the second reason, the third reason is the strong relationship between language and music and the findings of studies linking musical ability to AC morphology. Abdul-Kareem & Sluming (2008) argue that a single gyrus is the most frequent finding in both MRI and post-mortem studies and duplications only occur in 24-33% of cases, multiple gyri only in 6% of cases. The rarity of posterior duplications (CPDs) and multiple gyri makes their occurrence even more exciting, in particular paired with the findings that multiple gyri and duplications might be associated with outstanding musical abilities. As explained already in quite some detail, J. Benner et al. (2017) for instance, linked HG multiplications to musicianship, finding that 90% of musicians in their study exhibited them in either one or both hemispheres. Similar findings were reported by Schneider et al. (2005), in whose study the absolute size of HG correlated strongly with musical aptitude. Building upon the strong relationship between language and music and their shared origin, a link between musical abilities and right AC might be advantageous for language learning and processing as well. Brandt et al. (2012) wrote a whole article on why spoken language should be seen as a special type of music and that musical hearing and musical ability are essential prerequisites for language acquisition. Arguing that the acquisition of both language and music are equal in effort and speed, they concluded that without the ability to hear musically, it would be impossible to learn how to speak. I do not agree 100% with this theory since there is no reason to believe that music develops before language and it thus cannot be seen as a prerequisite of the other. Still, language and music are interconnected, and many studies have already shown that advanced musical abilities (whether learnt or predisposed), such as good rhythm or pitch perception, result in better speech perception (Schellenberg, 2015; Schön et al., 2004) and even in better speech production skills (Milovanov et al., 2008, 2010). Consequently, a region essentially involved in the perception and/or production of music might as well positively affect FLL processes in the brain.

To summarize, AC is one of the most researched and most frequently segmented brain regions to date. In other words, we know much more about the gross structural anatomy of AC than about any other brain region. Moreover, it has not only been found to correlate with high musical abilities or musicianship, but it is also the most important region when it comes to the processing of incoming auditory information. Leaving aside potential differences between left and right AC, structural variation in AC might prove advantageous for structural and functional processing of speech input. This is the basis of language aptitude as language learning first
requires the analysis of speech material. Owing to all these reasons, it seemed a promising approach to look into the relationship between AC morphology and language aptitude.

The significance of right AC for language aptitude shall be addressed from different angles in the following subchapters. First, the differences in processing between right and left AC will be dealt with, allowing for an interpretation as to how functions of right HG might be associated with language processing on a general level, and forms of language aptitude on a more specific level. Secondly, the role and significance of variation in HG and the HG morphotypes distinguished in the two experimental studies shall be addressed. This will be particularly relevant for the subsequent discussion on whether HG and its duplications comprise primary and/or secondary auditory processing areas and how this relates to the findings as presented in chapters 5 and 6. Last but not least, the stability of AC morphology will be addressed and it will be discussed how findings thereon might contribute to our understanding of language aptitude.

### 7.1.3. Functions of Right and Left Auditory Cortex

The first question that shall be addressed at this point is the following: Why did I find neuroanatomical correlates of foreign language aptitude only in the right hemisphere, but not in the left given the leftward lateralization of language? This topic has been touched upon already in the discussion of Study 2 (Chapter 6) but it shall be highlighted here again that the right hemisphere has been consistently found to be implicated in various language learning processes, most notably also in language aptitude. Therefore, a more detailed discussion will be provided.

Very recently, Novén, Schremm, Nilsson, Horne, & Roll (2019) found that cortical thickness of Broca’s area and specifically its right homologue were both related to aptitude for grammar learning and pitch discrimination. Especially pitch perception, a key to the perception and production of new phonological patterns (Posedel, Emery, Souza, & Fountain, 2012), has been argued to play an important role for FLL, and is thus hypothesized to be essentially involved in language aptitude. Focussing on structural properties of this area, Novén et al. found that the thicker the cortex in right IFG was, the worse subjects’ performance in a pitch change direction discrimination task was. In other words, less GM, that is a thinner cortex, was associated with high pitch discrimination ability, a potential prerequisite for the acquisition of
foreign languages. Several years earlier, H. Lee et al. (2007) found that GM density in right and left posterior SMG were proportional to the size of vocabulary learnt in their subjects. Adding further evidence, Kepinska and colleagues (Kepinska, de Rover, et al., 2017b, 2017a; Kepinska, Pereda, et al., 2017) related grammatical analytical abilities (measured by LLAMA F, a subpart of the LLAMA language aptitude battery that was also used in Study 2) to the right hemisphere as well. Looking into functional activation, Kepinska et al. found that those subjects with a high aptitude for the acquisition of a novel grammar, engaged the right hemisphere more frequently (AG, SGM among others). Furthermore, they observed that increasing proficiency in grammar learning was associated with a stronger local synchronisation within right-hemispheric brain regions. These few studies shall just highlight that the functional relevance of the right hemisphere for (foreign) language processing is still far from being fully understood. Even if a leftward lateralization of language cannot be denied, it is irrefutable that numerous regions in the right hemisphere (primarily homologous regions to the left-sided language regions) are essentially involved in various processes related to FLL and might consequently play an important role for language aptitude as well.

The second important question is: Why should right AC be so important for foreign language aptitude? Let us begin by pinpointing the vital role of AC in general, without considering a differentiation between left and right AC at this point. The AC is one of the most important regions for language processing since no language can be learnt without the correct perception and analysis of speech input. This happens in primary and secondary auditory areas. During first language acquisition, speech input is vital for a child to successfully master the task of acquiring their native language. Likewise, when acquiring a foreign language, the perception of speech and all speech-related cues is essential for becoming a fluent speaker of this language. Therefore, there is little doubt that learning a foreign language requires excellent auditory and phonological perception skills. Similarly, having an outstanding or high potential for FLL, such as was the case with the participants in our study, could easily be based on excellent auditory and phonological skills that facilitate all later occurring stages of language processing.

One explanation for the involvement of right AC in language learning could be deduced from its relevance for first language acquisition. Studies have reported that right AC plays an important role in first language acquisition (Perani et al., 2011). Based on their findings, Perani et al. argued that speech input to infants is more dependent on processing of the right primary
and secondary auditory areas. Likewise, speech-related acoustic cues like speech prosody, dominantly processed in the right hemisphere, are essential as a source of information for infants who acquire their native language (Homae et al., 2006). Homae et al.’s idea is that speech processing in the human infant brain develops from the analysis of pitch information and is based on the comparison and integration of information in speech sounds with acquired prosodic structures. It should not be forgotten that even later stages of language learning depend highly on the recognition and memorization of prosodic contours of an unfamiliar language, which still show a rightward asymmetry in the adult brain. These features are perceived and processed long before elements of the new language are explicitly used and manipulated. This idea goes hand in hand with Brand et al.’s (2012) theory of language learning being based on the processing of musical properties (namely, syllables as rhythmic entities). In their work, they argued that it is more productive to see spoken language as a special type of music, mainly from a developmental perspective. In their article, they suggested that children acquire language as a sequence of syllables that are learnt in a specific rhythm, i.e. language is learnt by each child as a sequence of rhythmic elements. When focussing on rhythm or speech prosody then, studies have shown that these features are already right-lateralized in 10-month old infants (Homae et al., 2006). In a review by Chiron et al. (1997) it is also emphasized that the right hemisphere matures earlier in foetal and postnatal stages, which could enable global feature extraction before more analytic skills are required. According to the authors, the blood flow generally shows a right predominance between age 1 and 3, mostly due to activity in posterior associative areas. The shift of right-to-left predominance occurs in the fourth year of life then. Therefore, also language is predominantly right lateralized in the early stages of language acquisition.

Moreover, differences in temporal integration windows could also explain the relevance and involvement of right AC for language aptitude. While left AC was frequently reported to possess a sensitivity for the processing of fast acoustic events and is therefore expected to be more sensitive to temporally conveyed fundamental pitch of complex tones (Zatorre et al., 2002), the right side appears to be specialized for spectral discrimination of different timbres of musical and vocal sounds (Schneider et al., 2005). This predisposes the right AC for the processing of finer resolution in the frequency domain (Hyde et al., 2008; Poeppel et al., 2004). Nonetheless, it has also been suggested that speech perception, construed as the integration of acoustic information into representations that enter the linguistic system, requires bilateral
auditory processing in early stages (Poeppel, 2001). This stands in contrast to the aforementioned hypothesis assigning different features (spectral vs. temporal processing) to the different hemispheres. In Poeppel’s model, both hemispheres extract information out of the acoustic signal, but they work in different temporal integration windows. This was supported by Poeppel (2003) finding that speech perception was mediated bilaterally. As such, left AC was required for the analysis of short segments (25-50ms) and right AC for long segments (200-300ms; see also Boemio et al., 2005; Poeppel, 2001). Considering these different, distinct temporal integration windows, it seems that right AC is more likely predisposed for the analysis of slow musical rhythms and melodic contours, but not single speech sounds. Regarding language, however, also syllabic structures and prosodic modulations of spoken language require the analysis of longer segments (Leong & Goswami, 2014), which are usually processed in right AC. Consequently, right AC could be involved in language aptitude due to its significance for processing longer segments, such as rhythms and melodic contours, as well as other suprasegmental features.

To put it in a nutshell, the right hemisphere very generally and especially right AC have been found to be not only essentially involved in first language acquisition, but also in the acquisition of foreign languages. Despite the leftward lateralization of language, particularly auditory processing is highly reliant on the right hemisphere and it thus seems little surprising that a variety of language learning skills, have recently been reported to be associated with right-hemispheric language regions. Likewise, right AC is responsible for the processing of numerous features that are not only relevant for music, but also for language. Consequently, it might be more important than left AC for the processing of unknown speech material, particularly for longer segments that are being acquired (e.g., syllables, words, sentences) and suprasegmental features like melody and prosody that are expected to play a vital role in FLL.

7.1.4. Individual Variation of Heschl’s Gyrus and the Significance of Gyrification Patterns

Before starting an interpretation, it has to be clarified that the types of HG that were distinguished in the neuroanatomical analyses are not linked to a functional distinction, nor is one type superior to another on any level. Based on the analyses and results found by P. Schneider and his research group, three to four types of HG are clearly distinguishable from
one another, even for a non-expert. The four types are single gyrus, CSD, CPD and multiple gyri (in which case there are more than two complete gyri). While some studies distinguished between the three types only (Da Costa et al., 2011; as is the case in Study 1, Chapter 5), other studies distinguished between the four types (J. Benner et al., 2017; Seither-Preisler et al., 2014; Serrallach et al., 2016; as is the case in Study 2, Chapter 6).

So far, neuroscientists are far from fully understanding why HG presents such high inter-individual and inter-hemispherical variation. In other words, we do not know why some people have CPDs, others CSDs, and why the types of HG differ between left and right AC in the very same individual. To date, only very few studies have addressed the topic of how the different forms of HG are related to functional processing of auditory input and if and to what extent differences in the gross anatomy of AC or GM volumes thereof are potentially related to white matter differences or differences in functional connectivity within brain areas. Overall, HG is known to exhibit very high anatomical variability not only in size but also in gyrification across hemispheres and individuals. It is particularly interesting that left HG shows considerably less variation than right HG. This has not only been shown in the present studies but also in a large study by Marie et al. (2015). In their study, they distinguished between three types of HG, like in Study 1 presented in the present thesis, and looked at 430 ACs of individuals with various backgrounds. They found that regardless of the other variables considered (e.g., musical experience, linguistic background, sex), the right hemisphere showed considerably larger variation than the left hemisphere.

The first claim that HG in the left and right hemispheres are structurally not equal came from von Economo and Horn (1930), who suggested that HG was often longer and wider on the left, but more frequently duplicated on the right side. They concluded that the PAC in both hemispheres appeared equal given that HG was larger on the left and occurred in the form of duplications on the right. The most systematic microscopic investigation was undertaken by Rademacher et al. (1993) who concluded that left HG was over 10% larger than right HG in 6 out of 10 cases (not statistically proven, however). Penhune et al. (1996) found a consistent left-over-right asymmetry in total PAC volume, mostly resulting from increased cortical white matter on the left. In the latter studies, however, only the first gyrus was considered as PAC and included in analyses. Cytoarchitectonic (Fullerton & Pandya, 2007) and functional (De Martino et al., 2015), as well as myelin-based studies (Wasserthal, Brechmann, Stadler, Fischl, & Engel, 2014) have shown that PAC is localized in the anterolateral two-thirds of HG (see
also (Hackett, Preuss, & Kaas, 2001; Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002). Da Costa et al. (2011) argued that this assumption that PAC occupies only the first, anterior division of HG duplications was wrong. In their study, da Costa et al. confirmed that PAC spans both divisions of duplicated HG, not only the first division as earlier presumed. Providing further structural and functional evidence, (Zoellner et al., 2019) found that PAC was represented by areas of reduced cortical thickness and spanned across HG duplications. Consequently, possessing multiple gyri in right or left AC could prove to result in differences in primary and secondary auditory processing of speech.

In the first study I distinguished between three types of gyri (single, CSD, CPD) only because almost all of the participants had one of the three, only a minor percentage displayed multiple gyri in either hemisphere. Therefore, putting these occurrences into a separate group would have severely limited the statistical power of the analyses and the results would have been negative due to the small group size for that category. In the second study, however, a large number of children had multiple gyri in their AC, which is why multiple gyri were counted as a separate category. Since there is no evidence as to whether it is functionally or structurally highly relevant to possess two complete gyri (i.e. a CPD) or three complete gyri (i.e. multiple gyri), one could argue that it is both plausible to count both as separate categories or combine them as a ‘more than one gyrus’ category, at least from a structural perspective. In the analyses of the second study, the results were also highly significant when the analyses were performed with multiple gyri and CPDs in a single category, but the effect was mainly due to the occurrence of multiple gyri, not of CPDs.

To my knowledge, hardly any studies have looked into functional connectivity and activation differences between subjects who possess CPDs and/or multiple gyri in either hemisphere. Tzourio-Mazoyer et al. (2015) were some of the only studies looking into the structure-function relationship between the anatomy of HG and speech-hemispheric lateralization in a large sample. They found that duplications were associated with a decrease in surface area of anterior HG and functional activity. Left-handers were less likely to possess right HG duplications and smaller surface areas of the right first HG. They further found that subjects with a single gyrus in left AC had leftward anatomical and functional asymmetry of HG, while those with a duplication in left AC had anatomical and/or functional leftward asymmetries. The authors concluded that they were the first to show an association between
functional asymmetries in HG and inter-individual variability of inter-hemispheric gyrification patterns.

On the other hand, several studies have reported that specific patterns of duplications are related to certain disorders or musicianship. Wengenroth et al. (2010) found differences in functional activation, accompanied by 2.2-fold increased volume of left HG in subjects suffering from Williams Syndrome. HG duplications were also found in children with learning disabilities (Leonard et al., 2001). In their 2001 study, Leonard et al. searched for anatomical markers that could restrict variability in the reading disability phenotype. 11 out of their 15 individuals with reading disability had a major phonological deficit. Overall, only four anatomical measures distinguished controls from reading-disordered, one of them being a large duplication of HG on the left side. In a later study, Leonard et al. (2002) found that the size of a single left HG was significantly associated with reading disability. They further developed an anatomical risk factor index for phonologically based learning disability, concluding that phonological deficits and language impairment were qualitatively different disorders highly linked to anatomical deviations. Following that argument, HG duplications could be seen as markers for phonological-based learning disability, associated mostly with reading disorder. Considering all these results, it seems most likely that differences in HG morphotype could lead to consequences in phonological processing, the next step after primary auditory analysis. If we take a closer look at the auditory pathway\(^7\), we see that any form of incoming information from the major auditory tract passes through the medial geniculate body to reach the transverse gyri of Heschl, i.e. HG. Then, the signal reaches the PT and the STG, meaning that all these three areas (HG, PT, STG) are instantly connected during auditory analysis processes (Rauschecker & Scott, 2009; Webster, 1992; for the specific role of duplications, see 7.1.5.).

The more difficult part now is to explain how these differences in gyral patterns could result in functional or structural differences and consequently lead to variation in language aptitude. Structural connectivity depends on the arcuate fasciculus that links auditory to frontal regions. Comparable to differences in GM in HG, this major fibre tract involved in speech and

---

\(^7\) The auditory pathway as described in this chapter is a very simplified one and does not claim to be comprehensive at all. For an excellent summary and overview of all steps between the inner ear and HG, please see the chapter entitled *An Overview of Mammalian Auditory Pathways* in Webster, Popper and Fay (1992).
language processing has been linked to schizophrenia (De Weijer et al., 2011), auditory-verbal hallucinations (Geoffroy et al., 2014; Hubl et al., 2004; McCarthy-Jones & Longden, 2015) and musical ability (Halwani, Loui, Rueber, & Schlaug, 2011). Moreover, specifically the left arcuate fasciculus has also been linked to word learning (López-Barroso et al., 2013), autism (Fletcher et al., 2010; Wan, Marchina, Norton, & Schlaug, 2012), dyslexia (Langer et al., 2017; Vandermosten et al., 2012, 2015) and aphasia (Fridriksson, Guo, Fillmore, Holland, & Rorden, 2013). Therefore, structural connectivity seems to have a large impact on the development of a large variety of symptoms and disabilities, although it is unclear whether it is the consequence or cause thereof. It seems very likely that functional structural connectivity in the arcuate fasciculus could be a result of structural variation in AC. Assuming that structural variation in HG is a biologically determined trait, it should precede the functional connections built to make up the arcuate fasciculus, which are expected to develop much later. In this case, the specific morphotypes of HG could lead to differences in the strength and directedness of the fibre tracts relating auditory regions to other language-related areas. However, studies exploring the relationship between the arcuate fasciculus and structural variation in AC are needed to confirm this hypothesis. From the present thesis it is unfortunately impossible to shed further light onto this topic.

7.1.5. **Duplications – Primary or Secondary Auditory Areas?**

Until now, the AC has been presented and treated as one single entity in the brain, one area responsible for a specific process, namely analysing incoming auditory input. This was done on purpose in order not to confuse the reader, but a thorough discussion requires a distinction between primary and secondary auditory areas. Since one of the main arguments of this thesis is that right AC morphology might potentially be involved in foreign language aptitude, as suggested by the findings of the two studies, it seems worthwhile to closely investigate differences in function and structure of primary and secondary auditory regions in the human brain.

The central auditory system in the human brain comprises several structures, including the brainstem, the midbrain, the thalamus and the cerebral cortex. The early auditory system is known to decompose complex sounds, like speech, into elementary time-frequency representations, before higher level phonetic, phonological and lexical processing takes place.
Auditory areas comprise primary processing areas (core region in the superior surface of the temporal lobe) and secondary areas located in the parabelt region, as found in research on non-human and human primates (Galaburda & Sanides, 1980; Langers, van Dijk, & Backes, 2005). This distinction has originally been proposed based on differences in cytoarchitectonic features (i.e. differences in cell types; Brodmann, 1909; Rademacher et al., 2001), with functional studies providing further evidence. When the PAC perceives any kind of auditory input, it directly forwards it to the secondary areas, although secondary or parabelt areas do not follow the hierarchical and tonotopic organization of primary areas (Langers et al., 2005). Regarding speech and language, it is worth mentioning that no analysis of the specific type of input or meaning of the perceived sound or word happens at the primary level. Only acoustic features like loudness, frequency, timbre and the like, as well as information on the specific location of and distance to the sound are analysed in primary auditory areas. All further language-related processing happens at the secondary level, that is in secondary or parabelt regions, which could potentially comprise HG duplications.

It is not clear to date whether HG and the few morphotypes in which it occurs are associated with either primary or secondary auditory processing. Recent studies have shown that HG comprises a posteromedial primary and an anterolateral secondary part, that is not the whole surface of HG is responsible for primary auditory processing. Moreover, cytoarchitectonic studies have also found that PAC and HG do not exactly overlap at the microscopic level (Morosan et al., 2001). Quite in contrast, secondary areas have been shown to be vital for the larger integration of more complex auditory patterns and their recognition, which is why secondary auditory areas have been particularly emphasized in language and music research (Leaver & Rauschecker, 2010). While the focus of left-sided secondary areas was on language, primary and secondary areas in the right hemisphere have been linked to music more frequently. As proposed in reviews by (Zatorre, 2013; Zatorre, Fields, & Johansen-Berg, 2012), numerous reports so far have described associations between the anatomical characteristics of HG and long-term auditory learning for music and speech, or even specific expertise in one of the two domains. Furthermore, primary and secondary areas have been found to show a characteristic hemispheric specialization (Schirmer, Fox, & Grandjean, 2012; Zoellner et al., 2019), discussed in a large variety of studies. In other words, the specific morphotypes investigated in this study are at the intersection of primary to secondary auditory areas.
What is known for sure is that the additional gyri that were found in highly gifted adults and children are not necessary or functionally relevant for the primary analysis of auditory information, although studies by Schneider and colleagues (Zoellner et al., 2019) have shown that primary analysis of auditory input might take place more posteriorly in individuals with HG duplications. This seems to be especially the case for specific populations, e.g., children with learning disabilities or ADHD (Seither-Preisler et al., 2014; Serrallach et al., 2016). A major problem is that many of the studies looking into the relationship between specific abilities and AC only included the first HG and ignored the occurrence of duplications. This was the case in Wong, Perrachione, & Parrish (2007), who related left HG volume to linguistic pitch learning, and Golestani et al. (2007, 2002) who explored the relationship between foreign speech sound learning and HG anatomy. Therefore, these studies did not really take into account the variation in which HG occurs, but mainly looked into variation of the first gyrus of HG.

Considering that each AC is a hub for numerous neural circuits to other language-relevant brain areas involved in multimodal processed, motor activation and speech production (Friederici & Gierhan, 2013), the specific role of these gyri could be related to functional and structural advantages in foreign language processing. Doubtlessly, more efficient and faster processing and integration of larger speech units (not only single sounds but also syllables and whole words) could be advantageous for the processing of unknown speech material. Considering the results of the two studies, higher GM volumes in secondary auditory areas of the right hemisphere might provide better structural links to other language-related areas, e.g., the temporo-parietal junction or the IPL and be relevant for functional or structural connectivity between these areas. However, it is not entirely clear whether HG duplications, in particular in the case of multiple gyri, belong to secondary AC or might already be related to multimodal processing starting most likely with the PT. Moreover, there is no concrete boundary between the temporo-parietal junction and secondary auditory areas, meaning that the occurrence of multiple gyri might as well be of relevance for our interpretation of the PT and the temporo-parietal junction.

My personal interpretation is that duplications of HG could result in an advantage for phonological processing in language learners if the duplications are functionally and structurally linked to other language-relevant regions. Phonological processing is a vital process during SLA/FLL and could thus lead to higher efficiency and better performance in a
variety of tasks related to language learning. While every individual relies on these regions in the left hemisphere, using the right hemisphere in this case could make the difference between an average language learner and a high aptitude language learner. However, if the duplications are not well linked on the structural and functional level, they might result in a variety of symptoms or deficits (e.g., dyslexia, ADHD). This is a very simple approach and much more research is needed to confirm these hypotheses. I would like to mention that clearly defining the roles of HG duplications was not a primary goal of the present studies and the topic is therefore only addressed briefly at this point.

To summarize, it is unknown to date if HG duplications fulfil only primary auditory or secondary (i.e. phonological and lexico-semantic) auditory functions. Studies have shown that duplications are implicated in primary auditory processing, but this does not mean that they cannot be involved in secondary auditory and phonological processes as well. From the present study, I cannot infer on the functional or structural relevance of the anterior STG or deduce the concrete functions of HG and its duplications, in particular regarding its involvement in foreign language processing. More studies are needed to explore this highly interesting topics, in particular studies combining functional and structural investigations to primary and secondary auditory areas would be ideal to shed further light on this topic.

7.1.6. Right Auditory Cortex and Musicality

The present research only partly succeeded in confirming a relationship between high musical ability and the gross structures of AC in the right hemisphere. Only in the adult population, musical ability was significantly correlated with the occurrence of CPDs in right AC. In the child study, in contrast, where the same test was used, no significant relationship between musicality and AC morphology could be found.

The whole project for this PhD thesis was based on findings relating musical ability to GM volume and structural variation in right AC (see Benner et al., 2017; Schneider et al 2005). Although it was not a major aim to test this relationship in the present studies, it was a by-product as musicality was tested alongside with language aptitude and the same neuroanatomical analyses that had been applied in previous studies were used. In the adult group, those individuals who possessed CPDs, that is more than one single HG, in their right hemisphere, also scored highly on the AMMA test assessing rhythmic and pitch perception. In
the second study, however, none of the statistical tests found a relationship between AC morphology and musicality, neither with HG shape nor with GM volumes in right or left AC. Even more surprisingly, none of the behavioural links with musicality were found in the younger population.

Why did I not find a relationship between AC morphology and musicality in both studies? There are several possible explanations for the lack of findings in the second study. The first option is that the results differed due to age, the most obvious explanation. In the first study adults were assessed, compared to children and teenagers in the second study. Another option is that the weak link in the first study could have been a false positive that was not found in the larger sample for study 2. Another methodological constraint could be the test I used for assessing musicality, the AMMA test. It might have been a good choice in adults, not in the child group, however.

Although age might sound like a promising explanatory factor, it is very unlikely that the link between AC morphology and musicality is dependent on age. Since one of the main arguments of the present thesis is that AC morphology is a stable trait, it is very unlikely to change over time. In particular, the oldest children of Study 2 were 16 years, and the youngest subjects of Study 1 were 20 years of age. Therefore, there was not such a large age gap. The only reason why age might have nonetheless played a role is due to the last reason presented in the previous paragraph. The AMMA test is usually used for children older than 12/13 years and some of the children of Study 2 were 10-13 (only four altogether). Results of the AMMA test might have varied slightly in the younger population with older children having fully developed their musical ability and thus receiving higher scores. Adults, on the other hand, as tested for Study 1, should have already reached a stable level of musical ability, meaning that age would not have made any difference to their scores on the AMMA test. This might be a possible explanation as to why age might have played a role, but it is still very unlikely to have influenced the results.

Another explanation is that the observed link might have been a false positive and thus the result of a statistical problem. Even though this might provide a better explanation than the two other presented options, it does not explain why other studies (Schneider et al., 2005; Benner et al., 2017) found a correlation between AC morphology and musical ability. In Schneider et al.’s study, AC morphology was linked to one of the AMMA subtests only and in
Benner et al. (2017) AC morphology was related to musicianship. Therefore, the links that have been found varied substantially between studies, with either musical abilities (pitch perception) or professional musicianship being related to AC morphology.

In sum, AC morphology did not correlate with musical ability in the younger population of children and teenagers, only in the adult study. However, the link between AC morphology and musicality in the adult study was much weaker than the link between language aptitude and structural variation in HG. It thus requires further research to confirm the role of AC morphology for musicality and/or musicianship as suggested by earlier studies.

7.1.7. Auditory Cortex Morphology as a Stable Neuroanatomical Trait

The present thesis presented two studies that strongly support the claim that language aptitude might be partly biologically determined and thus innate, at least on the neural level. The results will be briefly interpreted with this respect and it shall be discussed what an innate perspective on language aptitude and AC morphology might look like and what the findings of the present study might contribute to that.

I have already marginally dealt with the possible standpoints one could take when talking about innate aspects of behavioural findings, like the intricate relationship between WM and language aptitude. No matter who you talk to, be it laymen, researchers, physicians or psychologists, opinions will diverge on this matter. Opinions diverge not only because of the inconsistent findings of studies, but also because of the possibly human problem of accepting that one’s fate might be destined before birth and is thus not easily changeable. However, this is not what biologically determined or predisposed truly means (see Gagné’s model of giftedness in 2.1) and the reader should by now be aware that an innate trait only develops under the right circumstances and with environmental support.

Since I started working on my PhD project, I have been confronted with the fact that people generally still struggle with the concept of ‘innateness’. I think that it is partly because of the term itself and because it is hard to accept that there are things you might be unable to change but I think it is high time that we admit that humans, gifted in different domains as they might be, are not capable of achieving everything, no matter how much we want to believe that. With many concepts, it has already been accepted that they are to a great extent unalterable and cannot be enhanced through training or experience, that is they are innate. Intelligence,
instance, is a good example for a comparison here. When applying different intelligence tests, we will always find differences between individuals and there is nothing wrong with that. With language aptitude, the picture is actually quite similar. At some point, we have to accept that a potential can be innate and at the same time also to a certain extent dependent on environmental influences and personality traits. As Gagné (2000, 2004, 2010) very clearly displayed in his model, an aptitude in a domain cannot develop without environmental stimulation and developmental processes.

If a trait is described to be genetically determined and thus ‘innate’, it is expected to be determined before birth (pre-natally) or during the very first days and weeks of an infant’s life. In other words, external influences should not have had a considerable influence on that trait, meaning that something must have been there before environmental stimulation begins. Some people would probably even argue that the first days and weeks of an infant’s life are already too late, and it all has to happen before birth or at latest in the last weeks of gestation. Neuroscientific research on the development of the human brain has shown that basic structures, like AC, are already developed before the birth of a child. Even if the auditory system is not fully mature in utero, the developing auditory system enables response to sound already at gestational weeks 25-27 (Galambos & Hecox, 1978, for a good summary on the development of the ear and hearing functions, refer to Litovsky, 2015). Chi et al. (1977) described temporal development of the cerebral hemispheres in their work, reporting that STG and HG develop after gestational week 28, the PT and AC after gestational week 31. The earliest primary fissure, on the other hand, appears already at the eighth gestational week. Few studies to date have explored the longitudinal developments of structures in the brain over a lifetime. In the works of Schneider & Seither-Preisler (2015) and Seither-Preisler et al. (2014), the authors reported that HG and GM in AC remained stable (99.1%) in a group of 109 children over 3.6 years (on average). This is particularly striking given that the brain undergoes so much change functionally from birth on. Some examples thereof would be alterations and modulations of white matter (structural connectivity between brain areas), cortical thickness (the thickness of the GM layer over a gyrus) or functional connectivity (which regions in the brain are connected to each other and how). When looking at this substantial difference between high stability and complete instability, we might see why the human brain is so complex and possibly also why it is capable of holding in place so much information, from memories to skills, emotions and the like.
To summarize, the human brain is plastic and can change on many levels due to maturation, experience and training. The neuroanatomy of the human brain, on the other hand, is less likely to change during life, only cortical thickness and GM surface volume seem to undergo major changes. The focus of the two studies presented in this thesis was on one area of the brain that is developed very early on and has been shown to be stable over long periods of time, namely HG. While the present studies cannot contribute anything to the stability of structural variation in AC, it seems very likely that the correlation between AC morphology and language aptitude was not due to maturational processes or experience, but was based on an a priori physiological trait.

7.2. The Concept of Language Aptitude Revisited

I hoped that the results of the two studies would give clear indications as to whether the research hypotheses should be rejected or could be accepted with minor adjustments. However, when writing up the results and working on the discussion sections on the behavioural correlates of language aptitude, I realized that interpreting the results accordingly was a quite complex and challenging process. In other words, answering if musicality or WM were implicated in foreign language aptitude was easy (yes or no according to the results) but explaining why they are or are not implicated therein and how this may alter the overall concept of language aptitude was a completely different story. This was further complicated by the simple fact that there is no agreement on the specific components of the construct of language aptitude. Therefore, I decided to dedicate a small subchapter to my own perception and ideas of the concept of language aptitude to enable a discussion on the behavioural correlates of language aptitude and the potential changes induced by the findings.

The concept of language aptitude has undergone changes in the past years and as already touched upon in Chapter 2, the term ‘aptitude’ has often been replaced by the plural form recently. As already clarified, there is no doubt that various related skills make up language aptitude and they do not necessarily overlap or interact at all. Still, there have been a lot of inconsistencies, with many recent publications from well-known aptitude researchers still talking about ‘language aptitude’ as one construct (e.g., S. Li, 2018; Long & Granena, 2018; Winke, 2018), while others have been using the plural form for years. Possibly, the results of the present studies can provide further evidence as to the relation between various concepts.
and skills related to language aptitude and how a satisfying, general framework can be developed. In order to do so, let us now investigate what the tests I introduced as measuring language aptitude in the two experimental studies looked like and in how far they interacted or overlapped. In the first study, an English pronunciation task, a Hindi speech imitation task, and three tests of the MLAT were chosen as language aptitude tests. In the child study, the four subtests of the LLAMA and the Hindi speech imitation test, as well as the Hindi perception of the MULT/AP were chosen as language aptitude tests.

In Study 1, the phonetic memory task (MLAT III) was not linked to the other two language aptitude tasks of the MLAT, parts IV (grammatical sensitivity) and V (vocabulary learning), while the latter two were significantly linked to each other. It has been suggested that MLAT IV is a measure of meta-linguistic awareness and not directly language aptitude, providing a possible explanation for the lack of correlation with any of the other tests (Robinson, 2007). However, it was also not linked to the number of languages learnt with an individual or age, partly disproving this claim made by Robinson. In the adult population, the Hindi task was not associated with the MLAT tests, but it showed a strong relation to the English pronunciation score, indicating that a certain giftedness in speech imitation might interact with or lead to good pronunciation skills in a language. One could also argue, on the other hand, that the excellent pronunciation in English (proficiency) led to better Hindi speech imitation through language-learning experience.

In Study 2, the parts of the LLAMA were linked to each other before correction for multiple comparisons but only one link remained significant thereafter. LLAMA E (sound-symbol association) was significantly linked to LLAMA B (vocabulary learning). There was also a trend (after correction) for LLAMA D (phonetic memory) to correlate with LLAMA B and LLAMA E, while LLAMA F only showed a trend to correlate with LLAMA E. It might thus be suggested that LLAMA F measures a very different aspect or facet of language aptitude to that measured by LLAMA B and D, for instance, although LLAMA E could be the mediating variable between all these tests. The major problem is that neither of the tests, be it the MLAT or the LLAMA, only measure one specific component of language aptitude. Therefore, it would not be wise to say that LLAMA F is completely independent of the other three and measures a separate form of language aptitude. Also, the Hindi test, a measure for WM and speech imitation ability, was thought at first to capture good perception, good memory and good motor control for speech imitation. Nonetheless, the Hindi speech imitation test did not correlate at
all with LLAMA D, in which foreign speech input has to be retained in phonetic memory, or non-word span. In the child study, I undertook an extra testing to assess the relationship between productive and perceptive skills. Therefore, the MULT/AP part assessing Hindi speech perception was administered as well. The main aim behind assessing both the perception and production of an unknown language (Hindi, in this case) was to test whether there was a relation between the two suggesting that production and perception are really strongly linked abilities. Surprisingly, there was no significant correlation (no tendency, not even the slightest hint) between children’s performance on the Hindi speech imitation and the Hindi speech perception test (MULT/AP). Moreover, the Hindi speech perception results were not linked to results on the LLAMA test, neither to any of the WM scales, except for a weak link to digit span backward. Regarding the concept of language aptitude, it seems plausible to argue that production and perception are largely distinct abilities that do not necessarily overlap. One of my initial arguments why the Hindi speech imitation task could be correlated with AC morphology in the right hemisphere was because of the fact that speech imitation requires excellent perceptive skills as a basis for the imitation. However, in this case speech imitation and speech perception would have to show a strong relationship, which was not the case.

Based on the findings of both studies, one could argue that not one single concept of language aptitude was measured in the two experimental studies. Rather, it seems that a plethora of different abilities and aptitudes were assessed through the variety of administered tests. Without explicitly providing support for the existence of each of the following categories, I would argue that the different scales were related to the following abilities: (1) facets of speech perception (including the idea of the component ‘phonetic coding ability’; assessed by MLAT III, LLAMA D, MULT/AP), (2) speech production (not really included in any concept of language aptitude; Hindi speech imitation, English pronunciation), (3) vocabulary learning (MLAT V and LLAMA B), and (4) implicit (grammatical) language learning including semantic, phonetic and syntactic knowledge of a language (related to language analytic ability and possibly highly dependent on implicit learning aptitude; assessed by LLAMA E, F, MLAT IV). While this clearly sounds more realistic than claiming to measure the overall concept of language aptitude, there are some major limitations to that idea as well. First, not all tests assessing speech perception showed a correlation with each other at all (MULT/AP could not be linked to LLAMA D and measures between adults and children could not be compared
directly to verify potential links). Second, the fourth category is vast and seems to heavily rely on analytic skills, which do not necessarily have to be language-specific in this case. Another problem with these categories is also that they might be differently dependent on other cognitive abilities (like WM), relate to each other to various extents and change according to the learning context and stage, which needs to be addressed in a new model of language aptitude.

Doughty (2019) highlights that language aptitude is seen as a componential construct and each component (independently or synergistically in complexes) underlies different aspects of FLL. These different aspects comprise noticing differences in input, mapping form to meaning and functions, pattern internalization and the like (as proposed by Skehan, 2002). A model of language aptitude or language-related aptitudes should thus be capable of answering all concerns associated with these points. Recently, it has further been suggested that there are distinct cognitive aptitudes for explicit and implicit learning, two essential aspects of FLL (Granena, 2013, 2016; Linck et al., 2013). Whereas the first refers to the process of consciously learning rules (e.g., grammar), the second refers to learning regularities and rules without awareness or intention (Hulstijn, 2005). The second is also closely associated with statistical learning, a mechanism operating on many levels, like word learning (K. Smith, Smith, & Blythe, 2011), speech segmentation (Saffran, Aslin, & Newport, 1996) or phonetic learning (Maye, Werker, & Gerken, 2002). Doubtlessly, implicit and explicit learning are measured in both the MLAT and the LLAMA but in different extents in the different tasks. Furthermore, it remains unanswered whether the two learning abilities, explicit and implicit learning aptitudes, interact, overlap or are related to teach other at different levels and stages of learning. Such domain-general abilities, also including WM, need to be considered when talking about foreign language aptitude and the various abilities underlying this umbrella term.

Reiterer (2019) nicely summarizes the current state by claiming that:

1) Yes, there is such a thing as aptitude for learning a second language, and yes, it is important; 2) it is not everything, i.e. quantity and quality of practice, for instance, are also very important factors; 3) different aptitudes play a different role for different elements of language under different circumstances and at different stages […].
To summarize, the present work can only marginally help us deepen our knowledge and understanding of the concept of language aptitude. In the present thesis, various language-related abilities were assessed, among them speech perception, speech production or pronunciation, vocabulary learning, and an ability to acquire rules of syntactic, phonological or semantic nature in a language. While some tests correlated with each other and/or contributed to the classification of high and/or low aptitude individuals, many of the administered tests failed to show any form of relation to other abilities. In other words, although all tests were chosen as measures of language aptitude(s), they measured very different facets or notions of language aptitude, if not even completely separable constructs. In other words, the two presented studies confirm that more research on the various abilities related to FLL in different contexts and at different stages of FLL is needed to further advance our theoretical models of language aptitude. It will further be essential to investigate the interplay between these abilities and in how far they depend on other skills, such as WM capacity or explicit/implicit learning.

7.3. Behavioural Correlates of Language Aptitude

The neuroanatomical correlates of foreign language aptitude were discussed extensively in 7.1. and it is high time to move on to the behavioural correlates of language aptitude that were found in the two studies. This subchapter is divided according to the most significant relationships between assessed scales and language aptitude. These were WM capacity, musicality and other behavioural correlates (e.g., arithmetic skills, intelligence, school performance). First, let us briefly look at the research hypotheses again before starting an interpretation:

**Research Hypothesis 2:**

While working memory is expected to show the strongest link to foreign language aptitude given its importance for overall learning processes, also musicality and musical experience will show striking correlations with language aptitude due to the positive transfer between these abilities.
Research Hypothesis 3:

In children and teenagers between 10 and 16 years of age, there should be a strong association between academic achievement, self-reported aptitude, arithmetic competence and performance on the language aptitude battery.

First, the relationship between language and music shall be looked at and findings of the two studies will be discussed accordingly. This is important since the link has been established in the literature but could only marginally be confirmed in one of the two experimental studies. The discussion on the association between musical ability and language aptitude will be followed by an exploration of the interconnectedness and interplay between WM capacity and language aptitude, including a discussion on the potential consequences of the present findings for a concept of language aptitude. In the last part of the behavioural subchapter, the focus will lie on the findings of the younger population, in which academic achievement, arithmetic competence and linguistic background were included in the behavioural assessment. The major aim therein will be to explain the strong connections between these concepts and language aptitude.

7.3.1. The Relationship between Music and Language Aptitude

Previous studies have highlighted the strong, positive relationship between musicality or musical experience and FLL, a finding that cannot be confirmed in the present studies. In a recent review addressing the strong link between the two, Turker, Sommer-Lolei, & Christiner (2018) not only thoroughly investigated their relationship on various levels, but they also focused on the potential of including musical training in language learning classrooms given the potential high benefit for learners. The present studies, however, do not corroborate findings from earlier studies and question the often-suggested intense transfer of musical skills and training on the acquisition of foreign languages. Doubtlessly, it could be that factors like test choice, sample size, correction of statistical results and the like, could have led to null-findings. Nonetheless, the results of the two studies deserve to be addressed separately and in comparison, including a discussion of the option that language and music do not always positively interact like earlier suggested.
In Study 1, the expected relationship between language aptitude, musicality and musical experience/training could only be partly confirmed. For the correlational analyses, five tests which were supposed to measure language aptitude were included, namely the Hindi task (measuring speech imitation), MLAT III (measuring phonetic ability), MLAT IV (measuring grammatical sensitivity), MLAT V (measuring vocabulary learning) and the English pronunciation assessment (measuring foreign language proficiency in a second language). According to the hypothesis and the earlier suggested link between musicality and language learning, a significant correlation between all language ability measures and the number of instruments, as well as with the two measures of musicality, rhythm and pitch perception (AMMA score) was expected. The correlational analyses only confirmed that the number of instruments correlated highly and significantly with the Hindi speech imitation score and MLAT IV. In other words, the more instruments subjects had learnt in the course of their lives, the higher their results on the Hindi speech imitation task and the higher their scores on the vocabulary learning task. Apart from that, the number of instruments also correlated significantly with AMMA tonal and rhythm scores as expected. To conclude, musicality and musical training seem to go hand in hand, suggesting that either musically gifted individuals chose to play more instruments and gain more experience in the musical domain, or those with musical training had better rhythm and pitch discrimination skills. Through the results of the present study it is unfortunately not possible to draw any causal conclusions, i.e. neither of the two competing options can be confirmed through the correlational analyses.

Regarding the link between language aptitude and the number of instruments, it is possible that the correlation is due to the fact that musical training, considered in the number of instruments in this case, positively influenced speech imitation and vocabulary learning, although the analyses failed to detect a relationship between the two latter. On the one hand, it seems logical that musical training enhances speech perception and motor control, especially since singing in choirs or professionally was included in the number of instruments. Consequently, a link between speech imitation and musical experience seems highly likely. The association with vocabulary learning, in contrast, is more difficult to explain in simple terms. One explanation is that vocabulary learning reflects school and academic performance, meaning that those with generally better grades at school and university are just better and more trained in memorizing and learning, explaining the association with vocabulary learning. Differences in musical training could also reflect socio-economic status and family
background, which could easily be linked to higher academic performance or intelligence. However, these variables were unfortunately not assessed in the first study and only those with an IQ below a certain percentile rank were excluded from participation.

The correlations with musical experience were only present for two administered language aptitude tests. This is particularly surprising given that many of the subjects played and had learnt multiple instruments, had sung in a choir and their scores on the musical ability tests were expected to confirm this (supported by the significant correlation between AMMA results and the number of instruments). The results of the correlational analysis could not confirm a strong relationship between any language-related scale and tonal perception, while the English score at least correlated with the rhythm perception task. This is not a highly surprising finding since very good pronunciation skills have been suggested to depend heavily on speech perception and auditory-motor integration.

Regarding the neuroanatomical results, it is particularly surprising then that the speech imitation task, the MLAT overall results and musical ability, combining both perception tasks, were associated with structural variation in right-hemispheric AC. Given that there was no behavioural overlap, one would hardly expect to find such a large overlap in the neuroanatomy of a specific brain region. Of course, one might argue that the limited sample size could have led to the lack of significance in the correlational analyses. Potentially, if I had included more participants in the first study, many other variables might have correlated with music-related scales. On the other hand, evident relationships like those between musical ability and musical training, or speech imitation and pronunciation skills, were also significant with that small sample size. Additionally, I would at least have expected a tendency in other results (e.g., rhythm perception and phonetic coding as measured by MLAT III) that might have become insignificant due to the correction for multiple testing. This was not the case. Therefore, we have to assume that even with a larger sample size results would have been similar. To summarize the results of the first study, there is no evidence that musical ability (tonal and rhythm perception) is significantly related to different facets of language aptitude as measured by the MLAT or the Hindi speech imitation task. Rhythm perception, in contrast, could be linked to English pronunciation skills. Musical experience was only linked to speech imitation skills and vocabulary learning.
In Study 2, no relationship between any language- and music-related variable could be found. Neither in correlational analyses, nor in the comparisons between subjects in the low and high aptitude group. This was only mildly surprising given the results of the first study, but it was more surprising given how many of the children had started to receive musical training at a very young age. Many of the children had been part of the longitudinal JeKi ('Jedem Kind ein Instrument' – translated to English as: an instrument for each child; a longitudinal study led by the research group of P. Schneider in Heidelberg) study, in the course of which they received extracurricular musical training and could choose to learn instruments at very young ages already. It was expected that the early musical training might have provided an additional advantage for children with musical experience since musical training, even if just for a couple of months, would lead to very profound changes in the brain of a child and lead to advantages in auditory and phonological processing, positively influencing FLL processes. Nonetheless, there were no significant relationships between any music- and language-related scores and there was no link of musicality with AC morphology on the neural level. Consequently, the often suggested highly positive transfer of musical abilities to language aptitude could not be corroborated by the findings of the second study with children and teenagers.

As already discussed in the discussion section of Study 2, various reasons could be responsible for the lack of findings. The AMMA musicality test is suitable for both teenagers and adults. However, a minor limitation might be that the AMMA test is suggested to be used with children aged 13 years or older (Gordon, 2004) but a handful of children were between 10 and 13 in the second study. Thus, one might argue that the AMMA might not have been the ideal choice for the younger children. I chose to neglect this issue for the sake of being able to compare results between the two groups and because of the very small number of children younger than 13 years of age. Therefore, the AMMA test was also administered to the children between 10 and 13 years since the IMMA (Intermediate Measures of Musical Audiation; Gordon, 1984) is generally slightly easier, scores are interpreted differently and a comparison with the other children who performed the AMMA would have been very difficult. In retrospect, administering the AMMA to all children was most likely the best choice, even if the test choice might marginally have affected results of the second study.

In the adult group, the number of instruments correlated with the Hindi speech imitation score and the vocabulary learning test. Adults played between zero to three instruments, while
two instruments were already the exception rather than the rule in the child group, where children mostly played no or just one instrument. This suggests a large difference in musical experience, which is further supported by the fact that adults had mostly played the instruments for very long periods of time, years or even decades. Children, on the other hand, had spent very little time playing instruments or had only tried them for shorter periods of time (less than a year). This could explain why there might be potential differences concerning the results with musical training and experience. Nonetheless, not only musical experience was limited, but also performance on language-related tasks was different. The children overall performed much worse on the Hindi task and consequently, there was minimal variability. In other words, not only the lack of instruments but also the lack of variation in the Hindi score could have been responsible for the null finding regarding the relationship between linguistic and musical abilities. The second significant correlation of the adult study linked rhythmic ability to English pronunciation. In the child study, English pronunciation skills were not assessed, primarily because children were not highly proficient in English and their pronunciation would have been expected to vary considerably due to school type, years of training, experience abroad, number of lessons per week and the like. In the child group, pitch and rhythm perception did not show any relationship to any of the other tasks assessing phonetic memory, sound-symbol learning, vocabulary learning and grammatical sensitivity. Given the much larger sample size in the behavioural tests (64 children and teenagers), obvious relationships would be expected to show very striking results, even after correction for multiple comparisons.

To get back to the main topic, there are three possible explanations for the observed lack of association between language and music in the present studies: (1) the two skills are more or less independent from each other with potential, but rather coincidental overlap (e.g., Stokes, 2001, failed to find correlations between music training and foreign language acquisition); (2) musical training or high musicality do not necessarily lead to a positive transfer to linguistic skills, meaning that it can enhance the learning of foreign languages but it is not needed for the development thereof; (3) there is a link between the two domains that just could not be corroborated in the present studies.

In the present studies, I can partly disprove the first option due to the results of the first study but I have to consider the possibility that they are independent due to the results of the second study. It could well be that the ability to acquire foreign languages is not necessarily associated with an individual’s music ability and/or musical training. Most studies looking into
the positive transfer of musical abilities and language learning studied the perception of speech- or music-related cues, like pitch perception (Besson et al., 2007; Moreno et al., 2009; Schön et al., 2004) or the processing of non-native speech sounds (Pastuszek-Lipinska, 2008), or segmental and tonal contrasts (C. Marie, Delogu, Lampis, Belardinelli, & Besson, 2011). Also the production of speech, that is pronunciation skills, in another language have been linked to musical ability (Milovanov et al., 2008, 2010; Posedel et al., 2012). The tasks administered to the children, however, comprised very different abilities compared to the aforementioned. According to the second option, it could have been that the highly gifted language learners of the second study did not need musical experience or musicality to develop the outstanding language skills but might possibly have profited from them if they had any. I want to emphasize that it is possible that there are relationships between language ability and musical ability and experience that I was unable to detect in the second study due to methodological constraints or for other reasons.

To summarize, language and music have been long known to be intricately linked, with musical training and high musical ability usually leading to better FLL skills both in the receptive and productive domain. The present studies, however, could not corroborate these findings. In the first study, music and language were moderately linked, also on the neural level, while there was no association between the two in the second study. Various explanations have been discussed and but the possibility that there is no link had to be explored as well. This is necessary given that there is a considerable publication bias and a replication ‘crisis’ (see 7.3.1.1).

7.3.1.1. Replication Crisis and Publication Bias

With regard to the negative findings concerning the relationship between language and music in the present studies, I would like to point out two essential points. In the past years, many well-known researchers stated on platforms like twitter that they were upset with the publication bias (a topic, that has been documented in the literature for decades) and the so-called ‘replication crisis’. Both are worth discussing given the large impact they have on topic and research choice, pressure on researchers and the so-claimed truth of what is being published.
The publication bias refers to the actual state that mostly only (positive) results are encouraged to be published by journals, editors and reviewers. In other words, if your study does not have any (positive) findings providing the world with new, detailed information on a subject (e.g., there is no link between language and music), you will struggle to find a journal with a high impact factor willing to publish your study. The publication bias is also caused by researchers who feel that their negative findings are not worth publishing and are therefore less likely to attempt publication. Joober, Schmitz, Annable, & Boksa (2012) explain the existence of the publication bias as follows:

In large part, it is the highly competitive environment for funding and career promotion that incites researchers to submit predominantly positive results for publication, knowing that they are more likely to be considered for publication by editors, more favourably reviewed by peers and, once published, more likely to be cited.

A consequence of the publication bias is that only (positive) findings are well-published, and thus more frequently read and cited. As a result, similar studies that could not corroborate these findings are more likely to be neglected and discouraged from publication, especially in well-read journals. Looking at the relationship between language and music, for instance, you can find thousands of articles written about their strong link, but what about the potential thousands of studies that could not find such a link and either neglected these findings in their publication or did not succeed in publishing at all? I cited numerous studies highlighting that musical training and experience have positive effects on first and FLL. Theoretically, it might well be the case that for the number of studies that revealed a positive relationship, the exact same number of studies was not published or published in unknown and less-read journals finding no relationship between music and language at all.

The same applies to the so-called ‘replication crisis’, which basically just describes the current sate that originality and innovation are demanded, not only when writing grant applications, but also for publishing in good journals. In other words, if you do something that has already been done a couple of times, regardless your results, it is sometimes not considered original enough to be well published or funded. This is a major issue since no single study can reveal the truth and if there are other studies using exactly the same methodology and coming to different results, we should challenge the validity of and the assumptions made by the first. However, given that the second was not applying new methods and unravelling new facts, it
might be published in a less known journal with a minimal impact factor and thus will not reach such a wide audience.

Both topics need to be addressed at this point since the second study I conducted is partly a replication study and has null-findings with regard to the relationship between language and music. My point in raising awareness for these two issues here is that academic research, as promising as it often sounds, is also subject to major issues, such as the two aforementioned. Clearly something needs to be done to promote the presentation and publication of null-findings and encourage researchers to replicate studies in other to find out whether their already achieved results are replicable and thus valid.

7.3.2. Working Memory as the Most Important Correlate of Language Aptitude

The results of the first study with the adult population showed very weak links between WM and language aptitude scores in the form of Hindi speech imitation ability, English pronunciation and the MLAT. Digit span forward correlated highly and significantly with the Hindi score, not with any of the MLAT scores, however. Digit span backward also correlated with the Hindi score, although the correlation became only a trend after applying correction for multiple comparisons. Non-word span, the last WM scale, only correlated with the Hindi score but not any other score measuring language aptitude. Overall, the link between WM scales and the Hindi speech imitation test were not surprising as the latter was very similar to a non-word span test. Still, there were no significant correlations with overall English pronunciation ability or the MLAT tests, which surely came as a surprise.

Interestingly, WM scores in the adult study also did not correlate with one another. This is astonishing given that they measure very similar components of WM. The results of the PCA, on the other hand, clearly showed that the tests assessed similar, connected constructs. The three components as extracted from the PCA were: musical ability (AMMA tonal, AMMA rhythm and the number of instruments loaded heavily onto this component; number of languages, grammatical sensitivity and non-word span loaded weakly onto this component), language aptitude (all MLAT tests and the English pronunciation score loaded heavily onto this component; the Hindi score loaded moderately onto this component) and WM capacity (the Hindi score, non-word span, digit span forward and digit span backward loaded heavily
onto this component; the number of instruments loaded weakly onto this component). According to the PCA, all three WM tasks and the Hindi task measured the same underlying construct, namely that of WM capacity. Considering the results of the PCA and the lack of correlations in the statistical analyses, it might well be that case that the small sample size limited the power of statistical analyses.

In the second study, statistical analyses pointed towards a very strong association of language aptitude with WM. In the sample of children and teenagers, the three WM scales were all significantly linked to each other and two of them correlated with language aptitude as measured by the LLAMA. Before correction for multiple comparisons, all LLAMA tests correlated with digit span forward and three out of four with digit span backward. Unfortunately, after statistical correction these results only remained tendencies with p values slightly above .05. The only WM scale that failed to show any correlations with LLAMA B, D, E and F was non-word span. This was particularly surprising since the Hindi speech imitation score was positively correlated with overall language aptitude and was thought to be very similar to the non-word span test. While there is a high similarity between the Hindi speech imitation task and a general non-word span test (see results of study 1), they might measure language aptitude and WM to a different extent. Specifically, the lack of increasing complexity and the fact that the syllables that have to be repeated in the Hindi test do not follow the phonotactic rules of German, could have led to the observable difference between the two. Also, and this has been discussed extensively in the discussion of Chapter 6, the Hindi test focused on native-like pronunciation which is rated by native speakers in comparison to non-word span, in which the accuracy of pronunciation is not as strictly considered as in the Hindi rating. Apart from these findings, digit span forward also correlated with self- and parent-reported aptitude, meaning that the higher parents and children rated children’s ability to learn foreign languages, the higher their performance on digit span forward was.

The results of the PCA in Study 2 are partly different since mathematical scores were also included in the analysis. In the PCA, all assessed scales (sub-tests included as separate entities) were included and the analysis extracted three major underlying components as measured by all 17 scales. These three components were mathematic abilities, language aptitude/WM and musical ability. This finding was particularly interesting since the second component combined two separate components from the first study. However, all WM tasks and all language aptitude tasks, including the Hindi score, loaded heavily (loadings >.493) onto
the second component, clearly pointing towards an intricate link between all of them. Furthermore, when comparing the differences in performance on measured test scales between high- and low aptitude children, all three WM scales differed significantly according to group status. Thus, it seems that WM is much more closely linked to language aptitude in the younger population (see also Hu et al., 2013, who found that WM was more involved in less proficient learners) than in the adult group, a result that is worthy of discussion.

The exact same tests for WM were used in the studies with children and adults, only the language aptitude tests used were different. While the same stimuli were used for the Hindi speech imitation (with the only difference being a lower number of raters for the rating procedure), the LLAMA test is slightly different from the MLAT. Although it was built in analogy to the MLAT (Meara, 2005; Rogers, 2017), it assesses language aptitude language-independently and therefore does not require knowledge in any language as prerequisite for being able to perform well on the test. The limitation of the MLAT was certainly that it was in English, but this should not pose a major problem as good English pronunciation is expected to be a result of high language aptitude. To conclude, if we agree that the slight differences between the language aptitude tests might have only played a minor role, it seems worthwhile to discuss the option that WM is differently related to language aptitude in the two investigated populations.

It has been suggested that WM increases from child- to adulthood and peaks at around 30 years of age (Alloway & Alloway, 2013). This difference was further tested (not presented in the results section, however) with the nonparametric equivalent of an independent t-test, the Mann Whitney U test, with all participants of the two studies (30 in the adult study, 64 in the child study) confirming that all three WM scales differed significantly between age groups (for an easier understanding, the mean scores are provided as well). There was a significant difference in performance in digit span forward (child group: $M = 7.3 \pm 2.2$; adult group: $M = 9.6 \pm 1.9$; $U = 396, p < .001$), digit span backward (child group: $M = 6.8 \pm 2.0$; adult group: $M = 8.8 \pm 2.1$; $U = 439, p < .001$) and non-word span (child group: $M = 4.9 \pm 1.6$; adult group: $M = 7.6 \pm 1.7$; $U = 222, p < .001$) between adults and children/teenagers. The results confirmed that WM increased with age and combining this information with the findings present above, it could be concluded that WM might play a different role according to the level of the learners or the stage of learning. When skills are developing, such as in childhood and adolescence, WM could be much more important than at an older age when potentials are expected to have
already developed into talents. Therefore, an explanation why WM was much more closely associated with language aptitude in the child group might well be that it is of greater importance in younger learners, explaining the lack of correlations found in the adult study. This is in accordance with suggestions from Wen (2016), reporting that effects of phonological WM might be most obvious in the case of either very young subjects, less proficient learners or adult learners in early stages of the FLL process (refer to Hu et al., 2013; Hummel, 2009; O’Brien, Segalowitz, Collentine, & Freed, 2006). Consequently, the influence and importance of WM might be dependent on learners’ age and proficiency. In the adult study, all subjects were experienced language learners, explaining why WM might not have been that important at their stage of learning. Possibly other abilities, like meta-linguistic awareness come into play in later stages of learning and partly replace the significance of WM for foreign language learning then. In the younger population, on the other hand, WM could be much more important fostering other abilities like implicit learning, attention and much more, explaining the significance of WM in early foreign language learners.

Many researchers have argued that it is quite a challenge to integrate a major component like WM into the overall concept of language aptitude while still upholding the view that language aptitude is mostly innate and cannot be enhanced through training and language experience. Findings as to whether WM trainings can really enhance other cognitive abilities, like intelligence or executive functioning, are mixed (Melby-Lervåg & Hulme, 2013; Melby-Lervåg, Redick, & Hulme, 2016). Shipstead, Redick, & Engle (2012), for instance, claim that WM trainings need to be applied with a wider range of tasks to guarantee that results are not merely due to task specific learning, a limitation encountered by numerous studies. Therefore, it is not fully understood and confirmed to date how and if WM really improves with training and has effects on other cognitive abilities through this enhancement. In models of first language acquisition (see Wen, 2016 for a summary), WM has gained increasing importance as a key mediating factor in various frameworks. Wen summarizes that phonological WM has been shown to be involved in major acquisitional and developmental aspects of first language acquisition (vocabulary, possibly also morphosyntax), while executive WM has been more closely related to processing-oriented activities (e.g., pronoun detection or ambiguity resolution). Apart from first language acquisition, the involvement of WM in L2 learning has been proved in numerous studies, suggesting that it has even greater effects on the acquisition of foreign languages and their processing.
7.3.2.1. A New Concept of Language Aptitude?

The two studies revealed that high WM is linked to high language aptitude in children and teenagers who are learners at an early stage of foreign language acquisition, while a strong link between the two could not be fully confirmed in the adult population. What are now my conclusions as to the significance of WM for language aptitude?

In the most recent book on language aptitude edited by Wen, Skehan, Biedroń, Li, & Sparks (2019), the authors argue that the contemporary research agenda focusses on disentangling and elucidating potential cognitive and neuropsychological aptitude constructs that are supposed to facilitate foreign language acquisition, real-time processing, and long-term linguistic development. Future directions include the development of aptitude tests that measure different aptitudes/components of aptitude, including implicit and explicit learning and suitable for different stages of language development (dealt with in the chapter of DeKeyser). Furthermore, they explain how individual differences might be implemented in the overall theoretical framework, possibly informing teaching paradigms (dealt with in the chapter by Vuong and Wong). Looking at these future research agendas, it becomes clear that despite the joint efforts undertaken by language aptitude researchers all over the world, there are numerous unanswered questions. One of the major discussion points and current debates is the involvement of WM in a construct of language aptitude and the lack thereof. In Skehan's (2002) processing stages, WM is essentially involved in various levels of L2 processing, suggesting that there is no way to change the current theoretical framework of language aptitude without taking WM as a component into consideration, which has been strongly supported by Wen and Skehan in following publications.

Skehan and Wen have consistently pointed out that language processing in the human brain requires WM by linking language aptitude to second language learning processes. Still, high WM alone does not make a gifted language learner and conversely, high language aptitude does not automatically require high WM. Additionally, there is still no consensus on the specific components of WM and how they are integrated in the second language learning process. In his parallel architecture, for instance, Jackendoff, (2007, 2010) proposed that linguistic WM consists of three subparts, namely a phonological, a syntactic and a semantic WM component. While this was an interesting proposal, Jackendoff failed to give hints as to how to operationalize the components in the overall framework. The major problem we face is
that from our current knowledge, we cannot determine the exact amount of variation in individual learning behaviour that might be explained by WM (Juffs & Harrington, 2011).

Based on the findings of the study and my own conceptions and ideas of language aptitude, I would like to point out two obvious solutions to the dilemma of including WM as a major component of language aptitude. The first would be to consider WM a prerequisite for all learning processes, meaning that it does not only operate on the linguistic level, but is required for any form of learning – which is most likely already the case. As such it would not necessarily have to be included in the concept(s) of language aptitude(s) as it is not specific to language aptitude(s) but operates on a very different level. The second option would be to consider WM an integral part of SLA/FLL and include it either as a separate component, equal in status to other components like phonetic coding ability or language analysis ability. Another possibility for this inclusion would be to make WM an integral aspect of each component according to the role it plays for each of the components. For instance, it could be a highly significant part of the ability to decode speech input (phonetic coding ability), while it would be only one of many factors relevant for language analysis ability (along with meta-linguistic awareness or implicit learning, for example). A major shortcoming of the second option is that if we allow WM to become part of a construct like language aptitude, I am afraid that the same has to be done for all other learning-related constructs, from arithmetic skills to general intelligence to executive functions. Likewise, knowing that WM is differently involved at different learning stages, the concept(s) of language aptitude(s) could not be entirely the same for early or late stage learners then. Consequently, an integration of WM would be a quite challenging undertaking. It remains a domain-general skill that is doubtlessly substantially involved in FLL through phonological and executive WM aspects, but there is a reason why researchers have been fighting for the inclusion of WM in a model of language aptitude with only partial success to date.

To put it in a nutshell, WM is doubtlessly one of the most important human cognitive capacities that makes it possible for us to learn languages, play instruments, speak, sing, do calculations, solve problems and so much more. WM has been at the core of so much research in the past decades, specifically in the context of developmental disorders. Doubtlessly, it deserves the attention it has been getting but, in my view, WM is a highly significant prerequisite and not a core component of language aptitude. For me, language aptitude is what the brain makes out of the endless possibilities we have of storing and creating speech and
language in- and output. It is a successful combination of various different cognitive abilities, from problem-solving to executive functions, analytical ability and so many more – not just WM alone. All these certainly play a role in the development from language aptitude to language talent. Still, we need to conduct more research to shed light on the influence of phonological and executive WM on different aspects of language aptitude in different learning contexts and at different learning stages before finalizing our view of WM as an integral part of language aptitude.

7.3.3. School Performance, Arithmetic Skills and Self- and Parent-Reported Aptitude

Children who performed significantly better in the language aptitude testing also had better school grades in English, their second language, and German, their first language. Moreover, they were learning more languages at the time of testing and considered themselves as better language learners. Since the results were already discussed in detail in Chapter 6, and no comparisons can be made to the adult study, I will only briefly comment on the findings again and discuss them in the overall context of language learning.

In the child study, children and their parents were asked to report on their perception of their own (done by children) and their child’s (done by a parent) ability to acquire foreign languages. It was striking that the relationship between parents’ and children’s estimations was higher than any other correlation found in the whole study, although parents and children were asked independently from each other and no interaction on that topic had occurred before the interviews. The participants and their parents had only been informed that language-related tests would be included in addition to the musical tests and the MRI session, but they had not received any information on the specific tasks that would be applied. The correlations were confirmed in the group comparisons, which showed that those children with a high aptitude for learning foreign languages were also those who had reported themselves as more gifted and those who had been reported as more gifted by their parents. These group comparisons suggest that children and parents did not only report their aptitudes based on the school grades they had achieved, but most likely to their individually felt potential for learning foreign languages, which was then reflected in their performance in the LLAMA test battery.
Why did I include self- and parent-reported measures in the younger population? My main intention behind including these questions was to see whether children were aware of their ability for acquiring foreign languages. Secondly, I was interested in seeing whether their estimations would also reflect on their performance in the language aptitude battery and in their performance in language subjects at school. Last, I wanted to verify whether children’s perceptions and those of their parents would match. The results with these two variables suggest that the LLAMA language aptitude battery is a good measure of language aptitude since those with high aptitude scores in the battery also thought that they possessed a high potential for the acquisition of foreign languages. Additionally, we can learn from the findings that school grades, even if not the best measure for assessing language ability, do reflect language aptitude and are related to children’s perceptions of themselves. In other words, it seems that school grades, even if highly dependent on the school, the teacher and the classroom, are not such a bad indicator for language learning potential since those children who think of themselves as better learners also have better grades.

There was also a link between the number of foreign languages learnt and the language aptitude scores. This highly positive relationship between the number of foreign languages being learnt and the achieved language aptitude scores can be explained by two hypotheses: (1) Children who (know that they) learn languages effortlessly and quickly, know about their language learning potential through the ease with which they have learnt English, for example, and therefore they make the decision to learn more foreign languages (be it at school or through extracurricular activities); (2) Children who have acquired or learnt more languages might have profited from the language classes to an extent that their meta-linguistic awareness was increased and therefore they scored higher in the language aptitude testing. Both possibilities are equally likely, although I would personally speak in favour of the first. However, the reader shall be reminded that these are mere hypotheses based on correlational findings and no causal evidence can be deduced.

One interesting finding with regard to academic achievement is that it correlated with self- and Parent-reported aptitude, but not with any of the measures of language aptitude. In fact, school performance did not correlate with any language aptitude measure, which is surely surprising. The biggest surprise being that school grades and the number of languages learnt by a subject did not show any correlation with LLAMA D, the vocabulary learning task. One would expect that the more languages you know, the more likely you are to have developed a
good strategy for vocabulary learning and the better your grades in English are, for instance. This was not the case, however. Still, the comparison between high and low aptitude learners revealed that high aptitude learners have generally significantly better grades in English and German.

The correlational analyses of the younger population further revealed moderate links between WM scales, measures of language aptitude and arithmetic skills, which were supported by the results of the comparison between high and low aptitude learners. Arithmetic fluency was linked to vocabulary learning and there were tendencies of a relationship with LLAMA D and E as well. The comparison between high and low aptitude learners further revealed that more gifted language learners performed significantly better in the arithmetic fluency test and both skills were significantly linked to WM. The connection between performance on the arithmetic fluency test and the language aptitude battery stands in some contrast to the common belief that language and mathematical skills do not necessarily go hand in hand. The most likely explanations for the found connection between arithmetic skills and language aptitude, as well as with WM, is that WM is an essential shared foundation of these two abilities. This explanation is in accordance with intelligence research that has shown that high WM capacity is a vital part of domain-unspecific general intelligence (Neubauer & Stern, 2013). Given these findings, it is thus not surprising that WM could be a fundamental factor associated with both skills as they are highly dependent on overall learning. Another possibility is that both arithmetic skills and language aptitude show large overlap, possibly even beyond the mere contribution of high WM capacity to both. Overall, analytic abilities could also aid both skills and contribute to the development of linguistic and mathematical skills.

One very interesting aspect that has only been marginally addressed was that the high and low language aptitude groups differed neither in intelligence, nor in socio-economic status. Although these variables (IQ, socio-economic background) were only taken into account to exclude the possibility that high language aptitude merely reflects high IQ or high socio-economic status, the results deserve some discussion. The overlap between intelligence and language aptitude has been a matter of debate (see 2.3) and some researchers have suggested that the overlap largely stems from the fact that intelligence tests often include tasks that test language-analytic knowledge (Sternberg, 2002). Others, however, found that there was a large overlap between language aptitude and intelligence, also in the absence of those tasks. The second study suggests that high aptitude learners are not excellent language learners because
of their high intelligence, nor do they have an advantage in the form of a high socio-economic status. Therefore, the study clearly supports that claim that language aptitude is largely independent from general intelligence (Sáfár & Kormos, 2008; Kiss & Nikolov, 2005) and it further proposes that socio-economic status is not critical for high language aptitude. While certainly a large variety of environmental factors contribute to FLL, socio-economic status does not seem to be one of them.

To sum up, more gifted language learners had better grades at school, thought of themselves as more gifted, were rated as more gifted language learners by their parents, and possessed higher WM capacity and better arithmetic skills. The correlational analyses, on the other hand, further revealed that the number of languages was significantly associated with some of the language aptitude measures, as well as with self- and parent-reported language aptitude. Last, neither intelligence nor socio-economic status could predict high language aptitude.

**7.4. Implications of the Present Studies for Education and Science**

While the implications and future research directions of the present thesis are more or less clear for science, it is difficult to say whether education can profit from the findings as well. Therefore, I will begin with the easier task, namely briefly mentioning future directions for research. Subsequently, some potential implications and future direction for education and language learning shall be given.

Future directions for research in the cognitive neurosciences are to address the functional and structural connections of AC to frontal and parietal regions to further explore the role of HG duplications. It is also vital to further determine the specific involvement of HG duplications in phonological and auditory tasks, also with the aim of specifying whether and to what extent they are involved in primary and secondary auditory processing. Additionally, we need to explore the roles of right and left HG in larger groups of subjects with high or low language aptitude(s) or with regard to specific FLL-related skills. The argument that HG morphology is a stable neuroanatomical trait can only be confirmed through a longitudinal
study that continuously measures GM volumes and investigates HG morphotypes from child-
to adulthood in the very same individuals.

I am afraid that the practical applications gained from my research are minimal but I hope that the theoretical knowledge certainly helps us better understand language aptitude and the significance of WM for the development of language aptitude to language talent. Looking at the results of the LLAMA test, it becomes clear that certain children and teenagers are far better language learners than others without even receiving explicit instruction of how to learn. It could be that these children’s analytical abilities and their implicit learning aptitude are much higher than of those children who received very low scores in the language aptitude test. LLAMA overall results were neither associated with intelligence, nor socio-economic status or age. This means that older and smarter (i.e. with higher general intelligence) children, as well as those with an outstanding socio-economic background, did not perform better compared to those who were younger, had a lower IQ and a lower socio-economic status. This is a rather new finding and supports the validity of the present study highlighting that language aptitude is independent of many factors, such as socio-economic status and general intelligence. Apart from these conclusions, I argue that implicit learning mechanisms are highly important for all learning processes and need to be fostered as soon as possible in children. Future research should be aimed at looking more into implicit learning mechanisms underlying language aptitude and confirming that language aptitude is not dependent on IQ and high socio-economic status – potentially this also applies for FLL on a more general level.

Doubtlessly, developing a high proficiency or an expertise in a specific domain is highly dependent on an individual’s motivation to learn. There has been little research so far exploring to which extent motivation depends on predetermined or environmental influences, but research in the psychology of motivation has shown that the wish or desire to expand one’s competence (intrinsic, positive motivation) is far stronger than extrinsic, negative motivation, e.g., fear of failure or critique (Winner, 1996). It is often assumed that motivation is a personality trait that strongly impacts learning behaviour, although the question of the interaction between aptitudes and stamina, as well as motivation, remains unanswered to date (Seither-Preisler and Schneider, 2015). However, research so far has shown that children with an aptitude in a specific domain are often highly motivated regardless of their environment, and their motivation leads to hard work, resulting in better performance, leading to more hard work (Winner, 1996). In their neurocognitive model on musical aptitude, Seither-Preisler and
Schneider (2015) assume that aptitude leads to heightened intrinsic motivation to learn, resulting in learning-induced plasticity. According to their model, a predisposed aptitude in the form of structural properties in the human brain is coupled with increased neuronal efficiency, and only together and with the right environmental stimulation, a talent develops. Since I very certainly agree with this idea and the model itself, my conclusions are very simple in this regard – if children are intrinsically motivated to do something, it is very likely that there is a reason why they are so motivated. I would thus suggest to foster intrinsic motivation through giving children the option to spend their time with whatever they are most interested in, be it singing, playing an instrument, learning a language, doing sports or being creative (just some of the many options). It would also be highly interesting to look into these specific intrinsic motivations and when they develop.

Moreover, I think that people generally need to update their views on learning, and everything associated with learning. Just recently, for instance, a paper has been published on general beliefs about learning styles – concluding that 95% of people believe that people learn best according to the learning style that suits them best, which is particularly interesting given that no experimental research so far has shown that learning styles are really fruitful at all. Thus, I highly recommend reading the article “Maybe they’re born with it, or maybe it’s experience: Toward a deeper understanding of the learning style myth” (Nancekivell, Shah and Gelman, 2019). The authors claim that there are numerous myths about cognition that need to be addressed. If you have children yourself, you might agree instantly with this claim, especially since there is a lot of pressure involved in educating your child to its full potential nowadays. This means that they must learn instruments as early as possible, attend extra-curricular activities, sports classes or learn as many foreign languages as early as possible because these activities have been claimed to be associated with the one or the other cognitive advantage (e.g., better executive functions, higher WM). However, many of these cognitive myths could only be partly confirmed in research so far and further studies need to be conducted to further investigate their truth. Especially, studies need to include larger samples and well-known studies need to be replicated with clean methodologies to explore the validity of many studies.

Regarding language aptitude, it will be essential to invest more years of intense research to determine the core abilities and cognitive traits linked to the concept. Of course, this includes more research on the involvement of WM components in language aptitude, or for different
other overall cognitive abilities related to language aptitude. It is very hard to propose research related to language aptitude if there is still disagreement on what language aptitude actually includes and whether there are multiple aptitudes that need to be set in a relation in a model. I am afraid that I have no answer to that problem, but I hope that more experienced and wise aptitude researchers will come up with excellent ideas about how to update our current concept of language aptitude(s) with regard to the most recent findings in the field.

To conclude, the findings of the present thesis can only tell us that we should scrutinise our beliefs on cognition and the brain and only believe in results that have been replicated in studies with large sample sizes and clean methodology. They also highlight that language aptitude is linked to arithmetic skills and school achievement, while at the same time not being dependent on high intelligence and/or high socio-economic status. What is vital, is an individual’s motivation to learn, be it because of a genetic predisposition, external influences or many more. There is no learning without motivation and if we look at the way schools function, there is often very little space for developing a true motivation. Future research should address many problems that have been pointed out continuously in this thesis, such as the involvement of WM in a model of language aptitude or the concrete functions of HG duplications, just to name two examples.
8. Conclusion

The present studies were conducted with the aim of detecting potential neuroanatomical correlates of foreign language aptitude in primary and secondary ACs of children and adults. Another major aim was to further unravel the behavioural relationships between WM capacity, musical ability and an individual’s potential for the acquisition of foreign languages. The results of the two presented studies suggest that right AC morphology, i.e. gyrification and GM volumes of right AC, are linked to language aptitude with the occurrence of single gyri being linked to low aptitude and the possession of more than one single gyrus (CPDs and multiple gyri) being an indicator for high language aptitude. Although the specific functions of duplications have not been thoroughly explored to date, possessing multiple gyri or complete posterior duplications in your AC could play an important role in structural and functionally connectivity linking auditory areas to frontal-motor and inferior parietal-multisensory regions responsible for phonological, semantic and syntactic analyses, as well as speech production. The presented studies are the first to suggest a strong link between gyrification in right AC and GM volume and language learning skills.

On the behavioural level, the results indicate an essential involvement of WM in the acquisition of any foreign language, in particular in children and teenagers still at the earlier stages of language learning. The interplay between WM and language aptitude in adults, on the other hand, was only moderate as other variables such as meta-linguistic awareness could contribute to performance in FLL situations at later stages. With regard to the overall results, it seems plausible to integrate phonological WM in a new language aptitude(s) model and further explore the extent to which WM shapes early language learning in contrast to stages of high achievement or proficiency. While language aptitude and WM were intricately linked, the previously suggested and often-confirmed interconnectedness between language learning and musical ability, on the other hand, could not be corroborated through the findings of the two studies. Overall, only musical training was linked to speech imitation ability in adults. Especially in the younger population, musicality and musical training failed to show correlations with any other assessed skills and were not involved in the differentiation between high and low aptitude learners. Language aptitude, on the other hand, was not only significantly correlated with arithmetic competence, but also with school performance in L1 (German) and
L2 (English) and self- and other-reported language aptitude. Moreover, the acquisition of foreign languages at school and in extra-curricular contexts was positively linked to performance on language aptitude tests and contributed to children’ and parents’ positive view of their/their child’s capability for foreign language learning.

To summarize, further studies need to be conducted in order to verify the suggested link between foreign language aptitude and right-hemispheric AC morphology, which has not been explored in studies to date. From the present results, it seems that only right AC but not left is structurally involved in the manifestation of foreign language aptitude. There is further evidence that the neuroanatomical markers found in the present thesis (HG morphotypes and GM volumes in AC) are highly stable neuroanatomical traits that do not change due to experience or training, and could thus be genetically predisposed neuroanatomical markers for language aptitude. Behaviourally, it seems worthwhile to integrate WM into the concept of language aptitude and further investigate the link between language aptitude and arithmetic competence and school performance, which was highly significant in the present thesis. Last but not least, the often-cited link between language and music could not be corroborated in the present studies, suggesting that language aptitude and musicality or musical experience were not linked in the participants of the two experimental studies.
9. References


230


243


252


256


257


267


List of Figures

Figure 1 The development of natural abilities, termed ‘gifts’, to competencies, so-called ‘talents’, through environmental influences (catalysts) and developmental processes (taken from Gagné, 2005)..............................................................12

Figure 2 A neurocognitive model of competence development described by Schneider & Seither-Preisler (2015; English version used at conferences and presentations provided by Seither-Preisler in personal communication). The model attempts to capture the interaction between dispositional factors (i.e. a potential profile), natural maturation (right side) and training-induced plasticity (left side). Additionally, pedagogic intervention including socio-economic factors come into play and can positively influence motivation and practicing behaviour........................................15

Figure 3 Aptitude complexes as described by Robinson (2002). The most inner circle consists of 10 core abilities that contribute to higher order aptitude factors (aptitude complexes; figure taken from Robinson, 2002). The hierarchical order goes from abilities to aptitude profiles, task aptitudes and finally pragmatic/interaction abilities and traits.................................................................26

Figure 4 LLAMA B is a vocabulary learning task. Words are given in the middle (in written form) and each little figure denotes a different piece of vocabulary. In two minutes, participants have to remember as many words as possible by clicking on them and remembering the association between the new word and the symbol.....................34

Figure 5 LLAMA E tests the acquisition of sound-symbol correspondence. Each combination of a letter and a digit (e.g., 0ì, 3ì) is pronounced differently in the given, new language and subjects have to understand how the system works and learn the associations between two sounds and two symbols..........................................................36

Figure 6 LLAMA F tests language analysis ability. Each grey square presents a sentence and a picture, the sentence describing in the new language what is present and happening in the picture. ...........................................................................37

Figure 7 Anatomical areas involved in the dual stream model proposed by Hickok and Poeppel (figure adapted from Hickok, 2012). The left hemisphere is presented on the left, the right hemisphere on the right. Blue: dorsal stream (articulatory network and sensorimotor interface), pink: ventral stream (combinatorial network and lexical
interface), green: spectrotemporal (auditory) analysis, yellow: phonological network. 


**Figure 8** Lateral view of the left hemisphere of the superior temporal lobe. Colours used in the figure correspond to rough cytoarchitectonic parcellations into primary (grey), secondary fields (blue) and the multimodal integration cortex (pink; taken from Hickok & Poeppel, 2000). \textit{AG} angular gyrus, \textit{MTG} middle temporal gyrus, \textit{STG} superior temporal gyrus, \textit{pSTP} posterior superior temporal plane, \textit{aSTP} anterior superior temporal plane, \textit{H} Heschl’s gyrus, \textit{I} Insula, \textit{FO} frontal operculum, \textit{PCG} precentral gyrus, \textit{PO} parietal operculum.

**Figure 9** AC/HG from different viewpoints in the author’s brain. (a) Sagittal view of left HG, (b) axial view of right and left STG including HG, (c) 3D reconstructed AC and STG of both hemispheres; right hemisphere marked in blue, left hemisphere marked in red.

**Figure 10** Model of dual auditory processing and connections of auditory areas to the sensory system. The model was designed to close the loop between speech perception and production (red: postero-dorsal stream, green: antero-ventral stream; taken from Rauschecker and Scott, 2010). \textit{PMC} premotor cortex, \textit{CS} central sulcus, \textit{IPL} inferior parietal lobe, \textit{IFG} inferior frontal gyrus, \textit{STS} superior temporal sulcus.

**Figure 11** Schematic representations of the four main HG morphotypes: a) single gyrus, b) CSD, c) CPD and d) MD (adapted from Benner et al., 2017). \textit{aHG} anterior HG, \textit{PP} planum polare, \textit{FTS} first transverse sulcus, \textit{SI} sulcus intermedius, \textit{HS1} first Heschl’s sulcus, \textit{HS2} second Heschl’s sulcus, \textit{cHS} first complete Heschl’s sulcus.

**Figure 12** The three-component WM model proposed by Baddeley and Hitch (proposed in 1974; taken from Baddeley, 2003a), consisting of the visuospatial sketchpad, the central executive and the phonological loop interacting with each other while temporarily maintaining any incoming information.

**Figure 13** The multi-component model of WM (Baddeley and Hitch, 2000; taken from Baddeley, 2003a): the central executive controls the phonological loop, the episodic buffer and the visuospatial sketchpad, which are then linked to LTM, more specifically to visual semantics, episodic LTM and language.
Figure 14 Wen’s phonological-executive WM model explaining the relation between phonological and executive WM with regard to language and tasks associated with the two (taken from Wen, 2019).

Figure 15 According to Baddeley, the regions most involved in WM are the right cerebellum, the right dorso-parietal cortex, the inferior frontal cortex bilaterally, the fusiform gyrus in the left hemisphere and the left SMG (taken from Baddeley, 2003a). AR articulatory rehearsal, CE central executive, PS phonological storage, VC visual cache/storage, IS inner scribe (spatial rehearsal).

Figure 16 Detailed steps of the semi-automated manual segmentation in BrainVoyager QX. First, the whole brain data is pre-processed (a), and then STG including AC is manually segmented in each slice of the brain (b). After eliminating all areas outside of STG (c), AC is reconstructed as a 3D mesh displaying the whole STG including PT and HG (d).

Figure 17 The three types of HGs distinguished in the present study: single gyrus, common stem duplication (CSD), complete posterior duplication (CPD; from left to right). aSTG anterior superior temporal gyrus, HG Heschl's gyrus, PT planum temporale.

Figure 18 Factor scores for the three components of the PCA analysis (musicality, language talent and WM) compared in subjects with a single gyrus, CSDs and CPDs. Factor scores range from -1 to +1.

Figure 19 Differences in WM capacity in adults with high or low speech imitation ability. Scores range from 0 to 14 in all three tests.

Figure 20 Comparison of high and low speech imitators concerning their results on the various language aptitude scores: MLAT III, IV and V.

Figure 21 Averaged AC surfaces in the left and right hemispheres in individuals of the two aptitude groups (high Hindi speech imitation score on the left, low Hindi speech imitation score on the right). Left hemisphere (L) in blue, right hemisphere (R) in red.

Figure 22 AC morphology of all individuals classified as either high or low aptitude individuals according to the Hindi speech imitation score. High aptitude subjects are in the top row, compared to low aptitude individuals in the bottom row. Each specific type is also given for each HG. Left hemisphere in blue, right hemisphere in red.
single gyrus, *CSD* common stem duplication, +CPD complete posterior duplication

**Figure 23** Mean total Hindi score (between 0-10) for all three types of AC morphology, namely single gyrus, CSD, and CPD in right AC (A) and left AC (B) of all subjects. CPDs are clearly associated with higher mean scores in the Hindi speech imitation test, as supported by results of the statistical analyses.

**Figure 24** The relationship between HG types (single, CSD, CPD) and the overall MLAT raw scores (ranging from 0 to 120). In right AC (A), higher scores in the MLAT were linked to the possession of CSDs and CPDs. No relationship was found in the left hemisphere (B).

**Figure 25** Comparison of the mean total AMMA scores (ranging from 0 to 80) for individuals displaying the three distinguished HG characteristics, namely single gyrus, CSD and CPD. Evidently, the mean score for CPDs was significantly higher than that for the other two types.

**Figure 26** Four types of HGs distinguished in Study 2: single gyrus (Single), CSD, CPD and multiple gyri (Multiple). \(aSTG\) anterior superior temporal gyrus, \(HG\) Heschl’s gyrus, \(PT\) planum temporale.

**Figure 27** A comparison of performance of the high and low aptitude groups on all three WM tasks and the Hindi speech imitation task. Raw scores on the three WM scales ranged from 0 to 14, while the Hindi speech imitation scores ranged from 0 to 10. All comparisons are significant except for performance in the non-word repetition.

**Figure 28** An exemplary illustration of four HGs in the high and low aptitude groups respectively. For details of the four designated types, see the legend. The right hemisphere is marked in blue, the left in red. \(CSD\) complete posterior duplication \(CPD\) complete posterior duplication.

**Figure 29** Averaged HG surfaces of subjects in the high (left panel; \(N = 21\)) and low aptitude group (right panel; \(N = 21\)). High language aptitude is associated with CPDs and multiple gyri in the right hemisphere, while low language aptitude results are associated with a single gyrus in the right hemisphere. Left hemisphere (L): blue; right hemisphere (R): red.
**Figure 30** Language aptitude scores (LLAMA; range: 0-100%) for different HG types in (A) right AC and (B) left AC. From left to right: single gyrus (single), CSD, CPD and multiple gyri.

**Figure 31** Results of the linear regression analyses showing the relationship between LLAMA overall performance mean scores (ranging from 0 to 100%) and GM volumes (mm³) in right (A) and left (B) AC for all tested individuals. 21.4% of variance in language aptitude are explained by right-hemispheric HG GM volume.
**List of Tables**

**Table 1** Carroll’s components of language aptitude (taken from Dörnyei & Skehan, 2003), including phonetic coding ability, grammatical sensitivity, inductive language learning ability and associative memory and descriptions of the four components. 22

**Table 2** Skehan’s model of language aptitude focuses on SLA stages and their corresponding L2 cognitive processes. The three stages of SLA are language input, central processing and creating language output. The cognitive processes underlying these stages are segmentation<, noticing, pattern recognition, complexification, handling feedback, error avoidance, automatization, creating a repertoire and lexicalization (taken from Skehan, 2016; see Wen, 2016). ................................................................. 24

**Table 3** The five subtests of the MLAT (Carroll and Sapon, 1957), their names, and what is being tested in each part of the test. For their relations to aptitude components, see Table 4. .................................................................................................................. 31

**Table 4** The four components of language aptitude as defined by J.B. Carroll and what each part of the MLAT measures well or weakly (taken from Rysiewicz, 2008). ............. 32

**Table 5** The different subtests of the LLAMA language aptitude battery, LLAMA B, D, E and F, aiming to measure the construct of language aptitude (Meara, 2005). ............. 33

**Table 6** Important neurological terms with their corresponding definitions necessary for an understanding of neuroscientific studies (Bear, Connors, & Paradiso, 1996, pp. 23-25 & 163–175, 2015). ............................................................................................................. 50

**Table 7** Mean scores and standard deviations for all assessed scales for all participants. Scores given with two decimal places. AMMA Advanced Measures of Music Audiation, MLAT Modern Language Aptitude Test, n° number ................................................................. 107

**Table 8** Pearson correlation coefficients (rounded to three decimal places) for all assessed scales. digit span b digit span backward, digit span f digit span forward; instr. instruments, lang. languages *** significant at p<.05 after FDR correction (Benjamini and Hochberg, 1985); * significant at p<.05 before correction .......... 109

**Table 9** Rotated component matrix displaying three core components measured by all scales in the PCA: Component 1 (Musicality), Component 2 (Language Aptitude) and Component 3 (Working Memory). No loadings below .3 are shown..................... 111
Table 10 Results of students’ t-tests. Differences in mean scores (M ± SD) between subjects in the high and low speech imitators (mean scores and standard deviations are rounded to one decimal place). t-values, p-values and correlation coefficient r (effect size; three decimal places) are provided. ................................................................. 114

Table 11 Results of students’ t-tests. Differences in mean scores (M ± SD) between subjects in the high and low aptitude groups according to the Hindi speech imitation score (mean scores and standard deviations are rounded to one decimal place). t-values, p-values and correlation coefficient r (effect size; three decimal places) are provided. .... 117

Table 12 Overall distribution of HG types in right and left hemisphere (RH/LH), as compared to distribution in high and low aptitude groups according to the Hindi score. .... 118

Table 13 Results on all assessed scales, including school grades in German (first language) and English (second language). Mean scores and standard deviations provided rounded to two decimal places. L1 first language, L2 second language. ................. 150

Table 14 Spearman correlation matrix for all assessed scales. Correlation coefficients (r) and significance values (p) are provided for each correlation, rounded to three decimal places. ........................................................................................................ 152

Table 15 The results of the varimax rotated component matrix for the PCA including all 17 variables resulted in three components. All r-values below .35 were discarded in the analysis (and thus do not appear in the table). AF arithmetic fluency. ................. 154

Table 16 Differences in performance on measured test scales between high-aptitude and low-aptitude children (Mann Whitney U Test). Z- and p-values, as well as effect size (correlation coefficient r, Z-value divided by the square root of the total sample size; Field, 2013) are displayed. ................................................................. 157

Table 17 The distribution of HG types, namely single, CSD, CPD, multiple gyri, in both hemispheres with percentage values given as well. ......................................................... 160