



universität
wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

„Musicality and Second Language Acquisition: Singing and
Phonetic Language Aptitude“

verfasst von / submitted by

Mag. Markus Christiner

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktor der Philosophie (Dr. phil.)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on the
student record sheet:

A 792 327

Dissertationsgebiet lt. Studienblatt /
field of study as it appears on the student record
sheet:

Sprachwissenschaft

Betreut von / Supervisor:

Univ.-Prof. Mag. Dr. Wolfgang U. Dressler

Mitbetreut von / Co-Supervisor:

Assoz. Prof. Dr. Susanne M. Reiterer

Content

1. Introduction.....	10
1.1. Research question 1	13
1.2. Research question 2	14
1.3. Research question 3	14
1.4. Research question 4	14
2. Aptitude and language aptitude	18
2.1. Defining Aptitude	20
2.2. Historical overview of language aptitude research	24
2.3. Language aptitude tests	25
2.3.1. The Modern Language Aptitude Test (Carroll & Sapon 1959).....	25
2.3.2. Pimsleur Language Aptitude Battery (Pimsleur 1966).....	26
2.3.3. The Defense Language Aptitude Battery (Petersen & Al-Haik 1976)	27
2.3.4. VORD (Parry & Child 1990).....	28
2.3.5. Cognitive Ability for Novelty in Acquisition of Language-Foreign (Grigornko et al. 2000).....	29
2.3.6. LLAMA (Meara 2005).....	30
2.3.7. High-level Language Aptitude Battery (Doughty et al. 2008).....	31
3. Aptitude and musicality	33
3.1. Musical aptitude versus achievement	36
3.2. Measuring musical abilities	37
3.3. Historical overview of musical aptitude	40
3.4. Musical aptitude tests and measurements	41
3.4.1. The Seashore test (Seashore 1919a, 1919b).....	42
3.4.2. Measures of Musical Abilities (Bentley 1966)	43
3.4.3. Gordon’s music aptitude tests	43
3.4.4. The Musical Ear Test (Wallentin et al. 2010).....	45
3.4.5. Profile of Music Perception Skills (Law 2012; Law & Zentner 2012).....	46
3.4.6. Goldsmiths’ Musical Sophistication Index (Müllensiefen et al. 2012)	47
3.4.7. Performance tests	48
3.4.8. Discussion	48
4. Singing	50
4.1. Singing – a musical subcategory.....	51
4.2. Music and speech: perception and structural research.....	52
4.3. Singing and the general population.....	53

4.4.	Singing as a hybrid of music and speech: memory.....	53
4.5.	Research about non-professional and professional singing.....	54
4.6.	Singing research.....	55
4.6.1.	Singing development – childhood.....	56
4.6.1.1.	Caretaker input and lullabies.....	57
4.6.1.2.	Infancy stage and early childhood	58
4.6.1.3.	Middle and late childhood.....	59
4.6.2.	Voice experts (professional singing).....	60
4.6.3.	Non-professional singing	62
4.7.	Measuring Singing	65
4.8.	Terminology.....	70
4.9.	Discussion	72
5.	Working memory and (phonological) short-term memory.....	74
5.1.	Baddley`s multicomponent model	75
5.1.1.	The phonological loop	75
5.1.2.	The visuo-spatial sketch-pad.....	75
5.1.3.	The central executive	76
5.1.4.	The episodic buffer	76
5.2.	Measuring working memory, short-term memory and terminology.....	76
5.3.	Working memory and the language aptitude construction	79
5.4.	Discussion	81
6.	Performance anxiety	81
7.	Questionnaire and construction.....	84
7.1.	General features	85
7.2.	Wording and writing of questionnaire items.....	86
7.3.	The main questionnaire.....	87
7.4.	Multi-item scales, singing behaviour during childhood and performance anxiety.....	88
7.5.	Questionnaire: education and singing self-estimation	88
7.6.	Design of the questionnaires	89
7.7.	Reliability analysis of multi-item concepts.....	89
7.7.1.	Reliability analysis: singing behaviour during childhood.....	90
7.7.2.	Reliability analysis: performance anxiety.....	90
8.	Statistical analysis	91
9.	The development of the MULT AP (Multilingual Aptitude Test)	93
9.1.	Test development of the MULT AP	95

9.2.	Developing the procedure	99
9.3.	About the tasks.....	99
9.4.	Final test design	100
9.5.	Testing Procedure	104
9.6.	Participants.....	105
9.7.	Language background	105
9.8.	Educational status	106
9.9.	Musical background.....	106
9.10.	Statistical analysis	107
9.10.1.	The scoring of the main MULT AP variables.....	107
9.10.2.	Correlations.....	108
9.10.2.1.	Main scores:	110
9.10.2.2.	Subscores of the PL perception score	110
9.10.2.3.	Gender differences	111
9.10.2.4.	Highest level of education and age	111
9.10.3.	Group comparisons	112
9.10.3.1.	Age groups and speech perception.....	112
9.10.3.2.	Musical instrument playing and non-musicians.....	114
9.10.3.3.	Different types of musical instruments and speech perception ability	115
9.11.	Discussion	117
10.	Empirical investigation:methodology.....	119
10.1.	Overview of the empirical investigation.....	119
10.2.	Testing procedure and participants	120
10.2.1.	Language background	122
10.2.2.	Educational status	122
10.2.3.	Musical background.....	123
10.3.	Methods: Measuring language aptitude	123
10.3.1.	Speech imitation tasks: about the speakers	124
10.3.2.	Preliminary testing	125
10.3.3.	Procedure for the participants	125
10.3.4.	Additional rating scales: self-awareness concepts	125
10.3.5.	Raters	127
10.4.	Singing ability	128
10.4.1.	Familiar song singing.....	129
10.4.2.	Song learning (imitation)	129
10.4.3.	Raters	130

10.4.4.	Reliability analysis: singing the new song and “Happy Birthday”	130
10.5.	Technical support: recording/mastering/ratings.....	131
10.6.	Advanced Measures of Music Audiation (Gordon 1989).....	131
10.7.	Short-term-memory.....	132
11.	Results and data of the empirical investigation	132
11.1.	Normal distribution, gender and age.....	132
11.2.	Factor analysis: Aptitude measurements and components.	132
11.2.1.	Factor 1: Singing measurements	135
11.2.2.	Factor 2: Speech imitation	137
11.2.3.	Factor 3: Perception	138
11.2.4.	Discussion	140
11.3.	Correlational analysis and regression models: phonetic aptitude measured by speech imitation ability	141
11.4.	Phonetic aptitude as measured by imitation tasks.....	160
11.4.1.	Verbal short-term memory.....	160
11.4.2.	Singing predictors	160
11.4.3.	Musical perception ability (tonal aptitude)	161
11.4.4.	Performance anxiety, self-estimation concepts and speech perception	162
11.4.5.	Comparisons of predictors with previous research	163
11.4.6.	Hindi, Tagalog, and Mandarin	164
11.4.7.	Stress-timed, syllable-timed and mora-timed languages	166
11.5.	Speech imitation total.....	166
11.5.1.	Phonetic aptitude, and the phonological short-term memory and foreign languages.....	170
11.5.2.	Phonetic aptitude, singing and memorable	171
11.5.3.	Discussion	172
11.6.	Perception versus production	174
11.6.1.	Group differences.....	174
11.6.2.	Results production.....	175
11.6.2.1.	Singers versus instrumentalists	179
11.6.2.2.	Singers versus amateurs	179
11.6.2.3.	Singers versus the non-musicians	179
11.6.2.4.	Instrumentalists versus amateurs	179
11.6.2.5.	Instrumentalists versus non-musicians.....	179
11.6.2.6.	Amateurs versus non-musicians.....	179
11.6.3.	Results Perception.....	180

11.6.3.1. Singers versus instrumentalists	184
11.6.3.2. Singers versus amateurs	184
11.6.3.3. Singers versus non-musicians	184
11.6.3.4. Instrumentalists versus amateurs	184
11.6.3.5. Instrumentalists versus the non-musicians.....	184
11.6.3.6. Amateurs versus non-musicians.....	184
11.7. Instrumentalists and vocalists	185
11.7.1. Vocalists and phonetic aptitude	187
11.7.2. Instrumentalists and phonetic aptitude.....	188
11.7.3. Performance anxiety	189
11.7.4. Singing behaviour during childhood.....	192
11.8. Discussion	193
12. Final discussion.....	194
12.1. Aptitude tests.....	194
12.1.1. Speech imitation as a language aptitude component	194
12.1.2. Singing and music perception ability as a component of language aptitude batteries	195
12.2. Predictors of phonetic imitation aptitude.....	196
12.3. Dissociation between perception and production	198
12.4. Performance anxiety and singing behaviour during childhood	200
13. Outlook	201
13.1. Early-acquired effects, training related influence and innate predispositions ..	201
13.2. Language aptitude perception and production measurements	202
13.3. Education and aptitude.....	203
13.4. Foreign language and music education.....	204
13.5. Ethical discussion.....	204
14. Conclusion	205
15. References.....	207
Appendix 1. Questionnaire development (Screenshots)	239
A1.1. Participants.....	239
A1.1.1. Factual questions (selected screenshots).....	239
A1.1.2. Singing self-estimation	244
A1.1.3. Singing behaviour during childhood and performance anxiety.....	244
A1.2. Ratings speech.....	245
A1.3. Rating singing	248
Appendix 2. MULT AP development	250

A2.1.	Overview Stimlines.....	250
A2.2.	Overview Multap Variables	262
A2.3.	Histograms	263
A2.4.	Q-Q Plots.....	267
Appendix 3.	Empirical investigation.....	272
Appendix 4.	Key variables	273
A4.1.	Perception.....	273
A4.2.	Production	274
A4.3.	Phonological STM and further criteria	275
Appendix 5.	Behavioural concepts.....	276
A5.1.	Concept singing during childhood	276
A5.2.	Concept Performance anxiety	277
Appendix 6.	Reliability analysis – Raters	278
Appendix 7.	List of abbreviations	285
Appendix 8.	Zusammenfassung	290

Abstract

The main aim of this thesis is to investigate the relationship between foreign language aptitude and musical abilities, in particular with regard to singing ability. In addition, the influence of singing behaviour during childhood, performance anxiety, phonological short-term memory capacity and educational background on unfamiliar speech perception and speech imitation ability was examined. To this end, new testing methods for language perception and production were developed.

The analysis is divided into three main parts. First, speech imitation ability was measured by means of eight different languages which were unintelligible to the participants. Results have revealed that 10 different predictors are involved in explaining the variances of the speech imitation performances. The most important of these are phonological short-term memory, singing ability (rhythm and melody) and tonal perception ability.

Second, all eight languages were used together as an entire speech imitation aptitude test to simulate a polyglot condition. A regression model revealed that the number of foreign languages spoken, the phonological short-term memory, how well songs can be learnt and how memorable the languages occurred to the participants account for 60 percent of the variances of the entire speech imitation measurements. Surprisingly, individual differences in abilities which could predict specific languages have not been detected, except for tone languages. Thus, it can be suggested that imitation tasks as used in this investigation are less dependent on typology, and represent a more general phonetic imitation aptitude.

Third, comparisons between the unfamiliar speech perception and production tasks have shown that both can be dissociated, which also resembles findings of singing versus pitch perception. Results have furthermore demonstrated that the internal relationship of the different unintelligible speech perception and speech imitation tasks is remarkably different. While the imitation tasks are highly interrelated, the individual speech perception tasks are not. This also suggests that individual differences may be more easily observed in speech perception ability.

The analyses also suggest that language aptitude batteries would benefit from the inclusion of language imitation tasks that measure phonetic imitation ability. Furthermore, music perception ability and singing are also important aspect of measuring language aptitude as both can predict oral foreign language capacity. This thesis also provides evidence that singing and musical abilities facilitate oral foreign language acquisition processes in multiple domains. In addition, this thesis contributes to linking together areas of knowledge in the field of linguistics and musicology.

Acknowledgement

Above all, I wish to express my gratitude to my supervisor, Univ.-Prof. Mag. Dr. Wolfgang U. Dressler, who has offered me great advice and tremendous support in every respect. His valuable feedback was very inspiring and I have been deeply impressed by his profound knowledge. I also want to thank Assoz. Prof. Dr. Susanne M. Reiterer, who has supported me from the very beginning of my research career.

My gratitude also goes to the DOC-team Awarding Committee for trusting in my research abilities and, of course, to the Austrian Academy of Sciences, which has made all this possible. I would like to thank the 914 people who supported me in this project, either as participants, or as raters. I wish I could name you all.

I also want to express my gratitude to my Doc-Team colleague Sabine Sommer-Lolei, who has also become one of my best friends. I also want to thank Lucia Heiss, who is one of my oldest friends and who has known me all my life. My gratitude also goes to Angelika Wukovits for believing in me since the very first day we met.

I also want to thank my team and friends in Heidelberg who welcomed me on the very first day. Thank you all, Jan, Peter, Alex, Torte, Tine, Renate, Katharina and Jeannine. I also want to express my gratitude to Annemarie Seither-Preisler for discussions about statistics. My gratitude also goes to Sophie Appl for the constant support whenever I needed it. It was a real honor for me to work with Maria Ferstl I am deeply impressed by your kindness and support.

My gratitude also goes to Angelika Rumpler, who is one of the honest people I know, and one of the most reliable friends one can have. In the same vein, I would also like to express my gratitude to the staff and former colleagues of Sta. Christiana Frohsdorf for everything they did for me.

I also want to thank Luke Green for the language support, Wolf Wiedermann for all the effort, work load, knowledge, support and programming. Finally, I thank Matthias Gebert for appreciating my work and supporting me in every respect.

1. Introduction

Music has often been described as a universal language with which feelings can be expressed across cultures (Balkwill & Thompson 1999; Fritz et al. 2009). It has unifying characteristics and is associated with social bonding (Patel 2007), which shows that music is deeply rooted in society, as it is part of most ceremonial activities. In particular singing activities have a special position. The combination of melody and text facilitates passing on cultural knowledge and customs (Tsang et al. 2011), seems to affect memory, and also reveals basic emotions such as fear, sadness and happiness (Fritz et al. 2009). As can be seen, the list of functions and roles of music is diverse.

Music expresses ideas (Limb 2006), reveals identity and cultural belonging and has many functions which are associated with, and ascribed to language. However, the relationship between languages and music is more diverse and complex than mentioned above and can be observed in multiple domains. Both share a set of characteristics and features. Speaking and singing require the integration of similar sensory-motor and acoustic information. Yet, despite their similarity, language and music can also go in different directions. While speaking and writing generally demand the avoidance of ambiguity, in singing this is often achieved on purpose (Limb 2006). Taken together, it is not surprising that in the past decade, musical abilities and the link to language functions have gained considerable scientific interest in multiple disciplines, which all have started comparing both faculties on multiple dimensions.

In the second half of the 20th century, the first recognisable work to compare music and language by means of using tools was conducted by Bernstein (1976). He made an “attempt [...] to bring the methodology of transformational generative grammar to the study of music” (Bernstein 1976: 883). His early attempts to apply linguistic methods to the study of music were less successful (Honing 2011). In the past ten years music has successfully been integrated in the field of linguistics and has been considered by many reputable researchers (Barbo 2019). The reason for this is evident as music and language share a set of properties (Limb 2006) which are all important in research in both musicology and linguistics. Language and music consist of tonal properties, temporal aspects, and are based on hierarchical structures. Comparisons between both faculties have become diverse. Researchers have compared rhythmic and tonal aspects of music and speech (Krumhansl & Keil 1982; Patel 2007; Patel & Daniele 2003). Pitch structures of music and language have been analysed by Jackendoff and Lerdahl (2006) on a semiotic level. Linguistic syntax has been taken to

describe a musical counterpart which is governed by combined discrete structural elements (Patel 2003; Honing 2011).

Recent trends have also shown that interdisciplinary research more than ever has started to integrate knowledge of other disciplines into their own for good reasons. Increasingly more evidence has been provided that a single discipline cannot adequately address questions which overlap with many other disciplines. As music has been integrated into linguistics, a third discipline ought to be considered which is worth being analysed from a linguistic point of view: singing research.

Like poetry, singing is a hybrid construction which consists of elements of speech and music. While poetry has a long tradition in language studies and language research, singing has largely been neglected. This is surprising as evolutionary hypotheses have speculated about a musical protolanguage (Masataka 2009; Livingstone 1973) which is said to have been song-like for a long time. This view includes crucial information on singing as a scientific field, which belongs to both linguistics and musicology by their nature. Song usually consists of lyrics and melody and, depending on the musical genre, places one of the two main components into the foreground or background.

More recently, research has started to examine the relationship between singing and language learning. In particular, singing as a learning tool has been used to facilitate foreign language learning. This has many advantages, some of which have already been listed above. One is that melody of song has been said to facilitate verbal recall (Purnell-Webb & Speelman 2008), and second, singing includes the generation of vocabulary (Ludke et al. 2014). Studies have revealed that singing new words is a very effective way to learn them (Ludke et al. 2014). Other approaches are rare and only few investigations have started comparing singing and language abilities directly in the context of language aptitude (Christiner & Reiterer 2013; Christiner et al. 2018). Research on singing ability offers many opportunities. The most promising one is that examining singing abilities may be one of the most effective approaches to understanding shared and distinct mechanisms of music and language.

This thesis attempts to represent a variety of different approaches which all seek to provide insights into the relationship between phonetic aptitude, musical abilities, singing, short-term memory capacity, education, singing behaviour during childhood, performance anxiety, foreign language ability, musical instrument playing and environmental influences. It aims to reveal which capacities contribute to individual differences in phonetic aptitude, which will be measured by speech imitation tasks. The original idea to this thesis is based on

several previous investigations, in which issues or findings remained largely unaddressed, unexplained, or too vague to draw general conclusions.

In former research, vocalists of different proficiencies have been hired and tested for their ability to imitate unfamiliar speech, which turned out to be related to their singing ability. The latter has been found to be a better predictor for speech imitation ability than any of the musical perception measurements which were included in the research design. Consequently, it has been concluded that vocal flexibility, intensive vocal training and probably singing aptitude facilitate the ability to reproduce unusual new sound combinations of verbal material more effectively than music perception ability (Christiner 2013; Christiner & Reiterer 2013).

In a follow-up study, this was retested in a different form by hiring singers, instrumentalists and non-musicians, who were tested for their speech imitation abilities and music perception abilities. Results revealed that vocalists performed best, followed by instrumentalists who were still significantly better than the non-musicians. Evidence was mounting that singing may be a better predictor for speech imitation ability than music perception ability, at least for specific groups such as singers who train their vocal motor apparatus. Additionally, it has also been suggested that enhanced musical perception, as is the case with instrumentalists, also improves phonetic aptitude (Christiner & Reiterer 2015).

However, those studies included only music perception measurements, but not speech perception tasks which would also have allowed direct comparisons between speech imitation and speech perception ability. This is why this thesis aims to compare production and perception of music and speech in several ways as well. This should provide proper information regarding on which basis singing and speech imitation are related, and in turn, reveal whether the relationship between music and language perception is similar to that of their production counterparts: speech imitation and singing ability. Achieving this required a similar test design for music and speech perception as it had already been created for the singing and speech imitation tasks so that the perception designs for music and speech were also comparable (see section 9).

Another issue which emerged in the course of time was that only a single language has been used in the research design, and it was not discernable whether the results were transferable to other languages, which required extending the research design and the number of languages included. First results indeed revealed different predictors for Mandarin and Tagalog imitation performances. While the variances of the performances of Mandarin could be explained by phonological short-term memory (STM) capacity, tonal discrimination ability

and singing ability, for Tagalog it was the rhythmic ability and the phonological STM (Christiner et al. 2018). This has led to the assumption that individual differences in musical ability probably predict language-dependent preferences for specific sound structures. Most obviously, Mandarin, a tone language, could be predicted with singing and tonal discrimination ability.

Further research demand additional predictors which all could be involved in explaining individual differences in speech imitation ability, such as performance anxiety, age factors, singing acquisition processes, other related cognitive capacities and environmental factors. At the same time, the nature of speech imitation tasks of individual languages should be evaluated with the aim to at understanding their ecological validity. A speech perception test as an equivalent to the production test would reveal further information about individual differences. Additionally, a test design which consists of multiple languages would make it possible to observe individual differences dependent on languages. This was the starting point to this thesis, in which the focus lay on improving previously established test designs and developing new tests to address all the unsolved issue. This has led to the following hypotheses and research questions.

1.1. Research question 1

Phonetic aptitude research that includes singing is rare and the few investigations have to be retested and conceptually improved. This illustrates that if findings of previous investigations could be replicated, it would give additional evidence that specific abilities such as singing contribute to phonetic aptitude as measured by language imitation ability.

Furthermore, it could also be assumed that if more languages and a stronger test design were to be developed, additional evidence for its importance and reliability would be provided, and the nature and ecological validity of imitation tasks as a test design could be discussed.

In the same vein, it could also be observed whether different languages are related to different musical abilities or not. Following traditional theories, languages vary on their foot rhythm and are classified into stress-timed, syllable-timed and mora-timed languages. Mandarin or Thai have been categorised as tone languages with strong musical components. By using multiple different languages, it could also be shown whether there may be different transfer effects from music to language.

1.2. Research question 2

Aptitude measurements are diverse and multiple measurements have been developed. The nature of language imitation tasks as used in previous investigations is not well understood. Using various language imitation tasks of different languages would reveal information about the nature of imitation tasks and their relationships to each other, as well as to similar concepts such as measurements of short-term memory, speech perception, singing and musical abilities. Contrasting perception and imitation of unintelligible speech would also show whether both are related differently or similarly to each other. The knowledge generated could be used to improve current aptitude measurements and test batteries or to integrate new ones.

1.3. Research question 3

Language research has noted that foreign speech perception and production can surprisingly be dissociated on a behavioural and partially also on an anatomical basis. Those who performed better in foreign speech perception, for instance, did not necessarily pronounce speech better. Similar phenomena have also been reported in music research. Pitch discrimination and pitch production ability as in singing could be dissociated. Additionally, it has also been demonstrated that speech production and music production (singing) were more closely related than both were to music perception. Testing all the possible combinations would provide further evidence whether production and perception processes can be dissociated or not.

1.4. Research question 4

A number of aspects have been addressed less often in the context of phonetic aptitude. These are performance anxiety, singing behaviour during childhood, and further self-estimation concepts such as how memorable and difficult to imitate unfamiliar languages are to individuals. Similarly, the question as to how familiar unintelligible languages are to participants has also not been included in aptitude research.

This thesis is divided into 13 parts which give an overview of the knowledge of aptitude, the current knowledge of each research area, approaches and testing methods which are used for measuring language and musical aptitude, singing ability, working memory (WM), short-term memory (STM), long-term memory (LTM) and performance anxiety. This is followed by theoretical chapters that focus on questionnaire construction, statistical analysis and test development including the results of the development of the speech perception test.

The final chapters present the results and findings of the empirical investigation. The list of abbreviations can be found in appendix 7.

As the first chapter has introductory purposes, second 2 focuses on language aptitude and testing methods which are commonly used to assess language abilities, followed by the same for musical aptitude in chapter 3. This aims at demonstrating their differences and similarities, which is necessary in order to understand their relationship. Chapter 4 presents two branches of singing research, including measurements and additionally focusing on singing acquisition processes. This shows that musical abilities should be differentiated and places emphasis on acknowledging musical aptitude and singing ability as distinct and similar capacities.

In the following chapter WM capacity and STM will be discussed, for three reasons. First of all, the difference between WM and STM will be shown as both terms have been used interchangeably in the literature, which has led to much confusion which should be avoided. Second, STM is one of the most important predictors for speech imitation ability, which has been revealed in several previous investigations (Christiner 2013; Christiner & Reiterer 2013; Christiner et al. 2018). Third, background knowledge of WM and STM design delivers crucial information on test development, which is necessary for the speech perception test which has been developed. Chapter 6 illustrates performance anxiety and aims at demonstrating that this concept should be integrated into aptitude measurements, especially if performances are evaluated and group comparisons on musicians versus non-musicians are conducted.

The following two chapters (7 and 8) focus on theoretical aspects of test development and statistical analysis. As this thesis includes the development of certain tests and questionnaires, chapter 7 presents questionnaire constructions, conception and analysis, followed by chapter 8, which outlines the choices of the statistical analysis for the main two studies. The next chapter (9) describes the development of the speech perception test, including a short analysis which should illustrate the nature of the test and reveals the relationship between speech perception and musical instrument playing.

In chapter 10, the methods and measurements will be discussed. Chapter 11 presents the results of the empirical investigation. Most importantly, the ordering of the analysis is deliberate. Three main statistical approaches have been selected: factorial, regression and group analysis.

First of all, the nature of measurements has been investigated by performing factorial analysis (see chapter 11.2). This facilitates understanding the nature and relationship of the measurements included, and provides information about meaningful scoring for the later

analysis. Additionally, it allows discussions of the test designs and how they may be improved in future research.

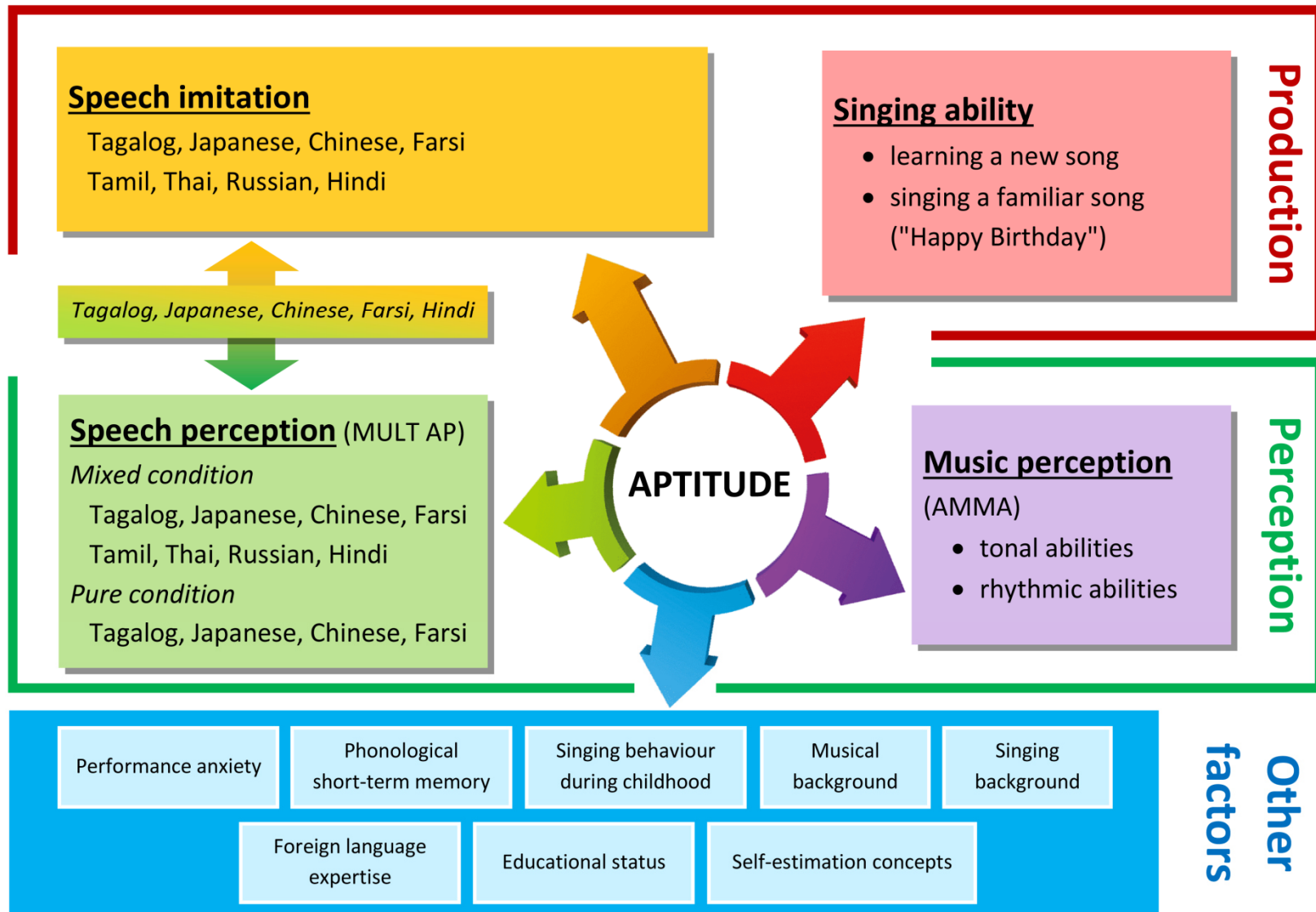
The second analysis reveals the most important predictors which can explain the variances of the performances of all language imitation tasks included in the research design (see chapters 11.3 and 11.4). It discusses the main predictors which can explain the variances of the speech imitation tasks in eight languages, including the presentation of new predictors, compares them with previous research and provides shared and distinct findings. Furthermore, the eight language imitation tasks together have also formed the basis for simulating a polyglot measurement which will be discussed (see chapter 11.5).

The third analysis focuses on group comparisons which aim at revealing whether differences in abilities can be observed when compared to previous research (Christiner & Reiterer 2015). This section places emphasis on perception and production measurements of this investigation in both music (music perception and singing) and speech (speech imitation and perception) and demonstrates individual differences based on four different groups: vocalists, instrumentalists, amateurs and non-musicians (see chapter 11.6).

One hypothesis which has largely been rejected is as follows. Traditional theories have suggested that languages vary on their foot rhythm and are classified as stress-timed, syllable-timed and mora-timed languages respectively. Considering classifications in linguistics, one aim was also to analyse whether participants perform differently according to the rhythmic dimension of the language they imitate. This, however, turned out to be problematic as the statistical analysis soon revealed additional information about the nature of imitation tasks. Therefore, a theoretical part that focuses on timing has been omitted, and only a short passage in section 11.4.7 will be provided.

In the opening of this thesis, a collection of all measurements has been provided in figure 1. It states all aspects which will be investigated, analysed, discussed and considered in this thesis. Figure 1 shows four main aspects speech imitation, speech perception, singing ability and music perception. Speech imitation ability, which is also the measurement associated to revealing phonetic aptitude, is the most important parameter, with which all the others involved will be compared. As can be seen, there is a division into production and perception variables, which illustrate to which category each of them belongs. In addition, five languages, Tagalog, Japanese, Mandarin, Farsi and Hindi, have an imitation and perception part, which allows direct comparisons between speech production and perception. The last category describes other factors and measurements which have been included to gain an impression of components which should be addressed in aptitude research.

Figure 1. Overview



2. Aptitude and language aptitude

Foreign language aptitude can easily be observed as there are always people who outperform others on foreign language tasks, even though they do not necessarily show to have outstanding abilities in other capacities. These people are considered to have high aptitude compared to others. This is a simple observation, but understanding, measuring and defining the origin of aptitude remains a complex and comparatively poorly understood endeavour. This, however, would form the basis of what is necessary to study the effect of language aptitude on learning foreign languages (Stansfield & Reed 2019). Instead, aptitude has very often been neglected since it has been considered to be a rather stable trait which is less likely influenced by the environment (Sparks et al. 2011) and is therefore less attractive for educational purposes which focus on positive interventions.

As language aptitude has been said to consist of multiple general cognitive abilities, specific verbal abilities but also personal and motivational factors (Linck et al. 2013), it seems to be very difficult to define aptitude and to make use of it since it is a complex rather than a unitary trait (Dörnyei 2005). Integrating aspects of aptitude in second language (L2) or foreign learning contexts would mean that primarily the difficulty of measuring language aptitude and what may be predicted with such measurements should be fixed first in order to facilitate abilities that foster foreign acquisition processes. This is very complex but one of the reasons why aptitude research is still discussed and criticised. The main issue concerns the fact that the aptitude construct has still not been specified (Skehan 2019), while another issue is in how far aptitude for foreign languages can predict achievement, as longitudinal studies that would show the role of aptitude in language success are very limited or do not exist (Wen et al. 2017). This makes it very difficult to grasp measures of aptitude because most of them focus more or less on potential to achieve high L2 proficiency, which includes uncertainty about whether potentials can predict achievement.

Furthermore, L2 learning processes in early childhood may be different from later stages of L2 learning during adulthood by its nature (DeKeyser et al. 2010). This requires different abilities to be put in the foreground. High attainment in foreign languages has primarily been suggested to rely on cognitive and perceptual abilities (Doughty et al. 2008; Linck et al. 2013). However, motivation, interest and other environmentally imposed factors influence high achievement as well (Dörnyei 2014). This makes it very difficult to measure cognitive abilities precisely and to differentiate them from acquired factors as they are intertwined to a certain extent.

In L2 and foreign language research learning style, learning strategies, personality, motivation, anxiety, aptitude, intelligence quotient (IQ) and age have been found to explain individual differences (ID) in foreign language capacities (Dörnyei 2006; Skehan 1991; O'Malley & Chamot 1990). At this stage it is worth mentioning that age, motivation, interest, learning style, learning strategies and IQ also contribute to language aptitude but are not part of this investigation as this would go beyond the scope. Some statements about them will be made here.

Age of acquisition plays a major role in first and second language learning processes (Hernandez & Li 2007) and in some instances aptitude has also been related to age. Some researchers have suggested that for adult foreign language learners aptitude is a critical factor, while it is not for children who learn by exposure (DeKeyser et al. 2010). As in this investigation language aptitude is observed among adolescents and adults, age is of less importance as it is controlled by participants' selection. Referring to learning strategies and cognitive styles, good language learners have been found to use different strategies and cognitive styles than poor learners do (O'Malley & Chamot 1990). The role of learning strategies and cognitive styles in relation to language learning and aptitude, however, has been criticised as there is little evidence that both improve foreign language learning (Skehan 1991), as well as whether they measure cognitive styles or abilities (Tiedemann 1989). IQ has been said to be related to the analytical abilities of language aptitude (Sasaki 1999) and to a number of other cognitive abilities which are also associated with language aptitude such as musical abilities (Banai & Ahissar 2013; Lynn et al. 1989) or working memory capacity (Colom et al. 2005; Mogle et al. 2008; Fukuda et al. 2010). Other well-researched L2 predictors such as motivation and personality have been shown to predict foreign language success (Ehrman & Oxford 1995; Dörnyei 2006). Relationships between L2 proficiency and personality have been explained by preferences for what individuals feel comfortable with (Krashen 1985). Research that focuses on personality include a number of characteristics such as extraversion, introversion, self-esteem, inhibition, risk-taking and anxiety (Zafar & Meenakshi 2012: 643). The theory of motivation and its relationship to foreign language success may be established best by consulting the work of Dörnyei and colleagues (Dörnyei 2014; Dörnyei et al. 2015). Anxiety has also been discussed and reviewed in the context of foreign language proficiency, for instance, by Sparks and Ganschow (2001). A low anxiety condition is crucial for successful adaptations of new language input (Horwitz et al. 1986: 127) and measures of anxiety correlate negatively with measures of foreign language achievement (Gardner Robert 1990). In foreign languages research it is important to consider

that anxiety can have an effect on language performances. Anxiety and language abilities will be discussed in chapter 6 in more detail.

To turn to aptitude research, in the last three decades there have been multiple, increasingly more diverse disciplines that have focused on aptitude, including neuroscience, psychology, cognitive science, linguistics and educational science. This requires looking at aptitude from a more general point of view independent of any language domain first.

2.1. Defining Aptitude

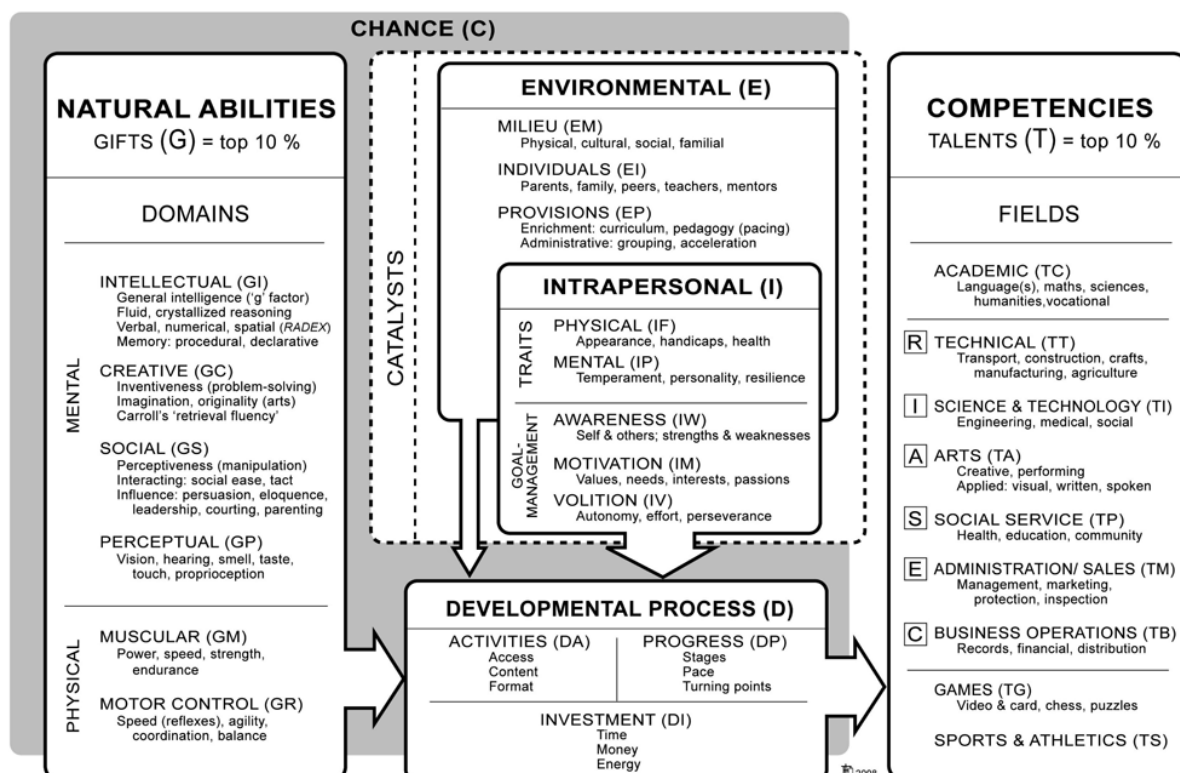
Referring to psychology, aptitude has been described as an ability that places individuals above their peers in a specific domain (Gagné 2004) such as in musical expertise, or talent for learning foreign languages. People who possess high talent and giftedness are usually found in the upper 10 percent of the normal distribution (Gagné 2008). Defining aptitude is rather complex and difficult, but essentially it is regarded as an exceptional capability in a specific domain or as a potential which facilitates development in particular fields (Jørgensen 2008). Aptitude is considered to be a pure ability which is largely free of any influence by nurture and can be compared to high potential to achieve something (Christiner & Reiterer 2015). Nardo and Reiterer (2009: 214) postulate four main notions which can be considered and referred to when speaking about aptitude:

1. as an innate gift that is separated from practical implications,
2. as a potential or propensity, the implications of which include a certain orientation towards something,
3. as an exceptional or outstanding ability for something,
4. as a special skill, capacity or ability.

In the field of psychology, two models on aptitude have had a tremendous influence on how aptitude is presented, viewed, measured, and understood. The first model, the *Three Ring Model*, consists of three basic components: above-average ability, high levels of task commitment, and high levels of creativity that are necessary to develop aptitude in specific domains (Renzulli 1978). For the author, giftedness is more of a process, whereby the potential has to be distinguished from the performance (Renzulli 2016). This means that great potentials which are manifested turn into superior performance. Identifying aptitude or giftedness, therefore, can be achieved by using “ [...] assessment procedures that readily identify potential for superior performance in [a] particular domain” (Renzulli 2016: 58).

The second model, the *Model of Giftedness and Talent (DMGT)*, focuses on two main concepts, in addition to the two environmental ones. They are giftedness (aptitude) and talent (Gagné 2004, 2008). The former refers to an outstanding natural ability, while talent refers to systematically developed abilities. According to the author (Gagné 2008) gifts are turned into talents when a person engages in activities which are only partially influenced by the genetic endowment as natural abilities are not innate to Gagné. Furthermore, he suggests that gifts manifest themselves and stabilize more easily in younger ages, which shows a similarity to language acquisition theories that suggest the earlier the language is acquired, the better the achievement (Johnson & Newport 1989; Lenneberg 1967). It is, however, noteworthy to mention that Gagné (2008) also says that aptitude can still be observed in adults by the speed of how they acquire something, which resembles notions about aptitude being a rather stable trait. Gagné's *DMGT* consists of three basic components: giftedness, talent and talent development process, and two subcomponents: environmental catalysts, and the intrapersonal catalysts as illustrated in figure 2 (Gagné 2008):

Figure 2. Gagné's Model of Giftedness and Talent (DMGT)



If the aforementioned mentioned explanations of aptitude and the two models are taken together, aptitude is defined as an ability or skill that is elaborate, predicts high

achievement and is recognized as an exceptional capability compared to others in a particular domain. Furthermore, the definition of aptitude is highly dependent on developmental processes when aptitudes stabilize by environmental influences and become visible by excellent superior performances in a particular domain. This implies that aptitude and talent are highly intertwined, which is also one of the major problems in aptitude research. It is still quite unclear in how far aptitude measures can predict high achievement as measuring aptitude most often includes the learning of new tasks and therefore equates to measuring potentials. This allows the observation of the speed at which individuals learn, but cannot be informative in terms of how advanced those individuals will be in achieving the newly learnt ability in their later lives i.e. reaching high performance in a particular domain. This issue shows that mainly beginner stages have been investigated in order to understand which capacities may facilitate high achievement. One way out of this problem is to investigate individuals' abilities who have already reached a high "end state" (Linck et al. 2013), which means that they appear to be an exceptional foreign language performer. The ideal way to observe aptitude and its relationship to achievement would be by conducting longitudinal studies that integrate the beginning of learning and dynamic processes which include measuring the speed of learning within several time points, and the end stage of high performers – a test design which is difficult to construct.

At this stage, ability, aptitude, potential, achievement, and language success should be defined. Ability and aptitude are most often used synonymously (Dörnyei 2005), which is also the case in this investigation. Carroll (1981: 86) refers to aptitude as "[...] an individual's initial state of readiness [...]", a capacity or potential which is rather stable and in the event that it is facilitated by proper input (environmental influence) is predictive to foreign language success. In terms of achievement and language success, it is difficult to give precise definitions as researchers usually talk and refer to the upper 10 percent of people who perform excellently but, depending on research, such individuals are very often not included in the samples. Therefore, achievement and language success more or less refer to individuals with excellent foreign language abilities but not necessarily to the high attainment group which could precisely be equated with native speakers.

Interestingly, there are contradictory explanations regarding the origin of aptitude. While some argue aptitude is not innate and only partly influenced by the genetic endowment (Gagné 2008), others suggest it to be of innate origin (Nardo & Reiterer 2009). Increasingly more evidence from brain research and genetic studies show that training effects cannot fully account for explaining individual differences, neither for elaborate speech perception and

production (Golestani et al. 2011; Golestani & Zatorre 2009), nor for excellent musical abilities (Pulli et al. 2008; Mosing et al. 2014; Butkovic et al. 2015). This illustrates that individual differences may partly be based on predispositions of inherent origin, or rather early acquired abilities.

Transferring aptitude to the field of linguistics, various terms have been used to describe foreign language aptitude such as second language aptitude, linguistic giftedness, language talent or special propensity (Dörnyei 2006). Thus foreign language aptitude has been considered to be a talent whereby people learn new languages more precisely and much faster than their peers (Carroll 1981, 1989, 1990, 1993). Wen (2012) also postulates the same notions as Carroll and argues that foreign language aptitude is based on the observation that individuals learn faster, perform more accurately and quickly and thereby achieve better results than their peers, while Skehan (2019) adds that language aptitude is a special ability which speeds up foreign language learning processes. Other researchers also found that foreign language aptitude is critical for adult learners in marked contrast to children, which suggests that learning processes of both children and adults are different in nature, depending on age factors (DeKeyser et al. 2010). This notion refers to critical period effects, which says that foreign language learning becomes significantly more difficult after puberty (Moyer 2014).

Foreign language capacities range considerably in the population and depend on multiple aspects. One of the most desired abilities is to speak foreign languages like native speakers do. Individuals who possess foreign language aptitude are often those who have the ability to speak and hear a foreign language on a native, near-native or native-like level (Dollmann et al. 2019). These have to be differentiated from bilinguals who are associated with having learnt two languages simultaneously, or before the critical age. Individuals with high foreign language aptitude have often achieved elaborate foreign languages skills after the critical age period. The ability to achieve high language proficiency comparable to native speakers is rather limited and only 10 to 15 percent of adult speakers are said to reach this high target (Reiterer et al. 2011).

Within the language aptitude category high-level aptitude has additionally been differentiated from more general aptitude by some researchers who suggest special capacities that lead to high level-attainment (Schneiderman & Desmarais 1988). What most researchers agree to, however, is that apart from motivational, personality factors and IQ that seem to affect foreign language aptitude, it is based on elaborate cognitive abilities (Carroll 1981; Doughty & Long 2003), a kind of hybrid construction (Dörnyei 2005). The developers of the

Hi-LAB aptitude battery, for instance, claim that high level attainment in second language proficiency largely relies on cognitive and perceptual abilities (Linck et al. 2013). Most frequently, cognitive abilities such as working memory capacity (Christiner & Reiterer 2013; Wen & Skehan 2011; Loosli et al. 2012; Juffs & Harrington 2011; Linck et al. 2014; Hummel 2009), logical reasoning (Pimsleur 1966), grammatical sensitivity, coding ability and inductive reasoning (Carroll 1981) are in the foreground in foreign aptitude research.

However, in the last two decades, other cognitive abilities deriving from language's neighbouring faculty of music such as musical abilities (Fonseca-Mora et al. 2015; Milovanov & Tervaniemi 2011) and singing ability (Christiner & Reiterer 2013, 2015; Turker et al. 2018; Ludke et al. 2014) have shown to predict language aptitude and influence foreign language success as well. Despite the current knowledge, the intensity of aptitude research has fluctuated remarkably over the last century. Therefore, a historical overview about language aptitude research will illustrate various research directions.

2.2. Historical overview of language aptitude research

Language aptitude research originated in the 1920s when the first large-scale experiments (Hunt et al. 1929) on foreign language aptitude were carried out (Sparks & Ganschow 2001). These early attempts already included investigating personality factors, intelligence, learners' background/knowledge, gender differences and, for the first time, individual features such as pitch perception (Nakamura 1988). In the 1920s and 1930s, the first aptitude tests were developed, which were referred to as prognosis tests and included most often two distinct approaches, the analytical and the synthetic approaches (Spolsky 1995). The latter focused on artificial or foreign language lessons, while the former involved "us[ing] items that tapped specific hypothesized cognitive abilities, usually through the first language, such as memory or vocabulary or some other aspect of verbal intelligence" (Spolsky 1995: 29). These tests aimed at predicting the variances of the performances in foreign language learning situations. The prognosis tests, however, did not meet researchers' expectations of examining foreign language learning issues at that time. The tests were originally developed to predict foreign language success to keep unfit students away from language classes (Spolsky 1995).

During World War II, language aptitude research regained considerable interest for military purposes to establish language programmes that were both cost-effective (Sasaki 2012) and predictive to achievement. These language programmes were based on previous results of the prognosis tests but did not generate proper results as they were also aiming at

pre-selecting learners with obviously more talent for particular languages than others (Sparks & Ganschow 2001). After World War II, language aptitude research took a new direction and was, and presently still is, highly influenced by Carroll, who started to work on aptitude testing batteries in the 1950s and 1960s (Wen 2012). First attempts to create a testing battery revealed that grades of instructors could be related to four subtests: “[...] artificial language learning, phonetic association, words in sentences, and paired associates [...]” (Sparks & Ganschow 2001: 43). The Modern Language Aptitude Test (MLAT) by Carroll & Sapon (1959) was the result of intensive research and testing and is still one of the most influential aptitude tests that has been developed so far. The MLAT has had a tremendous influence on which abilities have been found to be measureable.

Presently, the MLAT is still one of the most widely used language aptitude tests, even though further language aptitude tests have been developed and published. The most influential ones are the Language Aptitude Battery (PLAB) by Pimsleur (1966), the Defense Language Aptitude Battery (DLAB) by Petersen and Al-Haik (1976), the VORD test by Parry and Child (1990), the CANAL-F test by Grigorenko et al. (2000), the LLAMA Language Aptitude Test by Meara (2005) and the Hi-LAB by Doughty et al. (2008). In the following section the most influential aptitude tests will be presented in a chronological order.

2.3. Language aptitude tests

2.3.1. The Modern Language Aptitude Test (Carroll & Sapon 1959)

Carroll defines four aptitude components which according to him can be measured. Those are (Carroll 1990: 14):

1. phonetic coding ability,
2. grammatical sensitivity,
3. memory abilities/rote learning ability,
4. inductive language learning ability.

Phonetic coding ability has been defined as the “ability to identify distinct sounds, to form associations between these sounds and symbols representing them, and to retain these associations” (Carroll 1981: 105). Memory abilities refer to the ability to create associative links to memorise, which mainly tests rote learning ability and should therefore not be equated with a measure of general memory (Carroll 1990). Inductive language learning ability refers to the capacity to use words from a pool to form new sentences (Dörnyei & Skehan 2003; Wen et al. 2019), while grammatical sensitivity means the ability to understand the function of words within sentences (Dörnyei & Skehan 2003).

The MLAT consists of five different parts that measure Carroll's four language aptitude components (Stansfield & Reed 2019: 17–18) see table 1.

Table 1. MLAT

MLAT	
Part I	The <i>Number Learning</i> task where an artificial language based on English sounds needs to be learned
Part II	The <i>Phonetic Script</i> where phonetic symbols need to be remembered and retained
Part III	The <i>Spelling Cues</i> where the correspondence between sounds of speech and symbols are measured
Part IV	<i>Words in Sentences</i> that measures the ability to understand grammatical structures
Part V	<i>Paired Associates</i> where new vocabulary is learnt in a new language

2.3.2. Pimsleur Language Aptitude Battery (Pimsleur 1966)

The Pimsleur Language Aptitude Battery (PLAB) aptitude test is targeted at English native speakers, especially for students that wanted to study a foreign language.

According to Pimsleur et al. (1962), a number of variables had been found to be related to foreign language achievement such as intelligence, pitch discrimination and verbal ability. In addition, study habits, order of language study as well as personality factors and motivation were also mentioned as driving forces that influence language success by him and his colleagues (Pimsleur et al. 1962). In his later work he especially put emphasis on interest motivation and verbal intelligence as main contributors to language achievement, even though pitch discrimination, word fluency and reasoning also seemed to have an effect on language success in his studies (Pimsleur 1963). The PLAB is a test that makes it possible to observe individual differences in foreign language aptitude of students and may be useful for counselling students and understanding their preferences and cognitive abilities (Stansfield 1988). The test is a multiple choice test and is comprised of six parts (Stansfield 1988: 439–440), see table 2.

Table 2. PLAB

PLAB	
Part I	Students are instructed to fill in their grades for social studies, mathematics, English and science
Part II	Interests in foreign language studying (Likert scales where possible answers are predefined)
Part III	Tests vocabulary knowledge where students are instructed to find appropriate synonyms
Part IV	A language analysis task where students have to show their sensitivity to understand the grammatical system of a completely new language in Kabardian
Part V	A sound discrimination task where students are instructed to listen three times to almost identical sounding words in Ewe. The words have to be remembered and discriminated in a follow-up session of 30 sentences
Part VI	A sound symbol association task. Nonsense words based on English phonology are provided. Students have to match the nonsense words they hear with its written form (multiple choice design)

2.3.3. The Defense Language Aptitude Battery (Petersen & Al-Haik 1976)

The DLAB was developed for the Defense Language Institute (DLI) that offered foreign language courses in more than 50 languages (Petersen & Al-Haik 1976). At the DLI audiolingual training was favoured, including studying intensively up to six hours a day (Petersen & Al-Haik 1976). The original idea of the DLAB was to establish an aptitude test that showed higher predictability than previous language aptitude tests such as the MLAT or the PLAB mentioned before. The development of the DLAB was highly influenced by previous findings of Carroll, who participated in the Harvard Language Aptitude Project, and presented originally seven novel aptitude components (*Associative Memory, Inductive Language Ability, Sound-Symbol Association, Grammatical Sensitivity, Verbal Knowledge, Linguistic Interest, and Speed of Association*) in his early work (Carroll 1958). The DLAB contains two main sections. The first part, Horne's Assessment of Basic Linguistic Abilities (HABLA), developed by Colonel Kibbey, is a picture task from which language concepts from an artificial language need to be derived (Petersen & Al-Haik 1976). The second and main part, Al-Haik's Foreign Language Auditory Aptitude Test (AFLAAT), developed by Al-Haik in 1972 focuses on foreign sounds, sound-symbol associations and sensitivity for grammar (Petersen & Al-Haik 1976).

Table 3 illustrates the subtests of the DLAB (Petersen & Al-Haik 1976):

Table 3. DLAB

DLAB	
<i>Part I</i>	<i>HABLA</i>
	Picture task from which language concepts have to be derived
<i>Part II</i>	<i>AFLAAT</i>
	Foreign language sounds
	Section I Utterance Identification
	Section II Recognition of Vowel Patterns
	Section III Recognition of Stress Patterns
	Foreign language writing or sound-symbol association
	Foreign language grammar consists of four sections that deal with
	Section I Noun and adjective agreement
	Section II Possessive forms
	Section III Sentence structure
	Section IV Applying the grammatical rules acquired in the sections before

2.3.4. VORD (Parry & Child 1990)

The VORD test makes use of an artificial language structurally similar to Turkish, which is typologically quite different to other Western European languages (Parry & Child 1990). The original idea of this language aptitude test was that Turkish is a language that is quite unfamiliar to language students. The test consists of 32 items and is divided into three main subtests, even though a fourth text completion task had also been developed. Comparing MLAT and VORD results, it has been suggested that the former is still the test that predicts success in language learning more effectively. The authors have concluded that non-Indo-European languages may require different instruments for examining language predictions (Parry & Child 1990).

The subtests of the VORD are illustrated in table 4 (Parry & Child 1990: 35–36).

Table 4. VORD

VORD	
Subtest I	Nominal morphology (suffixes need to be added accordingly)
Subtest II	Verbal morphology (a listening multiple choice test where strings of correct verbal forms need to be identified)
Subtest III	Phrase and sentence-level syntax (measures analytical skills to apply grammar rules)
Subtest IV	Text completion

2.3.5. Cognitive Ability for Novelty in Acquisition of Language-Foreign (Grigornko et al. 2000)

The theory behind the development of the CANAL-F test is that foreign language learning requires being able to cope with novelty and ambiguity (Ehrman & Oxford 1995) and is based on cognitive theory. For the developer of this test, five knowledge acquisition processes play a central role in foreign language learning, these being selective encoding, accidental encoding, selective comparison, selective transfer and selective combination (Grigornko et al. 2000: 392). The knowledge acquisition processes for foreign language learning operate at different levels, namely the lexical, the morphological, the semantic and the syntactic levels. The processing of the four levels is further divided into oral and visual input and output. The oral input is the basic task for evaluating listening and speaking abilities, whereas the visual one is fundamental for learning to read and write. In the final stage the new language material needs to be encoded, stored and retrieved (Grigornko et al. 2000: 392). The novelty of this test should be in a more dynamic process that makes it possible to acquire new language material while taking the test and thus resembles a situation where learning takes place naturally and gradually. The CANAL-F consists of five sections of immediate recall, followed by four identical sections of the former subtests except for delayed recall (Grigornko et al. 2000). The subtests are presented in Table 5 (Grigornko et al. 2000: 394).

Table 5. CANAL-F

CANAL-F	
Part I	Learning meanings of neologisms from context (immediate and delayed recall)
Part II	Understanding the meaning of passages (immediate and delayed recall)
Part III	Continuous paired associate learning (immediate and delayed recall)
Part IV	Sentential inference (immediate and delayed recall)
Part V	Learning language rules

2.3.6. LLAMA (Meara 2005)

The LLAMA test has been developed in order to have a short aptitude test which can be conducted within 20 minutes in marked contrast to other aptitude tests which take considerably longer (e.g. MLAT). The second reason for the creation of this aptitude test was that the first language does not play a role, in contrary to the MLAT, which was originally designed for English native speakers (Rogers et al. 2017). The LLAMA test has received great attention in the last decade as it is easily accessible and very convenient in its usability. The current test consists of four different sections that measure “[...] vocabulary acquisition, sound recognition, sound-symbol correspondence and grammatical inferencing (Rogers et al. 2017: 49)”. The subtests are shown in table 6 (Meara 2005).

Table 6. LLAMA

LLAMA	
Section I	LLAMA B measures the ability to learn a relatively large sample of vocabulary in a Central American language in a very short period of time. Participants have to learn symbols and their assigned names
Section II	LLAMA D a sound recognition task where short stretches of spoken utterances in a British Columbian Indian language need to be recognized after being exposed to them previously
Section III	LLAMA E a sound-symbol correspondence task where relationships between the sounds and symbols need to be drawn
Section IV	LLAMA F a grammar task where the grammatical rules of an unknown language need to be acquired

2.3.7. High-level Language Aptitude Battery (Doughty et al. 2008)

The Hi-LAB has been developed to understand which cognitive and perceptual mechanisms are involved in attaining high-level proficiency and especially focuses on non-instructional settings. In this way it differs significantly from the MLAT, the DLAB and the PLAB test, which focus more on initial learning processes (Linck et al. 2013). High-level aptitude has been said to differentiate from common notions of language aptitude (Schneiderman & Desmarais 1988). This test has been designed to measure exceptional language performers who have already reached a high “end state” (Linck et al. 2013). The percentage of second language learners who achieve native-like proficiency is rather low. In the literature it has been estimated that around 5 to 15 percent of learners achieve native language proficiency (Selinker 1972; Hu et al. 2013). This is the target group which the Hi-LAB battery focuses on. The Hi-LAB is the most recently developed aptitude battery and includes various additional subtests compared to previously established ones.

Table 7 shows the tests and subtests of the Hi-LAB (Linck et al. 2013: 539–544).

Table 7. Hi-LAB

Hi-LAB	
Part I	Language History Questionnaire
Part II	Eleven Cognitive and Perceptual Tests
	Running Memory Span Test
	Antisaccade Test
	Stroop Test
	Task Switching Number Test
	Letter Span Test
	Non-Word Span Test
	Paired Associates Test
	Available Long-Term Memory Synonym Test
	Serial Reaction Time Test
	Phonemic Discrimination: Hindi, English Pseudo-Contrastive Test
	Phonemic Categorization: Russian Test

Table 8 represents the Hi-LAB’s constructs and measures that should reveal high-attainment aptitude (Linck et al. 2013: 535).

Table 8. Hi-LAB constructs and measures

Construct	Measure
Working Memory	
Executive Functioning	
Updating	Running Memory Span
Inhibitory Control	Antisaccade
	Stroop
Task Switching	Task Switching Numbers
Phonological Short-term Memory	Letter Span
	Non-Word Span
Associative Memory	Paired Associates
Long-term Memory Retrieval	ALTM Synonym
Implicit learning	Serial Reaction Time
Processing Speed	Serial Reaction Time
Auditory Perceptual Acuity	Phonemic Discrimination: Hindi, English
	Pseudo-Contrastive
	Phonemic Categorization: Russian

As shown in this chapter aptitude tests and research have become quite diverse and have taken different directions over the last century. Tests and measures that occur focus repeatedly on memory, novel and artificial language learning, applying grammar rules, language perception and production tasks, on learning processes, visual cues and sound relationships, to name but a few.

The tests available focus on three main approaches towards measuring aptitude, testing initial stages of acquisition processes (e.g. PLAB, DLAB), dynamic processes in which learning occurs while taking the test (e.g. CANAL-F), and end stages (e.g. Hi-LAB) where elaborate foreign language skills have already been observed. Carroll (1981) argues that in some cases measures of achievement can be seen as measures of aptitude. Whether measuring aptitude in initial stages of foreign language learning, in dynamic ways, or in the end stage will reveal the same aptitude factors may be a matter of debate. Possibly those aptitude factors that appear constantly in all stages may be the primary aptitude factors, while those that occur in specific stages may be in the foreground in particular stages of foreign language learning. To illustrate this notion, while in initial stages of phonetic language learning (novel speech) working memory capacity predicts phonetic foreign language aptitude (Christiner & Reiterer 2013; Christiner et al. 2018; Christiner & Reiterer 2018; Christiner 2018), working memory capacity is not predictive for already acquired phonetic language

material when content is available to individuals (Coulmeil et al. 2019). Even though this does not necessarily mean that working memory does not predict achievement, it may not be found by testing individuals' abilities in other stages. This may be one more crucial aspect as to why it is quite difficult to understand the underlying mechanisms of aptitude.

Furthermore, language typology has additionally been said to influence aptitude results (Parry & Child 1990). The system of the first language also influences acquisition processes of the second language. This suggests that the aptitude constructs may also vary according to L1 and its transfer to the L2s learnt, which would require systematic testing and constructions of all possible combinations.

What has largely been ignored so far is the integration of measurements of music in language aptitude batteries. The sound system of music essentially consists of pitches and timbres, while that of language consists of vowels, consonants, and pitch contrasts (Patel 2007). Both music and language have temporal aspects (e.g. rhythm) and hierarchical structures (Limb 2006). Pitch perception has been related to language achievement by early work (Pimsleur et al. 1962), but since then has received little attention in the development of language aptitude batteries. In the past two decades musical aptitude and musical proficiency tests have very often been borrowed by language experts and included in language aptitude measurements (Fonseca-Mora et al. 2015; Christiner & Reiterer 2013, 2015; Christiner et al. 2018; Christiner & Reiterer 2018; Milovanov & Tervaniemi 2011). Results shows that especially phonetic aptitude and musical aptitude/abilities (Christiner & Reiterer 2015) as well as singing ability are interrelated (Christiner & Reiterer 2013; Ludke et al. 2014). Musical aptitude, abilities and experience are related to language functions and seem to predict parts of language aptitude as well. Additionally, music acquisition processes as well as individual differences in musical abilities are linked to that of language abilities, which is why musical abilities, aptitude, experience and achievement will be looked at next.

3. Aptitude and musicality

In the past two decades the scientific community has shown considerable interest in musical aptitude due to many reasons. One, for instance, is that the origin of aptitude is still poorly understood, while another is that musical aptitude seems to be related to multiple other cognitive domains. This may be because auditory signals do not only form the basis for musical aptitude, but also for other cognitive abilities (Moreno 2009). Therefore, investigations in musical abilities and aptitude seem to be a rich source for finding new ways

to facilitate learning processes or to develop new therapies that have positive effects on neighbouring faculties such as language. The findings of recent research facilitate the notions mentioned before as studies have increasingly shown that musical abilities go beyond the musical domain (Moreno 2009). Positive transfer effects of musical abilities and expertise have been related to working memory capacity in adults and children (Strait et al. 2011; Christiner & Reiterer 2018; Hansen et al. 2013; Christiner & Reiterer 2013; Christiner et al. 2018), a capacity also crucial for phonetic language aptitude (Christiner & Reiterer 2013, 2015; Christiner et al. 2018; Hu et al. 2013; Reiterer et al. 2011; Milovanov 2009; Milovanov & Tervaniemi 2011). Musical training, however, is also linked to reading ability (Fonseca-Mora et al. 2015), to duration perception in speech (Chobert et al. 2014), to speech segmentation (François et al. 2013), to speech processing (Besson et al. 2011), to pitch discrimination ability in speech (Moreno 2009), to verbal memory (Ho et al. 2003; Moreno et al. 2011), to IQ (Schellenberg 2011; Banai & Ahissar 2013; Lynn et al. 1989), to intellectual development (Schellenberg 2005), and synchronisation of brain activity (Wengenroth et al. 2014; Seither-Preisler et al. 2014; Serrallach et al. 2016). Even though the positive transfer from music to language is diverse, it also depends on research design and musical measurements which have been used. Thus it can be suggested that musical sub-abilities may contribute to particular language functions more likely than to others.

As shown, investigations on musicality and its relationship to other cognitive abilities are diverse and terms such as musical abilities, musical aptitude and musical experience are commonly used to refer to the same concept, this being musicality (Levitin 2012). Musicality, depending on the perspective, is not well understood, as especially nature and nurture related individual differences in musical performances have lead to a controversial debate over innate and acquired musical capacities involved in both, achieving mastery or high expertise (Levitin 2012), and in differentiating innate musical aptitude (which is synonymously used for musical ability) from training or experience related effects (Seither-Preisler et al. 2014).

To recap, in chapter 2 aptitude was defined as an ability that is elaborate, can predict high achievement and is recognised as an exceptional capability in a particular domain. Gagné (2008) has argued that aptitude is influenced by genetic endowment but is not necessarily innate, while other researchers (Nardo & Reiterer 2009) talk about aptitude as an innate gift. In foreign language research the entire aptitude complex is not well understood (Skehan 2019), which is quite similar for musical aptitude research. However, the interest in research into musical aptitude started relatively early compared to the field of language, if we consider

early work on musicality by Michaelis (1805), or Billroth (1895). It has to be noted that, apart from elaborate perception ability, the aesthetic and cultural dimensions were in the foreground. This has changed since the late 1960s, with studies and musicality tests developed by Gordon (Seither-Preisler & Schneider 2015).

In research on musical aptitude and achievement, the discussions about training related effects versus innate musical abilities are the logical consequence as knowing the origin of individual differences in musical performances is the driving source for scientists to understand both aptitude and achievement. Researchers have quite similar definitions for musical ability, aptitude and achievement. Gordon (1967b, 1997) and Law and Zentner (2012) say that musical abilities can be compared to potentials that can be seen as a kind of readiness to learn. These notions for musical abilities and aptitude are comparable to those of language abilities and aptitude defined by Carroll (1981, 1989, 1990), by Dörnyei (2005), by Wen (2012), or by Skehan (2019). Musical experience, proficiency and training are associated with achievement and mastery and are also similarly used in the context of music science. Achievement can essentially be regarded as what has been learned or adapted, but should not be equated with proficiency, which is easily tested and measured at any time and is a composition of both aptitude and achievement.

Gordon (1997: 42) also claims that the use of the terms “ability, talented, gifted and musical” are rather misleading and confusing and he argues that instead the concepts of achievement and musical aptitude need to be differentiated. “Whereas music achievement is intellectual and primarily in the brain, music[al] aptitude is spontaneous and primarily in the cells and genes, that is, in the entire body” (Gordon 1997: 42). According to him, musical aptitude consists of two main components, the first being innate ones, seen as potentials, and the second being based on environmental influences including formal as well as informal music (self-) education (Gordon 1997: xiii). Similarly, Sloboda (2008) argues that musical proficiency is comprised of the interactions between both innate and acquired musical capacities. Researchers therefore clearly suggest that innate predispositions or early acquired capacities are responsible for musical aptitude, which has been supported by several other investigations in different fields more recently as well (Schneider et al. 2002; Park et al. 2012; Oikkonen et al. 2015). Furthermore, proficiency is a combination of aptitude and achievement which, in turn, shows that it is very difficult or almost impossible to separate the two precisely. Another term, *audiation*, coined by Gordon (1997) should be the basis for both, musical aptitude and achievement. Audiation has been described as a process where music is comprehended in the mind some time ago (Gordon 1997: 5). This term should not be

compared to types of imitation such as rote memory learning, which is a product, while “audiation is a process” (Gordon 1997: 9). Audiation will not be discussed as it is irrelevant for this investigation.

3.1. Musical aptitude versus achievement

Following the discussion about musical aptitude and achievement research seeks to find ways to contrast both more precisely. Neuroscientific investigations attempt to come closer to fixing this conundrum by trying to differentiate training related effects from innate ones. Musicians’ brain structure, anatomical endowment and complexity have repeatedly been reported to differ from non-musicians in many respects (Schneider et al. 2002; Schneider et al. 2005; Macnamara et al. 2014; Manzano & Ullén 2018; Imfeld et al. 2009; Baumann et al. 2005; Bangert et al. 2006; Altenmüller 2008; Brown et al. 2013; Halwani et al. 2011; Koelsch et al. 2005a; Benner et al. 2017) whereby it is not entirely clear where these differences come from. This has led to two opposing branches. The first favours the idea that individual differences in musical performances may be more related to innate and early acquired abilities (aptitude), while the second branch sees training-related effects as driving forces.

Brain research has been able to show that anatomical differences can be related to training effects (Macnamara et al. 2014; Platz et al. 2014; Manzano & Ullén 2018). A recent study has shown that brain plasticity and its relation to training effects can explain around 21 percent of the variances in the performances of experts (Macnamara et al. 2014), while another reported this to be able to explain 36 percent (Platz et al. 2014). Twin studies, for instance, have revealed that the cortical thickness in the auditory-motor network is more complex on the left hemisphere in musically trained individuals compared to their counterparts (Manzano & Ullén 2018). This is also reflected in the volume of the grey matter in the left cerebellar hemisphere, as well as the white matter microstructure in both hemispheres and the corpus callosum, which were more developed in twins that were musically trained (Manzano & Ullén 2018). Musical training has an impact on the white matter architecture of the corticospinal tract (Imfeld et al. 2009) and stimulates non-auditory regions such as pre-motor and supplementary motor areas as well as the cortex of the sensorimotor, the parietal and the dorsolateral prefrontal motor areas (Baumann et al. 2005; Koelsch et al. 2005a; Bangert et al. 2006; Zatorre et al. 2007; Altenmüller 2008; Chen et al. 2008; Brown et al. 2013). Singing training improves the volume and complexity of the white matter tracts of the arcuate fasciculus and, in turn, the interplay between sound production and perception (Halwani et al. 2011).

Increasingly more evidence, however, shows that individual differences in musical abilities and aptitude are also partly based on inherent predispositions (Park et al. 2012). Neurophysiological and morphological changes of the brain structure that stem from genetic predispositions have already been discussed two decades ago (Schneider et al. 2002). Recently, genetic studies have shown that several genes affect inner ear development and the auditory pathway, which in turn can also be related to musical aptitude (Oikkonen et al. 2015). Multigenerational studies have also reported that musical aptitude is related to predisposing genes (Pulli et al. 2008). Gene-related predispositions, for instance, determine the structure of grey and white-matter (Peper et al. 2007). On average individuals of high musical expertise possess 130 percent more grey matter volume compared to non-musicians (Schneider et al. 2002; Schneider et al. 2005). It is suggested that the underlying structure of Heschl's gyrus located in the superior temporal gyrus that is considered to be the core region for auditory processes, is remarkably stable in size and shape from childhood on (Seither-Preisler et al. 2014). Generally speaking, Heschl's gyrus occurs most likely as a single gyrus (Rademacher et al. 2001; Yoshiura et al. 2000), while multiplications of Heschl's gyrus appear less often at around 30 percent in the general population (Morosan et al. 2001; Leonard et al. 1998). Multiplications of Heschl's gyrus are functionally relevant for musical practice and interestingly, occur much more frequently in musicians (Benner et al. 2017)

Current research facilitates the suggestions of Sloboda (2008) or Gordon (1997) and leaves no doubt that the proficiency someone achieves is a product of both aptitude and acquired musical abilities. Neuroscience has had a tremendous influence on contrasting aptitude with achievement. Brain research alone, however, cannot reveal information about musical aptitude and achievement and furthermore relies on behavioural testing and musical measurements which will be dealt with in the following sections.

3.2. Measuring musical abilities

To recap, understanding the musical aptitude complex requires distinguishing musical aptitude and achievement. Aptitude, the pure, untrained ability, and achievement, the more training related component are unfortunately difficult to measure separately (Christiner & Reiterer 2015) and there is no general agreement on how musical abilities can be measured as objectively as possible (Law & Zentner 2012: e52508). Indeed, measuring musical aptitude requires considering multiple factors such as the cultural influence on measurements, individual differences and requirements for playing different musical instruments, instrument playing versus singing, defining which musical abilities should be measured, participants'

selection, performance anxiety, and environmental influences such as (self-) education, social status of musical aptitude, motivation or (parental) support. Furthermore, these measurements need to be separated and defined in terms of whether they belong to, aptitude or achievement.

Aptitude tests can be targeted at participants who are untrained or have never been exposed to what is being measured (Christiner & Reiterer 2015). This is a difficult approach in the field of music as everyone listens to music, sings occasionally, and is at least trained to a certain extent perceptually. Others additionally learn to play one or more musical instruments early in life, or are surrounded by musicians as family members, while some are not. The selection of participants is of integral importance in order to come closer to an understanding of aptitude and to distinguish it from achievement. Researchers also work with group comparisons and select professional musicians (instrumentalists and vocalists), semi-professional musicians, and non-musicians to compare and contrast their abilities (Christiner & Reiterer 2015; Hansen et al. 2013; Gaser & Schlaug 2003; Zuk et al. 2014). However, in music research participants can also be classified as musical sleepers or sleeping musicians. The term musical sleepers refers to people who possess musical skills and if they get enough time and the opportunity to learn, reach high proficiency, while sleeping musicians are people who no longer train or play musical instruments despite having trained for years (Law & Zentner 2012). This causes difficulties in classifying or defining group membership.

Other researchers such as Seashore believed that measuring aptitude may be achieved best by testing children who are not influenced by education (Seashore 1919b) – an approach that is also used in current research (Christiner & Reiterer 2018) but has its limitations as well. Musical abilities of children are rather unstable before the age of nine (Sloboda 2005) since children generally lack the ability to produce proper pitches up to the age of three (Krumhansl & Keil 1982). Five-year-old children are sensitive to contour and rhythm (Costa-Giomi 2003; Peretz et al. 2004) and six to seven-year-old children are sensitive to the musical harmony of their particular culture (Schellenberg et al. 2005). This illustrates that it is difficult to measure children's full aptitude repertoire if it has not fully been developed at the time tested. Following Gordon (2013) musical aptitude is dependent on musical input, that is to say that, a rich environment will facilitate musical aptitude. This indicates that aptitude is developmental, which supports the Model of Giftedness and Talent by Gagné (2008) which says that aptitude turns into talent by training. The above discussion has shown that participants' selection alone does not settle the discussion about aptitude versus achievement yet.

Culture specific differences are other factors that influence measurements. Music is culture related with specific rules and norms (Sloboda 2005) which can be inflected in the rhythmic organisation (Brown & Jordania 2013), in the tonal system (Patel 2007), or chord progressions (Patel 2007). Isochronous versus non-isochronous metres are difficult to be detected if non-isochronous metres are not part of the musical system of someone's culture (Hannon & Trainor 2007). Tonal systems or chord progressions are also culture specific (Patel 2007) and have been referred to as a kind of "musical grammar" that listeners are familiar with (Jackendoff & Lerdahl 2006). Indeed, musical systems vary across cultures and differences and similarities should be considered in music measurements (Trehub et al. 2015; Law 2012). Cultural differences and musicality, however, are not discussed in detail here, but are relevant for the selection of the participants since they should be aware of the musical system the test is based upon.

As cultural-specific differences lead to different music rules and norms in all musical domains, musical abilities can be subdivided more specifically into abilities such as sense for melody, rhythm or harmonic perception, to name the most obvious ones. Early research postulated that those abilities can be measured independently and separated from each other (Seashore 1919b), while most researchers favour that musical sub-abilities are also intertwined (Gordon 1965). The measurements of musical aptitude, therefore, need to include several dimensions and sub-dimensions which have to be assessed and evaluated.

Furthermore, the musical aptitude of individuals varies considerably. There is already evidence to suggest that playing specific musical instruments has an influence on music perception and may even influence preferences for choosing musical instruments (Schneider et al. 2005) or vice versa. Pitch perception preferences of musicians have been related to instrument families showing that "[f]undamental pitch listeners played predominantly percussive or high-pitched instruments, whereas spectral pitch listeners preferred lower-pitch melodic instruments and singing" (Schneider et al. 2005: 390). Indeed, playing the piano is based on different skills compared to playing the flute. While the first one requires the ability to use the fingers to generate music, playing the flute additionally requires controlling breathing. Different transfer effects from music to language have also been observed among vocalists and instrumentalists (Christiner & Reiterer 2015).

Another aspect that should be considered is performance anxiety as it occurs not only frequently in musicians and causes considerable distress (Fehm & Schmidt 2006), but presumably also in non-musicians. Studies have reported that one third of musicians suffer from performance anxiety (Fehm & Schmidt 2006), which illustrates that such effects can

have a tremendous influence on music studies. Unfortunately, studies on musical aptitude often ignore the fact that musical performances such as humming of melodies or singing may cause distress for the participants. Even though it has been reported that musicians also suffer from performance anxiety, they may be advantaged in music performance, singing and language imitation tasks as many of them are used to performing while non-musicians may be more nervous or even ashamed. Therefore, measurements that integrate performance anxiety are highly relevant and need to be integrated in test designs.

As shown, multiple dimensions have to be considered when measuring musical aptitude. The development of aptitude tests, as well as how aptitude was understood in the past and is understood nowadays may give an impression as to what has already been demonstrated so far.

3.3. Historical overview of musical aptitude

The study of musical aptitude has undergone several changes since its origin. At the beginning of the 19th century musical talent was discussed by Michaelis (1805), who argued that musical taste and musical discrimination skills were the predominant indicators for musical talent. During this time the aesthetic dimension played a significant role in defining musical talent. Almost a century later, Billroth (1895) wrote about elaborate perception skills when referring to musicality. This may mark the starting point in the tradition of research in music perception which is still one of the most dominant ways of measuring musical aptitude, abilities and musical achievement at the present moment. Even though early attempts to define musicality, musical talent and musical aptitude have been of scientific interest in the 18th and 19th centuries, measuring one of the aforementioned aspects was first in practice with the emergence of the first musicality test by Seashore in 1919 (Grashel 2008). His test, which originally consisted of five subtests, was the only one during that time and had a tremendous influence on measuring musical aptitude and test development later on. Seashore (1919a, 1919b) believed that measuring musical abilities during childhood is important as they are not influenced by educational input that may bias aptitude measurements. For Seashore, musical aptitude was a gift given by God. This included that musical aptitude was believed to be predefined by nature, while nurture was responsible for the level someone could achieve (Gordon 1981). Interestingly, Seashore's view that musical aptitude was inherent and thus stabilises at birth led to a huge controversy over nature versus nurture issues in the field of music science in which especially music educators criticised that this would mean that education were useless if this view would held true (Gordon 1981). Gordon, decades later,

proposed that musical aptitude is also developmental (Gordon 2013) especially before the age of nine (Sloboda 2005), which is still the most dominant view presently and closely related to general models of aptitude (Gagné 2008).

Another well-known musical aptitude test was created by Wing in 1948. This test also belongs to one of the early developed musical aptitude measurements. It came into use for children at secondary-school age and attempts to measure “[...] acuity of musical hearing and sensitivity to performance” (Wing 1962: 39). The first tests such as those of Seashore and of Wing focus more on musical achievement (Seither-Preisler & Schneider 2015), while a new trend was instigated in the 1960s by Gordon. He then put more emphasis on musical aptitude, on complex music processing and on observing individual differences. For Gordon potentials and achievement rely on audiation. His notions are still favoured in research on musical aptitude (Seither-Preisler & Schneider 2015). Even though Gordon’s musicality tests are widely used in the field of science, and for music education, they have also been criticised recently. The AMMA test, for instance, has been criticised for its complexity of the tasks (Law & Zentner 2012; Law 2012), or the PMMA for showing deviations from the norms represented by the manuals (Gouzouasis et al. 2007). Newer musical aptitude measurements such as the Profile of Music Perception Skills (PROMS) test developed by Law and Zentner (2012) focus on less complex stimuli than Gordon’s tests and aim at being more culture neutral (Law & Zentner 2012). More recently self-report inventories that are designed to reveal musical abilities (Müllensiefen et al. 2012), or music imitation tests (Wallentin et al. 2010) are available as well. Singing ability is most often tested by imposing performance tests of new or well-known songs (Christiner & Reiterer 2013, 2015; Christiner et al. 2018) or by measuring pitch accuracy (Salvador 2010). In the following section a number of well-known published musical ability and aptitude tests and their measurements are presented.

3.4. Musical aptitude tests and measurements

A number of musical aptitude or ability tests have been developed which are based on different assumptions or vary according to the musical abilities measured. Well-known musicality tests include the Seashore test (Seashore 1919a, 1919b), the Wing Standardised Tests of Musical Intelligence (Wing 1948), Measures of Musical Abilities (Bentley 1966), Musical Aptitude Profile (Gordon 1965), Primary Measures of Musical Audiation (Gordon 1979), Intermediate Measures of Musical Audiation (Gordon 1982), Advanced Measures of Musical Audiation (Gordon 1989a) and Audie (Gordon 1989b). More recent test batteries are, for instance, the Musical Ear Test (Wallentin et al. 2010), or the Profile of Music Perception

Skills (Law & Zentner 2012). Apart from musicality tests, a number of tests have also been developed for clinical purposes, such as the Clinical Assessment of Music Perception test (Nimmons et al. 2008), or self-report assessments of musical skills, such as the Goldsmiths' Musical Sophistication Index (Müllensiefen et al. 2012). A selected sample is shown in the following section.

3.4.1. The Seashore test (Seashore 1919a, 1919b)

Seashore (1919a, 1919b) was fully aware of individual differences in the capacity to learn musical instruments or to sing. Therefore, he believed that musical talent is best analysed early in life as to him musical capacities should be measured before they are influenced by music education (Seashore 1919a, 1919b). The test consists of six sections as shown in table 9 below (Seashore 1919a, 1919b).

Table 9. The Seashore Test

The Seashore Test	
Age	10 to 16
Part I	measures the <i>Sense of Pitch</i>
PartII	measures the <i>Sense of Intensity</i>
PartIII	measures the <i>Sense of Time</i>
PartIV	measures the <i>Sense of Consonance</i>
PartV	measures <i>Tonal Memory</i>
PartVI	measures the <i>Sense of Rhythm</i>

The *Sense of Pitch* condition is a discrimination task where participants have to decide whether the second of two tones is higher or lower than the first. The *Sense of Intensity* condition focuses on loudness or strength where participants have to discriminate “whether the second is weaker or stronger” (Seashore 1919a: 12). The *Sense of Time* condition is based on hearing three clicks, whereby the two intervals between the clicks are of interest. The second interval is either shorter or longer. For the *Sense of Consonance* condition, participants listen to two different tone combinations where they need to decide which combination sounds better. In the *Tonal Memory* condition “a series of tones [is] played twice” (Seashore 1919a: 15). The second time the series is played, one note is different and has to be identified. The *Sense of Rhythm* condition is a discrimination task where the patterns in each pair are different or the same (Seashore 1919b).

3.4.2. Measures of Musical Abilities (Bentley 1966)

To measure musical abilities Bentley (1966) developed a battery that consists of four separate tests. They were designed to measure abilities that are fundamental for music making for both instrument playing and singing (Young 1973). For Bentley the melodic phrase was “[t]he most elemental form of music [...]” (Young 1973: 74) and his interest in understanding individual differences in musical aptitude originated from his observation that some children remain monotones while others are not, even though their environments or homes are comparable (Bentley 1969). The subtests are shown in table 10 (Bentley 1969: 41–42).

Table 10. Measures of Musical Abilities

Measures of Musical Abilities	
Age	between 7 to 14 years of age
Part I	<i>Pitch Discrimination</i> subtest
Part II	<i>Tonal Memory</i> subtest
Part III	<i>Chord Analysis</i> subtest
Part IV	<i>Rhythmic Memory</i> subtest

In the *Pitch Discrimination* subtest, participants need to discriminate between two pitches and indicated whether the second pitch is the same, higher or lower. In the *Tonal Memory* subtest, participants are instructed to count the number of pitches in a melodic fragment. This melody is repeated a second time, and the participants have to indicate the number of the one pitch that has been altered. The *Chord Analysis* test requires the participants to identify the number of pitches in each chord presented. The *Rhythmic Memory* subtest consists of rhythmic patterns that are repeated. Participants have to indicate whether the patterns are the same or different. (Bentley 1969: 41–42; Young 1973: 75).

3.4.3. Gordon’s music aptitude tests

A series of music aptitude tests have been developed by Gordon. In chronological order, the first test, The Musical Aptitude Profile (MAP) was created in 1965, followed by Primary Measures of Musical Audiation (PMMA) in 1979, Intermediate Measures of Musical Audiation (IMMA) in 1982, Advanced Measures of Musical Audiation (AMMA) in 1989 and Audie in 1989. The latter is a test for three-to-four-year-old children and consists of tonal and rhythmic subtests which should be carried out at different times (Gordon 1989b).

The first of Gordon's tests, the MAP, consists of three main parts and is split up into seven subtests. It was originally developed to inform parents of their children's musical aptitude, to identify musically talented students, to observe individual differences and to provide students with proper materials so that they improve (Gordon 1967a: 1). Music education is no criterion to take the test which mainly consists of subtests where participants have to compare musical passages and indicate them as being either correct or incorrect, while in the preference tests students are instructed to decide which of the paired musical performances are more tasteful (Gordon 1967a). Table 11 shows Gordon's MAP sections below.

Table 11. Musical Aptitude Profile (MAP)

Musical Aptitude Profile (MAP)	
Age	between 9 to 19 years
Part I	<i>Tonal-Imaginary: Melody and Harmony</i>
Part II	<i>Rhythm-Imaginary: Tempo and Metre</i>
Part III	<i>Musical Sensitivity: Phrasing, Balance, and Style</i>

The PMMA and the IMMA were developed for children. Both consist of two main sections that measure tonal and rhythmic ability where the participants have to listen to two musical statements and rate them as to either the same or different (Gordon 1979, 1982). The individual subtests constitute separate sections. The PMMA test has especially been criticised for showing deviations from the norms published (Gouzouasis et al. 2007; Stamou et al. 2010). The test is also very long for children at this particular age, which makes it very difficult to be run within a single sitting (Christiner & Reiterer 2018).

Table 12 and table 13 show the measurements of the PMMA and the IMMA.

Table 12. PMMA

Primary Measures of Musical Audiation (PMMA)	
Age	Kindergarten through Grade 3
Part I	measures <i>Tonal Ability</i>
Part II	measures <i>Rhythmic Ability</i>

Table 13. IMMA

Intermediate Measures of Musical Audiation (IMMA)	
Age	Grade 1 through 6
Part I	measures <i>Tonal Ability</i>
Part II	measures <i>Rhythmic Ability</i>

The AMMA test consists of rhythm and tonal discrimination tasks and is similar to the PMMA and IMMA except for the test design. The paired musical statements are embedded in one single test design where either rhythmic, tonal or no changes may occur. For the tonal discrimination tasks notes are modified in the different condition when the melody is played for the second time. In the rhythm subtest, tempo, metre or duration can be altered in the different condition (Gordon 1989a). Table 14 shows the measurements of the AMMA.

Table 14. AMMA

Advanced Measures of Musical Audiation (AMMA)	
Age	Adolescents and adults
Condition I	measures <i>Tonal Ability</i>
Condition II	measures <i>Rhythmic Ability</i>

3.4.4. The Musical Ear Test (Wallentin et al. 2010)

The Musical Ear Test is quite similar to Gordon's AMMA and has a rhythm and melody subtest. The difference is that both subtests are done separately in marked contrast to the AMMA test where both conditions "[...] are embedded in the same melodic context [...]" (Wallentin et al. 2010: 189). Table 15 shows the measurements of the Musical Ear Test.

Table 15. The Musical Ear Test

The Musical Ear Test	
Age	presumably adults
Part I	<i>Melody Discrimination</i> task (including pitch and contour)
Part II	<i>Rhythm Discrimination</i> task

Within the *Melody* subtest melodies range between 3-8 tones. In the different condition one pitch violation occurs. With (Law & Zentner 2012) in the violations half of them

“[...] constitute contour violations [...]” (Wallentin et al. 2010: 189). In the *Rhythm* condition pairs of “rhythm sequences [that] contain 4–11 wood block beats” (Wallentin et al. 2010: 189) are presented and need to be discriminated. In the different condition one rhythmic change has been introduced.

3.4.5. Profile of Music Perception Skills (Law 2012; Law & Zentner 2012)

The original test consists of nine subtests and lasts around 60 minutes. The stimuli are not complex but instead focus on basic sound properties. The reason for this is that the authors raise criticism in testing musical ability if tested by more complex tasks such as in Gordon’s AMMA test. Complex musical stimuli measure a combination of skills and may be less informative. However, there are also a number of other reasons for the development of the PROMS such as overcoming a lack in valid statistical analysis of other music aptitude batteries, or presenting a more culture neutral version (Law & Zentner 2012). Table 16 shows the measurements of the PROMS test.

Table 16. PROMS

Profile of Music Perception Skills (PROMS)	
Age	general adult population
Part I	<i>Melody</i>
Part II	<i>Standard Rhythm</i>
Part III	<i>Rhythm-to-Melody</i>
Part IV	<i>Accent</i>
Part V	<i>Tempo</i>
Part VI	<i>Pitch</i>
Part VII	<i>Timbre</i>
Part VIII	<i>Tuning</i>
Part IX	<i>Loudness</i>

The *Melody* subtest is a discrimination task where paired musical statements are either indicated to be the same or different (Law 2012). The test becomes increasingly difficult by conditions of atonality and note intensity. The standard *Rhythm* condition is a paired discrimination task which contains three levels of difficulty where listeners have to state whether the paired samples are same or different (Law 2012). The *Rhythm-to-Melody* condition aims at “[...] targeting listeners’ ability to recognize rhythmic patterns when it is no

longer provided in its original form [...] but embedded in a melody” (Law & Zentner 2012: 4). The *Accent* “[...] subtest assesses skills in discerning the relative emphasis given to certain notes in a rhythmic pattern. As such, it is related to the concepts of meter in music [...]” (Law & Zentner 2012: 5). This task also consists of three levels of difficulty. In the *Tempo* subtest participants have to distinguish whether paired musical stimuli have the same or a different tempo. The *Pitch* test requires the participants to identify pitch changes by indicating whether the second pitch presented is lower, higher or the same (Law 2012). For this condition, particular attention has been paid to the test design by using pure tones to avoid that pitch changes are detected “as a result of harmonics [...]” (Law & Zentner 2012: 6). In the *Timbre* subtest chords were produced “[...] with the possibility for making very subtle changes (Law & Zentner 2012: 6)”. Here there are also three levels of difficulty. In the *Tuning* condition two chords are presented where listeners have to decide which of the two is more “discordant” or “mistuned” (Law 2012). In the last condition *Loudness* has to be defined by indicating whether the second tone of the two presented is the “same”, “louder”, or “softer”

3.4.6. Goldsmiths’ Musical Sophistication Index (Müllensiefen et al. 2012)

The Goldsmiths’ Musical Sophistication Index (Gold-MSI) is a test that aims to measure musical sophistication and is a self-report questionnaire inventory. The developer carefully reviewed existing literature that focuses on musical behaviour and abilities and decided to select five major dimensions, “namely resource allocation to music, music making, functional use of music, ability to verbalise musical experiences, and perceptual-cognitive skills” to write items that should reveal those concepts (Müllensiefen et al. 2012: e89642). After multiple preliminary testing and dimension reductions the Gold-MSI resulted in five factors that ought to assess musical skills, see table 17.

Table 17. Gold-MSI

Goldsmiths’ Musical Sophistication index (Gold-MSI)	
Age	any age
Subscale I	<i>Active engagement</i>
Subscale II	<i>Perceptual abilities</i>
Subscale III	<i>Musical training</i>
Subscale IV	<i>Singing abilities</i>
Subscale V	<i>Emotions</i>
General factor	<i>General Musical Sophistication</i>

3.4.7. Performance tests

In addition to standardised tests that measure musical abilities, a number of performance tests have been developed which are rated by experts or non-experts. These include, for example imitation tests where participants are instructed to reproduce rhythmical sequences by using hand-claps, to repeat melodies or to sing (Wallentin et al. 2010). One advantage of this test is that musical proficiency is revealed quite easily. One disadvantage is that ratings can vary, which requires the rating of the performances by multiple people, while another is that the tasks developed vary according to the type of research and are incomparable (Wallentin et al. 2010).

3.4.8. Discussion

There are multiple musical aptitude tests available with creative designs that aim to measure individual musical abilities and provide information about musical aptitude, musical experience and sophistication. At this stage several aspects need to be discussed. The first discussion concerns the large number of perception tests that exist, and will include a summary of their measurements, the second one is a critical reflection of why perception tests are favoured over production tests, and the third concerns the question as to what else should be considered in measuring musical aptitude.

Most measurements include same or different conditions that are discrimination tasks focusing on rhythm or melody (Gordon 1979, 1982, 1989a, 1989b; Wallentin et al. 2010). Musical aptitude and ability tests include measurements such as consonance/dissonance, pitch contour perception, harmonic perception, style recognition, melody memory, loudness or tone pitch and pitch intensity tasks.

Generally speaking, most musical aptitude measurements can be classified as perception tests, in which parts of music stimuli are not imitated or repeated, while production tests represent the opposite. For instance, singing or pitch discrimination have often been classified as production and perception test respectively, forming an opposing pair. In phonetic aptitude research, speech perception and speech imitation of unfamiliar languages represent similar pairs which allow the separation of perception from production. Even though production is not free of perceptual influence, being unfamiliar with the language perceived or imitated at least facilitates that the linguistic content is reduced. Production measurements of both music and speech are more difficult to analyse as they are often rated and evaluated by experts and non-experts.

The latter may also be one reason for the larger number of music perception tests as they are more easily executed. Furthermore, they can be computerised, which makes them more easily accessible, test conditions can be equalised, and the tests can be validated and standardised.

Musical aptitude, however, is not only associated with music perception ability, but also with performance. Presumably, it could be suggested that those who possess high perception skills are additionally more advanced in music performance or singing. Even though this is generally the case, research has also shown that accurate perception does not always predict accurate performance in phonetic language tasks and in music (Christiner & Reiterer 2015; Golestani & Pallier 2007). Excellent performances require further aspects which are of a more kinaesthetic and somatosensory nature, aspects that are crucial for singing (Kleber et al. 2010). Research has shown that optimisation of performance while training leads to an increase “in the volume of the primary motor cortex” (Nielsen & Cohen 2008: 65) and affects motor skills positively. Furthermore, language studies have revealed that vocal motor skills influence the accurate perception of language (Nasir 2009). This may indicate that testing music perception alone may not be enough to show the musical aptitude of individuals.

Moreover, tests such as the Gold-MSI show that musical abilities can also be measured by using questionnaires (Müllensiefen et al. 2012). However, it is doubtful whether using questionnaires without musical perception or performance tests can reveal a multidimensional impression about the musical abilities of individuals. In the same vein the background, e.g. music education, socio-economic status and environmental influence need to be measured as well in order to understand the influence of individuals’ surroundings. Further aspects such as performance anxiety (Fehm & Schmidt 2006), motivation (Sichivitsa 2007) and personality traits (Kawase 2016) also play a major role in how someone performs in perception or performance (production) tests. This indicates that essentially three primary factors are essential to measure: pure musical abilities (musical perception and performance tests), environmental influence (e.g. music education, socio-economic status, background) and the mechanisms that influence the performance while testing takes place, e.g. performance anxiety.

4. Singing

Singing is a very complex human vocal behaviour that has multiple functions in society and can be found across all cultures (Mithen 2007; Hutchins & Peretz 2012). The origin and role of song has often been discussed in evolutionary hypothesis (Livingstone 1973; Cross 2001; Hockett et al. 1964; Newman 1905; Darwin 1871; Fitch 2006; Huron 2001) and song is still a source for multiple speculations about the origin of speech and the role of music. The origin of song has been suggested to derive from emotional speech (Newman 1905), and its function has been related to “territorial songs [that] preadapted the hominids to both speech and symboling” (Livingstone 1973: 26). Song may have played a significant role in human courtship in ancient times (Darwin 1871), originally may have served to strengthen social ties (Arbib 2005) and have facilitated cementing social bonds (Patel 2007). Although most evolutionary notions are merely speculations, as no fossil recordings and hardly any evidence of the functions are available, song is still rooted in most ceremonial activities across cultures today (Mithen 2007). Singing is seen as a universal human vocal behaviour (Welch 1994) and is a crucial device to pass on cultural knowledge and customs (Tsang et al. 2011). It can reveal basic emotions such as fear, sadness and happiness (Fritz et al. 2009) and group music-making most likely results in a “shared mood state” (Patel 2007: 370) that has uniting characteristics.

Song is not only a human vocal behaviour but occurs in animal species such as apes and whales as well (Patel 2007; Masataka 2009). Even though its functions for animals may be different, animals are also capable of creating new ones (Patel 2007). Song is omnipresent in everyday life and, even infants’ vocalisation starts out more musical compared to adult speech, resembling singing more than speech (McMullen & Saffran 2004; Tsang et al. 2011). The relationship between singing, speaking and music making is very obvious on multiple grounds as they require the integration of similar sensory-motor and acoustic information. Thus it is quite surprising that interdisciplinary investigations in singing processing, generation and its influence on other cognitive domains have largely been neglected in the field of science.

This may depend on multiple factors. Some of the reasons that are important for this investigation will be explained in the following section. Other singing research directions which are of minor importance will be mentioned as well but not further discussed here.

Going back to reasons as to why singing research and why interdisciplinary research is less frequently conducted the following five main areas are discussed below. First, singing is considered to be a musical subcategory and is not always distinguished from other musical

phenomena, meaning that it is often not explicitly mentioned. Second, music research often focuses on music perception instead of production, which creates the impression that musical aptitude is more or less an elaborate perceptual skill. Third, singing has often been considered to be a privileged skill which requires instruction. This has led to the idea that singing research on non-professionals may be concerned with difficulties and deliver results of minor importance. Fourth, singing as a hybrid category has to integrate language and musical components, for instance, in song memory, which requires the ability to work in an interdisciplinary way from multiple angles. Fifth, singing research can additionally be split up into singing research on non-professionals and professional singing which are often investigated by different disciplines. This includes different approaches, criteria, emphasis and measurements within research on singing.

4.1. Singing – a musical subcategory

Apart from librettology, in which literary and musical aspects of a text are differentiated, singing is often implied in music research as a musical “subcategory”, despite the high loads of speech elements which are included in singing.

This is why singing ability is often not a separate category in music research. On the level of acoustic perception, singing is closer to music, whereas on the level of production, it is closer to speech (Christiner & Reiterer 2013; Christiner & Reiterer 2019). Basic fundamental musical abilities are required for both singing and instrument learning. Singing and playing musical instruments both require similar perceptual abilities that are usually discussed in the context of musical aptitude. Tonal and rhythmic discrimination abilities are in the foreground (Gordon 1989a, 1997; Law & Zentner 2012) but with regards to motor ability, instrumentalists learn a new skill, while singers train their vocal apparatus (Kleber et al. 2010). Interestingly, well-known musical aptitude tests such as the Seashore test have also been used to assess and to speculate about singing ability (Seashore & Metfessel 1925), although they predominantly measure perceptual musical abilities as opposed to song production. Pitch perception ability has also often been used to assess singing accuracy (Phillips & Aitchison 1997a, 1997b; Phillips et al. 2002), which of course questions whether this will reliably provide information about singing ability.

Despite the closeness of speech and song in terms of their production, singing is astonishingly very seldom discussed as a language phenomenon, while, in marked contrast, poetry, with its music-like rhythmic structure, is always discussed in the context of language. The reason for this may be that our perceptual system is defined to place anything which

carries melody into the category of music. This can easily be shown as basic emotions of musical pieces are still interpretable, even if the content of the lyrics is not understood (Balkwill & Thompson 1999). Lyrics are less important in song than melody, while it is arguably the opposite in language and poetry.

4.2. Music and speech: perception and structural research

Cognitive scientists, psychologists and musicologists, on the other hand, often focus on general musicality and prefer to evaluate perception abilities. As a matter of fact, perceptual musicality tests are often computerised, comparable, executable, can more easily be evaluated and statistically analysed, in marked contrast to performance tests (see also chapter 3). The way musical aptitude is measured and the large amount of research in this particular field creates the impression that perception ability is the primary source that denotes musical aptitude.

A large corpus of research that focuses on overlaps between music and speech processing does exist (Peretz et al. 2015; Wong et al. 2007; Rogalsky et al. 2011; Patel 2011; Ding et al. 2017; Asaridou & McQueen 2013; Perrachione et al. 2013). For instance, overlaps between music and speech perception have been discussed by Patel (2011), who introduced the “OPERA” hypothesis. This hypothesis consists of five categories (overlap, precision, emotion, repetition, and attention) that are involved in musical training and improve not only exclusive reading abilities, but other linguistic abilities as well.

Researchers also put emphasis on structural elements of shared and distinct features of music and speech (Krumhansl & Keil 1982; Patel 2003; Patel & Daniele 2003; Patel 2007; Jackendoff & Lerdahl 2006; Koelsch et al. 2005b; Limb 2006). Both music and speech consist of tonal properties, vocabulary, hierarchical structures, temporal features (Limb 2006: 437) and are compared with each other on this basis. Comparisons between language and music based on their properties such as pitch, timbre rhythm as well as their structural similarity have been discussed by Patel (2007) and Jackendoff and Lerdahl (2006), to name but a few.

Approaching from these directions is more static compared to work on singing performance, which is a highly dynamic process, complex and difficult to assess objectively. The perceptual and the structural approach to comparing music and speech both ignore, or do not necessarily have to include, the production part and consequently, findings are not easily transferable and applicable to singing. While perception ability with regards to music and

language is included in this investigation, comparisons of shared and distinct structural features of music and speech will mostly be omitted as they are irrelevant for this thesis.

4.3. Singing and the general population

Generally speaking, it has often been suggested that singing requires training, which has led to assumptions that without vocal instructions singing ability is limited (Dalla Bella et al. 2007), even though quite the opposite has been shown and non-professional singers remember pitch and tempo of familiar songs astonishingly well (Amir et al. 2003). The ability to carry a tune has also been shown to be quite similar among both professional and non-professional singers (Dalla Bella et al. 2007), which demonstrates that singing is a vocal behaviour that is accessible to the majority. The notion that singing has been equated with being an extraordinary skill may be one reason as to why researchers find it difficult to hire participants. Moreover, researchers probably did not expect to get meaningful results from their studies.

4.4. Singing as a hybrid of music and speech: memory

Another aspect concerns the integration of several disciplines. As singing is a hybrid category (Christiner & Reiterer 2013) and consists of two main components, text and melody (Crowder et al. 1990), it requires the constant integration of different disciplines. Song memory is the best example to show that singing demands for looking at both, music and language processes.

Several suggestions regarding memory of song have been put forward and especially lesion studies on aphasics seem to give interesting findings. Depending on their impairment, some abilities are spared while others are damaged. For instance, the observation that aphasics can sing words with high linguistic accuracy while they are unable to reproduce the same utterances in speech was already been observed decades ago (Gerstman 1964; Geschwind 1971; Keith & Aronson 1975; Hébert et al. 2003). However, how song is stored and memorised has still not been clarified, as familiar song singing is often retained in aphasics, while new production of speech is impaired. Various notions do exist and lesion studies have given rise to speculations as to how song is stored.

It has been said that the storage of song is likely a dual, separate system for lyrics and melody (Peretz et al. 2004; Stahl et al. 2011), while other researchers speculate about the existing of a special storage of song (Peretz et al. 2004). Others have argued that music and language processes are not simply aligned with each other in song (Peretz et al. 2004), and

notions that song is comprised of integrated speech, text or melody (Steinke et al. 2001; Peretz et al. 2004) have been discussed as well. A key role has also been ascribed to melody which may serve as a mnemonic with which utterances are stored in the long-term memory (Gordon et al. 2010).

As a consequence of the relationship between music and speech in singing, musical therapeutic interventions for regaining the ability to produce spontaneous speech in aphasics have been developed. The Melodic Intonation Therapy, for instance, is based on musical perceptual and generational input and consists of four crucial components, the inner rehearsal, the auditory-motor feedback training, the intonation (singing) and the left-hand tapping (Norton et al. 2009). Investigations have come to different conclusions about why this therapy seems to positively affect spontaneous speech production. While it has been argued that intonation ability (intoned singing) facilitates speech recovery (Norton et al. 2009; Schlaug et al. 2010), it has also been opposed by findings of other researchers that concluded that the left-hand tapping may be the key element (Stahl et al. 2011). A very recent study tested aphasics' singing ability and concluded that even though aphasics show a high fluency in familiar song singing, this benefit is not transferable to singing of unfamiliar lyrics (Kershenbaum et al. 2019).

Lesion studies on aphasics are sources for illustrating the complexity and special position of song, which requires researchers to put emphasis on music and speech domains to understand which processes are intertwined and which ones are separable.

4.5. Research about non-professional and professional singing

Research on singing can be divided into two main branches, the first that investigates non-professionals, and the second that focuses on professional vocalists. Both focus on song production and perception, but song production is more emphasised in the field of professional singing, which is also reflected in the nature of training. For both the groups production of song is the central tool to measure singing ability. Professional singers, however, focus on more precise criteria and include more parameters that denote voice quality (Sundberg 1988), which has been said to improve with training of the vocal motor apparatus (García-López & Gavilán Bouzas 2010).

Contrarily, pitch accuracy has predominantly been related to singing accuracy in non-professional singing research (Demorest & Pfordresher 2015; Pfordresher et al. 2015; Dalla Bella 2015; Goetze et al. 1990; Dalla Bella et al. 2007; Demorest et al. 2015), even though contradictory results about their relationship have been reported as well (Berkowska &

Dalla Bella 2009). This relationship has led to the fallacy that pitch perception is the measurement for singing. Opposing this, increasingly more evidence has been provided that accurate perception does not always predict accurate production of singing ability (Loui et al. 2009; Tremblay-Champoux et al. 2010; Pfordresher & Mantell 2014; Berkowska & Dalla Bella 2009; Hutchins & Moreno 2013), which shows that in the past ten years more and more alternative explanations have been approved. A similar phenomenon has also been reported in phonetic aptitude in language research (Christiner & Reiterer 2015; Golestani et al. 2011; Christiner & Reiterer 2019). This leaves no doubt that more research is needed in order to understand the role of the production processes.

To recap, as has been shown, research into singing requires a very precise approach as to what should be analysed. This includes that language and music processes cannot be ignored as both are involved in singing. Interdisciplinary work, where singing ability should additionally be compared to other cognitive domains therefore needs to be aware of the hybrid status of singing, of the differences in professional and non-professional singing criteria, and should not assume that musical aptitude is always transferrable to singing ability. This is the first step towards understanding individual differences in singing performance. However, the next step entails that all the other parameters which are discussed in individual differences in (foreign) language aptitude and in musical aptitude (e.g. motivation, performance anxiety, personality, learning style, IQ, environmental influence, etc.) need to be consulted in singing research as well and added where they are meaningful.

4.6. Singing research

In previous chapters, terminology for aptitude and achievement for both language and music have been defined quite thoroughly and precisely before going into detail. In the field of singing it is much more complicated, which is why this chapter does not continue with definitions of aptitude, achievement and proficiency, nor does it give an historical overview of singing research. Definitions and terminology will be given and discussed at the end of this chapter.

In the paragraphs before it has been demonstrated that singing research is quite complex and requires the integration of multiple dimensions which should be related to the type of investigation. As this thesis mainly aims to reveal the relationship between singing ability and phonetic language aptitude, the generation of song and of utterances in foreign languages are in the foreground and aptitude is the core term.

This predefines the following three sections and shows as to why it is worth putting emphasis on production processes in singing and to coin a term such as “singing aptitude” which is distinct from musical aptitude at the end of this chapter.

The first section focuses on the acquisition of singing and music perception and production, which illustrates how song and music is learnt. This reveals information about abilities that can be separated and show analogies to acquisition processes in speech.

The second section is based on the idea of language aptitude research as illustrated in chapter 2. There it has been demonstrated that understanding aptitude can also be approached from analysing high attainment groups, that is to say, people who possess of high language aptitude, who have already reached a high proficiency in one or more foreign languages (Linck et al. 2013; Doughty et al. 2008). Similarly, this can also be done for singing research, which has to include the perspective of singing experts and professionals, the high end state of singing, including their vocal training strategies and measurements to get an impression of elaborate singing abilities and performances that stem from aptitude for singing.

The third section concerns singing measurements and criteria for evaluating the singing performance of children, professional and non-professional singers to illustrate the various ways singing ability is rated and assessed. Looking at shared and distinct approaches towards analysing singing acquisition processes, professional and non-professional singing research will provide information about which singing abilities may be involved in elaborate singing and consequently reveal abilities that are outstanding and which may differ from non-professional singing.

4.6.1. Singing development – childhood

While research into language acquisition processes is highly diverse, music acquisition (Schellenberg et al. 2005) and singing acquisition processes during childhood are underrepresented in the literature (Tsang et al. 2011). Following Cohen (2008), shared and distinct processes of language and singing acquisition processes as well as the question as to whether there may be sensitive periods in which singing ability develops in a better and more elaborate way, can reveal interesting information which may also give insights to understand other domains.

Language acquisition processes are said to include critical and sensitive periods that are age related (Hernandez & Li 2007; Birdsong & Vanhove 2016), which is similar for music (Gordon 2013; Sloboda 2005; Patel 2007) and is presumably also true for singing acquisition processes (Cohen 2008; Tsang et al. 2011). This is not surprising since the latter develop

alongside language acquisition processes and are intertwined in many respects. Generally speaking, most singing acquisition research focuses on pitch discrimination ability and singing accuracy, even though contradictory results about their relationship have been presented for children (Goetze et al. 1990; Salvador 2010). One study reported a significant relationship between vocal accuracy and pitch discrimination abilities for primary school children (Stauffer 1986), while another study showed that vocal accuracy and pitch discrimination did not yield correlations in nursery school children (Apfelstadt 1984). More recently, research on pre-school children has revealed that singing ability is related to pitch but not to rhythmic discrimination ability (Christiner & Reiterer 2018). Following the discrepancy of results, a systematic research that focuses on singing acquisition processes is needed. Even though considerably more research on singing acquisition is available, compared to research on singing ability of non-professionals, a general lack of studies on singing may also be related to the fact that there are hardly any unified rules on how singing ability is measured (Demorest et al. 2015).

Understanding the mechanisms of singing and music acquisition processes requires looking at caretakers' input, lullabies, the development of vocalisation of infants, and early and late childhood.

4.6.1.1. Caretaker input and lullabies

Music and language acquisition processes do not simply start at birth but are influenced prenatally (Kisilevsky et al. 2003; Lecanuet et al. 1987; Groome et al. 1999; Welch 2006). The prosodic features of the mother's voice (Welch 2006) as well as certain musical elements (Hepper 1991) are familiar to infants at birth. This kind of preconditioning leads to the fact that infants prefer and are more attentive to maternal singing as opposed to other voices (Welch 2006). Lullabies occur in all cultures and are based on similar concepts (Trehub & Trainor 1998). Maternal singing is one way to strengthen the bonds between infants and their mothers (Patel 2007; Tsang et al. 2011), which may also be related to the fact that children imitate female vocals more accurately compared to male vocals (Green 1990). Song targeted at infants has multiple functions such as calming down (Trehub 2001), strengthening emotional ties (Trehub et al. 1997), but also learning to distinguishing vocal timbres that evoke negative or positive feelings (Welch 2007).

To recap, language and music acquisition processes take place simultaneously and the input infants receive shapes their vocal production beyond infancy (Welch 2006). Language targeted at infants resembles song as it is slower, carries longer pauses and includes

more pitch variations compared to adult speech (McMullen & Saffran 2004). Maternal singing is also very simple, slow and contains higher pitch levels compared to songs directed at adults (Welch 2006). Studies have revealed that mothers produce acoustically more extreme language features (Kuhl et al. 1997) which facilitate developmental processes of the infants. Singing ability is age-related (Davies & Roberts 1975; Goetze et al. 1990; Geringer 1983; Goetze 1989) and requires the integration of multiple abilities that depend on physical growth and are developmental (Sergeant & Welch 2008).

4.6.1.2. Infancy stage and early childhood

In the infancy stage, distinguishing vocalisations which belong to speech or song is highly challenging as “[...] our adult perception of such vocal acts is often blurred by a relative linguistic and musical immaturity and by our own desire to make sense of the vocalisations” (Welch 1994: 3). The first year of infancy is a period in which vocal activities are rather diverse. Evidence has been provided that at around two months of age infants’ vocalisation can contain musical aspects (Tafuri & Villa 2002), musical babbling starts at the age of five to six months (Moog 1976), followed by the development of fundamental aspects of vocal motor skills (Locke et al. 1995) which are required for both singing and speaking. The babbling song phase has been said to be the starting point of singing development (Welch 1994). The first songs are generated at around the age of one (Dowling 1999). Infants in their first year can detect deviations from standard melodies (McMullen & Saffran 2004), perceive melodic contours and temporal information (Costa-Giomi 2003) and their abilities to perceive musical elements seem to be ahead of their ability to reproduce any vocalisations which develops alongside motor control. Singing development after the first year is influenced by a variety of different singing behaviours of the particular culture in which children grow up (Welch 2006).

Children between the ages of two and three still lack the ability to produce proper pitches (Costa-Giomi 2003; Krumhansl & Keil 1982) and spontaneous singing is characterised by limited pitch and vocal range as well as instability of key (Welch 2007). Studies about voice quality at this particular age are still rare; in fact, with the exception of some studies that conclude that quality of voice is age- and experience-dependent (Leighton & Lamont 2006; Rinta & Welch 2009), such studies do not exist. This may largely rely on the fact that physical growth is not yet completed at this point (Tsang et al. 2011). Quality of voice for professional singing has also been suggested to improve by training and can last many years (Omori et al. 1996).

Comparing singing acquisition processes to those of language acquisition, it is obvious that at the age of three children's linguistic development seems to be more elaborate (McMullen & Saffran 2004). This may also be the case as in this developmental stage the vocalisations of children shift from more associated singing to a higher frequency of speaking (Welch 2006).

The development of the singing voice has been described by Welch (1986, 1994, 1998). In Welch's (1994) model on the developmental sequences of singing behaviours he claims that the primary starting point is the linguistic typology in which infants grow up and are culturally shaped by mother tongue and the particular music. Then, infants start learning words and fragments of songs and acquire the rhythmic surface, followed by pitch contour and finally develop key stability (Welch 1994: 4–5). In an earlier work, Welch (1986) distinguishes five stages before the singing voice is prepared to succeed in more complex singing activities. In a later publication, the model is revised to four stages (Welch 1998).

The main aspects of the developmental continuum of singing ability is presented in table 18 (Welch 1986: 299–300).

Table 18. Developmental continuum of singing ability

Developmental continuum of singing ability	
Stage 1	Words are in the initial centre of interest.
Stage 2	Some pitch variations occur which may mark the beginning of a certain kind of awareness for pitch control.
Stage 3	Active attempts to control vocal pitch are made. Vocal range expands.
Stage 4	Fine tuning of pitches are mastered apart from uncomfortable pitches which lie outside the vocal range.
Stage 5	Pitch and melodic errors are rare. The ability to match pitches has been developed, and vocal range has increased.

4.6.1.3. Middle and late childhood

Referring to age of acquisition in singing and specifying a point in an infant's development when singing ability is comparable to adults is very difficult. At the age of five to six children's singing abilities become more stable in many respects (Welch 1994), but are still not comparable to those of adults. An understanding for the structural organisation of song may be developed early at the age of five to six (Tsang et al. 2011). The vocal range is

around one and a half octaves at the age of seven (Welch et al. 2011) and children are already sensitive to the harmony of the music of their particular culture (Hannon & Trainor 2007; Costa-Giomi 2003; Schellenberg et al. 2005).

In the literature it has been shown that inaccurate singing has been reduced considerably at the age of ten to eleven where the rate of inaccurate singing has been reported to be around 12 to 18 percent (Gould 1969; Davies & Roberts 1975). This resembles observations concerning adult occasional singers where around 10 to 15 percent face difficulties in singing in tune (Pfordresher & Brown 2007; Dalla Bella et al. 2007).

Voice quality in children's singing has largely been ignored (Rinta & Welch 2008). However, it has been suggested that at the age of ten children have one voice for both singing and speaking (Rinta & Welch 2009). Concerning gender differences between seven-to ten-year-olds, girls turned out to be better than boys, although this decreases with age (Welch et al. 2011). Vocal range of ten-year-olds is around two octaves (Welch et al. 2011), which is comparable to adults without vocal training. This suggests that at around the age of ten to eleven, depending on the individuals, vocal range, quality of voice and singing accuracy seem to be developed at a level almost comparable to non-professional adult singing. This is considerably later compared to what has been said about the ability to make musical judgements similar to those of adults. This should be possible at the age of nine (Gordon 2013).

The ability to perceive music, song and speech is developed earlier than the ability to reproduce speech, melodies or songs which of course develops alongside motor control (Iverson 2010) and is not surprising. This, however, shows that perception and production can be dissociated in the developmental stage with perception being developed earlier. The question remains as to whether this can be reversed as well. Indeed, in some instances this is possible as changes in motor commands of the vocal apparatus can alter the perception of words (Nasir & Ostry 2009), and individual cases where non-professional singers were tone-deaf but still sang accurately have also been reported as well (Berkowska & Dalla Bella 2009).

4.6.2. Voice experts (professional singing)

In marked contrast to language and music science, perception research of professional singing is rare (Sundberg 1999) as singing requires very precise vocal motor abilities. Voice experts often put more emphasis on generational processes including their refined abilities, and effects which are achieved by specific commands of the vocal motor

apparatus (García-López & Gavilán Bouzas 2010). Singing is also often compared to speaking (García-López & Gavilán Bouzas 2010; Rutkowski 1996; Rinta & Welch 2009; Rinta & Welch 2008; McClellan 2011) and song learning or vocal instruction include speech to song approaches (Roberts & Davies 1975). Such a strategy includes training people to sing a note within the speaking range (Roberts & Davies 1975) and puts emphasis on the effortlessness of speaking, which should be implemented in singing.

As a result, various speech related approaches to teaching singing such as the *Speech Level Singing* method have emerged. As the name suggests, this aims at training singing on the speech level which requires a stable median laryngeal position (McClellan 2011).

Voice experts, however, also claim that singing and speaking cannot be compared on all grounds as singing is restricted by the duration and pitch system. This basically means that singers do not “simply import the code used in speech when they sing” (Sundberg 1999: 208). Intelligibility of vowels is also reduced with the increase of pitch height (Sundberg 1977), and breathing while singing is more active in exhalation than in speaking (García-López & Gavilán Bouzas 2010), to name but a few differences.

Referring to similarities between singing and speaking, the most obvious one is that both are based on the same principles such as body posture, breathing, resonance, and articulation (García-López & Gavilán Bouzas 2010: 443) and rely on the same physical aspects (Christiner & Reiterer 2013). Sound for speech and song is generated in the larynx (Berke & Long 2010). The vocal organ is essentially comprised of three crucial “[...] components: a) the respiratory system, [...]; b) the vocal folds,[...]; and c) the vocal tract with which vocalization is realized. The vocal folds are modulated by the airstream from the lungs and the vocal tract[...] gives each sound its final characteristic spectral shape and thus its timbral identity” (Sundberg 1999: 172).

Professional vocal training has many effects on generational processes of vocalisation. For instance, singers are able to modify vowel quality in singing and speech more precisely, which may be achieved by the “[...] widening of the pharynx required for the generation of the formant and the typically associated lowering of the larynx” (Sundberg 1999: 184). Vocal timbre can also be altered by lowering and raising the larynx, which allows singers to achieve different effects of the voice (Ekholm et al. 1998). Vocal techniques focus on larynx positions, correct glottis closure, breathing technique and body posture, and stresses the ability to coordinate the tongue, the soft palate, and the intrinsic and extrinsic muscles of the entire vocal motor apparatus (García-López & Gavilán Bouzas 2010). On this basis trained vocalists can be compared to athletes.

Even though much more can be said about singing on this basis, the above discussion demonstrates that vocal training methods make singers aware of the generational processes involved. This aims at refining the interplay between the respiratory system, the vocal folds and the vocal tract to maximise the quality of their voice (Dalla Bella 2015), which is the marker for distinguishing a professional from a non-professional voice (Sundberg 1988). In other words, quality of voice refers to “[...] the distinctive attributes that describe the singing voice. The same term is also often used in an evaluative sense to denote the degree to which a particular vocal production meets professional standards of excellence” (Ekholm et al. 1998: 182).

The following twelve criteria have widely been accepted to denote quality of voice: intonation accuracy, freedom throughout vocal range, appropriate vibrato, evenness of registration, resonance/ring, flexibility, intensity, dynamic range, legato line, efficient breath management, colour/warmth and diction (Wapnick & Ekholm 1997: 432). As can be seen, these criteria are very specific and precise in nature and contrast to non-professional and developmental singing measurements, as the latter are more concerned with pitch matching or pitch accuracy.

Research on professional singing was also conducted by a few neuroscientists who found that intensive vocal motor training seems to leave a deep imprint in the brain structure, which can even be differentiated from more general musical effects. Investigation discovered that intensive vocal motor training leads to “an increase in volume and microstructural complexity of specific white-matter tracts connecting regions that are fundamental to sound perception, production, and its feedforward and feedback control [...]” – an effect which has not been detected by instrumentalists (Halwani et al. 2011: 1). In another study, classical singing has been related to “[...] increased functional activation of bilateral primary somatosensory cortex, representing proprioceptive feedback from the articulators and the larynx, in concert with increased involvement of the cerebellum and implicit motor memory areas at the subcortical level” (Kleber et al. 2010: 1151). The findings suggest that vocal training improves the kinaesthetic system.

4.6.3. Non-professional singing

Research on non-professional singers is less frequently conducted (Berkowska & Dalla Bella 2009) and most often focuses on poor singing ability (Dalla Bella 2015; Dalla Bella & Berkowska 2009; Dalla Bella et al. 2009; Tremblay-Champoux et al. 2010; Ayotte et al. 2002; Pfordresher & Mantell 2014; Sloboda et al. 2005). Inaccuracy in singing is often

associated with pitch perception deficits (Tremblay-Champoux et al. 2010; Dalla Bella et al. 2009; Hutchins et al. 2014), and vocal pitch matching tasks and various ways to measure pitch perception have been used (see 4.7). Results are contradictory and some researchers show relationships between singing ability and pitch perception abilities (Estis et al. 2009), while others do not confirm any relationships between the two at all (Moore et al. 2008).

Musical disabilities such as congenital amusia have been studied to understand the interaction between music processing and production. Congenital amusia is a lifelong deficit in perceptual musical ability (Tillmann et al. 2009) that occurs without other cognitive or hearing disorders (Peretz 2008) and cannot be explained by a lack of music exposure (Ayotte et al. 2002). It affects all domains of pitch discrimination ability (Hyde & Peretz 2004) and memorisation of tone sequences (Tillmann et al. 2009; Williamson et al. 2010b), and consequently also music production (Williamson et al. 2010b). Amusics are poor singers (Hutchins & Moreno 2013), which especially creates the impression that perception deficits in general cause poor singing ability.

Increasingly more evidence has also been provided that pitch production and pitch perception can be dissociated, which suggests separate neural pathways for both (Loui et al. 2009). Even though poor singing ability has often been related to pitch perception ability (Sloboda et al. 2005), more recently it has been illustrated that the inability to implement motor abilities more likely cause poor singing (Tremblay-Champoux et al. 2010), which especially contradicts early singing research. Around 4 percent of the population suffer from congenital amusia as a perceptual deficit (Sloboda et al. 2005), while around 10 to 15 percent of adults sing out of tune (Dalla Bella & Berkowska 2009). This suggests that there may be other mechanisms that cause poor singing ability as well. Results have, indeed, revealed that poor perception abilities cannot fully account for all cases of reported poor singing.

In the following part, three alternative explanations are given, all of which favour the notion that kinaesthetic deficits to be involved in poor singing abilities. First, poor pitch singing has also been defined as a sensorimotor translations deficit of the vocal system which specifically affects (vocal) imitation (Pfordresher & Mantell 2014). This should mainly be caused by the inability to plan vocal-motor actions.

Second, other researchers have promoted the so-called *Vocal Sensorimotor Loop Model* that attempts to explain singing processes. It is based on observations of singing research on the general population by Berkowska and Dalla Bella (2009). They have presented a schema which aims to explain processes involved in familiar song singing. In this

model the authors put emphasis on the interplay between perception and production. According to them:

Singing from memory of well-known melodies requires the retrieval of pitch and temporal information from long-term memory and fine motor planning/implementation. In addition, the ongoing vocal production is fed back to the system (i.e. perception), compared with the intended melody, thus eventually influencing motor planning (e.g. through error correction) for subsequent note to be produced (Berkowska & Dalla Bella 2009: 77).

Similar processes have also been suggested for singing new songs, which requires, instead of long-term memory, higher loads of working memory capacity (Berkowska & Dalla Bella 2009), an observation which has also been made subsequently several times (Christiner 2013; Christiner & Reiterer 2013; Christiner & Reiterer 2019).

Third, another idea has been put forward by Hutchins and Moreno (2013), the *Linked Dual Representation Model*, which aims at demonstrating a model of vocal perception and production. The authors present three main ideas before they explain their model. The first one, in which perception precedes the production of vocals, is the easiest. This observation has already been illustrated in section 4.6.1. The second view includes the reverse, where the motor-relevant features of vocal input are perceived first, while the third model would be in favour of a dual route where perception and production are largely independent (Hutchins & Moreno 2013: 1).

An alternative to conventional dual-route models, the linked dual representation model of vocal perception, proposes that it is possible that “vocal-motor representation can be mediated by the symbolic representation” (Hutchins & Moreno 2013: 8). In other words, it is also possible to encode motoric representations directly. The model suggests

[...] that vocal information can be processed in two ways. First, it can be encoded as a symbolic representation, such that we gain conscious knowledge of the identifiable features of vocal stimulus. This process, which is what we normally equate with conscious perception, allows us to determine whether a tone is higher or lower than another, the same or different from another, and allows us to make identification categorization judgements. Second, vocal information can be encoded as a motoric representation, such that it enables reproduction, imitation, or generative production. (Hutchins & Moreno 2013: 7)

The authors claim that this model is potentially applicable to all vocal behaviour regardless of whether it belongs to speech or song and claim that this entails that singing may improve speaking and imitation ability. If such a model could be sustained it would also allow the conclusion that there are at least two ways in which singing or possibly also other vocal

behaviours can be approached. The motoric and direct way may be more effective for generational processes, which are crucial for professional singing. Importantly, research on singing puts emphasis on alternative ways to explain singing ability by highlighting the role of motor abilities.

4.7. Measuring Singing

Assessing singing ability is very challenging as “singing is defined by a complex web of interacting factors embracing perception, cognition, physical development, maturation, society, culture, history and intentionality” (Welch 1994: 3). Music educators rank pitch intonation as the key element to refer to singing ability (Watts et al. 2003), even though other vocal measurements such as vocal range, or quality of voice are also crucial elements (Tsang et al. 2011). The latter are especially important in order to distinguish professionals from the non-professional voice (Sundberg 1988), as has been discussed previously. Pitch and intonation accuracy has been ranked to be one of the most important parameters to assess untrained singers (Cao et al. 2008).

Testing singing ability includes measuring sung responses, pitch accuracy, rating scales for evaluating singing performances (Salvador 2010) but also measurements of vocal range (Rutkowski 2013) or quality of voice (Omori et al. 1996; Gupta et al. 2018; Ekholm et al. 1998; Wapnick & Ekholm 1997). The latter is comprised of multiple dimensions and is based on music theory (Cao et al. 2008).

Singing achievement is most often measured by pitch-matching tasks or rating scales where the singing performances are evaluated and assessed (Salvador 2010). Similarly, computerised methods are also very commonly used to assess pitch accuracy (Price et al. 1994). Referring to the stimuli used to assess singing ability and achievement most likely songs have to be sung, or pitches have to be repeated or discriminated. Songs can be hummed, sung on syllables or with lyrics. Text can be reduced or completely omitted. The flexibility of the speech elements of song is based on the fact that lyrics are less important since they are usually not interpreted like in language, but rather fulfil aesthetic purposes (Limb 2006). Researchers test singing ability by using either familiar (Dalla Bella et al. 2007, 2009; Christiner & Reiterer 2013; Christiner et al. 2018) or new melodies/songs (Christiner 2013; Christiner & Reiterer 2013; Christiner & Reiterer 2019). Goetze and colleagues (1990) mention further singing tasks such as singing single tones, singing only melodic patterns, repeating pitches of male or female voices, and imitating pitches played by musical instruments. Additionally, they mention tasks where pitch is self-selected or prescribed.

In general, as already mentioned, singing tasks are subdivided into two main categories: imitation tasks (repeating new, unfamiliar melodies or songs) or familiar song singing. Singing familiar songs such as “Happy Birthday” is often targeted at non-musicians (Dalla Bella et al. 2007, 2009; Christiner & Reiterer 2013; Christiner et al. 2018). Some researchers prefer using songs with lyrics, while others argue that singing the well-known song on a single syllable such as on “la” reduces memory loads (Berkowska & Dalla Bella 2013). Memory is what is most different when comparing singing new or familiar songs. While the melody and temporal information of already learned songs are retrieved from long-term memory, the imitation of new pitch sequences requires higher loads of working memory capacity (Dalla Bella et al. 2007; Christiner & Reiterer 2013; Christiner 2013; Christiner & Reiterer 2019).

Imitation tasks, however, can also include pitch tasks where short melodies or single pitches are repeated (Dalla Bella 2015). In such cases, the analysis can also be performed by pitch-matching programmes as carried out by Hutchins et al. (2010) , Price et al. (1994), or rating scales (Rutkowski 1996; Rutkowski & Snell Miller 2002).

As acoustic measurements of singing are often computerised programmes that evaluate pitch accuracy (Cooper 1995; Salvador 2010), this has some implications that lead to difficulties in measuring singing performances. First, for most people it is possible to imitate simple melodies as well as carry a basic tune (Dalla Bella et al. 2009; Dalla Bella et al. 2012; Dalla Bella et al. 2007; Pfordresher & Brown 2007), only around 10 to 15 percent of occasional singers face difficulties in singing in tune (Pfordresher & Brown 2007; Dalla Bella et al. 2007). This means that individual differences are hardly ever observed on this basis alone, especially if trained singers’ abilities are assessed. They will most likely sing in tune. Second, some who sing a song perfectly well, but based on different pitches, would be evaluated worse, even though they may sing without difficulty as such measurements would evaluate in how far “the sung pitches differ from the criterion pitch” (Salvador 2010: 41).

Another issue is that single tone singing and its relationship to the singing of whole melodies is not entirely understood (Hutchins et al. 2014). The assessment of singing ability is usually dependent on the ability to sing whole longer sequences of melodies. However, individual differences have also been observed in pitch-matching ability where participants had to match their own singing voice or synthesised vocal tones showing different results (Hutchins & Peretz 2012).

At this stage it is noteworthy to turn to expert singing, which focuses on different approaches to training voices and to assessing the trained voice. To recap, singing ability is

highly dependent on vocal motor ability, especially from the point of view of a singer, as already discussed. Even for non-singers the ability to imitate pitch accurately largely relies on motor ability and is related to two basic components: “[...] phonatory motor control (muscle movements used to regulate vocal fold tension) and perceived pitch height” (Pfordresher et al. 2015: 242). Singing requires motor control of the laryngeal muscles to produce accurate pitch either retrieved from long-term memory as in familiar song singing, or by imitating new songs (Pfordresher et al. 2015). Poor singing has been related to the inability to control sensorimotor translations (Hutchins & Peretz 2012; Berkowska & Dalla Bella 2009) and to perceptual deficits in individuals that suffer from congenital amusia (Sloboda et al. 2005). However, only 4 percent of the general population is said to suffer from congenital amusia, it is very likely that, as suggested by Hutchins and Peretz (2012), poor singing is explained by poor motor control rather than by purely perceptual deficits. Vocalists train perceptual and motor components to improve their interplay, which is the most crucial difference to non-singers. This does not only improve pitch accuracy but also vocal range and quality of voice, the latter of which is the marker for professional singing (Sundberg 1988).

Voice quality has been said to be highly developed in trained singers after around four years of instruction (Omori et al. 1996), which indicates that it is the marker for elaborate singing expertise. Concerning solo voice performances in experts, twelve criteria that measure the quality of voice have been mentioned by Ekholm et al (1998: 183). They are: “appropriate vibrato, resonance/ring, colour/warmth, intensity, dynamic range, efficient breath management, evenness of registration, flexibility, freedom throughout vocal range, intonation accuracy, legato line, and diction” (Ekholm et al. 1998: 183).

Singing research has shown to work on three dimensions which are measured:

1. First, assessing the singing voice from a developmental perspective.
2. Second, analysing singing accuracy and poor singing ability in ordinary people.
3. Third, evaluating singing ability of professional singers/vocalists.

In the following section some ways of measuring singing abilities and suggestions regarding what singing measurements should include are shown. The assessment of the singing voices of children is diverse and multiple rating scales have been developed (Hornbach & Taggart 2005; Rutkowski & Snell Miller 2002; Rutkowski 2013). A well-

known rating scale, the Singing Voice Development Measure (SVDM) developed by Rutkowski (2013: 206) is illustrated in table 19.

Concerning the singing accuracy of non-singers, researchers have criticised that a lack of core tasks that measure singing makes it very difficult to compare research on singing across studies (Demorest et al. 2015). The Seattle International Singing Research Symposium in Seattle in October 2013 attempted to find a more unitary approach towards measuring singing accuracy (Demorest et al. 2015) and should facilitate that future studies may be comparable. A list of tasks which have been proposed are shown in table 20 (Demorest et al. 2015: 267–268).

Table 19. Singing Voice Development Measure SVDM

Singing Voice Development Measure SVDM	
1	"Pre-singer" does not sing but chants the song text.
1.5	"Inconsistent Speaking Range Singer" sometimes chants, sometimes sustains tones and exhibits some sensitivity to pitch but remains in the speaking voice range (usually A2 to C3).
2	"Speaking Range Singer" sustains tones and exhibits some sensitivity to pitch but remains in the speaking voice range (usually A2 to C3).
2.5	"Inconsistent Limited Range Singer" wavers between speaking and singing voice and uses a limited range when in singing voice (usually up to F3).
3	"Limited Range Singer" exhibits use of limited singing range (usually D3 to F3)
3.5	"Inconsistent Initial Range Singer" sometimes only exhibits use of limited singing range, but other times exhibits use of initial singing range (usually D3 to A3).
4	"Initial Range Singer" exhibits use of initial singing range (usually D3 to A3).
4.5	"Inconsistent Singer" sometimes only exhibits use of initial singing range, but other times exhibits use of extended singing range (sings beyond the register lift: B3-flat and above).
5	"Singer" exhibits use of extended singing range (sings beyond the register lift: B3-flat and above).

Table 20. Seattle Singing Accuracy Protocol (SSAP)

Seattle Singing Accuracy Protocol (SSAP)	
Comfortable Range	
task a)	Count to find the comfortable speaking pitch which is related to the singing pitch
task b)	Sing a familiar song. The mean pitch is compared to that of task a
task c)	Sustain a pitch on "oo"
Singing Imitation Tasks	
task a)	Imitate a pitch sung by a human
task b)	Imitate a single pitch played on a piano
task c)	Pattern task. Imitate six four-pitch patterns
Two Song-Singing Tasks	
task a)	Sing a familiar song
task b)	Sing a song on a neutral syllable such as "doo"
Adaptive Pitch Discrimination Task (Elicits perceptual deficits)	
Musical Background Questionnaire	

In marked contrast to non-professional singers, assessing the professional voice focuses on more elaborate vocal skills. The evaluation of the trained singing voice is most often rated by singing experts and rating scales. In contrary to singing on the general population, singing accuracy is not the primary category which is observed. Table 21 shows what a study on professional singer's criteria should include (Ekholm et al. 1998: 185–186).

Table 21. Rating scale of professional singing

Rating scale of professional singing - Score range from 1 (poor) to 7 (excellent)	
Criterion 1	colour/warmth
Criterion 2	resonance/ring
Criterion 3	appropriate vibrato
Criterion 4	clarity/focus
Criterion 5	overall score

In a study by Gupta et al. (2018), the evaluation of singing perception of the professional singing voice, the authors' assessment of singing quality is based on music theory. They analysed seven singing criteria and their relationship to the overall score of the singing quality. The measurements used are presented in table 22 (Gupta et al. 2018: 3–5).

Table 22. Perceptual measurements to assess singing quality

Perceptual measurements to assess singing quality	
Criterion 1	Intonation accuracy
Criterion 2	Rhythm consistency
Criterion 3	Appropriate vibrato
Criterion 4	Volume
Criterion 5	Voice quality
Criterion 6	Pronunciation
Criterion 7	Pitch dynamic range
Criterion 8	Overall score

Measuring singing ability and achievement requires careful consideration of multiple perspectives and has to be selected accordingly to what should be measured. Various methods have been shown which evaluate and assess the singing voice. In marked contrast to musical aptitude tests such as the PMMA (Gordon 1979), the IMMA (Gordon 1982), the AMMA (Gordon 1989a) or the PROMS (Law 2012), singing performances are most often rated by experts, even though acoustic measurements exist as well. Although researchers have criticised that there is no unitary method to evaluate singing ability (Demorest et al. 2015), it has to be stressed that researchers have to adapt their criteria depending on the participants and the type of information they want to generate. Although professional and non-professional singing measurements differ, there are no reasons why singing criteria of professional singing should not be applied to non-professionals, especially if the discussion about perception and production is considered.

4.8. Terminology

In previous chapters on language and musical aptitude, terms such as potential, aptitude, and achievement were distinguished and discussed. Terminology will be provided in this section. Generally speaking, it could be assumed that the same terms do exist and can be used for singing, as previously discussed in the fields of language and musical aptitude. This, however, is only partly the case and is more complex, as has been presented throughout this chapter. First, research on singing is limited; second, singing ability is very often not distinguished from other musical abilities; and third, singing research on professional and non-professional singing needs to be contrasted and compared.

In the context of singing science the terms: singing accuracy, singing achievement, singing ability, poor singing ability and musical aptitude are most frequently used. While singing achievement can be compared to language and musical achievement, for the other terms it is more difficult.

In research musical aptitude and singing achievement are compared (Salvador 2010), instead of singing aptitude and singing achievement. This suggests musical aptitude as starting point which develops into singing talent and achievement. This has probably emerged due to the fact that measuring singing ability is often equated with pitch accuracy (Price et al. 1994; Phillips & Aitchison 1997a; Demorest & Pfordresher 2015; Dalla Bella 2015; Dalla Bella et al. 2007; Goetze et al. 1990), which is a more general musical ability and not a single criterion that defines singing, if all non-expert and expert singing criteria presented before are considered. Strictly speaking, the term singing aptitude does not exist or is not used in the scientific literature.

Instead, the term singing accuracy is predominantly mentioned and is the main target of scientists in singing research (Demorest & Pfordresher 2015; Pfordresher et al. 2015; Dalla Bella 2015; Goetze et al. 1990; Dalla Bella et al. 2007; Demorest et al. 2015), especially if investigations do not focus on expert singing, but instead on singing ability of non-singers (Gupta et al. 2018). Singing accuracy has been related to four special skills by Goetze et al. (1990: 19):

Pitch discrimination, including the abilities to tell one pitch from another and to recall successions of pitches.

Pitch production, vocalizing over a wide range of pitches.

Pitch monitoring or the ability to attend to vocal pitch.

The motivation to attempt to sing.

Furthermore, in singing research aptitude obviously has a similar meaning to talent. For instance, it has been claimed that “[t]alent is a fragile entity and must be nurtured into maturity by the ministrations of individuals who are knowledgeable about, and dedicated to, the task” (Hollien 1993: 197). Researchers say that singing talent has been described “as a special natural ability to produce musical modulations of the voice, whereby sounds vary over a wide range of frequencies and are in tune with each other or where such sounds are melodious” (Watts et al. 2003: 299). Strictly speaking, if following Gagne’s model of DMGT (Gagné 2008: 1) talent is not a natural ability but rather “[...] designates the outstanding mastery of systematically developed abilities, called competencies (knowledge and skills), [...] that places an individual at least among the top 10 % of age peers who are or have been

active in that field”. The fuzziness of how terms are used in singing research leads to a severe problem in comparing research and in distinguishing singing abilities from more general musical aptitude(s).

Moreover, while a large amount of literature has been published on innate musical aptitude, there is none available in singing research. One of the few statements which could be found is that “[t]he assumption that the ability to sing is an innate talent rather than a trainable skill is suspect” (Rutkowski 1996: 354). This may be, as already mentioned, due to the fact that the starting point from which singing matures is generally referred to musical aptitude or musical ability (Salvador 2010). Following research the impression is created that high expertise in singing and instrument learning originates from the same source, namely musical aptitude.

As seen in section 4.6.3, there is growing body of evidence that alternative ways of singing accurately may exist which are not necessarily based on elaborate perception ability but on internal models, fine motor planning and implementation (Berkowska & Dalla Bella 2009; Pfordresher & Mantell 2014; Hutchins & Moreno 2013). Similarly, the discussion about professional singing research also shows that an awareness of motor abilities is the most essential aspect of singing training. This contradicts general ideas about musical aptitude, which is most likely associated with enhanced perception abilities. To exemplify, as discussed in chapter 3, innate morphological differences in the brain structure (Schneider et al. 2002; Schneider et al. 2005) and genetic pre-dispositions of the inner-ear (Oikkonen et al. 2015) have an impact on auditory processes which are associated with musical aptitude.

Following the discussion about perception and production within this chapter, it is surprising that a similar counterpart for the kinaesthetic system does not exist. Consequently, it can be concluded that accurate singing is the result of the interplay between musical perception ability and enhanced motor control. If perception and production can be separated, it is very obvious that singing ability or singing aptitude is the combination of both.

4.9. Discussion

Research on singing, especially interdisciplinary research on singing and its effects on neighbouring faculties are underrepresented in the literature due to multiple reasons. Therefore, this chapter has included various research directions to make attempts to show the rich sources which could facilitate understanding of the relationships between song, music and speech.

This chapter has shown that singing ability is a controversial topic. Singing reveals information about shared and distinct processes of music and language. It has also been demonstrated that the dissociation between perception and production needs to be reflected in the terminology, especially if we refer to musical aptitude and a possible equivalent “singing aptitude”. Evidence has been provided that motor ability plays a central role in the acquisition of singing. In addition, findings suggest that apart from perceptual deficits such as congenital amusia, the inability to implement motor ability can explain poor singing ability as well – a finding which facilitates the enormous impact of the kinaesthetic system involved in singing processes and possibly also for language learning. The same applies to professional singing, which especially integrates vocal motor training.

As singing consists of speech and music elements, it is also interesting to compare whether the enormous impact of motor ability is also relevant for language abilities. Indeed, such phenomena have been observed. In previous investigations, the imitation of unfamiliar languages have always yielded stronger correlations to singing ability rather than to any perceptual musicality measurement (Christiner 2013; Christiner & Reiterer 2013; Christiner & Reiterer 2019). In another experiment it was also revealed that singers and instrumentalists could be distinguished on their ability to imitate unfamiliar language material, although this could not be replicated on their perceptual skills where they have been found to be equal (Christiner & Reiterer 2015). This also suggests the close relationship between singing ability and phonetic aptitude in terms of motor ability but not necessarily in terms of perception. Interestingly, a dissociation between speech production and perception has also been reported by other researchers (Golestani & Pallier 2007). This raises the question as to whether the findings of singing ability and the suggestions made about the dissociation between perception and production can also be transferred to speech –an interesting question which will be addressed by the research design in this investigation that compares speech perception and production in order to understand their connections.

In addition, the entire discussion of this thesis has also shown that interdisciplinary research on singing requires integrating music and language criteria that influence individual differences in performances. In previous chapters on musical and language aptitude, factors that influence individual differences have been shown as well. These were learning style, learning strategies, personality, motivation, anxiety, aptitude, intelligence quotient (IQ) and age in foreign language capacities (Dörnyei 2006; Skehan 1991; O'Malley & Chamot 1990). For music the cultural influence on measurements, individual differences and requirements for playing different musical instruments, considering instrument playing versus singing,

which musical abilities should be measured, participants' selection, performance anxiety, working memory, and environmental influences such as (self-) education, social status of musical aptitude, motivation or (parental) support have been mentioned. All these criteria also play a central role for singing research and have to be considered as well.

The following sections will show further aspects that are necessary to understand the relationship between singing ability and phonetic aptitude.

5. Working memory and (phonological) short-term memory

The concept of working memory (WM) has reached considerable attention in the field of science and has consequently become one of the most influential psychological constructions. WM capacity has been studied in multiple disciplines and has been discussed in neuroscientific, clinical, educational or psychological scientific contexts, to name but a few. Complex WM capacity has been said to have an impact on multiple domains such as on mathematical ability (Schmader & Johns 2003), intellectual ability (Conway et al. 2003; Conway et al. 2002; Engle et al. 1999), academic achievement (Swanson & Alloway 2012; Gathercole et al. 2007; Rogers et al. 2011), influences mastery of first and second language capacity (Wen & Skehan 2011; Wen et al. 2019; Dörnyei & Ryan 2015; Majerus et al. 2006) and can be limited by the arousal of negative anxious feelings (Figueira et al. 2017).

Despite the large scientific interest and research, there are three major aspects that need to be addressed and clarified. First, the concept of WM and their components require careful inspection as the interdisciplinary use and the change in models that represent WM capacity theory have led to some confusion about the two terms short-term memory (STM) and WM. This also has an effect on measurements of WM capacity and that of STM as it is not always clear which tests measure the former or the latter concept, which is the second issue that will be explained.

The third aspect is less important for this investigation, but should at least be mentioned. While WM capacity and STM capacity in the context of language are quite intensively studied, with music science it is different. There is huge controversy over whether a “tonal loop” for music as an equivalent of the phonological loop for language capacity exists or not. Studies which compare verbal and tonal stimuli directly are underrepresented in the literature, and do not deliver consistent results. For instance, while it has been suggested that there may be a different storage component for tonal and speech material (Salamé & Baddeley 1989), other researchers suggest that pitch sounds of speech and non-speech are not

differently stored (Semal et al. 1996). Brain research has concluded that verbal and tonal storage and rehearsal ability rely on largely overlapping neuronal networks (Koelsch et al. 2009), while more recent behavioural studies again concluded that the processing of musical and verbal sounds correspond only slightly in the auditory STM (Williamson et al. 2010a). Musical ability and its relationship to WM capacity will be ignored, as this would go beyond the scope of this investigation.

In the opening of this chapter the most common WM model, the multicomponent model, will be presented, followed by a discussion about terminology and measurements, before language aptitude and WM is discussed.

5.1. Baddley's multicomponent model

WM capacity is a system that enables the storing, maintaining and manipulating of temporary information (Baddeley 2003b) and is required for solving complex tasks such as learning or comprehension (Baddeley 2010). The multicomponent model is actually one of the most influential theoretical WM models and mainly consists of three components: the central executive, which is the control system that coordinates the two subsystems, the phonological loop and the visuo-spatial sketch-pad (Baddeley 2003b, 2010). In addition, a fourth dimension, the episodic buffer, as a more passive store operates as a kind of interface that is also accessible for the long-term memory.

5.1.1. The phonological loop

The phonological loop is most essential for language processes. Following Baddeley (2010: 138), it consists of two features, the first being a kind of phonological store in which memory traces are held before they fade, and the second, the verbal or subvocal rehearsal mechanisms that allow decaying memory traces to be refreshed.

5.1.2. The visuo-spatial sketch-pad

In marked contrast to the phonological loop, this system integrates “spatial, visual, and possibly kinaesthetic information into a unified representation which may be temporarily stored and manipulated” (Baddeley 2003a: 200). This subsystem may be less significant for language-related functions, but still is involved, for instance, in everyday reading (Baddeley 2003a).

5.1.3. The central executive

This system is mainly said to be the centre for “attentional control of the working memory” (Baddeley 2003a: 201). It has been suggested that executive processes are largely responsible for individual differences in WM span (Daneman & Carpenter 1980). This system is said to be involved in carrying out complex tasks and receives, coordinates and integrates information from the phonological loop, the visuo-spatial sketch-pad and from the long-term memory (Dörnyei 2005).

5.1.4. The episodic buffer

The episodic buffer is more or less seen as “[...] a storage counterpart of the central executive [...]” (Dörnyei 2005: 64). Both the central executive and the episodic buffer may be separated whereby the latter represents a more passive store (Baddeley et al. 2009). The episodic buffer is limited to four chunks and should probably be able of holding “[...] multidimensional episodes or chunks, which may combine visual and auditory information possibly also with smell and taste” (Baddeley 2010: 138). According to Baddeley, the episodic buffer functions as a kind of interface between the central executive, the phonological loop, the visuo spatial sketch-pad and the long-term memory (Baddeley 2000).

As the name suggests, the multicomponent model is not a unitary storage system but consists of certain subsections. In language science, the phonological loop and the phonological loop capacity have been studied intensively and may be the most important concepts that influence individual differences in (foreign) language learning. The emergence of the multicomponent model, however, also leads to confusion regarding terminology which will be discussed in the following section.

5.2. Measuring working memory, short-term memory and terminology

The terms WM and STM have led to much confusions and are still often used to refer to the same concept (Baddeley 2012; Cowan 2008). This has its roots and origin in early research (Baddeley 2010; Cowan 2008), but a lack of knowledge of what is being measured, WM capacity or STM, is another issue. The latter matter, together with using WM and STM to refer to the same concept, leads to difficulties in the interpretation of results in the current literature.

To recap, understanding the underlying differences between WM and STM requires looking at how the model of WM emerged. It was believed that STM is a unitary holding place which was later replaced by Baddeley’s multicomponent model (the phonological loop,

the visuo-spatial sketch-pad, the central executive and the episodic buffer), which is the most widely accepted concept of WM at the present moment (Cowan 2008). Since then the term WM has more likely been used to describe the entire model (Cowan 2008), while STM refers to a simpler mechanism.

Following Baddeley (2012: 4), STM refers “to the simple temporary storage of information, in contrast to [WM], which implies a combination of storage and manipulation”. Other researchers claim that “[...] STM is considered a subset of WM. STM is a simple storage component, whereas WM is that storage component as well as an attention component” (Engle et al. 1999: 310).

In the context of learning, interactions between STM and long-term memory (LTM) have to take place. According to the research, STM and LTM are not completely distinct capacities, and their relationship has been discussed in multiple different ways. Some researchers suggest that both constitute one single memory store (Cowan 1988), while others have argued that STM and LTM are separate storage systems (Baddeley 2003b). STM has been said to be influenced by various aspects of information that is stored in the LTM (Norris 2017). It has also been assumed that STM can be seen as activated LTM (Cowan 1988). Baddeley (2000) introduced the episodic buffer to link the central executive, the phonological loop, and the visuo-spatial sketch-pad to the LTM (Baddeley 2000) and indicated their relationship. Even though LTM is irrelevant for this research as only STM measurements are included, the relationship between STM and LTM form the basis for language aptitude, as it is based on the idea that language input must be stored in the LTM for later retrieval and, in a wider sense, relates to achievement.

Referring to STM and WMC, the different explanations of researchers are quite consistent. Consequently, it is all important to determine whether the STM or the entire WM capacity is being measured and tested. However, it has often been criticised that it is not well understood as to whether tasks require “additional processing and/or manipulation” (Schulze & Koelsch 2012: 229), or just simple rehearsing or storing of information.

Even though it may be less important to understand the differences between STM and WM in a variety of other research, it is important to understand whether a process requires complex or simple mechanisms in this investigation where test development also plays a central role (further illustration will be made in section 9). Understanding the nature of simple and complex WM span tests requires looking at the difference between WM and STM measurements.

Following Conway et al. (2005), the distinction between WM and STM measurements can be detected by looking at the history of WM span task development, where emphasis was put on an immediate memory system with limited capacity that only stores and rehearses information. This, however, represents a measurement of STM, while WM capacity includes the processing of additional information (Conway et al. 2005) or the performing of a secondary task (Turner & Engle 1989). Understanding the differences of both more precisely requires looking at measurements that have to be categorised as either complex WM or STM tasks.

WM span tasks are diverse and include counting spans, reading spans, operation spans, n-back tasks, running spans and keeping-track tasks (Conway et al. 2005). Generally speaking, designs of WM and STM measurements are quite diverse. There is a large number of reading span tests available that measure complex WM capacity (Daneman & Carpenter 1980; Alloway 2007), all of which have similar test designs. Most reading spans require participants to read sentences where they have to remember, for instance, the last word and after they have read some other sentences, those words have to be recalled in the correct order (Conway et al. 2005). Instead of words participants can also be instructed to remember letters. These test designs focus on WM capacity as they are more of a dual nature, which entails that information has to be processed and stored (Engle et al. 1999). Operational span tests require participants to solve mathematical operation strings, which consist of at least two operations in a row (Engle et al. 1999). Similarly, counting spans where shapes have to be counted, their numbers remembered and later recalled (Case et al. 1982; Alloway 2007) also belong to the complex WMC measurements.

Typical tasks that measure the STM are forward span tasks where for instance dissimilar or simple words need to be recalled in a correct serial order by writing them down (Engle et al. 1999), or non-word repetition which measures the (phonological) STM (Gathercole 2006). For the backward counterpart, participants are instructed to repeat words in reversed order (Engle et al. 1999). Presumably backward spans focus more on controlled attention which makes them more of a hybrid task, but still more evidence has been provided that they may be best categorised as STM tasks. A factor analysis revealed that both forward and backward tasks are components of the same factor (Engle et al. 1999), which could lead to the interpretation that backward spans demand a “mental transformation” as there is no a new stimulus being imposed (Conway et al. 2005).

Another subcomponent of the WM capacity is the phonological loop or the phonological STM, which both refer to the same concept. The phonological loop capacity is

one of the most frequently investigated components of WM capacity (Kormos & Safar 2008), especially in language science. Testing the phonological STM is achieved by simple tasks (Engel de Abreu & Gathercole 2012) such as by using non-word repetition tests of different lengths (Kormos & Safar 2008) or by digit spans that require the serial recalling of numbers (Thorn & Gathercole 2001; Engel de Abreu & Gathercole 2012). Digit spans are tests that usually require the participant to repeat sequences of digits in the same order as they hear them (Engel de Abreu & Gathercole 2012; Baddeley 2003b; Thorn & Gathercole 2001). For this reason STM tests or phonological STM tests are not always differentiated in the literature and are used to measure largely the same ability. Most WM and STM measurements are very valid and reliable (Conway et al. 2005).

As WM capacity is age-related, similar to language, music and singing acquisition processes, the number of items that can be remembered, retrieved and stored is limited in children (Loosli et al. 2012; Roman et al. 2014). However, the socioeconomic status can also have an impact on the phonological WM capacity in children. This is obviously related to the input they receive, as a significant correlation between the phonological WM and vocabulary has been found (Korecky-Kröll et al. 2019). This shows that WM capacity is also trainable (Klingberg et al. 2002).

Word length is another aspect which plays a crucial role. The longer the words, the more slowly they are rehearsed (Thorn & Gathercole 2001). This increases the chance that words get lost in the phonological store (Baddeley et al. 1984), which has been supported by cross-linguistic comparisons of digit span testing where the number of digits remembered can differ depending on the language – an effect which has been related to word length of the digits (Thorn & Gathercole 2001). The shorter the names for the digits are, the higher the number of items is that can be repeated. Generally speaking, around four chunks, or five words of unrelated material can be remembered (Baddeley 2010; Cowan 2001), while the memory has a capacity of around 15 words if the sentences are meaningful (Baddeley 2010).

As shown WM and the subcomponent STM capacity can be differentiated on a certain kind of basis. Unfortunately, it is difficult to understand whether research focuses on complex WMC or STM as long as the nature of the measurements used are not understood. This should be considered in the analysis to avoid confusion.

5.3. Working memory and the language aptitude construction

Language aptitude and its relationship to WM capacity and STM have been discussed in multiple investigations for first (Cain et al. 2004; Majerus et al. 2006;

Vulchanova et al. 2014; Hu & Schuele 2005) and second language acquisition (Wen & Skehan 2011; Wen et al. 2019; Dörnyei 2005; Dörnyei & Ryan 2015; Hummel 2009; Robinson 2005, 2002, 2019b). The reason for this is quite obvious since language input must be “encoded into WM [...] and then stored in long-term memory for later retrieval [...]” (Dörnyei & Ryan 2015: 54), which suggests that WM probably turns into long-term memory (Wen & Skehan 2011).

The role of WMC, however, has led to controversial debates. Researchers have concluded that the aptitude construct could probably be replaced by the verbal WM (Miyake & Friedman 1998), while others prefer to refer to it as one of the most important, but not exclusive components of language aptitude (Robinson 2002; Hummel 2009). The enormous impact of WM capacity on multiple cognitive domains is unquestionable. It is involved in a variety of complex higher cognitive abilities such as reasoning, comprehension, learning, cognitive control (Biedroń & Pawlak 2016), speech and music processes (Schulze & Koelsch 2012; Koelsch et al. 2009). This being said language aptitude is also related to other cognitive abilities aside from WM capacity (Robinson 2019a).

In the last decades, individual differences in language capacities and their relationship to WM and STM have been intensively studied in various disciplines (Sagarra & Herschensohn 2010; Robinson 2005; Gathercole et al. 1997; Gathercole 2006; Speciale et al. 2004; Mackey & Sachs 2012; Révész 2012; Wen & Skehan 2011; Wen et al. 2019; Conway et al. 2008; Payne & Whitney 2002). Primarily, language learning processes such as speech production, reading, comprehension, or vocabulary learning have been related to complex WM and STM ability (Gathercole & Baddeley 1993, 1990).

In this study phonetic aptitude will mainly be investigated by measuring oral foreign language performances. The latter has especially been related to be the measurement of the phonological loop (Payne & Whitney 2002; Christiner 2018) which is a “highly applicable conceptualisation of verbal STM” (Burgess & Hitch 2006: 629). Evidence for the role of the phonological STM comes from clinical studies where patients that suffer from impairments of the phonological loop have shown poor abilities to learn foreign languages (Baddeley 2010; Baddeley et al. 1998) and new unfamiliar phonological forms of utterances (Baddeley et al. 1998). Approaching from the opposite direction, high L2 performers such as polyglots possess an elaborate phonological STM, which is one of the most important predictors that contributes to the mastery of multiple languages (Papagno & Vallar 1995). This is why the phonological loop is also seen as a language acquisition device (Biedroń & Pawlak 2016). Non-word repetition is especially related to the ability to learn new phonological elements of

languages, but not the words' semantic interpretations (Gathercole & Baddeley 1990). Non-word repetition has also been associated with the imitation of unintelligible foreign language material, which leads to the assumption that both are quite similar concepts (Christiner & Reiterer 2013). The ability to imitate unintelligible speech has always yielded strong correlations to the ability to repeat digits in forward or backward order in adults (Christiner 2013; Christiner & Reiterer 2013; Christiner & Reiterer 2019; Christiner et al. 2018) and children (Christiner et al. 2018; Christiner & Reiterer 2018). Therefore, digit span tests will also be integrated in this investigation.

5.4. Discussion

WM and STM capacity are considered to be the two most important predictors for language aptitude of both first and second language mastery. As discussed, it has led to several debates about whether the language aptitude construction could be replaced by WM or not. Results are contradictory and it seems to be more plausible to assume that other cognitive capacities contribute to language aptitude as well.

Additionally, the discussion about the difference between WM and STM capacity has to be taken seriously as both are measured with different tests. Even though the distinction between the two is often ignored, it should be considered that complex WM tasks are dual processes in which “[...] a combination of storage and manipulation” (Baddeley 2012: 4), or storage and attention takes place (Engle et al. 1999: 310). In marked contrast, STM is a simple and temporary storage system and a subcomponent of WM.

In this investigation, only measurements of the STM are taken. Phonetic aptitude is especially related with the phonological loop capacity which requires STM measurements. The phonological loop store can hold some memory traces for some seconds before they fade (Baddeley 2003b: 830), which is especially important for remembering and learning new unintelligible language material (Christiner & Reiterer 2013; Christiner et al. 2018). The latter tasks will be introduced in section 10.3.1.

6. Performance anxiety

Performance anxiety is a phenomenon that occurs in many situations such as during testing conditions (Calvo & Eysenck 1996), in educational settings in which evaluation takes place (Horwitz 1986), or on stage for musicians where it is termed “stage fright” (Kokotsaki

& Davidson 2003). Anxiety, however, also has effects on cognitive domains which can influence performances remarkably (Calvo & Eysenck 1996).

Generally speaking, performance anxiety is associated with a very unpleasant negative feeling that can be described as a “psychological state in reaction to perceived threat concerning the performance of a task under pressure” (Cheng et al. 2009: 271). In previous chapters on language and musical aptitude, anxiety was defined as a crucial factor that influences individual differences in performances. In this short chapter, just some aspects will be addressed that should raise awareness of the inclusion of such measurements in behavioural research since anxiety may have an undesired side effect on the interpretation of results.

Individual difference in performance could vary as a result of performance anxiety. The impact of the latter may also depend on the measurements or tasks participants have to fulfil. This means that in some instances integrating anxiety measurements may be more or less important. More precisely, whereas direct oral performances such as the imitation of new utterances or singing may be more influenced by anxiety, perceptual measurements, which are often executed with headphones alone, may be less affected by unpleasant feelings. Since in this research a variety of performances are included in the empirical investigation, it is necessary to consider this concept as well.

Research on performance anxiety is highly diverse and complex, and multiple conceptualisations of how anxiety emerges have been discussed (Cheng et al. 2009). Measuring performance anxiety is most often realised by targeting questionnaires at participants and the content varies according to the particular field. For instance in the foreign language classroom, the *Foreign Language Classroom Anxiety Scale* (Horwitz et al. 1986; Speciale et al. 2004), has been developed for assessing individual differences in anxious behaviour, or the *Spielberger's State-Trait Anxiety Inventory* (Spielberger et al. 1983), which has been used to assess the anxiety of singing students (Kokotsaki & Davidson 2003), to name just two.

As already mentioned, performance anxiety can emerge in all kinds of situations. In language research, personality traits that include a number of characteristics such as extraversion, introversion, self-esteem, inhibition, risk-taking and anxiety (Oz 2014; Dörnyei 2006; Hu et al. 2013; Zafar & Meenakshi 2012) have been investigated for good reason as they influence both language achievement and performances.

In the foreign language classroom, performance anxiety plays a significant role. A state of low anxiety is crucial for successful learning of new language input as language

learners who are anxious face difficulties in concentrating and are not receptive to the input they get (Horwitz et al. 1986: 127). This may depend on the fact that anxiety is like a filter that has effects on other cognitive domains. For instance, the subsystems of the WM capacity, such as the phonological loop, can be affected by anxiety (Calvo & Eysenck 1996). Similar effects have also been found for the concept of worrying, where the phonological STM and the central executive seemed to be limited, but not the visuo-spatial sketch-pad (Rapee 1993). Bearing previous discussions in mind, WM capacity, particularly with regard to the phonological STM, has been associated with (foreign) language aptitude and it is very obvious that measures of anxiety are negatively related with measures of foreign language achievement (Gardner Robert 1990). Anxiety has been associated with foreign language learning for grammar, reading and writing (Baran-Łucarz 2016), but particularly seems to affect oral language performance (Phillips 1992, 1991; Woodrow 2006).

Similarly, performance anxiety is also very prominent in music studies. Musical performances do not only affect less experienced musicians but can also cause considerable distress for trained musicians (Fehm & Schmidt 2006). Studies have reported that one third of musicians showed signs of suffering from performance anxiety (Fehm & Schmidt 2006), which shows that such effects can have a tremendous influence on music studies. By this it is meant that especially group comparisons of musicians and non-musicians should be treated cautiously. Studies have revealed that music students that are more advanced will also perform remarkably better under conditions of increased anxiety (Hamann & Sobaje 1983; Kokotsaki & Davidson 2003). When comparing musicians and non-musicians, the latter may suffer more from unpleasant and anxious feelings than trained musicians do. If this then were to lead to better performances in general, it would create the impression that musical abilities have an effect on other cognitive domains such as on language abilities, but in fact they would be biased by performance anxiety.

As a result, anxiety seems to have an impact on certain dimensions which are crucial for this investigation, such as language and musical performances, and on cognitive domains such as the phonological STM that can cause considerable distress in testing conditions.

To prevent individual performances from being biased by performance anxiety, a questionnaire has been developed that focuses on general vocal behaviour and asks about how comfortable the participants feel in singing and in the imitation of new unfamiliar languages. In addition, speaking as well as singing behaviour in public will be tested in order to get an impression of whether performance anxiety plays a significant role in this investigation or not. The questionnaire was introduced to the participants at the end of the testing procedure to

prevent the questions from causing performance anxiety. The development and reliability of the questionnaire will be discussed in chapter 7 on questionnaire development.

7. Questionnaire and construction

Questionnaires are very simple and effective tools to collect a large amount of data in a very short period of time. Even though it seems that questionnaires are easily prepared and written, there are multiple aspects which have to be considered. These include, for instance, simplicity, instructions, design, length, reliability, anonymity, statistical analysis and the question as to which information should be generated (Dörnyei & Taguchi 2010). In addition, as in the past 20 years questionnaires have been computerised and increasingly more often are executed online, they require careful planning in order to get meaningful and reliable results. On this basis even a simple and very basic questionnaire becomes a very challenging task, which goes beyond the design and has to include when, how, where and in which order questionnaires are presented to informants.

First, it has to be mentioned that established questionnaires do already exist for all the parameters measured in this investigation. These include musical, singing and language expertise questionnaires, which are also suitable for collecting data for this research. Using existing questionnaires would of course reduce the workload, but would cause other problems, for example concerning length; the reduction and shortening of unnecessary elements would weaken the reliability and validity of the tests which have been built upon using the entire questionnaire. Some questionnaires, such as the Goldsmith's Musical Sophistication Index (Müllensiefen et al. 2012), a self-report assessment questionnaire that provides information about musical skills, are simply too long and precise, while others do not explicitly ask for the specific information which is required for the tasks of this research. Therefore, a main questionnaire has been designed and developed that is comprised of different sections which interact with other behavioural measurements.

Although questionnaire construction can be described at length, only the most relevant information concerning why and how this questionnaire was written and realised will be provided. For more details in the field of second language acquisition, it may be best to consult Dörnyei and Taguchi (2010), but questionnaire design and development is relatively independent from the field. Thus it is worth consulting other works by Lietz (2010) or Schwarz and Oyserman (2001), for instance. For this investigation certain aspects are crucial.

They are types of questions, general features, wording and writing of questionnaire items, sensitive topics, multi-item scales, and reliability.

7.1. General features

The format and length of a questionnaire can have a significant effect on the results obtained (Schwarz & Oyserman 2001). In terms of format or layout, it is crucial that the questionnaire looks short, has an appropriate density, is orderly and well structured (Dörnyei & Taguchi 2010). In terms of length, it has been suggested that the shorter a questionnaire is the better the results, even though contradictory results have been presented in this regard (Rolstad et al. 2011). Following Dörnyei and Taguchi (2010), questionnaires that take more than thirty minutes to complete, and are over six pages, are too long. This is also considered in this investigation.

Another aspect concerning general questionnaire features is that question designs have different effects. The distinction between open-ended and closed-ended question formats is of integral importance (Schuman & Presser 1981). A closed ended design provides participants with “response options” from which they can choose the box that describes their response best (Dörnyei & Taguchi 2010). This has the advantage that responses are given according to the researchers’ intentions and main objectives (Schwarz & Oyserman 2001) and can therefore easily be used for statistical analysis.

In marked contrast, the open-ended question format allows participants to give their own response. This increases data richness (Connor Desai & Reimers 2019) but at the same time makes it very difficult to analyse the data which has been generated (Dörnyei & Taguchi 2010). Another reason for favouring closed-ended questions is that there is a tendency that they are answered faster than open-ended ones since options are provided (Connor Desai & Reimers 2019).

Rating scales are most frequently used where participants are required to give their response and judgment. Rating scales such as Likert scales or true or false items are the most commonly used (Dörnyei & Taguchi 2010). Likert scales are used to indicate in how far participants agree or disagree with a statement. This type of question is quite effective in quantitative research, while true or false items generate more limited options in terms of analysis.

Generally speaking, there are three types of questions in questionnaires: factual, behavioural and attitudinal (Dörnyei & Taguchi 2010)(Dörnyei 2010). Factual questions are straightforward and rather descriptive, and ask for background information such as gender,

age and foreign language proficiency, which suggests that they are very likely to be answered correctly (Dörnyei & Taguchi 2010: 23). For the other two types of question, it is more complex. While behavioural questions ask about what informants do, attitudinal questions ask what people think. Both question types are not always answered reliably. It is thus crucial to be aware of what type of question is being asked. Perhaps the best examples for this are sensitive or embarrassing questions, which are very often answered incorrectly (Tourangeau & Smith 1996). Sensitive questions are considered to raise “[...] concerns about disapproval or other consequences (such as legal sanctions) for reporting truthfully or if the question itself is seen as an invasion of privacy” (Tourangeau & Smith 1996).

It has generally been suggested that sensitive questions should be avoided (Dörnyei & Taguchi 2010). However, it is not always clear whether or not a certain topic is particularly sensitive for the informant, since this may depend on factors such as gender or cultural differences, or simply on the individual. While it may be more obvious that questions that focus on respect for traditions of other cultures will be positively answered, their reliability may be questioned. Participants may be forced to give an answer that is culturally expected (Dörnyei & Taguchi 2010). Less obviously sensitive questions could be those asking for age, which can be perceived as invasion of privacy. Similarly, asking for someone’s own self-estimation about how well she or he sings is a sensitive topic as in Austria, Germany and many other countries, people are expected to underestimate their own abilities. Being aware of sensitive questions and questions that facilitate culturally expected answers may help to write a reliable questionnaire. Solutions may be to avoid them entirely (Dörnyei & Taguchi 2010), or write items which are indirectly linked to the target concept.

7.2. Wording and writing of questionnaire items

Wording plays a significant role in how questions are answered. One key element is simplicity (Foddy 1993) or avoiding wordings that tend to elicit socially desirable answers (Stocké & Hunkler 2007). There is also controversy over including negatively worded questions. It has often been recommended to exclude them for the reason that these are more difficult to comprehend (Foddy 1993), and are more often incorrectly understood and answered (Dudycha & Carpenter 1973). However, it could also be suggested that two similar questions, one with a negative and one reversed with a positive meaning can function as a control and attention task.

Word choice is another serious issue which requires consideration. Adverbs of frequency are most often used and difficult to be interpreted since the meaning of adverbs can

differ remarkably for individuals (Lietz 2010). For instance, the interpretation of “usually” in questions can depend on the content, the kind of activity as well as on the participant who has to answer them (Foddy 1993) and words and phrases such as “very” or “not too” can be rather vague (Bradburn & Miles 1979). For these reasons, adverbs of frequency were largely avoided in questions. Wording and very little changes of wording can have a remarkable impact on the results (Schwarz & Oyserman 2001; Dörnyei & Taguchi 2010; Lietz 2010). This poses a problem; the solution to this can be found in writing multi-item scales. This concept is quite promising as it includes items that are worded differently but refer to the same target concept. This has many advantages. First of all, single questions may not be efficient tools to gather information of the target content. It can be more easily achieved by more questions that ask for the same content (DeVellis 2012). It also makes it possible to address more aspects that are associated with certain target content which cannot be addressed with a single question (Dörnyei & Taguchi 2010). For effective multi-item scales, it has been recommended to have at least four questions for each component (Dörnyei & Taguchi 2010).

The advantage of multi-item scale concepts is that statistical analysis is quite effective. A certain number of question items is crucial for the statistical analysis which could be performed by calculating a Cronbach’s alpha coefficient. The concept should have a minimum score of about 0.7 in order to be considered reliable (Field 2009). Using multi-item scale concepts also enables researchers to delete ineffective questions or statements that weaken the concept after the research has been completed.

7.3. The main questionnaire

For this investigation certain questionnaire types have been developed. Factual questions about the background of the participants were selected based on a questionnaire which was computerised and performed online.

The following aspects were the subject of the investigation: a) age and gender b) nationality and residence, c) educational background of the participants and their parents, d) number of first language and foreign languages, e) average hours of singing within a week, f) singing self-estimation, g) number of instruments played. Furthermore, multi-item scales were developed for singing behaviour during childhood and performance anxiety.

In addition, rating scales were also used to evaluate the imitation tasks where the participants had to indicate how memorable each language sounded to them, how difficult they found the imitation of unfamiliar language and how familiar the language imitated occurred to them (see section 10.3.4). Rating scales were also developed for evaluating the

individual performances of the foreign language imitation tasks and the singing tasks, which were rated by native speakers of the respective languages (see section 10.3.5) and voice experts for singing (see section 10.4.3).

7.4. Multi-item scales, singing behaviour during childhood and performance anxiety

Singing acquisition processes were discussed in chapter 4, which revealed that singing acquisition processes share multiple aspects with language acquisition processes. Following Cohen (2008), it is possible to reveal interesting information by looking at shared and distinct processes of language and singing acquisition, as well as by considering whether there may be sensitive periods in which singing ability develops better and more elaborately. Such information may also provide insights to better understand other domains. Even though little information can be revealed by testing adults, their singing behaviour during childhood can be asked in a questionnaire. In previous research singing behaviour turned out to be related to singing ability (Christiner 2013). Thus it would be interesting to retest this observation and to look at the singing behaviour during childhood and its relationship to other abilities, such as the imitation of new languages or the ability to discriminate music or language.

Another multi-item concept which has been developed aims to reveal the effects of performance anxiety on music and language performances. This is a crucial concept as it could be assumed that performance anxiety plays a role in measuring singing ability or the imitation of unfamiliar language. Understanding the influence of performance anxiety on singing and new language imitation can show which effects anxiety has on the performances – an aspect which has often been underrepresented in aptitude research, as discussed in chapter 6.

7.5. Questionnaire: education and singing self-estimation

To control the educational background in order to avoid educational biases, the educational background of the parents and that of the participants was carefully selected and grouped accordingly into 10 different categories. 0 indicates “no school education”, 1 “primary school”, 2 “main general secondary school”, 3 “apprenticeship, technical and vocational school”, 4 “secondary academic school (general qualification for university entrance”, 5 “post-secondary non-tertiary education”, 6 master craftsman, mastership examination”, 7 “Bachelor or equivalent level”, 8 “Master or equivalent level” and 9 “Doctoral degree or equivalent level” (see illustration appendix A1.1.1, Step 5 / education).

The scale is largely based on the recommendations of the European Qualifications Framework.

In addition, Likert scales have been used where participants have to indicate their self-assessment of singing ability. The self-assessment scale includes “very poor”, “poor”, “average”, “good”, or “very good”.

7.6. Design of the questionnaires

All questionnaires were programmed on a computer for various reasons. First, data collection could be standardised and transfer errors could be reduced; second, participants could fill in some of the questionnaires at home; and third, administration was quite easy and the error rate of data transfer dropped to almost zero. This, however, had an influence on the test design. As suggested, questionnaires should not be longer than six pages and should not take longer than 30 minutes to complete. To follow the recommendations, all questionnaires could be filled in within 30 minutes, but were split up into different meetings anyway. Each question was presented individually on the screen and the programme prevented participants from returning to previously answered questions. To avoid incorrect answers, control variables were included. For instance, participants were asked about their mother tongue and whether they had a second mother tongue they grew up with. Both required participants to select the languages by choosing the language from a dropdown menu. The control variable followed next, where they were asked “whether they grew up with a second mother tongue or not”. This was a factual question where the participants had to indicate “Yes” or “No”. Then the participants were given the possibility to double-check their answers, for example to see if they corresponded to the two questions answered before. The questions of the multi-item scale concepts were presented to the participants in a randomised order.

7.7. Reliability analysis of multi-item concepts

After all participants finished the questionnaire ($N=86$), the reliability of the concepts *Singing during childhood* and *Performance anxiety* were assessed. To test whether the multi-item concepts were reliable, Cronbach’s alpha was applied. Both concepts show high reliability and were above the acceptable range of .7 and not a single question had to be removed to improve the concept. It is important to note at this point that the questions presented are an English translation of the original questions which were written in German (see appendix 5). When reusing the questions in another language, it is crucial to consider cultural aspects and of course to retest the internal reliability of the questions.

As will be shown, *Singing behaviour during childhood* and *Performance anxiety* show high reliability which indicates that they can be used for further analysis. The results of all questions and their relationship to the other variables of this investigation will be discussed in the analysis and results section in detail.

7.7.1. Reliability analysis: singing behaviour during childhood

The multi-item concept *Singing behaviour during childhood* consists of eight questions. The concept was tested for its reliability ($N=86$) which showed that all Cronbach's $\alpha = .82$ and therefore well-established. Table 23 shows the items used in the concept *Singing behaviour during childhood*.

7.7.2. Reliability analysis: performance anxiety

The multi-item concept *Performance anxiety* consists of eight questions. The concept was tested for its reliability ($N=86$) which showed that all Cronbach's α was $=.87$ and therefore well-established. Concerning the questions, the participants were familiar with the wording and terminology used, and knew what they mean. This is especially important for speech imitation.

Table 24 shows the items used in the concept *Performance anxiety*.

Table 23. The concept *Singing behaviour during childhood*

Concept <i>Singing behaviour during childhood</i>	
	Cronbach's α if item deleted
As a child I enthusiastically joined in with the singing at church whenever the possibility arose.	.79
As a child I never used to worry if I sang beautifully or not.	.84
As a child I enjoyed singing in a choir, with friends, at Christmas, birthdays, or the like.	.77
As a child I used to dream of being a famous singer.	.81
As a child I never reflected on my singing abilities. I just sang.	.81
As a child I used to sing whenever I could.	.77
As a child I was a member of our school choir.	.80
As a child I used to sing more often than my friends.	.77

Table 24. The concept *Performance anxiety*

Concept <i>Performance anxiety</i>	
	Cronbach's α if item deleted
I love imitating foreign languages.	.88
I do not get nervous when I am supposed to do a spontaneous imitation of something.	.87
I enjoy singing a song that has just been played to me.	.86
I am not afraid of performing a song in front of people.	.85
I enjoy singing and making music in front of people.	.85
I prefer singing alone at home when no-one can hear me.	.86
I don't mind giving a speech in front of people.	.86
I enjoy giving a speech in front of a larger audience.	.85
I enjoy being in the spotlight with other people listening to me.	.85
I am always nervous whenever I have to give a presentation.	.87
I enjoy being on a stage.	.85

8. Statistical analysis

Even though this investigation does not focus on how statistical analysis should be done, it is necessary to include a short passage that provides information about the main ideas of the statistical analysis of this investigation.

Generally speaking, before an analysis is carried out the data has to be checked for outliers and its normal distribution. Then correlational, regression, factorial or group analysis should be performed, which has also been performed in this investigation. Additionally, correlations and group comparison should also be corrected for multiple comparisons.

Considering the literature, various ways of how the data can be analysed are presented, but hardly ever has a unitary approach within the scientific field been demonstrated. This may be the case since there are various ways of analysing the data. This is why it is very difficult to find proper methods and to interpret the results. To overcome this issue, it may be better to rely on knowledge based on professional statisticians and to be as conservative as possible.

A second crucial parameter which facilitates statistical analysis in a long run is the strength of the test design, which includes factors such as the number of samples used or the participants hired. A strong test design can increase the reliability of the results. This is why also as many samples and participants as possible have been included in this investigation to have robust data, but at the same time testing conditions, as described in chapter 7, have been kept as short as possible. Instead participants had to come a second time.

In the following short passage, the choice of the statistical methods will be discussed. Correlational, regression and factor analysis as well as group comparisons form the basis of the data analysis. The correlational analysis has only been applied in order to observe the relationships between the individual variables, but also to understand which variables should be entered in multiple regressions.

Regression analysis is a very common way “of predicting an outcome variable from one predictor variable [...] or several predictor variables [...]” (Field 2009: 198). There are different methods which can be used: hierarchical (blockwise entry), forced entry, or stepwise methods. The first one, the hierarchical method, assumes predictors being chosen based on past research whereby the most important variables come first. The forced entry method includes that all predictors are “forced into the model simultaneously” (Field 2009: 212), whereas the stepwise method is based upon solely mathematical aspects.

In this investigation, the stepwise method has been chosen for defining the predictor variables. Even though past research is available with similar predictors (Christiner 2013; Christiner & Reiterer 2013; Christiner & Reiterer 2019; Christiner et al. 2018), this approach seemed to be more objective. Regression models in this investigation have been used for all individual unfamiliar language imitation tasks.

Another option to analyse the data collected is to run factor analyses. This tool can be used to reduce the data set into smaller factors (Field 2009: 626). In language aptitude research factor analysis has been used quite frequently (Petersen & Al-Haik 1976; Grigorko et al. 2000; Dowling 1999; Doughty et al. 2008), where it has primarily been applied in order to understand whether specific measurements belong to one or more factors. However, factorial design has also been applied to compare individual components of different aptitude tests. This means that new aptitude measurements can be compared to previous ones to understand whether the same, similar or different dimensions are tested. This is also an important parameter in this investigation. First, phonetic aptitude as measured by unintelligible language imitation tasks may load onto one or different factors which could also depend on language typology. The same could be true for the speech perception tasks or the

musical measurements. Thus a factor analysis will reveal crucial information about the aptitude measurements of music and speech in this investigation.

A third parameter which has been included in the research design is a group comparison design. Group comparisons can be performed in several ways. The best option is for categories to be provided in which participants label themselves accordingly. Individual differences within the groups can be observed by performing ANOVAs, MANOVAs, t-tests for parametric analysis or Kruskal-Wallis tests and Mann-Whitney tests for non-parametric analysis. Other possibilities include creating extreme groups in which continuous variables are transformed into categorical ones. Even though the latter is a common practice in social sciences, it is questionable for several reasons. For instance, it results in the variability of the continuous variable being reduced, or participants potentially being misclassified (Cohen 1983). Alternatively, a taxometric analysis could be applied to understand the categorical structures of data (McGrath & Walters 2012). Extreme groups have thus been rejected for this research. Instead, in this investigation, participants had to label themselves as a vocalist, as an instrumentalist, as amateurs (participants who play musical instruments) and as a non-musician in order to overcome this issue.

In the following chapter the test development of a new way of measuring speech perception will be presented and discussed. For these descriptive elements of the results, correlational analysis and group comparisons are employed to get initial results.

The empirical investigation features a factorial design that includes all aptitude measurements to isolate different factor loadings. This should also facilitate the understanding of the relationship of all measurements, which in turn will reveal relevant information for the scoring and grouping of the measurements for the follow-up analysis. The latter includes correlational analysis, followed by regression models with which the main predictors that can explain the variances of performances can be discussed. Finally, group comparisons between singers, instrumentalists, amateurs and non-musicians will be discussed in section 10.

9. The development of the MULT AP (Multilingual Aptitude Test)

The language aptitude speech perception test MULT AP and its development will be discussed in this section. The original idea to develop the MULT AP was inspired by five main components: measurements of phonological STM capacity, perceptual musicality tests, language aptitude tests, previous ways of measuring phonetic aptitude (Reiterer et al. 2011; Christiner 2013; Christiner & Reiterer 2013, 2015; Christiner 2018; Christiner & Reiterer

2018; Christiner et al. 2018), and a reoccurring finding which suggests a certain kind of dissociation between perception and production in speech and music (music perception versus singing).

Studies that compare and contrast perception and production in the context of language aptitude are rare. Following traditional research in the 20th century, psycholinguistic models, such as the *TRACE Model of Speech Perception*, suggest that speech perception and production can be separated (McClelland & Elman 1986), while other models, such as the *Motor Theory of Speech Production*, propose that speech production is crucial for speech perception (Liberman & Mattingly 1985). In L2 research, the *Perceptual Assimilation Model* proposes that perception and production of speech share representations which rely on articulatory gestures (Best 1995). This model is similar to the *Motor Theory of Speech Production* and suggests that both perception and production are related. The latter notion has also been facilitated by recent research which has illustrated that listening to speech increases activation of brain regions which are typically associated with motor activity (Watkins et al. 2003).

Surprisingly, however, research has also noted the opposite. For instance, foreign speech perception and production can be dissociated on a behavioural and partially also on an anatomical basis (Golestani & Pallier 2007). In this study, the authors conclude that individuals who perceive foreign speech sounds faster do not necessarily pronounce them better. Another study has shown that the ability to alter perception can be imposed by vocal motor command changes (Nasir & Ostry 2009), which provides new perspectives on the primarily dominating view that perception precedes production (Hutchins & Moreno 2013). Obviously, the process can also be reversed. Similar phenomena have furthermore been reported in music research. Pitch discrimination and pitch production ability as in singing could be dissociated (Loui et al. 2009). These findings have led to notions about dual-route models that propose that aside from the process of conscious perception, alternatively motoric representations can directly be encoded in vocalisation (Hutchins & Moreno 2013).

Previous investigations by my colleagues and me have revealed similar results that suggest dissociations between production and perception, but the research design showed some limitations which will be revised in the present study. Thereby, sample collection and test design should be refined, adapted and extended to be able to draw some more generalisations about the dissociation of perception and production. Former studies included speech imitation, song production and music perception tasks that aimed to reveal their relationships (Christiner & Reiterer 2013, 2015; Christiner & Reiterer 2019). Findings

illustrated that music perception ability rarely turned out to predict high ability to reproduce new language material, or the ability to sing (Christiner 2013; Christiner & Reiterer 2013, 2015; Christiner & Reiterer 2019). To be more precise, it has been noted that vocal behaviour such as singing ability always yielded stronger relationships to unintelligible speech imitation ability (Christiner & Reiterer 2013; Christiner et al. 2018; Christiner & Reiterer 2019) and the ability to mimic foreign accents (Coumel et al. 2019) rather than perceptual musical measurements did.

Even though perceptual musical aptitude such as tonal or rhythmic ability also correlated with accent imitation and speech imitation ability, music perception played a minor role especially when professional singers were tested (Christiner 2013; Christiner & Reiterer 2013, 2015). In all of the follow-up research evidence was mounting that singing may be a better predictor for speech imitation than music perception ability, at least for specific groups such as singers.

However, the limitations of those studies were that only music perception tasks were integrated in test designs and not unintelligible speech perception tasks, which makes it possible to directly compare speech imitation with speech perception. This could additionally explain on which basis singing and speech imitation are related, and in turn, reveal whether the relationship between music and language perception is similar to that of the production counterpart: speech imitation and singing ability.

Achieving this required a similar test design for music and speech perception like it had already been created for the singing and speech imitation tasks, so that the perception designs for music and speech are also comparable by the nature of their test design. Therefore, a suitable speech perception test needed to be created which is comparable to musical aptitude tests (e.g. AMMA test) since an equivalent for language aptitude, which integrates these aspects, does not exist. The aim was hence to integrate the new way of measuring speech perception aptitude into a previous test design which will be shown in chapter 10.

9.1. Test development of the MULT AP

Existing validated language aptitude tests such as the MLAT or the LLAMMA test are great tools to measure language aptitude. However, they were not suitable measurements for this project as already mentioned. Even though the LLAMA test (LLAMA D) is a sound recognition task (Meara 2005), it has no equivalent part in which the same language should be imitated (see section 2.3.6).

In addition, traditional research would favour integrating more than one language for three main reasons. First of all, it cannot be assumed that every unintelligible language will be perceived similarly well. Second, using more languages increases the validity and reliability of the research, which was another reason for rejecting the LLAMA test, or any other available existing test. Third, people possess language aptitude such as polyglots will be able to perform better in more than one language, which would be a stronger indicator for measuring language aptitude. As none of the already established tests has offered multiple language testing, this was another reason for rejecting all of them.

As I wanted to compare the production tasks with the perception tasks of the same languages, the development of a new test was unavoidable. Therefore, other cognitive tests that measure similar concepts such as the phonological STM, music perception and other language aptitude tests have been analysed to develop a suitable speech perception test design.

First, the typology of the speech perception tasks also needed to be comparable to that of the speech production tasks, which required using the same languages. This included the selection of the languages used. The speech production tasks are based on the “Hindi test” developed by Reiterer and Dogil (2009). This is a speech imitation task that consists of words and phrases up to eleven syllables in length. As a syllable length of between nine and eleven syllables turned out to be most predictive, I developed similar imitation tasks using words of the same length in several other languages for testing speech imitation. On this basis seven languages were chosen depending on the following four main criteria. The first and most important one was that the languages should be unintelligible to the participants, which is why frequently spoken foreign languages in Austria or Germany such as French, Spanish or Italian were rejected. The second criterion was the syllable length of between nine and eleven syllables. The third parameter was that languages vary with regards to their foot rhythm and can be classified into stress-timed (Russian, Farsi), syllable-timed (Hindi, Tagalog) and mora-timed languages (Tamil, Japanese), the purpose of which was to ensure that there are typologically different languages are included in the research design. The fourth criterion was that two tone languages (Thai, Mandarin Mandarin) were selected as comparisons to music and singing measurements may reveal interesting findings. The languages Tagalog and Mandarin have already been successfully used in another study and turned out to be as predictive as the Hindi samples were (Christiner et al. 2018). For the development of the production tasks see section 10.3.1.

Second, WMC and the (phonological) STM can be measured in different ways. The STM is often measured by repeating strings of numbers in a numerical order (Baddeley 1992; Thorn & Gathercole 2001; Engel de Abreu & Gathercole 2012; Baddeley 2003a, 2003b), or by repeating simple, dissimilar (Engle et al. 1999), or non-words (Kormos & Safar 2008). Complex WM span tests can include picture matching tasks (Baddeley 1992), operational span tests, counting spans, or reading spans (Conway et al. 2005), which has been discussed in detail in chapter 5.

To recap, WM can be described as a temporary storage system in which information for complex cognitive tasks are stored (Baddeley 1992). The *Measurements of Adult Intelligence* by Wechsler (1939) includes a typical digit span sub-test where strings of numbers need to be repeated in the same order, and a backward span where the numbers have to be recalled in reversed order. As previously discussed, this represents a phonological STM measurement which is an important parameter that has influenced MULT AP design in terms of length of the utterances used to measure unintelligible speech perception.

The mean of the numbers which are repeated is seven for digits in German (Christiner & Reiterer 2013; Christiner & Reiterer 2019) and English (Gathercole & Baddeley 1993) for adults. In addition, the number of items remembered also depends on word length, which can vary across languages (Thorn & Gathercole 2001). For children, considerably lower number recall has been reported, such as three to four digits (Roman et al. 2014; Christiner & Reiterer 2018) or two to three digits (Gathercole & Baddeley 1993), depending on the type of research. Similar observations have been made for speech material, such as non-word memory spans which are repeated, for instance (Benner 2005).

Previous research has shown that the mean of the non-words remembered in the correct order is considerably lower compared to the digit span ranging between four and five (Christiner & Reiterer 2013; Christiner et al. 2018; Christiner 2013), which is consistent with other researchers who report the same numbers for unrelated words (Cowan 2001; Baddeley 2010). The length in both tasks plays a significant role and in this respect it is crucial to put emphasis on the design, which in both tasks includes the requirement that the numbers or the non-words are repeated. This further limits the memory span. Therefore, memory span differences between imitation and perception tasks have to be discussed in the context of perceptual musical aptitude tests in order to understand the rational idea about the length choice of the speech perception stimuli and responses.

Musical aptitude tests such as the PMMA, the IMMA, and the AMMA by Gordon (1979, 1982, 1989a) have already been shown in chapter 3. These tests require the participants

to listen to two melodies which are either the same or include a rhythmical or tonal change in the second melody played. While the PMMA and the IMMA are short as both are targeted at children, for the AMMA test rather complex and lengthy melodies were created. This provided additional evidence for the appropriate length for speech perception stimuli.

Comparing musicality tests to the test designs of the digit and non-word memory spans, the latter two STM measurements show one difference which is quite important. They require the digit or the (non-) words to be repeated, while the melodies of the AMMA test are only discriminated as mentioned above. Consequently, the rehearsal ability is thus not interrupted by the auditory input of one’s own voice, which lowers the risk that information is lost. Therefore, I could assume that sequences of a similar length to the lengthy melodies of the AMMA test, but at least longer than those of the production tasks, would be suitable for the speech perception test, except for one difference. Whereas the AMMA test plays musical statements twice, I decided to take one or more elements out of the string previously presented for the correct answer, or I played another example for the different condition. This is based on the language aptitude LLAMA D subtest by Meara (2005), which was discussed in section 2.3.6. The LLAMA D test consists of stretches of utterances which are presented first, followed by utterances which need to be remembered. It must be indicated whether they had been presented before or not (Meara 2005).

Looking at all the test designs, I assumed that strings of between four and six words for young children (the subtest for children is omitted in this investigation), and eight, ten and twelve different language constituents for adolescents and adults may be appropriate in terms of length, followed by the response which consists of one, two or three items depending on the difficulty.

Table 25 shows the original basic test construction for adolescents and adults.

Table 25. The original basic test construction

String	Responses: three possibilities of increasing difficulty
1 2 3 4 5 6 7 8	1 or 1 2 or 1 2 3
1 2 3 4 5 6 7 8 9 10	1 or 1 2 or 1 2 3
1 2 3 4 5 6 7 8 9 10 11 12	1 or 1 2 or 1 2 3

9.2. Developing the procedure

The first step included language selection and recruiting participants who were native speakers of specific languages, together with whom the samples were developed. To have enough languages, native speakers of twelve languages were selected, eight of which were taken for the final test. For each language, words, phrases and sentences were collected. The syllable range was from one to eleven syllables, with at least three different words, utterances or sentences being recorded for each syllable. The recording took place in the music studio with music software Steinberg CUBASE 6. All sound files were normalised for loudness by a professional sound technician. Once the collection of samples has been completed, the prototype test was designed, after which a website was created and a computer programme was developed which satisfied the demands of this language perception test. The next step included preliminary testing where the speech perception test designs for adolescents, adults and children were evaluated and refined. The children design is excluded in this investigation and the version for adolescents and adults is the same.

9.3. About the tasks

This perception task consists of either mixed (different languages) or pure (one single language) language strings of eight, ten or twelve different constituents or phrases depending on task difficulty. The syllable length of the constituents in each string varies from one to nine syllables, and participants, after they hear a string, have to listen to either one, two, or three comparative phrases and to indicate whether the phrase or phrases was/were included in the string or not. Additionally, they need to indicate whether they are very sure, rather sure, or unsure about their answers. For each correct answer, participants get a point. This is the main rational logic behind the test design (see 9.4 and appendix 2 for each task).

Some further criteria must be mentioned as well at this stage, namely the rational idea of deciding to include mixed and pure language strings in the test. By mixed language conditions it is meant that a single string consists of different languages, and pure conditions means that a string and response consists only of one language spoken by the same speaker. Mixed conditions (*ML perception*) will not provide information about the ability to perceive a particular language, but they will be easier than the pure conditions (*PL perception*) as participants can rely on different speakers and different languages.

In turn, it can be suggested that the pure conditions may be more difficult. Research has continually observed that speech and music perception are related (Schön et al. 2004; Peretz et al. 2015; Seither-Preisler et al. 2018; Seither-Preisler et al. 2014; Chobert et al. 2014;

François et al. 2013; Besson et al. 2011; Moreno et al. 2009; Oechslin et al. 2010). Therefore, it could be suggested that musicians who possess musical aptitude may perform better in the pure conditions rather than non-musicians.

Secondly, the individual languages of the pure conditions should be directly comparable to the language imitation so that it is possible to observe whether speech perception and production are related.

9.4. Final test design

The MULT AP for adolescents and adults consists of multiple items, and each item consists of a so called Stimline and a Stimcompare. A Stimline consists of a certain number (usually between 8 and 12) of Stims, where a Stim is a word in a foreign language with a certain number of syllables. A Stimcompare consists of 1, 2, or 3 Stims and is compared against the Stimline by the participant.

The participant listens to the Stimline first, whereby each Stim is separated from the other by a pause of 50ms. 2 seconds after listening to the Stimline the participant listens to the Stimcompare. Then the participant has to tell whether (all of the Stims in) the Stimcompare was/were included in the Stimline or not, and indicate how certain they are on a scale from 1 to 3 (“very sure”, “rather sure”, “can’t tell”). After the test is complete, a score is calculated according to the number of correct answers given. A familiarisation task at the start of the test is not counted, and the items are presented to the participants in randomised order. The 3 types of Stimcompares (1, 2 or 3 Stims) are categorised into subsections (A, B and D). The letter C is omitted to avoid confusion with the meaning of “correct” for the answer given.

Example:

The Stimline consists of 8 Stims and the Stimcompare consists of 2 Stims. These 2 Stims can either both included in the Stimline, one of them can be included in the Stimline, or none of them can be included in the Stimline. The participant is then asked whether the Stimcompare was fully included, and can answer either “yes” or “no”. The answer “yes” is only correct if all 2 Stims of the Stimcompare were included in the Stimline, and the answer “no” for all other cases.

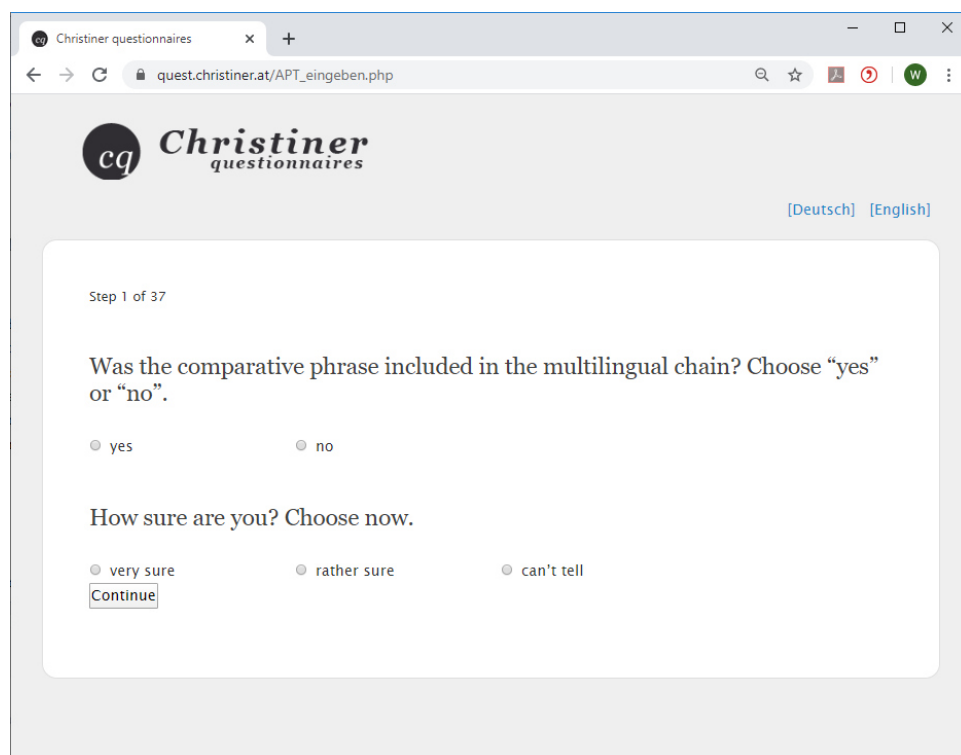
Figure 3 shows the screen where the participant has to provide an answer to the item.

Initially the prototype of the test consisted of 17 Stimlines with two different Stimcompares in each subsection (A, B, D) which resulted in 102 items. As the test turned out to take too long, causing participants to give up, it was shortened by applying Cronbach’s α

and removing items that were too good and too poor (easy or hard) the final test “Multap-VS” – where the “VS” stands for “very short” – resulted in 37 items.

Figure 4 shows the interpretation of the database-output describing all items, Stimlines, Stimcompares and Stims.

Figure 3. The screen where the participant has to provide an answer to the item.



The screenshot shows a web browser window with the address bar displaying "quest.christiner.at/APT_eingeben.php". The page header features the "cq Christiner questionnaires" logo and language links for "[Deutsch]" and "[English]". The main content area is titled "Step 1 of 37" and contains the following text: "Was the comparative phrase included in the multilingual chain? Choose “yes” or “no”." Below this text are two radio button options: "yes" and "no". Further down, the text reads: "How sure are you? Choose now." followed by three radio button options: "very sure", "rather sure", and "can't tell". At the bottom left of the form, there is a "Continue" button.

Figure 4. The database Output

Multap - VS

Stimline 13

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

⑥

⑦

⑧

⑨

⑩

①

②

③

④

⑤

① **S-Id:** Stimline/Stimcompare-Identifier

② **SCS:** Stimcompare subsection

The subsection of the Stimcompare; within the Stimline this value is 0, within the Stimcompares it describes the subsection (A, B, D).

③ **SCNR:** Stimcompare number

Number of the subsection of the Stimcompare

④ **Pos:** Position of the Stim within the Stimline/Stimcompare

⑤ **Language:** Language of the Stim

⑥ **Syll:** Syllable length of the Stim

⑦ **Stim-No:** Recording number of the Stim with the same syllables

⑧ **Identifier:** Database identifier of the stim

⑨ **f_c:** “false” or “correct”

Printing “c” or “f” in the Stimcompare section, see ③

If “false” the correct answer to the item is “not included in Stimline”, if “correct” the correct answer to the item is “is included in Stimline”.

⑩ **SL_Po:** the position of the Stim within the Stimline.

For the Stimline section this number is the same as in ④ but for the Stimcompare it shows whether the Stim is included in the Stimline – the position number in the Stimline is printed – or if the Stim is not included in the Stimline – there is no number printed.

Ⓐ **The Stimline section**

e.g. “SL13” read: Stimline 13

Ⓑ **The Stimcompare section**

e.g. “SC-13-A1” read: Stimcompare to Stimline 13, subsection A, number 1

Ⓒ **false or correct,** see ⑨

9.5. Testing Procedure

The MULT AP alone simply provide information about the discrimination ability of different languages and thus also includes a questionnaire that collects data on age, gender, first language, foreign language proficiencies, educational status, music and singing background. As the MULT AP has also been designed to compare perception with production, this test should be used together with musicality tests or other language aptitude measurements. The test has been programmed and takes on average around 25 minutes including the questionnaire.

The perception test has been designed for large-scale investigations and to reach a large number of people in order to get valid and reliable information. Participant selection played one fundamental role in this respect. In marked contrast to most investigations, here the conscious selection of people has been avoided as the participants should be as close as possible to a natural population. The selection for the final sample was based on the answers of the questionnaire where the participants reported whether one or more of the languages were intelligible to them or not. If they indicated that the languages were unintelligible to them they were included in the final analysis. Even though this meant that many of the participants had to be excluded from the final analysis, 434 participants fulfilled all criteria. The participants were selected according to their language proficiency, meaning that participants who spoke or understood the content of one or more languages were excluded from the research. One fundamental aspect of this test is that the language material should be unfamiliar to the participants so that they need to rely on the acoustic information only.

The test was accessible online, which was necessary in order to collect data of a large sample, which is required for the analysis and validity of a test. The participants received login details which allowed them to enter whenever they wanted. As controlling test conditions is very difficult if the experimenter is not present while the testing takes place, several control mechanisms were installed to ensure that conditions are as equal as possible for everyone.

First, this included very precise instructions of the test in order to make sure that everyone understood the task. Second, participants were given instructions concerning their computer speakers, which were tested in the familiarisation task before the start of the main test. The answers to the familiarisation task were provided before the test started. In the test a control function of the programme prevented participants from listening to sound files a second time. Participants had to fill in their responses first, before the programme allowed

them to continue. In the event that the internet connection failed or the participants stopped, the programme continued where the last response was given.

9.6. Participants

As already mentioned, participants were asked to perform the test online. The login details were provided. 767 participants made attempts to take the test. 661 of the participants finished at least the first half of the test. The final version, the MULT AP “VS” which consists of 37 items, was taken by 434 people who fulfilled all criteria to be included for the final analysis.

The mean of the age range of the participants was 21.88 (min. 11; max 68) years. 266 of them were female and 168 were male. The residence of the participants was mainly Austria ($N = 256$), followed by Germany ($N = 148$) and others.

9.7. Language background

The selection of the participants was dependent on whether they understood one of the languages integrated in the test design or not. Most of the participants' first language was German ($N = 387$), followed by Turkish ($N = 6$), Croatian ($N = 6$), English ($N = 5$), Serbian ($N = 5$), Bosnian ($N = 3$), Armenian ($N = 3$), and other languages appearing once each. 386 participants were monolingual, 39 participants were bilingual and 9 were trilingual, with the most common second language being German ($N = 21$), English ($N = 10$), Turkish ($N = 4$), Croatian ($N = 4$) and other languages appearing once.

Strictly speaking, as Russian was also included in the test design it could be suggested that the 14 participants whose first language is Croatian, Serbian and Bosnian may have understood fragments in Russian. All of them, however, reported that the Russian words used were unintelligible to them. Even if they had understood some Russian words, these occur just in the *ML perception* conditions, meaning that the potential influence on the language aptitude scores is extremely low and can thus be ignored.

The most frequently spoken foreign languages were English ($N = 388$), German ($N = 10$), French ($N = 7$), Romanian ($N = 2$), Dutch ($N = 2$), and other languages appearing once each. 18 participants spoke no foreign language, 141 participants stated that they speak one foreign language, 214 indicated that they speak two foreign languages, 44 participants spoke three foreign languages, 12 mastered four languages, 1 five languages and 5 participants' responses were incomplete.

9.8. Educational status

The participants' educational status varies considerably and was entered according to the educational status which was completed at testing time. 99 participants finished primary school only, 65 the main general secondary school, 18 the technical and vocational school, 156 secondary academic school (general qualification for university entrance), 10 the post-secondary non-tertiary education, 2 the mastership education, 27 Bachelor studies, 46 Master studies, and 11 had a Doctoral degree.

In addition, the educational status of the parents was also collected, and even though the variables will not be discussed since this would go beyond the scope of this investigation, will be shown. The educational background of the mothers is as follows: 2 of them received no school education, 5 finished primary school only, 66 the main general secondary school, 122 the technical and vocational school, 98 secondary academic school (general qualification for university entrance), 40 the post-secondary non-tertiary education, 6 the mastership education, 25 Bachelor studies, 61 Master studies, and 9 had a Doctoral degree.

The educational background of the father is as follows. 4 fathers received no school education, 8 finished primary school only, 54 the main general secondary school, 123 the technical and vocational school, 71 secondary academic school (general qualification for university entrance), 13 the post-secondary non-tertiary education, 39 the mastership education, 24 Bachelor studies, 70 Master studies, and 28 had a Doctoral degree.

9.9. Musical background

The participants reported the number of hours they sang on any occasion on average in a week. The instructions included any form of singing whether it was for professional or non-professional purposes. The latter also included singing in the car or the bathroom. On average participants sang around 1.8 hours a week. 15 participants reported that they are professional singers. The singing self-report questionnaire showed that 50 participants claimed to be "very poor" singers, 89 to be "poor" singers, 169 claimed to sing like the "average", 104 indicated they were "good" singers and only 22 indicated they were "very good" singers.

Furthermore, the type of instrument played was also collected. In addition, the participants had to indicate the first one they played best. The latter will form the basis for the group comparisons in which the individual instruments will be grouped together into three main categories: keyboard instruments, wind instruments, and plucking and string instruments (see section 9.10.3.3). 170 played no musical instrument, 83 played piano or keyboard as their

first instrument, 55 played the guitar, 30 played recorder, 15 played the violin, 15 played the drums, 15 played other wind instruments, 12 transverse flute, 9 trumpet, 7 played the saxophone, 7 played another plucking instrument, 6 other string instruments, 4 accordion, 3 clarinet, 2 other percussion instruments and 1 an electronic musical instrument.

32 participants reported that they were professional musicians. In addition, the participants reported the number of instruments they played. While 170 participants played no musical instrument, 145 played one, 79 played two, 26 played three, 9 played four, 3 played five, 1 played seven and 1 played nine musical instruments.

9.10. Statistical analysis

As the test should be used together with other aptitude measurements, the statistical analysis focuses on the basic elements of the test. Comparisons to other aptitude measurements will be provided in the empirical investigation in chapter 11. Thus the descriptives and the normal distribution will be shown in appendix A2.2 and the correlations to the variables of the questionnaire will be shown in table 27.

Finally, group comparisons will be undertaken with the aim of illustrating whether gender, age, musical training and the type of musical instrument played has an influence on speech perception ability. For the latter, the main scores, these being *Overall Aptitude perception score*, the *PL perception* and the *ML conditions*, will be analysed. The group comparisons are also corrected for multiple testing by applying a Benjamini-Hochberg correction.

9.10.1. The scoring of the main MULT AP variables

The main variables of the MULT AP have been used and are presented in the table below. Those are the *Overall Aptitude perception* which is comprised of all 37 language perception tasks, the *PL perception* and the *ML conditions*. The score consists of 5 tasks in Tagalog, Hindi, Mandarin, Japanese and Farsi and comprises 25 items in total. The *ML condition* score includes all eight languages and consists of 12 items. The individual five languages' scores consist of Mandarin, Tagalog, Japanese, Hindi and Farsi and their relationship will be discussed further in section 9.10.3. The scores for all variables range between 0 and 1. The descriptives of the variables and the normal distribution have been provided in the appendix A2.2.

Table 26 shows the scoring.

Table 26. Scoring MULT AP

Name	Constituents
Overall Aptitude score	37 all items
Pure language score	5 items in Hindi 5 items in Mandarin 5 items in Japanese 5 items in Tagalog 5 items in Farsi
Mixed language score	12 items (each string consists of different languages)

9.10.2. Correlations

The correlations have been corrected for multiple comparisons by applying a Benjamini-Hochberg correction. Variables that are still significant after correction are indicated by a +. Variables that do no longer correlate with the main speech perception scores and are non-significant after correction are not discussed in the following section.

Table 27. Correlations

	Overall Aptitude perception	PL perception	ML perception	Mandarin perception	Tagalog perception	Japanese perception	Farsi perception	Hindi perception	Highest level of education	Highest education of mother	Highest education of father	No of LIs	No of foreign languages	Singing regularly hours per week	Self estimation of singing capacity	No of musical instruments played	Age
Overall Aptitude perception	1	.82***	.60***	.36***	.43***	.49***	.37***	.36***	.18***	.20***	.21***	.02	.12*+	-.06	-.00	.16***	.17***
PL perception	.82***	1	.15***	.46***	.45***	.58***	.48***	.25***	.24***	.19***	.20***	.03	.10*	-.06	-.01	.13***	.24***
ML perception	.60***	.15***	1	.06	.14***	.10*	.02	.02	-.06	.08	.11*	.02	.03	-.02	-.00	.09	-.07
Mandarin perception	.36***	.46***	.06	1	.01	.06	.07	-.00	.09	.01	.02	-.02	.02	-.03	.00	.03	.10*
Tagalog perception	.43***	.45***	.14***	.01	1	.05	-.07	.14***	.07	.13***	.16***	.06	.13***	-.07	.06	.10*	.06
Japanese perception	.49***	.58***	.10*	.06	.05	1	.07	.06	.28***	.22***	.20***	.03	.14***	-.06	-.02	.09	.25***
Farsi perception	.37***	.48***	.02	.07	-.07	.07	1	-.01	.08	.04	.06	-.01	-.04	-.02	-.03	.02	.08
Hindi perception	.36***	.25***	.02	-.00	.14***	.06	-.01	1	.11*	.01	.01	.01	.08	.03	-.01	.09	.12***
Highest level of education	.18***	.24***	-.06	.09	.07	.28***	.08	.11*	1	.16***	.19***	-.12*	.44***	-.19***	-.06	.15***	.88***
Highest education of mother	.20***	.19***	.08	.01	.13***	.22***	.04	.01	.16***	1	.54***	-.05	.16***	-.04	.10*	.19***	.06
Highest education of father	.21***	.20***	.11*	.02	.16***	.20***	.06	.01	.19***	.54***	1	-.13***	.15***	-.06	.09*	.17***	.13***
No of LIs	.02	.03	.02	-.02	.06	.03	-.01	.01	-.12*	-.05	-.13***	1	-.11*	.04	.04	-.03	-.10*
No of foreign languages	.12*+	.10*	.03	.02	.13***	.14***	-.04	.08	.44***	.16***	.15***	-.11*	1	.02	.01	.20***	.33***
Singing regularly hours per week	-.06	-.06	-.02	-.03	-.07	-.06	-.02	.03	-.19***	-.04	-.06	.04	.02	1	.41***	.15***	-.23***
Self estimation of singing capacity	-.00	-.01	-.00	.00	.06	-.02	-.03	-.01	-.06	.10*	.09*	.04	.01	.41***	1	.28***	-.09
No of musical instruments played	.16***	.13***	.09	.03	.10*	.09	.02	.09	.15***	.19***	.17***	-.03	.20***	.15***	.28***	1	.08
Age	.17***	.24***	-.07	.10*	.06	.25***	.08	.12***	.88***	.06	.13***	-.10*	.33***	-.23***	-.09	.08	1

9.10.2.1. Main scores:

First, the *Overall Aptitude perception* score correlates with all speech perception tasks, which is not surprising as they are contained in the entire score. Interestingly, the *Overall Aptitude perception* is also related to the educational scores of the participants $r_s = 0.18$, p (two-tailed) < 0.01 , the education score of the mother $r_s = 0.20$, p (two-tailed) < 0.01 , and the father $r_s = 0.20$, p (two-tailed) < 0.01 . Additionally, there is a correlation to the *Number of foreign languages* $r_s = 0.12$, p (two-tailed) < 0.05 , the *Number of musical instruments* played $r_s = 0.16$, p (two-tailed) < 0.01 , and age $r_s = 0.17$, p (two-tailed) < 0.01 .

Second, the *PL perception* score also correlates with all speech perception tasks. Surprisingly, the *PL perception* is also related to the educational scores of the participants $r_s = 0.24$, p (two-tailed) < 0.01 , the education score of the mother $r_s = 0.19$, p (two-tailed) < 0.01 , and the father $r_s = 0.20$, p (two-tailed) < 0.01 . In addition, there is a correlation to the *Number of musical instruments* played $r_s = 0.13$, p (two-tailed) < 0.01 , and age $r_s = 0.24$, p (two-tailed) < 0.01 .

Third, the *ML perception* score correlates with the *Overall Aptitude perception* score $r_s = 0.60$, p (two-tailed) < 0.01 , the *PL perception* score $r_s = 0.15$, p (two-tailed) < 0.01 , and *Tagalog perception* $r_s = 0.14$, p (two-tailed) < 0.01 .

9.10.2.2. Subscores of the PL perception score

The subscores *Mandarin perception*, *Tagalog perception*, *Japanese perception*, *Farsi perception* and *Hindi perception* show only few correlations to each other or to other variables. All subscores correlate with the *Overall Aptitude perception* and *PL perception* scores, which is not surprising as they are all contained in the main scores. They will not be explained further. Apart from this, *Mandarin perception* and *Farsi perception* are not related to any of the other variables.

For *Tagalog perception* it is different. It also correlates with *ML perception* $r_s = 0.14$, p (two-tailed) < 0.01 , *Hindi perception* $r_s = 0.14$, p (two-tailed) < 0.01 , the education score of the mother $r_s = 0.13$, p (two-tailed) < 0.01 , the father $r_s = 0.16$, p (two-tailed) < 0.01 , and the *Number of foreign languages* $r_s = 0.12$, p (two-tailed) < 0.05 .

Japanese perception also correlates with the educational scores of the participants $r_s = 0.28$, p (two-tailed) < 0.01 , the education score of the mother $r_s = 0.22$, p (two-tailed) < 0.01 , the father $r_s = 0.20$, p (two-tailed) < 0.01 and to the *Number of foreign languages* $r_s = 0.14$, p (two-tailed) < 0.05 . *Hindi perception* also correlates with *Tagalog perception*.

The correlations show two interesting findings. First of all, it was not expected that education of the participants as well as that of their parents would be related to many of the speech perception variables which should purely measure aptitude. The effect, however, is rather low. Despite this, it should still be observed in order to understand the rationality of this finding (see section 9.10.2.4).

Another interesting finding is that the individual languages *Mandarin perception*, *Tagalog perception*, *Japanese perception*, *Farsi perception* and *Hindi perception* seem to be less related than expected. *Tagalog* and *Hindi perception* are the only two languages which correlate with each other as illustrated in table 27. This is a very surprising finding which is different for the imitation task of the different languages which usually show high correlations (Christiner et al. 2018). This is also true in this investigation and will be discussed in more detail in section 11.1.

9.10.2.3. Gender differences

In order to illustrate whether unfamiliar speech perception ability is distributed differently across gender, group comparisons of the *Overall Aptitude perception*, *PL perception* and *perception* tasks have been run. Results revealed that no significant difference could be observed for gender in all three conditions. No significant difference between female ($Mdn = 0.57$) and male ($Mdn = 0.57$) participants in the *Overall Aptitude perception* condition could be detected, $U = 21952$, $z = -.10$, $p > .05$. No significant difference between female ($Mdn = 0.59$) and male ($Mdn = 0.59$) participants in the *PL perception* condition could be detected, $U = 21843$, $z = -.19$, $p > .05$. No significant difference between female ($Mdn = 0.50$) and male ($Mdn = 0.51$) participants in the *ML perception* condition could be detected, $U = 21967$, $z = -.09$, $p > .05$. The effect size has not been calculated as there was no significant difference found between gender.

9.10.2.4. Highest level of education and age

The relationship of the *Highest level of education* of the participants to the main speech perception variables *Overall Aptitude perception* and *PL perception* is crucial and should be investigated. Generally, both conditions should largely be unrelated to education. In this respect age may play a significant role, as the age range is between 11 and 68 years. It could be suggested that the relationships between education and the two main aptitude scores, the *Overall Aptitude perception* and the *PL perception*, are age dependent. To test this, partial correlations method has been applied.

The original Spearman correlations between the *Overall Aptitude perception* condition correlates with *Highest level of education*, $r_s = 0.18$, p (two-tailed) < 0.01 . The *PL perception* scores also correlates with *Highest level of education*, $r_s = 0.24$, p (two-tailed) < 0.01 .

The partial correlations of the same variables are described below and reveal that when *Age* on the relationship between *the Overall Aptitude perception* and *Highest level of Education* is controlled $r = 0.11$, p (two-tailed) < 0.05 their relationship diminishes but is still significant. When *Age* on the relationship between *PL perception* and *Highest level of Education* is controlled $r = 0.12$, p (two-tailed) < 0.05 their relationship diminishes as well. This indicates that age has a slight effect on two of the main variables. This is why age groups were created, namely in order to illustrate at which age speech perception ability increases, which will be illustrated below.

9.10.3. Group comparisons

9.10.3.1. Age groups and speech perception

As shown before, the correlations between education and the two main scores *Overall Aptitude perception* and *PL perception* diminish when controlled for age. Therefore, age groups for adolescents and adults have been created with the aim of illustrating at which age group differences occur in the ability to discriminate foreign speech.

The adolescents have been divided into adolescents I (aged 11-12), adolescents II (aged 13-15), and adolescents III (aged 16-17). The age range of the adult group is 18 to 68. The reason for deciding to split the adolescents into three groups is due to the fact that significant differences to other groups were expected to be found for the young participants. The aim was to find at which particular age the participants performed significantly better.

As the data shows a non-normal distribution, a Kruskal-Wallis test, followed by Mann-Whitney tests have been applied. Table 28 shows the significance of the Kruskal-Wallis test and the Mann-Whitney tests for all groups.

Table 29 shows the descriptive and the confidence intervals of all age groups. Figure 5 shows the mean differences across age groups.

As expected, a significant difference in the two main scores *Overall Aptitude perception* and the *PL perception* have been detected between the adolescents I group and all the others (see table 28 and 29). The results suggest that thirteen-year-old adolescents possess the same capacity to discriminate unfamiliar speech material similarly to adults.

Figure 5. Age groups

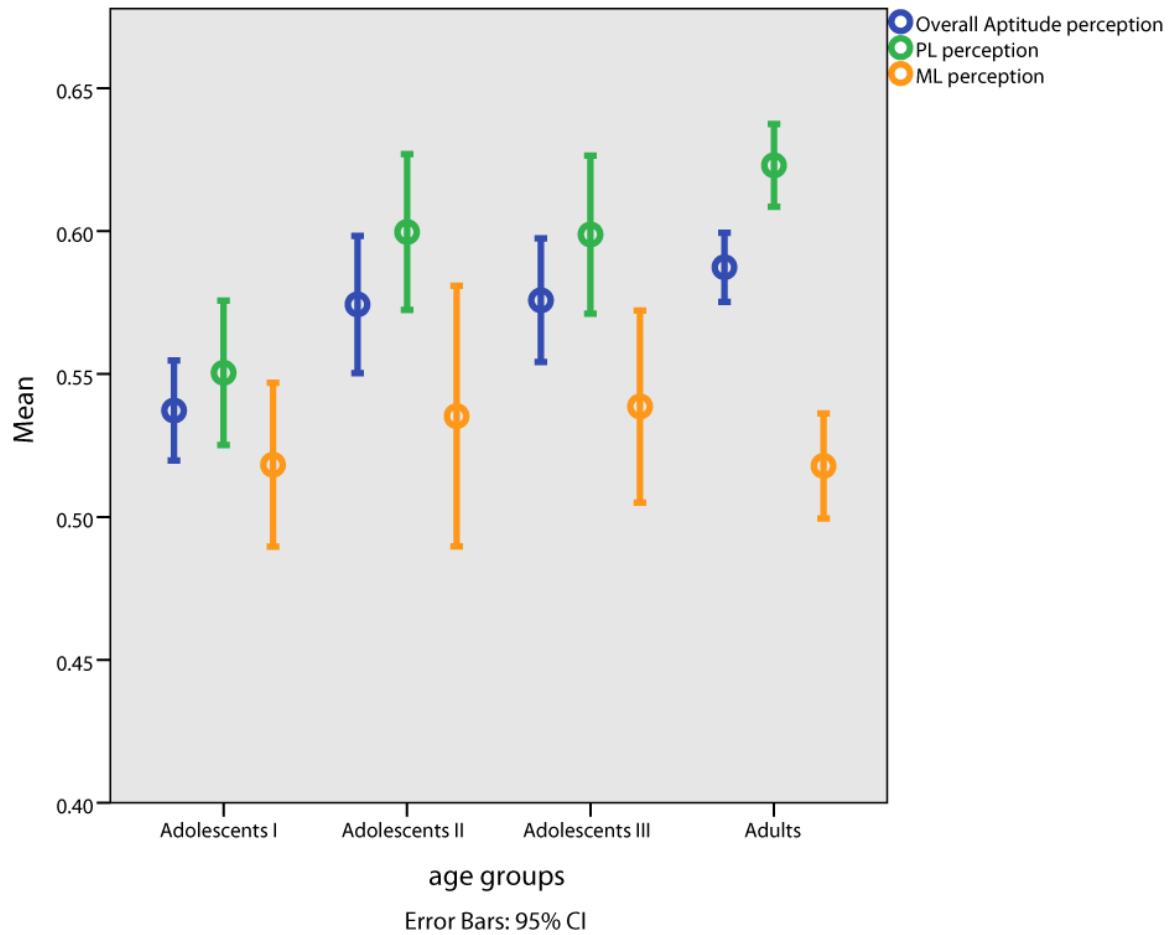


Table 28. Mean differences across age groups

	Overall Aptitude perception		PL perception		ML perception	
	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>
Kruskal-Wallis Test for 4 groups	0.00	sig	0.00	sig	0.69	not sig
Group comparison (Mann-Whitney)						
Adolescents I - Adolescents II	0.01	sig*	0.01	sig*	0.64	not sig*
Adolescents I - Adolescents III	0.01	sig*	0.01	sig*	0.43	not sig*
Adolescents I - Adults	0.00	sig*	0.00	sig*	0.84	not sig*
Adolescents II - Adolescents III	0.90	not sig*	0.92	not sig*	0.84	not sig*
Adolescents II - Adults	0.43	not sig*	0.13	not sig*	0.51	not sig*
Adolescents III - Adults	0.40	not sig*	0.19	not sig*	0.27	not sig*

* Significance after Benjamini-Hochberg correction ($p < 0.05$)

Table 29. Descriptives for age groups.

	Overall Aptitude perception	PL perception	ML perception
<i>N</i>	73	73	73
<i>M</i>	0.537	0.550	0.518
<i>SE</i>	0.009	0.013	0.014
<i>Median</i>	0.541	0.545	0.500
<i>min</i>	0.351	0.318	0.167
<i>max</i>	0.703	0.818	0.833
<i>95% CI Lower</i>	0.520	0.525	0.490
<i>95% CI Upper</i>	0.555	0.576	0.547
<i>N</i>	52	52	52
<i>M</i>	0.574	0.600	0.535
<i>SE</i>	0.012	0.014	0.023
<i>Median</i>	0.581	0.591	0.500
<i>min</i>	0.405	0.318	0.167
<i>max</i>	0.757	0.818	0.917
<i>95% CI Lower</i>	0.550	0.572	0.490
<i>95% CI Upper</i>	0.598	0.627	0.581
<i>N</i>	69	69	69
<i>M</i>	0.576	0.599	0.539
<i>SE</i>	0.011	0.014	0.017
<i>Median</i>	0.595	0.591	0.500
<i>min</i>	0.405	0.318	0.167
<i>max</i>	0.811	0.773	0.833
<i>95% CI Lower</i>	0.554	0.571	0.505
<i>95% CI Upper</i>	0.597	0.626	0.572
<i>N</i>	238	238	238
<i>M</i>	0.587	0.623	0.518
<i>SE</i>	0.006	0.007	0.009
<i>Median</i>	0.595	0.636	0.500
<i>min</i>	0.351	0.364	0.167
<i>max</i>	0.865	0.955	0.917
<i>95% CI Lower</i>	0.575	0.609	0.499
<i>95% CI Upper</i>	0.599	0.637	0.536

9.10.3.2. Musical instrument playing and non-musicians

In order to understand whether instrument playing has an impact on speech perception ability, a Mann-Whitney test was performed for the three main MULT AP variables, *Overall Aptitude perception*, *PL perception* and *ML perception*.

The results revealed that on average in all conditions those who played a musical instrument were better. Age had no effect on the musical measurements. In addition, for multiple testing, a Benjamini-Hochberg method was applied for the correction of the error rate for multiple comparisons. Even after correction, the difference remained significant.

The results revealed that the participants who played a musical instrument ($Mdn = 0.5870$) performed significantly better in the *Overall Aptitude perception* condition compared to those who did not play any musical instrument ($Mdn = 0.5583$). This difference was significant $U = 18594$, $z = -3.094$, $p < .01$, $r = -.15$. The *PL perception* condition also shows that participants who play a musical instrument ($Mdn = 0.6162$) performed significantly better compared to those who did not play any musical instrument ($Mdn = 0.5856$). This difference was significant $U = 19158$, $z = -2.659$, $p < .01$, $r = -.13$. The *ML perception* performance also shows that participants who play a musical instrument ($Mdn = 0.5334$) performed significantly better than those who did not play any musical instrument ($Mdn = 0.5092$) $U = 20046.5$, $z = -1.976$, $p < .05$, $r = -.10$. All, however, show a small effect only.

As expected participants who play a musical instrument performed significantly better in the ability to discriminate unintelligible speech compared to the non-musicians. This implies that playing musical instruments seems to have a positive effect on speech perception ability. Research has provided evidence for the effect of musical training on speech perception ability. Musical training facilitates the detection of speech incongruities (Schön et al. 2004) and improves the ability to perceive new language material (Chobert et al. 2014; François et al. 2013; Besson et al. 2011; Wong & Perrachione 2007; Wong et al. 2007).

One assumption that the *PL perception* condition may be more discriminative than the *ML condition* cannot be sustained, at least in this context. There is a slight trend that indicates that the difference in the *PL perception* condition is larger. Indirectly, this finding also implicitly shows that the language perception tests seem to be predictive in this respect as group comparisons reveal the same information as has been reported by other researchers.

9.10.3.3. Different types of musical instruments and speech perception ability

Research has shown that pitch perception preference influences musical instrument choices. For instance, spectral pitch listeners showed preferences for lower-pitched melodic instruments including singing, while fundamental pitch listeners clearly prefer percussive or high-pitched instruments (Schneider et al. 2005). Whether the type of musical instrument played also influences speech perception is still an unanswered matter of debate and as far as known has not been investigated.

Research in this area would require selecting musicians who play particular instruments or just one instrument to precisely categorise them. However, this is hardly ever possible as musicians often play multiple musical instruments, or change the musical instrument of their choice in the course of time. Even though the research design of this

investigation was not intended for testing speech perception ability and its relationship to the ability in playing particular types of musical instruments, some basic ideas have been demonstrated which should be integrated and investigated more precisely in future research.

The questionnaire required the participants who played a musical instrument to enter the instrument in the order of their proficiency. This at least indicates that the first instrument, if participants play more than one instrument, is the one which they play best. Therefore, group comparisons have been undertaken to determine whether there are individual differences in speech perception depending on the first and best instrument played.

As the participants played a variety of different instruments, running group comparisons with unequal numbers of participants would have not been an appropriate statistical approach. Thus, three main groups have been selected and categorised as 1) keyboard instruments consisting of piano, keyboard and accordion (N=81), 2) wind instruments consisting of clarinet, recorder, transverse flute, saxophone, trumpet and other wind instruments and, 3) plucking and string instruments comprising guitar, violin, and other string and plucking instruments. All other categories including non-musicians have been discarded from this analysis.

No significant difference between keyboard instrument players (Mdn = 0.62) and wind instrument players (Mdn = 0.57) in the *Overall Aptitude perception* condition could be detected, $U = 2737.50$, $z = -1.90$, $p > .05$. No significant difference between keyboard instrument players (Mdn = 0.64) and wind instrument players (Mdn = 0.59) in the *PL perception* condition could be detected, $U = 2772$, $z = -1.79$, $p > .05$. No significant difference between keyboard instrument players (Mdn = 0.58) and wind instrument players (Mdn = 0.50) in the *ML perception* condition could be detected, $U = 2741$, $z = -1.91$, $p > .05$. The effect size has not been calculated as there was no significant difference across groups.

No significant difference between keyboard instrument players (Mdn = 0.62) and plucking and string instrument players (Mdn = 0.60) in the *Overall Aptitude perception* condition could be detected, $U = 3346$, $z = -.83$, $p > .05$. No significant difference between keyboard instrument players (Mdn = 0.64) and plucking and string instrument players (Mdn = 0.64) in the *PL perception* condition could be detected, $U = 3413.50$, $z = -.62$, $p > .05$. No significant difference between keyboard instrument players (Mdn = 0.58) and plucking and string instrument players (Mdn = 0.50) in the *ML perception* condition could be detected, $U = 3341$, $z = -.85$, $p > .05$. The effect size has not been calculated as there was no significant difference across groups found.

No significant difference between wind instrument players (Mdn = 0.62) and plucking and string instrument players (Mdn = 0.60) in the *Overall Aptitude perception* condition could be detected, $U = 2830.50$, $z = -1.120$, $p > .05$. No significant difference between wind instrument players (Mdn = 0.64) and plucking and string instrument players (Mdn = 0.64) in the *PL perception* condition could be detected, $U = 2780$, $z = -1$, $p > .05$. No significant difference between wind instrument players (Mdn = 0.58) and plucking and string instrument players (Mdn = 0.50) in the *ML perception* condition could be detected, $U = 3341$, $z = -.85$, $p > .05$. The effect size has not been calculated as there was no significant difference across groups.

Results indicate that even though decisions for musical instruments are related to pitch perception preferences, a difference in speech perception which may be related to particular musical instruments played cannot be detected. However, this result must be taken with caution as cleaner group differences have to be defined and integrated into future research.

9.11. Discussion

As already mentioned before, the MULT AP should be used with other aptitude measurements as it is less informative as a single score and has been especially developed to compare speech perception, speech production and musical measurements. However, as the main questionnaire contains background knowledge group comparisons are possible. In this respect, the analyses of this thesis show group comparisons including gender, age, musicians versus non-musicians and group comparisons on the types of instruments played.

Additionally, a correlational analysis of the main variables of the speech perception tasks and the questionnaire was shown. This analysis was also performed to reveal information regarding the nature of the test for the follow-up study presented in the following chapters.

The correlations have provided important information about the relationship of the language perception tasks. The individual languages do not correlate with each other, apart from *Tagalog* and *Hindi perception*. This indicates that the speech perception tasks of each individual language seem to be rather independent of each other. This is contrary to individual language imitation tasks which are usually related to each other. The latter will be discussed in section 11.1. Interestingly, the correlations also indicate that there is a relationship between the main aptitude variables *Overall Aptitude perception* and *PL perception* condition and

education. This was not expected. Their relationship to education, however, diminishes when controlled for age.

Indeed, age has an effect on the speech perception performances. In order to improve the validity of the speech perception measurements four age groups have been created. The adolescents were split into three groups, while the adults remained a single group. Results revealed that the adolescent I group which is aged between eleven and twelve years, performs significantly worse compared to all other groups. This suggests that at the age of thirteen no significant difference to participants of older ages will be detected as shown in section 9.10.3.1. In contrast to age, gender differences were not found (see section 9.10.2.3).

Group comparisons also show that the participants who played a musical instrument were better at discriminating unfamiliar speech than the non-musicians. The ability to play a musical instrument can be seen as an indicator for musical training but not for aptitude. Research has shown that musical training improves speech processing and perception (François et al. 2013; Besson et al. 2011; Moreno et al. 2009; Schön et al. 2004; Thompson et al. 2004; Wong & Perrachione 2007; Kraus & Chandrasekaran 2010; Hu et al. 2013; Reiterer et al. 2011). This can also be observed in this investigation. However, one has to bear in mind that the effect was rather small. Even though the number of participants is relatively high, results should be taken cautiously and testing should be repeated.

Group comparisons have also been undertaken to determine whether there may be individual differences in speech perception depending on the first and best instrument played. Therefore, the musicians were grouped into keyboard instrument players, plucking and string instrument players and wind instrument players. The findings indicate that there is no significant difference in speech perception which could be related to instrument choice. As no cleaner group difference was possible, the results on the speech perception ability and its relationship to instrument choice should consequently be treated with some caution.

The MULT AP measures aptitude for retaining and perceiving unfamiliar speech. Thus comparisons to traditional music aptitude tests that measure music perception ability may additionally give information about the relationship between speech and music perception aptitude. This will be provided in the empirical investigation in chapter 11.

How to use these measurements may depend on the individual needs of the research but at the present moment it is clear that further research is required to understand the nature of measuring speech perception aptitude in this way. The first analysis has provided crucial information about the nature of the MULT AP which suggests that at the age of thirteen no

significant differences between adult and adolescents in the ability to discriminate unfamiliar speech material are expected to be observed.

10. Empirical investigation: methodology

10.1. Overview of the empirical investigation

The idea to this doctoral thesis is based on various directions of previous research in which the relationship between phonetic language aptitude, phonological STM capacity, musical abilities and singing abilities have been investigated (Christiner 2013; Christiner & Reiterer 2013, 2015, 2018; Christiner et al. 2018; Christiner & Reiterer 2019). The development of the entire research project has thus lasted a number of years and as a result has turned out to be highly complex and multifaceted. Therefore, figure 1 represents the entire project with its main tasks, aims and objectives.

It is comprised of five major parts

- a) Speech imitation
- b) Singing ability
- c) Speech perception
- d) Music perception
- e) Other factors

First, speech imitation refers to repeating unintelligible language material. This is the main phonetic aptitude measurement that tests the ability to generate unfamiliar speech. It consists of eight different languages and aims to determine whether similar and distinct results of previous investigation can be found when multiple languages are included in the test design.

Secondly, the speech imitation measurement also has an equivalent speech perception measurement. As already discussed and shown in chapter 9, the MULT AP is a speech perception aptitude measurement which can be compared to the speech imitation tasks and consists of the same eight languages with one difference. There is a mixed condition in which all languages can occur, and a pure condition in which individual languages are tested. The latter consists of only five languages, Tagalog, Farsi, Japanese, Hindi and Mandarin which can directly be compared to their imitation counterpart. This aims to reveal the relationship between unintelligible speech perception and production of the same languages.

Additionally, singing ability, which requires the learning of a new song and the singing of a familiar song (production) is also included. The AMMA test, which measures tonal and rhythmical music aptitude, is also used to assess the perceptual musical abilities which will be compared to speech perception, singing ability and speech imitation ability.

Further factors have been implemented to understand other mechanisms that could influence phonetic aptitude and produce undesired effects which bias the results. This includes the development of questionnaires which provide information about singing behaviour during childhood, performance anxiety, educational status, music, singing, language and foreign language background, and self-awareness of imitation ability. Furthermore, STM capacity has been measured as well.

In previous chapters the main idea of this investigation was explained including the development of the MULT AP findings of which have been shown before. The development of this test was necessary in order that the empirical investigation in which the new speech perception test is integrated could be conducted.

The empirical investigation aims at revealing the relationship of unintelligible speech production and perception, and their relationship to all other measurements which have been demonstrated in figure 4. Additionally, the test design used in previous investigations has been improved to be able to get a multidimensional impression of the ability to imitate unintelligible speech of typologically different languages and of their relationship to STM, singing, musical aptitude and other related components.

10.2. Testing procedure and participants

The testing procedure was crucial as not all participants were suitable for this investigation. In total, 121 participants were recruited and 86 completed the whole test. The procedure was divided into three steps. First, the questionnaire, which collected information on the background, details including language, music and singing background, was completed at home as every participant received login details.

This included a pre-selection phase where participants had to report whether they speak one of the languages of the speech imitation tasks (Hindi, Tagalog, Mandarin, Thai, Japanese, Tamil, Russian, or Farsi) or not. If participants spoke one of these languages, they were not invited to continue the entire test (which of course was possible if they wanted to, but they were excluded from the final analysis either way). In addition, the participants were also asked a second time whether they speak one of the languages or had ever stayed in a country where one of these languages is spoken. Hence, one assumption was that aptitude can

be measured best by testing the ability to imitate unfamiliar languages, which at the same time allows the conclusion that educational advantages could not bias the results (Christiner & Reiterer 2015).

Second, after the background information was provided, participants were instructed to perform the MULT AP (language perception test) at home including the questionnaire about singing behaviour. Third, the participants were twice invited to a music studio where they were tested for their language and musical abilities. This ensured that test conditions were always the same. During the first meeting, the musicality test (AMMA) was carried out, followed by the phonological STM test (Wechsler 1939). The second time, the participants were tested for their speech imitation and singing ability. Then the final parts of the questionnaires were completed.

The participant selection is based on two main criteria: musical proficiency, and on whether the participants understood or spoke one of the eight foreign languages which had been selected for the speech imitation and perception tasks (see also 10.3.1).

Strictly speaking, musical proficiency varies considerably which is why it is very difficult to create categories (see section 3.2). In this investigation all participants had to label themselves according to the instructions of the four categories provided (instrumentalists, vocalists, amateurs who had experience in musical instrument playing, and non-musicians). Being a vocalist or instrumentalist meant that they were a member of a band, music or singing student or teacher. Within this framework they were instructed to indicate whether they defined themselves as an instrumentalist or a vocalist. The other participants had to indicate whether they had some experience in musical instrument playing (amateurs) and had received at least one year of educational instructions. Furthermore, they had to indicate whether they were still able to play the musical instrument they had learnt. The non-musicians were those who, apart from some musical activities during primary school, never learned, or made any attempt to learn a musical instrument later in their lives, nor had they participated in a choir. Cleaner group differences were impossible. Some researchers also tend to use continuous variables which are transformed into categorical ones. This approach, however, is highly questionable and includes several statistical problems and has therefore been rejected (see chapter 8).

Additionally, the research design was chosen in order to get meaningful results about musical abilities and their relationship to all measurements included, irrespective of group membership. Therefore, a number of different musical measurements which allow the

observation of participants' musical abilities without categorising them are possible by running regression and factorial analyses.

The selection of participants included a small sample of vocalists ($N = 15$) and instrumentalists ($N = 15$). Additionally, amateurs ($N = 21$) and non-musicians ($N = 35$) participated in the investigation. For reporting additional information of the musical expertise of the participants, the number of musical instruments played, the hours participants sang a week and a questionnaire focusing on their singing behaviour during childhood as well as on their performance anxiety had been developed. 36 participants were male, while 50 were female. The age mean was $M = 34.53$.

10.2.1. Language background

One of the main criteria was that the participants did not speak one of the languages which have been selected. Most of the participants' first language was German ($N = 75$), followed by Turkish ($N = 4$), Croatian ($N = 2$), English ($N = 1$), Bosnian ($N = 1$), Italian ($N = 1$), Polish ($N = 1$) and Ukrainian ($N = 1$). 76 of the participants were monolinguals, 10 were bilinguals and 1 grew up with 3 languages. The most frequently spoken second language was English ($N = 68$), followed by German ($N = 5$), French ($N = 2$) and Dutch ($N = 1$). 10 participants spoke no second language, 28 participants reported to speak one foreign language, 36 indicated to speak two foreign languages, 7 participants spoke three foreign languages, 3 mastered four languages, and 2 spoke five languages. As five of the participants' first languages were Slavic languages (Croatian, Bosnian, Polish and Ukrainian), they were asked whether the *Russian imitation* tasks were intelligible to them or not. The Croatian and Bosnian participants reported to have not understood the content, while the Polish and Ukrainian participants reported to have understood fragments of the sentence material. Therefore, the two participants whose first languages were Polish and Ukrainian have been excluded for all *Russian imitation* abilities in the statistical analysis.

10.2.2. Educational status

Educational status is largely based on the recommendations of the *European Qualifications Framework*. The scale ranges between 0 and 9, whereby 0 indicates "no school education" and 9 indicates a "Doctoral degree". All precise descriptions are contained in section 7.5. In this investigation, the participants' educational status was entered according to the educational status which had been completed at testing time. 3 participants finished primary school only, 4 the main general secondary school, 13 the technical and vocational

school, 21 secondary academic school (general qualification for university entrance), 5 the post-secondary non-tertiary education, 1 the mastership education, 10 Bachelor studies, 25 Master studies, and 4 had a Doctoral degree.

In addition, the educational status of the parents was also collected. With regards to the mothers, 16 finished the main general secondary school, 32 the technical and vocational school, 7 secondary academic school (general qualification for university entrance), 15 the post-secondary non-tertiary education, 1 the mastership education, 5 Bachelor studies, 8 Master studies, and 2 had a Doctoral degree.

With regards to the fathers, 4 finished primary school only, 12 the main general secondary school, 24 the technical and vocational school, 15 secondary academic school (general qualification for university entrance), 3 the post-secondary non-tertiary education, 10 the mastership education, 2 Bachelor studies, 10 Master studies, and 5 had a Doctoral degree.

10.2.3. Musical background

The participants reported the number of hours they sang on any occasion on average in a week. The instructions included any form of singing whether it was for professional or non-professional purposes. The latter also included singing in the car or the bathroom, for instance. On average participants sang around 2.04 hours a week. 15 participants reported to be professional singers. The singing self-report questionnaire showed that 10 participants claimed to be “very poor” singers, 23 to be “poor” singers, 28 claimed to sing like the “average”, 21 indicated to be “good” singers and only 4 indicated to be “very good” singers.

The participants had also to indicate the first musical instruments they played best. 38 played no musical instrument, 15 played piano or keyboard, 11 played the guitar, 5 played recorder, 2 other wind instruments, 3 transverse flute, 2 trumpet, 2 saxophone, 2 accordion, 2 clarinet, 1 other percussion instruments, 1 other plucking instrument, 1 violin, 1 clarinet, and 1 drums.

15 participants reported to be professional musicians. In addition, the participants reported the number of instruments they played, while 38 participants played no musical instrument, 22 played one, 18 played two, 2 played three, 4 played four, 5 played five, and 1 played seven.

10.3. Methods: Measuring language aptitude

For this investigation phonetic aptitude tests have been developed and are measured in two ways. The first has already been shown in chapter 9, the MULT AP, which reveal

individual differences in the ability to perceive unfamiliar languages. This test will be compared to the ability to imitate the same languages.

The second way is the ability to imitate unfamiliar languages, which will be presented in this section. The original idea is based on the Hindi test developed by Dogil and Reiterer (2009), which consists of words and phrases of different syllable length. To develop similar tests, phrases and sentences in the seven other languages (the same which have been used for the MULT AP) have been collected which consist of 9 and 11 syllables. This syllable length turned out to be the most predictive. In marked contrast, for testing young children aged five to six an appropriate syllable length of 5 to 6 is recommendable (Christiner & Reiterer 2018).

10.3.1. Speech imitation tasks: about the speakers

For the production of language material, native speakers in Russian, Persian, Tagalog, Japanese, Tamil, Mandarin, and Thai were consulted. The reason for choosing those languages was that they are typologically different both in grammar and phonetically. In addition, Russian has practically no easily recognisable common lexical stock with the two other Indo-European languages, whereas only some exists for the two Indo-Iranian languages, and for this to be an issue there are too few Mandarin loanwords in Japanese. Therefore, the selection of languages represents a larger sample than used in previous research, and it could also be assumed that most participants do not speak one of these languages, which is a main criterion for measuring aptitude (for selection of participants consult section 10.2).

For the production of the imitation tasks, native speakers for each language were consulted to support the development of the language material. For this, the native speakers were also asked to exclude any loanwords. For instance, Tagalog incorporated many loanwords from Spanish, a language which is quite frequently spoken. For the language production, specific language features which do not exist in the German language were not explicitly selected, although there are some implied. For instance, the Hindi samples include retroflex consonants /ŋ/ and /ɭ/. The raters, however, were instructed to holistically rate the overall performance with no particular emphasis on specific language features (see 10.3.5).

The raters received a payment for their effort and they gave informed consent that the material collected could be used for this investigation. For Hindi, the already established samples developed by Dogil and Reiterer (2009) have been used, with the exception of one sample which was replaced by a different, nine-syllable-long other phrase.

For the other seven languages the native speakers were invited to a studio where they spoke a large number of phrases and sentences in their mother tongue. The best recordings and samples of the appropriate length (9 and 11 syllables) were taken for this investigation and targeted at the participants. Some sentences (in Mandarin and Tagalog) had already been used in another investigation before (Christiner et al. 2018).

10.3.2. Preliminary testing

The other phrases and sentences were preliminarily tested on 30 participants before. Preliminary testing revealed that the tone languages Thai and Mandarin were rather difficult to be imitate; especially if those two languages were the first ones, participants reported that they felt uncomfortable. Therefore, the ordering of the imitation tasks was highly important, and even though a randomised ordering is usually suggested, this has been rejected here. Instead, the order was always the same and started with the languages which were found to be easier according to preliminary testing. In this ordering, as a precaution, neither the two tone languages nor the two Indo-Iranian languages had an adjacent position. This turned out to be a very effective way of testing.

10.3.3. Procedure for the participants

The participants had to repeat four simple phrases or sentences (statements) of either 9 or 11 syllables for each of the eight languages in a single sitting. First of all, the procedure included a familiarisation task where the participants were instructed to imitate language samples which were unintelligible to them. For this task, four samples in the Slovak language were used. Each sample was played three times separated by a pause of 50ms. After this the participants had to repeat the sentence or phrase and try to imitate the accent as best they could. This was recorded and the next language sample was played. The participants wore headphones with an integrated microphone (Beyerdynamic, DT 290) while their imitation performance was recorded.

10.3.4. Additional rating scales: self-awareness concepts

After the participants finished the language imitation tasks, they were instructed to fill in a questionnaire in which they were asked to evaluate the imitation tasks they had performed before.

The speech imitation tasks were also a familiarisation procedure for this rating part. As the participants had already heard all samples before, they were briefly familiarised with

the language samples. The rating criteria also required participants to compare the individual languages to a certain extent so that they could differentiate between languages that appeared to be more memorable (*Memorable*), were more difficult to imitate (*Difficult to imitate*) or were less familiar (*Familiar*) than others that had been listened to before. *Memorable* refers to how memorable the samples they listened to appeared to them. *Familiar* differs from memorable in the sense that a language being familiar should indicate whether the participants believed to have heard this language before without understanding the content. Furthermore, the parameter *Difficult to imitate* should provide information about how challenging the individual imitation tasks were for the participants.

The participants did not know which language they were listening to, which helped to prevent that the respective language typology from being associated with positive or negative attitudes. Even though it could be assumed that the participants made suggestions about the languages, they were only answered after the rating was completed.

They were instructed to listen to language 1, which consisted of four language samples (e.g. the Russian imitation tasks), and after all four samples were presented, they were asked to indicate the following three parameters on a scale from one to ten:

- a) How memorable the language was for them (*Memorable*)
- b) How difficult the language was to imitate (*Difficult to imitate*)
- c) How familiar the language seemed to them (*Familiar*)

After they had indicated their responses, the next language was presented. The participants had to perform this for all eight different languages that had been imitated. The rating scale ranged from 0 (min) to 10 (max). For *Memorable*, 10 means the language is very memorable, while 0 indicates that it is not. This is also true for *Difficult to imitate* (0 indicates not difficult and 10 very difficult) and *Familiar* (0 indicates not familiar and 10 very familiar). This scaling method has turned out to be the most appropriate in preliminary testing.

In the analysis, the ratings should be read in the following way. The respective language is in initial position, followed by the type of rating. For example, *Farsi memorable* refers to how memorable Farsi appeared to the participants, *Farsi difficult to imitate* illustrates how difficult Farsi seemed to them, while *Farsi familiar* shows how familiar this language was to them.

In addition, the ratings will also be used as a composite score that consists of all eight language ratings. Those will be called *All Languages memorable*, *All Languages difficult to imitate* and *All Languages familiar* (see section 11.6.).

10.3.5. Raters

Every single recording of the participants' performances was cut, checked, its loudness normalised, and was subsequently uploaded to a rating platform, where each language was evaluated by native speakers. In sum, 38 native speakers evaluated the participants' performances. To be more precise, 4 raters were consulted for Russian, 6 for Japanese, 5 for Tamil, 5 for Hindi, 5 for Mandarin, 4 for Tagalog, 5 for Thai, and 4 for Farsi. All of the raters evaluated the performances online where they were introduced to their tasks (see appendix A1.2). All raters received a payment as a compensation for their effort. The rater requirement was being an adult native speaker for the language which should be rated.

The instructions were to holistically evaluate the ability to imitate the original language files with no particular focus on specific language features. The rating scale ranged from 0 "min" to 10 "max", and raters were instructed to evaluate the overall performance of the sentences/phrases they listened to.

For evaluation of the performances expert ratings, such as those by phoneticians, are not required since research has shown that native speakers can intuitively make comparable judgements (Bongaerts 1999). In previous research, ratings by native speakers have provided meaningful information about oral languages proficiency (Christiner & Reiterer 2013, 2015, 2018; Christiner et al. 2018; Christiner & Reiterer 2019). Comparisons of non-experts and experts have only revealed that experts are stricter, but this does not have any effect on the internal reliability of the ratings (Christiner 2013). However, ratings can also vary depending on language difficulty. For instance, ratings of tone language imitations (Mandarin) compared to non-tone language imitations (Tagalog) were always lower if the speakers' first language was German or Turkish (Christiner et al. 2018). The same has been found in this investigation for both tone languages, Mandarin and Thai (see descriptives A4.2). This is obviously dependent on the fact that non-tone language speakers find it more difficult to imitate tone languages. This, however, does not influence the reliability of the ratings. Despite the fact that native speaker ratings have shown high reliability in past research, the internal consistency has been calculated in this research. In the event that any raters' judgements were to deviate from the others, it would be possible to remove them.

For the ratings, a reliability analysis was performed for the language imitation tasks. To this end, the Cronbach's Alfa coefficient was applied recommending a minimum score of 0.7 in general (Field 2009). The inter-rater reliability was rather high for all ratings. The inter-reliability of the individual samples of each language which illustrate whether the internal

structure of the languages samples is consistent has been provided in Appendix 6. They show very high reliability.

All ratings for the individual languages showed high reliability as provided in Table 30 below.

Table 30. Reliability analysis speech imitation

<i>Reliability analysis speech imitation</i>			
<i>Language</i>	<i>Cronbach's Alpha</i>	<i>N of Items</i>	<i>N of Raters</i>
Russian	.90	16	4
Japanese	.93	24	6
Tamil	.93	20	5
Hindi	.88	20	5
Mandarin	.86	20	5
Tagalog	.89	16	4
Thai	.86	20	5
Farsi	.89	16	4

10.4. Singing ability

Singing ability was tested and measured in two different ways. One task was to sing the familiar song “Happy Birthday” as is usually targeted at non-professionals (Dalla Bella et al. 2007; Dalla Bella & Berkowska 2009; Christiner 2013; Christiner & Reiterer 2013; Christiner & Reiterer 2019; Christiner et al. 2018). The second singing task was more complex. It is an imitation task where an unfamiliar song should be learnt in a rather short period of time. The singing performances of the participants were rated and evaluated by singing experts, a test design which has successfully been used and tested in previous research design (Christiner 2013; Christiner & Reiterer 2013).

As stated in the measurement section 4.7, singing without lyrics reduces memory loads (Berkowska & Dalla Bella 2013). However, singing with lyrics makes it possible to observe more evaluation criteria (Larrouy-Maestri et al. 2013). First, quality of voice is less likely to be observed in singing melodies (Christiner & Reiterer 2013). Second, intelligibility of lyrics is another crucial factor that influences decisions made about one’s singing performance (Ibrahim et al. 2017) as intelligibility especially increases with pitch height (Sundberg 1977). Another reason for including lyrics is that participants were instructed to perform the AMMA test which includes a perceptual tonal ability assessment tool that, at least, could provide information about their pitch discrimination ability.

10.4.1. Familiar song singing

“Happy Birthday” reveals already acquired singing ability and should ensure equal chances for the singing performances. It was assumed that all participants knew this song quite well since their childhood and it can also be suggested that they would not face difficulties in performing this well-known song.

10.4.2. Song learning (imitation)

The second singing condition resembled a learning condition. For this an unknown song was selected to ensure that participants had no prior knowledge. The aim of this task was to actively engage the participants in a singing learning condition to measure their singing aptitude as well. The participants were further instructed to repeat the song in a key which they found comfortable, as key did not play a role in the final ratings.

The learning condition was split into two parts which became increasingly more difficult. The participants had to sing the original part of the song after they had listened to the original sound file for the third time. The original part of the song was supported by musical instruments. However, the participants had to sing the song for the recording without background music purely from their memory as accurately as possible. The participants were also instructed to use different words if they had forgotten the lyrics.

For the first part of the song the lyrics were *whenever I miss, whenever I miss, I miss your smiling*. The second part of the song was an extension of the first one and the lyrics were as follows: *whenever I miss, whenever I miss, I miss your smiling, whenever I try, I try to fake a little smile*.

At this stage, some issues have to be addressed. Strictly speaking, it could be assumed that the English proficiency of the participants relates to the singing performances if the singing tasks make use of English lyrics. However, in previous research it has been noted that the relationship between singing in English and the imitation of English sentence material was related in the same way as singing with lyrics in English was to Hindi imitation. While the former was spoken by all participants, Hindi was completely unintelligible. If English lyrics would have biased the results, a stronger relationship between English imitation and singing ability would have been expected than singing to Hindi imitation. This, however, was not the case (Christiner & Reiterer 2013).

Additionally, the singing ratings include three main parameters: melody, rhythm and voice quality. This requires the raters to focus on particular elements and less on the language.

The raters were also instructed to ignore whether lyrics were incorrectly repeated when rating the three main singing criteria as the rating of lyrics is a separate category (see section 10.4.3).

10.4.3. Raters

The recordings of the singing tasks were rated by professional voice experts who performed the ratings online. The voice experts were two male and two female raters. For compensation they received a payment. As the raters had to evaluate 258 singing performances, it was not possible to do within a single sitting. Therefore, they received three separate buttons for each song. The program did not allow ratings to be omitted. The voice experts were instructed to rate only the first three parameters for “Happy Birthday”, and for the new song the experts also rated how well the participants repeated the lyrics:

- a) Melody
- b) Rhythm
- c) Voice quality
- d) Lyrics

The raters had to indicate their response on a scale between 0 and 10 with 10 being the highest value that someone had received, and 0 the lowest value. In addition, raters were informed that key did not play a role as the participants were instructed to sing in a range they found comfortable for their individual voice. This meant that if participants sang the song quite well but based on different pitches, this should not have an impact on the assessment. The singing criteria *lyrics* has been discarded from the analysis and has only had a control function in case it showed to have any effect on the other variables, which was not the case. Additionally, the raters were instructed to include neither that the pronunciation in English nor incorrectly sung words in the singing parameters.

10.4.4. Reliability analysis: singing the new song and “Happy Birthday”

In order to evaluate whether the ratings were consistent and reliable Cronbach’s alpha was applied. The ratings for the singing tasks showed high inter-rater reliability. The most important ones include the *New song 1 total* Cronbach’s $\alpha = .96$, *New song 2* Cronbach’s $\alpha = .98$, *New song total* Cronbach’s $\alpha = .98$, *Happy Birthday* Cronbach’s $\alpha = .95$. Regarding the song learning task, in the main analysis the *New song* variables 1 and 2 are always composite scores, meaning *New song rhythm 1* and *New song rhythm 2* comprise the *New song rhythm*

score. The same applies to the other *New song* criteria, quality of voice melody and the total scores.

Table 31 illustrates the reliabilities of all singing parameters in detail.

10.5. Technical support: recording/mastering/ratings

The recordings of all sound files were performed with Steinberg Cubase 8 software and headphones with integrated microphone. All recordings of the singing tasks and the unfamiliar language imitation tasks were then normalised to have the same quality and loudness. The sound files were converted to mp3 and uploaded to a computer programme in a randomised order for the ratings. The ratings for all sound files were performed online (see appendix A1.3).

Table 31. Reliability analysis singing parameters

<i>Reliability analysis singing parameters</i>			
<i>Singing parameters</i>	<i>Cronbach's Alpha</i>	<i>N of Items</i>	<i>N of Raters</i>
New song 1 melody	.89	4	4
New song 1 quality of voice	.91	4	4
New song 1 rhythm	.82	4	4
New song 2 melody	.94	4	4
New song 2 quality of voice	.95	4	4
New song 2 rhythm	.90	4	4
Happy Birthday melody	.88	4	4
Happy Birthday quality	.90	4	4
Happy Birthday rhythm	.74	4	4
New song 1 total	.96	12	4
New song 2 total	.98	12	4
New song total	.98	24	4
Happy Birthday total	.95	12	4

10.6. Advanced Measures of Music Audiation (Gordon 1989)

The AMMA test measures the participants' potential to discriminate paired musical statements which are either different or the same. Participants have to choose between three different conditions such as whether the paired musical statements are the same, or include rhythmical or tonal change. This test is usually targeted at university music and non-music majors and high school students, and is an aptitude test. The test consists of 33 items, whereby the first 3 are familiarisation tasks and were excluded from the final analysis.

10.7. Short-term-memory

In order to test the STM capacity of the participants the Wechsler Digit Span (Wechsler 1939) was used. This well-known test is composed of a forward digit span and a backward digit span. The test was programmed online and participants had to listen to the numbers and repeat them in the correct order again. After the participant made two mistakes in a row the test ended.

11. Results and data of the empirical investigation

11.1. Normal distribution, gender and age

As expected age did not show to have any effect on the variables as the participants were all adults. The same is also true for gender, where no significant difference was observed. In order to determine whether the data is normally distributed or not, a Shapiro-Wilk test was applied for all variables under consideration. The sample is not normally distributed if the variables are significantly different from the normal distribution indicated by $p < .05$. Some of the variables show a normal distribution, while others do not. Appendix A4.2 and A4.3 include the normal distribution of all variables.

11.2. Factor analysis: Aptitude measurements and components.

The relationships of the individual aptitude measurements were analysed by performing a factor analysis. This aimed at understanding the components of each factor, which should facilitate choices for scoring and explain the rationality of the research design. In addition, the relationship between the individual tasks should illustrate their similarities. This provides information about selecting or omitting variables in future research and supports scoring of the follow-up analysis of this investigation.

In the following analysis the individual scores which were introduced in the method section have been entered, with the exception of some. The measurements included in the factor analysis are all presented in table 32 below. The *Overall Aptitude perception* score and the *PL perception* score have been excluded as both are composite scores and the individual sub-scores (*Tagalog*, *Farsi*, *Japanese*, *Mandarin* and *Hindi perception*) have been entered in the factor analysis only. The *ML perception* score has been discarded as it shows no correlation to any of the variables below. The scores *Familiar*, *Memorable* and *Difficult to imitate* have been ignored as their relationship to the speech imitation tasks is obvious. In additionally, the main interest in the concepts, *Familiar*, *Memorable* and *Difficult to imitate*

lies in whether or not they are involved in explaining the variances of the individual language imitation performances. The latter cannot be achieved in the factorial design, but rather by performing regression analysis, which will be run in section 11.3. Furthermore, the total scores of the *AMMA* test and the Wechsler digit span *STMT* have not been entered but always their sub-scores since the loadings of the individual scores may reveal crucial information and relationships to other variables. The multi-item concepts *Singing behaviour during childhood* and *Performance anxiety* have been included as a composite score since the reliability of their individual questions has already been provided in section 7.7.

A principal component analysis (PCA) was conducted on the 29 items with orthogonal rotation (varimax). The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, $KMO = .82$. Bartlett's test of sphericity $\chi^2(406) = 1661.60$, $p < .001$, indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. Eight components had eigenvalues over 1 and in combination explained 72.98 percent of the variance. Table 32 shows the factor loadings after rotation.

As could be shown in the factor analysis, three main factors, singing, speech imitation and perception can be detected. Factor loads over .40 have to be considered to belong to the factor.

The first factor *Singing* is comprised of all singing parameters (melody, rhythm and quality of voice) of both singing tasks including *Self-estimation of singing capacity* and *Singing behaviour during childhood* as expected. *Singing regularly hours per week* does not load onto this factor. Within the same factor the *Number of musical instruments* played, *Performance anxiety* and *Tagalog imitation* also load onto this concept.

The second factor *Speech imitation* consists of all language imitation tasks. They show extraordinarily high loads onto each other, apart from *Tagalog imitation* which shows lower loads but still is within the accepted range. Additionally, the *Number of foreign languages*, *STMF* and *STMB* measurements load onto the *Speech imitation* factor as well.

The third factor, *Perception*, includes the rhythmic and the tonal musicality aptitude measurements of the *AMMA test* and the two language perception tasks *Farsi* and *Tagalog perception*.

Table 32. Summary of exploratory factor analysis results ($N = 86$)

<i>Rotated Factor Loadings</i>			
<i>Item</i>	<i>Singing</i>	<i>Speech imitation</i>	<i>Perception</i>
New song quality	.87	.30	.13
Happy Birthday rhythm	.87	.14	.26
Happy Birthday quality	.86	.21	.24
Happy Birthday melody	.84	.19	.28
New song melody	.84	.31	.13
New song rhythm	.84	.32	.12
Performance anxiety	.67	.15	-.26
Self-estimation of singing capacity	.66	-.09	-.20
Number of musical instruments	.62	.17	.12
Singing behaviour during childhood	.59	.13	-.36
Singing regularly hours per week	.38	-.08	-.28
Tamil imitation	.11	.82	.03
Mandarin imitation	.10	.78	.02
Hindi imitation	.13	.77	.02
Russian imitation	.22	.74	-.13
Japanese imitation	.17	.73	.24
Farsi imitation	.13	.72	.15
Thai imitation	.19	.71	.14
STMF	.04	.56	.06
Number of foreign languages	.27	.55	-.04
STMB	.02	.51	.26
Tagalog imitation	.43	.44	.27
Japanese perception	.01	.31	.01
Hindi perception	.10	.28	.02
AMMAR	.34	.13	.74
AMMAT	.34	.22	.64
Farsi perception	.01	-.24	.55
Tagalog perception	.04	.28	.51
Mandarin perception	-.11	.06	.27

Note: Factor loadings over .40 appear in bold.

11.2.1. Factor 1: Singing measurements

The singing tasks of this investigation are subdivided into two main categories: imitation tasks (repeating parts of a new unfamiliar song) and familiar song singing. Singing familiar songs such as “Happy Birthday” is often targeted at non-musicians (Dalla Bella et al. 2007, 2009; Christiner & Reiterer 2013; Christiner et al. 2018). In this investigation, both singing tasks, familiar song singing and an imitation task, have been included.

The imitation task was subdivided into two different tasks of increasing difficulty where a new song is imitated and learnt after listening to parts of the song three times. This test design has already been used successfully (Christiner 2013; Christiner & Reiterer 2019; Christiner & Reiterer 2013). The analysed evaluation criteria included three different parameters which were rated by singing experts: voice quality, melodic and rhythmic ability.

Music educators tend to rank pitch intonation and accuracy as the most important aspects of singing, (Watts et al. 2003), a measurement which is often used to measure the singing abilities of non-musicians (Dalla Bella et al. 2007, 2009; Christiner et al. 2018), while voice quality is usually targeted at professional singers (Ekholm et al. 1998; Sundberg 1988; Omori et al. 1996). The results of this investigation revealed that the individual singing assessment criteria of the ratings show unexpectedly high loadings onto each other and are components of a single factor. This indicates that they measure very similar concepts, but also shows that the individual criteria may not be necessary as a global rating, or that probably tonal measurements, as often suggested and applied to non-singers would give similar results. This also has an impact on the nature of the test design. The results furthermore indicate that the difference between familiar song singing and the learning of a new song is not rather large. Therefore, it can be suggested that a single singing task as used in previous research (Christiner et al. 2018) would largely reveal the same results.

Alternatively, it could also be argued that for the validity of the singing ratings, the individual singing criteria should be sustained. Another reason for keeping individual singing criteria is due to the choice of participants. Most of the participants of this investigation are non-professional singers. However, research on professional singers has also demonstrated that the individual singing measurements such as voice quality have effects on phonetic aptitude (Christiner & Reiterer 2013) and should be included as separate singing criteria for vocalists. One singing criterion which has been omitted is vocal range as this would have required introducing a further singing task which would have exceeded the scope of the testing conditions. Vocal range may also be an interesting singing criterion for adults and should be integrated in future research.

Two more singing measurements, *Singing behaviour during childhood* and *Self-estimation of singing capacity*, are also components of the singing factor, as expected. In the same singing factor, non-singing components are also implied. One is the *Number of musical instruments* played, which is quite obvious as singing and instrument playing are of course related in many respects. It could, however, also be assumed that variables that measure instrument playing will load onto the same factor as the *AMMAR* and *AMMAT*, yet this is not the case. Both musical aptitude measurements create a separate factor together with two speech perception scores. Inaccuracy in singing is often associated with pitch perception deficits (Tremblay-Champoux et al. 2010; Dalla Bella et al. 2009; Hutchins et al. 2014), though this has been increasingly criticised as contradictory results between singing ability and pitch perception abilities have been found (Estis et al. 2009). Further evidence has also been provided that pitch production and pitch perception can be dissociated (Loui et al. 2009). The results of this investigation shows that tonal musical aptitude, as measured with perception tests, seems to be less connected to singing and supports the former notions.

Another component is *Performance anxiety* which loads highly onto the singing factor. This demonstrates that this concept should be an all-important aspect in singing and music research, especially when performances are evaluated as shown in chapter 6. Surprisingly, *Tagalog imitation* seems to be more related to the singing factor than any of the other speech imitation tasks are. It could be suggested that tone languages are more likely to be associated with one of the two musical factors (either singing or music perception). The reason why Tagalog loads onto the singing factor may be of different nature. This unexpected finding will be discussed in section 13.3.

Consequently, the analysis of the factor load singing demonstrates several aspects which need to be considered. First of all, the first factor may be considered a singing factor, which shows that most of the singing components have been well chosen.

Importantly, *Performance anxiety* has an impact on the singing or musical parameters and thus requires to be integrated into research where performances are included to understand in how far unpleasant feelings have an influence on the results (see chapter 6). It has also been demonstrated that tonal musical aptitude, as measured by perception tests, is less connected to singing ability which more likely resembles findings which suggest that pitch perception and singing can be dissociated (see 4.6.3). As all the individual singing evaluation criteria for both singing tasks, the imitation and the familiar song singing tasks, i.e. show extraordinarily high loads onto the same factor, it could also be argued that there is not a

large difference among them. This suggests that a single singing task, either a song imitation or a familiar song singing task, would probably deliver very similar results.

These findings also have an influence on the scoring of the singing parameters. Apart from using all singing criteria as separate variables, it occurs that all singing criteria (melody, rhythm and voice quality) represent a strong singing concept that could be used as a composite score as well.

11.2.2. Factor 2: Speech imitation

The second factor, speech imitation reveals two essential results which have to be discussed. First of all, the two phonological *STM* measurements forward and backward and, second, the speech imitation tasks which all load onto the same factor. As shown in the factor analysis, the *STMF* and the *STMB* are both components of the same factor. Concerning digit spans forward and backward there is some controversy over whether they measure the same concept or not. As already discussed, backward spans focus on controlled attention, which makes them more of a hybrid task (see section 5.2). However, evidence has been provided that they may be best categorized as STM tasks (Engle et al. 1999) since repeating words or numbers in the reverse order involves only a “mental transformation” and does not impose new stimulus (Conway et al. 2005), as would be necessary to classify backward spans as WM tasks. As shown in this investigation results are in concordance with research conducted by Engle et al. (1999) as both the forward *STMF* subtest and the backward subtest *STMB* are components of the same factor.

However, the *STMF* and the *STMB* also load onto the same factor as the speech imitation tasks do. This essentially demonstrates that speech imitation tasks of unintelligible languages are *STM* tests, that is to say verbal *STM* tests. This finding is not surprising as this way of testing participants shows similarities with non-word, dissimilar or simple word repetition which needs to be recalled in a correct serial order.

One finding, however, is very surprising. All language imitation tasks show high loads onto the same factor, except for Tagalog. This suggests that they measure similar aspects. At this stage of the thesis two more aspects have to be considered: the language imitation task as an entire test, and the individual languages that should provide information about individual differences in the ability to imitate typological different languages. First, speech imitation shows the individual languages to be highly interrelated, demonstrating that, if they are put together, language imitation tasks represents a very strong test conception, similar to the singing measurements (see section 11.2.1).

Second, the close relationship of the individual language imitation tasks also has implications on previous research and language typology. Presumably, the eight language imitation tasks measure a more general phonetic imitation ability that may be less dependent on language typology. Otherwise the imitation tasks would load onto different factors or would consequently show lower loads onto the same factor. Additionally, it would indicate that imitation ability in individual languages would largely share the same predictors that can explain the variances of the performances in regression models, as has been shown in former investigations (Christiner 2013; Christiner & Reiterer 2013; Christiner & Reiterer 2019; Christiner et al. 2018). This assumption will be tested and illustrated in the follow-up analysis in section 11.3.

This finding about speech imitation is very surprising and has not been suggested before, but still is very plausible. It may also explain the dissociation of perception and production of new and unintelligible language material. While imitation tasks represent a motor action that is largely independent of language typology, perception is probably not. Other researchers also suggest that enhanced perception ability is involved in explaining individual differences of foreign language aptitude (Linck et al. 2014; Linck et al. 2013). If speech perception of different unfamiliar languages were simply the counterpart to the imitation of unfamiliar languages, their internal relationships would be similar yet they are dissimilar, as will be shown in the following section.

11.2.3. Factor 3: Perception

While the individual different languages imitated show very high loadings onto one factor, the different language perception tasks do not. Two language perception tasks load onto the same factor where also the *AMMAR* and *AMMAT* musical aptitude scores load.

The findings suggest two possible explanations if this way of measuring speech perception is appropriate. The musical aptitude measurements, *AMMAR* and *AMMAT*, as well as *Tagalog* and *Farsi perception*, all show loadings onto the same factor, which explains their close relationship. *Tagalog* and *Farsi perception* are also subcomponents of the *PL condition* of the MULT AP which aims to measure speech perception aptitude as described in chapter 9. This shows that the third factor may be a pure aptitude factor, which demonstrates that music and speech perception aptitude are obviously intertwined, at least for particular languages.

Second, if the internal relationships of the speech perception and speech imitation variables are considered, it could be concluded that individual differences of language performances are more likely to be found in speech perception rather than in speech imitation.

The speech perception tasks do not belong to a single factor, while the speech imitation tasks do. The results would be in concordance with the developers of the Hi-LAB aptitude battery. They claim that high level-attainment in second language proficiency largely relies on cognitive and perceptual abilities (Linck et al. 2013). The language imitation measurements have been designed to be of relatively high ecological validity because they should simulate a foreign language situation in which new words or phrases are imitated. Therefore, eight different languages have been selected to reveal whether ID to other capacities can be observed depending on language typology. In this respect, the results of this investigation more likely indicate that from the imitation tasks alone, it is less likely to reveal as to whether individuals possess high aptitude for pronunciation in particular languages. Some tendencies in individual differences in the ability to imitate unfamiliar languages which may be related to language typology have only been found for tone languages in comparisons to non-tone languages (see section 11.4.6). Except for the latter, it more likely indicates that imitation tasks provide information about ID in imitating new words in more general. Taken together, language imitation tasks irrespective of the languages chosen largely reveal similar information which makes it possible to draw generalisations about the potential to pronounce words in foreign languages.

The idea that there may be a general imitation ability that affects vocal behaviour is also implied in the singing factor. Like for the *Speech imitation* factor, the singing tasks show high loads onto a single factor. Pfordresher and Mantell (2014) argue that poor singing can be a result of a deficit of sensorimotor translations that specifically affects (vocal) imitation. If this is the case, it may also explain why there is not a large difference between imitating different languages, which may represent a general phonetic imitation ability which is not highly affected by the kind of language imitated. The same effect has not been observed for the language perception tasks, and even though some of the individual language perception tasks *Farsi* and *Tagalog perception* load onto the same factor, they are far less connected than the imitation tasks are. Thus, this could be additional evidence that individual differences in language typology in the early stages in language learning may be more likely observed in perceptual abilities (Linck et al. 2013). Nevertheless, this does not mean that imitation tasks are not informative in terms of phonetic aptitude. Moreover, it shows that imitation ability may be one missing link which should be integrated in language aptitude testing as a general phonetic imitation aptitude measurement (see section 12.1.1).

11.2.4. Discussion

The factor analysis delivers interesting findings of the measurements which have been included in the entire research design. First of all, three factors have been found which leads to the observation that in combination they explain 72.98% of the variance, which is quite high. This shows that the factors singing, speech imitation and perception seem to indicate quite a strong research design. However, there are also unexpected findings such as for the speech imitation tasks, which more or less seem to measure a general phonetic imitation ability and may be less informative to measure individual differences in language typology. In marked contrast, speech perception ability seems to vary and individual languages seem to be less related than the individual languages of the imitation tasks are. The *STM* forward and backward also load onto the same factor as the language imitation tasks, which indicates that the imitation tasks also measure a similar concept and also seem to represent a verbal STM measurement.

For singing it has been illustrated that the individual parameters are highly intertwined. The results also have implications on the scoring of some of the measurements. The singing parameters rhythm, melody and quality of voice show that they can go together as a single score since they measure similar singing concepts. The same is true for the individual speech imitation tasks, which all load onto the same factor. Additionally, *STMF* and *STMB* also belong to the speech imitation factor. The *AMMAT* and the *AMMAR* belong to a separate factor, which also justifies a composite score, even though this does not necessarily need to be mentioned, as multiple investigations have also noted this before.

The only measurement which is more difficult concerns the speech perception tasks, which are not all related or load onto the same factor. Thus, a composite score, the PL perception condition, which was introduced in section 9.10.1 and includes all five individual languages, Tagalog, Farsi, Japanese, Hindi and Mandarin, is questionable because of the low connectivity among these languages. However, this can also be interpreted differently. If individual differences can more likely be found in speech perception tasks, and less likely in imitation tasks, a relationship between the speech perception variables may not be necessary or detectable. This would mean that a composite score would be a more reliable speech perception concept as it integrates a wide range of typologically different language perception abilities. As this cannot be clarified within a single investigation, it has to be mentioned that all individual speech perception scores will be directly compared to their speech imitation counterparts in the same language in the follow-up analysis.

11.3. Correlational analysis and regression models: phonetic aptitude measured by speech imitation ability

Phonetic aptitude has been measured by the ability to imitate unfamiliar language material in eight languages. The imitation performances will be compared to all variables introduced in chapters 9 and 10. For better clarity, this section is divided into two major parts. The first one includes the analysis of the eight languages as separate speech imitation tasks, while the second analysis focuses on the speech imitation tasks as an entire phonetic imitation aptitude measurement in section 11.5.

To analyse the individual language imitation performances, eight regression models were carried out, some of which are comparable to previous research (Christiner 2013; Christiner & Reiterer 2013, 2018). In general, three important aspects will be addressed in this analysis. First, as this investigation is also a reinvestigation and includes more predictors and languages, it aims to reveal whether similar results to those in previous research can be replicated, and second, whether new predictors also contribute to the variances of the individual language imitation performances. Third, information should be provided as to whether some languages have similar and distinct predictors which could depend on language typology.

For the first analysis, all sub-scores of all individual tasks are used. This is also true for all speech perception variables as introduced in chapter 9. This means that first of all only *Mandarin*, *Tagalog*, *Japanese*, *Farsi*, *Hindi* and *ML perception* are included. The reason for this is that speech perception and imitation of the same languages will be compared, for instance *Hindi perception* and *Hindi imitation*. In addition, all singing criteria, musical aptitude measurements, *STM* and questionnaires, *Singing behaviour during childhood* and *Performance anxiety* are included in the analysis.

Data analysis revealed that some variables did not show a normal distribution. Therefore, Spearman's correlation analysis was applied in order to show the relationship of the individual variables first. As the data is quite complex, the correlational analysis is run as follows. Each of the eight languages is in front position of the correlation tables (33, 35, 37, 39, 41, 43, 45, 47) where their relationships to all the other variables will be presented. All 86 participants are included in the analysis apart from two participants who were excluded from the analysis in Russian (see section 10.2.1).

After each correlation matrix, a multiple linear regression model was performed. In the regression models all variables that correlated to the language imitation task (the dependent variable) were entered into a multiple linear regression as independent variables.

The independent variables were entered in the multiple linear regression models only if a probability of F-change < 0.05 was given. As regression models also facilitated the reduction of the predictors that can explain the variance of the performances, the correlations will not be discussed separately but only the predictors of the regression models.

A good theoretical framework is necessary for performing regression models, which is given in this case as previous, comparable research, is already available. This would justify using hierarchical (blockwise entry) regression models based on past research. However, in this investigation I have decided against this method in favour of a stepwise one which is unbiased by former investigations and based on mathematical decisions. This is more objective and is possible since multiple languages have been used, which is an additional re-test that examines whether predictors occur repeatedly.

The Spearman correlation tables (33, 35, 37, 39, 41, 43, 45, 47) illustrate that * means significant at the 0.05 level (2-tailed) and ** means that correlations are significant at the 0.01 level. + indicates that after applying a Benjamini-Hochberg correction, the variables are still significantly correlated.

After each correlations table (33, 35, 37, 39, 41, 43, 45, 47), tables that represent the regression models (34, 36, 38, 40, 42, 44, 46, 48) and figures (6, 7, 8, 9, 10, 11, 12, 13) for illustration of the main predictors are provided for each of the eight languages.

Table 33. Correlations for *Russian imitation*

	Russian imitation	Russian memorable	Russian difficult to imitate	Russian familiar	STMF	STMB	AMMAT	AMMAR	Number of musical instruments	New song melody	New song quality	New song rhythm	Happy Birthday melody	Happy Birthday quality	Happy Birthday rhythm	Singing behaviour during childhood	Self-estimation of singing capacity	Highest level of education	Performance anxiety	Number of foreign languages
Russian imitation	1	.24*+	-.05	.36**+	.46**+	.36**+	.28**+	.30**+	.38**+	.46**+	.46**+	.47**+	.42**+	.40**+	.37**+	.22*+	.15	.46**+	.30**+	.57**+
Russian memorable	.24*+	1	-.29**+	.27**+	.29**+	.14	-.06	-.00	.17	.12	.18	.20*	.13	.25*+	.22*+	.27**+	.14	.13	.19*	.15
Russian difficult to imitate	-.05	-.29**+	1	.05	-.04	-.12	.22*+	.07	-.12	-.06	-.08	-.10	-.13	-.15	-.10	-.22*+	-.19*	.02	-.10	.02
Russian familiar	.36**+	.27**+	.05	1	.18*	.14	-.10	-.07	-.05	.12	.12	.19*	-.04	.10	.10	.12	-.01	.19*	.16	.26**+
STMF	.46**+	.29**+	-.04	.18*	1	.46**+	.12	.15	.14	.20*	.27**+	.24*+	.23*+	.33**+	.24*+	.09	.17	.30**+	.11	.24*+
STMB	.36**+	.14	-.12	.14	.46**+	1	.12	.26**+	.09	.15	.23*+	.21*+	.19*	.29**+	.18	-.09	-.11	.05	-.03	.08
AMMAT	.28**+	-.06	.22*+	-.10	.12	.12	1	.75**+	.38**+	.28**+	.22*+	.25*+	.36**+	.29**+	.34**+	.11	.07	.16	.04	.18
AMMAR	.30**+	-.00	.07	-.07	.15	.26**+	.75**+	1	.48**+	.37**+	.31**+	.36**+	.45**+	.39**+	.43**+	.07	-.00	.09	.14	.28**+
Number of musical instruments	.38**+	.17	-.12	-.05	.14	.09	.38**+	.48**+	1	.53**+	.54**+	.53**+	.61**+	.60**+	.62**+	.27**+	.38**+	.29**+	.43**+	.42**+
New song melody	.46**+	.12	-.06	.12	.20*	.15	.28**+	.37**+	.53**+	1	.93**+	.95**+	.75**+	.76**+	.73**+	.36**+	.47**+	.26**+	.52**+	.36**+
New song quality	.46**+	.18	-.08	.12	.27**+	.23*+	.22*+	.31**+	.54**+	.93**+	1	.91**+	.78**+	.83**+	.75**+	.40**+	.48**+	.30**+	.56**+	.36**+
New song rhythm	.47**+	.20*	-.10	.19*	.24*+	.21*+	.25*+	.36**+	.53**+	.95**+	.91**+	1	.76**+	.77**+	.74**+	.40**+	.51**+	.22*+	.50**+	.39**+
Happy Birthday melody	.42**+	.13	-.13	-.04	.23*+	.19*	.36**+	.45**+	.61**+	.75**+	.78**+	.76**+	1	.88**+	.86**+	.32**+	.42**+	.29**+	.43**+	.36**+
Happy Birthday quality	.40**+	.25*+	-.15	.10	.33**+	.29**+	.29**+	.39**+	.60**+	.76**+	.83**+	.77**+	.88**+	1	.89**+	.40**+	.44**+	.29**+	.53**+	.38**+
Happy Birthday rhythm	.37**+	.22*+	-.10	.10	.24*+	.18	.34**+	.43**+	.62**+	.73**+	.75**+	.74**+	.86**+	.89**+	1	.40**+	.46**+	.21*+	.56**+	.33**+
Singing behaviour during childhood	.22*+	.27**+	-.22*+	.12	.09	-.09	.11	.07	.27**+	.36**+	.40**+	.40**+	.32**+	.40**+	.40**+	1	.36**+	.07	.45**+	.19*
Self-estimation of singing capacity	.15	.14	-.19*	-.01	.17	-.11	.07	-.00	.38**+	.47**+	.48**+	.51**+	.42**+	.44**+	.46**+	.36**+	1	.11	.49**+	.19*
Highest level of education	.46**+	.13	.02	.19*	.30**+	.05	.16	.09	.29**+	.26**+	.30**+	.22*+	.29**+	.29**+	.21*+	.07	.11	1	.40**+	.53**+
Performance anxiety	.30**+	.19*	-.10	.16	.11	-.03	.04	.14	.43**+	.52**+	.56**+	.50**+	.43**+	.53**+	.56**+	.45**+	.49**+	.40**+	1	.36**+
Number of foreign languages	.57**+	.15	.02	.26**+	.24*+	.08	.18	.28**+	.42**+	.36**+	.36**+	.39**+	.36**+	.38**+	.33**+	.19*	.19*	.53**+	.36**+	1

All variables that correlated with Russian imitation, which is the dependent variable, were entered as independent ones. The results revealed that 31 percent of the variance can be explained by three predictors: *STMF*, *New song melody* and *Russian familiar*.

Table 34. Multiple regression *Russian imitation*

Multiple regression models explaining the variance in Russian imitation

	<i>R</i>	<i>R</i> ²	F Change	Sig. F Change	<i>B</i>	<i>SE B</i>	β	<i>p</i>
Model 1	.40	.16	13.81	0.000				
Constant					1.62	0.55		
STMF					0.27	0.07	.40	< .01
Model 2	.50	.25	9.36	0.003				
Constant					0.43	0.65		
STMF					0.22	0.07	.33	< .01
New song melody					0.26	0.09	.32	< .01
Model 3	.56	.31	6.15	0.015				
Constant					0.05	0.65		
STMF					0.21	0.07	.30	< .01
New song melody					0.24	0.08	.29	< .01
Russian familiar					0.11	0.04	.25	< .05

Dependent variable: Russian imitation

Figure 6. Variance in *Russian imitation* explained

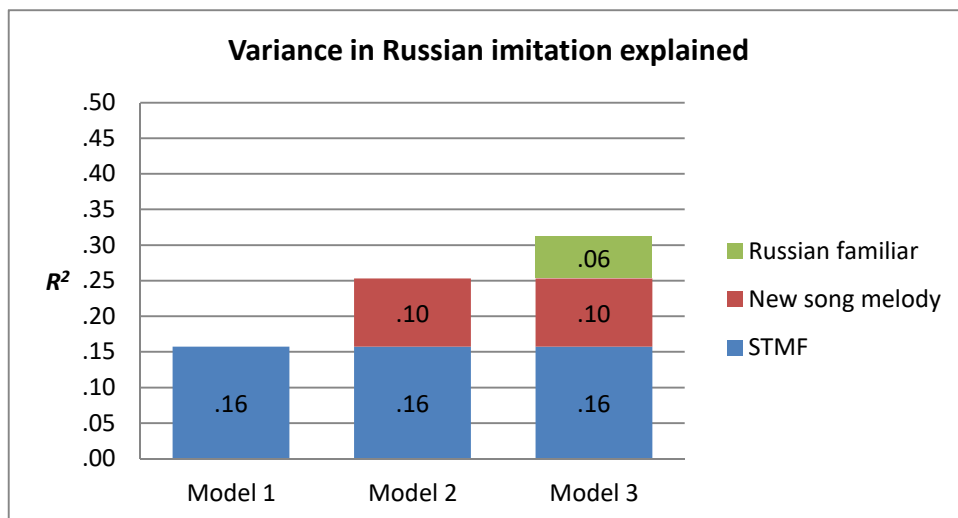


Table 35. Correlations for *Farsi imitation*

	Farsi imitation	Farsi perception	Farsi memorable	Farsi difficult to imitate	Farsi familiar	STMF	STMB	AMMAT	AMMAR	Number of musical instruments	New song melody	New song quality	New song rhythm	Happy Birthday melody	Happy Birthday quality	Happy Birthday rhythm	Singing behaviour during childhood	Self-estimation of singing capacity	Highest level of education	Performance anxiety	Number of foreign languages
Farsi imitation	1	.01	.12	.08	-.09	.46***	.31***	.29***	.25*+	.32***	.34***	.33***	.36***	.29***	.35***	.29***	.20*	.08	.28***	.20*	.37***
Farsi perception	.01	1	.23*+	-.19*	-.16	-.05	.04	.18	.22*+	.11	-.04	-.10	-.05	-.05	-.05	.08	-.19*	.01	-.03	-.08	-.13
Farsi memorable	.12	.23*+	1	-.28***	.39***	-.10	-.05	.13	.14	.25*+	.06	.07	.04	.05	.09	.11	-.04	.08	-.07	.05	.04
Farsi difficult to imitate	.08	-.19*	-.28***	1	-.09	.14	.17	-.06	.03	.05	.23*+	.20*	.24*+	.08	.10	.12	.23*+	-.03	.23*+	.12	.12
Farsi familiar	-.09	-.16	.39***	-.09	1	-.09	-.20*	-.02	-.05	.06	.05	.09	.04	-.01	.06	.02	.12	.02	-.08	.05	.14
STMF	.46***	-.05	-.10	.14	-.09	1	.46***	.12	.15	.14	.20*	.27***	.24*+	.23*+	.33***	.24*+	.09	.17	.30***	.11	.24*+
STMB	.31***	.04	-.05	.17	-.20*	.46***	1	.12	.26***	.09	.15	.23*+	.21*	.19*	.29***	.18	-.09	-.11	.05	-.03	.08
AMMAT	.29***	.18	.13	-.06	-.02	.12	.12	1	.75***	.38***	.28***	.22*+	.25*+	.36***	.29***	.34***	.11	.07	.16	.04	.18
AMMAR	.25*+	.22*+	.14	.03	-.05	.15	.26***	.75***	1	.48***	.37***	.31***	.36***	.45***	.39***	.43***	.07	-.00	.09	.14	.28***
Number of musical instruments	.32***	.11	.25*+	.05	.06	.14	.09	.38***	.48***	1	.53***	.54***	.53***	.61***	.60***	.62***	.27***	.38***	.29***	.43***	.42***
New song melody	.34***	-.04	.06	.23*+	.05	.20*	.15	.28***	.37***	.53***	1	.93***	.95***	.75***	.76***	.73***	.36***	.47***	.26***	.52***	.36***
New song quality	.33***	-.10	.07	.20*	.09	.27***	.23*+	.22*+	.31***	.54***	.93***	1	.91***	.78***	.83***	.75***	.40***	.48***	.30***	.56***	.36***
New song rhythm	.36***	-.05	.04	.24*+	.04	.24*+	.21*	.25*+	.36***	.53***	.95***	.91***	1	.76***	.77***	.74***	.40***	.51***	.22*+	.50***	.39***
Happy Birthday melody	.29***	-.05	.05	.08	-.01	.23*+	.19*	.36***	.45***	.61***	.75***	.78***	.76***	1	.88***	.86***	.32***	.42***	.29***	.43***	.36***
Happy Birthday quality	.35***	-.05	.09	.10	.06	.33***	.29***	.29***	.39***	.60***	.76***	.83***	.77***	.88***	1	.89***	.40***	.44***	.29***	.53***	.38***
Happy Birthday rhythm	.29***	.08	.11	.12	.02	.24*+	.18	.34***	.43***	.62***	.73***	.75***	.74***	.86***	.89***	1	.40***	.46***	.21*	.56***	.33***
Singing behaviour during childhood	.20*	-.19*	-.04	.23*+	.12	.09	-.09	.11	.07	.27***	.36***	.40***	.40***	.32***	.40***	.40***	1	.36***	.07	.45***	.19*
Self-estimation of singing capacity	.08	.01	.08	-.03	.02	.17	-.11	.07	-.00	.38***	.47***	.48***	.51***	.42***	.44***	.46***	.36***	1	.11	.49***	.19*
Highest level of education	.28***	-.03	-.07	.23*+	-.08	.30***	.05	.16	.09	.29***	.26***	.30***	.22*+	.29***	.29***	.21*	.07	.11	1	.40***	.53***
Performance anxiety	.20*	-.08	.05	.12	.05	.11	-.03	.04	.14	.43***	.52***	.56***	.50***	.43***	.53***	.56***	.45***	.49***	.40***	1	.36***
Number of foreign languages	.37***	-.13	.04	.12	.14	.24*+	.08	.18	.28***	.42***	.36***	.36***	.39***	.36***	.38***	.33***	.19*	.19*	.53***	.36***	1

All variables that correlated with Farsi imitation, which is the dependent variable, were entered as independent ones. The results revealed that 35 percent of the variance can be explained by two predictors: *STMF* and *AMMAT*.

Table 36. Multiple regression *Farsi imitation*

Multiple regression models explaining the variance in Farsi imitation

	<i>R</i>	<i>R</i> ²	F Change	Sig. F Change	<i>B</i>	<i>SE B</i>	β	<i>p</i>
Model 1	.52	.27	29.04	0.000				
Constant					-0.17	0.48		
STMF					0.34	0.06	.52	< .01
Model 2	.59	.35	9.59	0.003				
Constant					-1.89	0.72		
STMF					0.31	0.06	.48	< .01
AMMAT					0.07	0.02	.29	< .01

Dependent variable: Farsi imitation

Figure 7. Variance in *Farsi imitation* explained

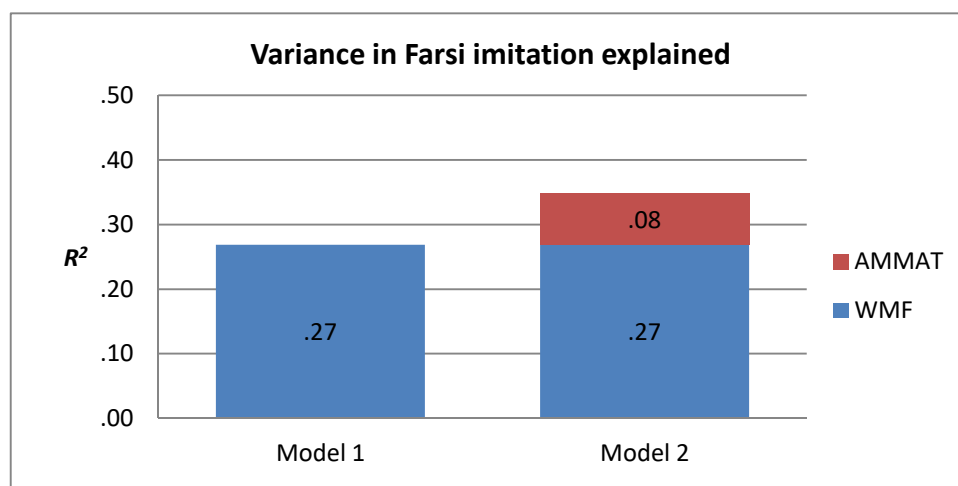


Table 37. Hindi imitation

	Hindi imitation	Hindi perception	Hindi memorable	Hindi difficult to imitate	Hindi familiar	STMF	STMB	AMMAT	AMMAR	Number of musical instruments	New song melody	New song quality	New song rhythm	Happy Birthday melody	Happy Birthday quality	Happy Birthday rhythm	Singing behaviour during childhood	Self-estimation of singing capacity	Highest level of education	Performance anxiety	Number of foreign languages
Hindi imitation	1	.32***	.24*+	.24*+	-.10	.43***	.44***	.27***	.28***	.18*	.38***	.38***	.41***	.28***	.35***	.26***	.16	-.00	.37***	.17	.38***
Hindi perception	.32***	1	.16	.07	.05	.13	.10	.14	.16	.13	.22*+	.18	.21*	.08	.05	.02	.13	.12	.13	.01	.18
Hindi memorable	.24*+	.16	1	-.44***	.29***	.21*	.01	.07	.15	.07	-.07	-.03	-.03	-.03	.02	-.04	.09	.04	-.03	-.01	.10
Hindi difficult to imitate	.24*+	.07	-.44***	1	-.23*+	.01	.11	.22*+	.24*+	.15	.34***	.26***	.33***	.13	.12	.21*+	.08	-.04	.17	.15	.25*+
Hindi familiar	-.10	.05	.29***	-.23*+	1	-.03	-.20*	.02	-.01	.05	-.01	-.00	-.02	-.02	-.05	-.06	.12	.11	-.10	.05	.12
STMF	.43***	.13	.21*	.01	-.03	1	.46***	.12	.15	.14	.20*	.27***	.24*+	.23*+	.33***	.24*+	.09	.17	.30***	.11	.24*+
STMB	.44***	.10	.01	.11	-.20*	.46***	1	.12	.26***	.09	.15	.23*+	.21*+	.19*	.29***	.18	-.09	-.11	.05	-.03	.08
AMMAT	.27***	.14	.07	.22*+	.02	.12	.12	1	.75***	.38***	.28***	.22*+	.25*+	.36***	.29***	.34***	.11	.07	.16	.04	.18
AMMAR	.28***	.16	.15	.24*+	-.01	.15	.26***	.75***	1	.48***	.37***	.31***	.36***	.45***	.39***	.43***	.07	-.00	.09	.14	.28***
Number of musical instruments	.18*	.13	.07	.15	.05	.14	.09	.38***	.48***	1	.53***	.54***	.53***	.61***	.60***	.62***	.27***	.38***	.29***	.43***	.42***
New song melody	.38***	.22*+	-.07	.34***	-.01	.20*	.15	.28***	.37***	.53***	1	.93***	.95***	.75***	.76***	.73***	.36***	.47***	.26***	.52***	.36***
New song quality	.38***	.18	-.03	.26***	-.00	.27***	.23*+	.22*+	.31***	.54***	.93***	1	.91***	.78***	.83***	.75***	.40***	.48***	.30***	.56***	.36***
New song rhythm	.41***	.21*	-.03	.33***	-.02	.24*+	.21*+	.25*+	.36***	.53***	.95***	.91***	1	.76***	.77***	.74***	.40***	.51***	.22*+	.50***	.39***
Happy Birthday melody	.28***	.08	-.03	.13	-.02	.23*+	.19*	.36***	.45***	.61***	.75***	.78***	.76***	1	.88***	.86***	.32***	.42***	.29***	.43***	.36***
Happy Birthday quality	.35***	.05	.02	.12	-.05	.33***	.29***	.29***	.39***	.60***	.76***	.83***	.77***	.88***	1	.89***	.40***	.44***	.29***	.53***	.38***
Happy Birthday rhythm	.26***	.02	-.04	.21*+	-.06	.24*+	.18	.34***	.43***	.62***	.73***	.75***	.74***	.86***	.89***	1	.40***	.46***	.21*	.56***	.33***
Singing behaviour during childhood	.16	.13	.09	.08	.12	.09	-.09	.11	.07	.27***	.36***	.40***	.40***	.32***	.40***	.40***	1	.36***	.07	.45***	.19*
Self-estimation of singing capacity	-.00	.12	.04	-.04	.11	.17	-.11	.07	-.00	.38***	.47***	.48***	.51***	.42***	.44***	.46***	.36***	1	.11	.49***	.19*
Highest level of education	.37***	.13	-.03	.17	-.10	.30***	.05	.16	.09	.29***	.26***	.30***	.22*+	.29***	.29***	.21*	.07	.11	1	.40***	.53***
Performance anxiety	.17	.01	-.01	.15	.05	.11	-.03	.04	.14	.43***	.52***	.56***	.50***	.43***	.53***	.56***	.45***	.49***	.40***	1	.36***
Number of foreign languages	.38***	.18	.10	.25*+	.12	.24*+	.08	.18	.28***	.42***	.36***	.36***	.39***	.36***	.38***	.33***	.19*	.19*	.53***	.36***	1

All variables that correlate with Hindi imitation, the dependent variable, were entered as independent ones. The results revealed that 41 percent of the variance can be explained by four predictors: *New song rhythm*, *STMB*, *Hindi memorable* and *Hindi difficult to imitate*.

Table 38. Multiple regression *Hindi imitation*

Multiple regression models explaining the variance in Hindi imitation

	<i>R</i>	<i>R</i> ²	F Change	Sig. F Change	<i>B</i>	<i>SE B</i>	β	<i>p</i>
Model 1	.46	.21	19.84	0.000				
Constant					0.95	0.49		
New song rhythm					0.31	0.07	.45	< .01
Model 2	.56	.31	10.89	0.001				
Constant					0.35	0.49		
New song rhythm					0.25	0.07	.36	< .01
STMB					0.13	0.04	.33	< .01
Model 3	.59	.34	4.22	0.043				
Constant					-0.11	0.53		
New song rhythm					0.26	0.07	.38	< .01
STMB					0.12	0.04	.31	< .01
Hindi memorable					0.09	0.04	.19	< .05
Model 4	.64	.41	8.03	0.006				
Constant					-0.67	0.55		
New song rhythm					0.20	0.07	.29	< .01
STMB					0.12	0.04	.30	< .01
Hindi memorable					0.15	0.05	.33	< .01
Hindi difficult to imitate					0.12	0.04	.31	< .01

Dependent variable: Hindi imitation

Figure 8. Variance in *Hindi imitation* explained

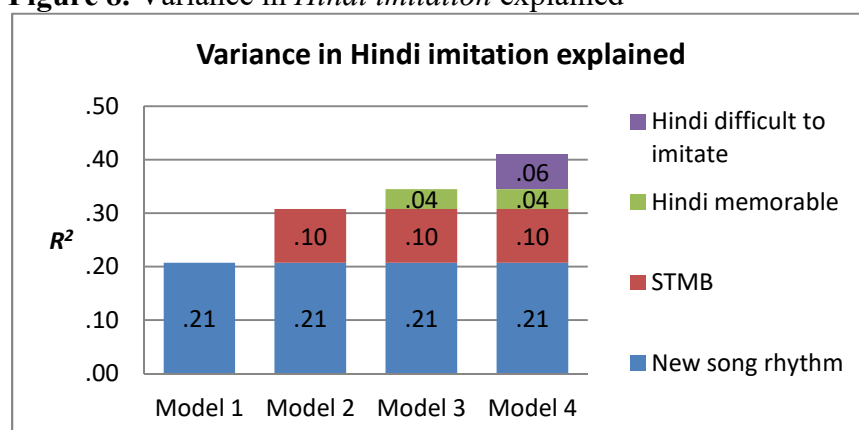


Table 39. Correlations for *Tagalog imitation*

	Tagalog imitation	Tagalog perception	Tagalog memorable	Tagalog difficult to imitate	Tagalog familiar	STMF	STMB	AMMAT	AMMAR	Number of musical instruments	New song melody	New song quality	New song rhythm	Happy Birthday melody	Happy Birthday quality	Happy Birthday rhythm	Singing behaviour during childhood	Self-estimation of singing capacity	Highest level of education	Performance anxiety	Number of foreign languages
Tagalog imitation	1	.11	.10	.11	-.09	.39**+	.26**+	.35**+	.45**+	.45**+	.52**+	.48**+	.52**+	.50**+	.54**+	.54**+	.35**+	.16	.36**+	.37**+	.44**+
Tagalog perception	.11	1	.11	-.08	-.13	.25*+	.23*+	.27**+	.34**+	.19*	.18	.14	.17	.26**+	.20*	.22*+	-.14	.11	.15	.05	.29**+
Tagalog memorable	.10	.11	1	-.26**+	.25*+	-.04	-.04	.21*	.32**+	.27**+	.15	.10	.11	.04	.01	.06	.03	.09	.03	.14	.06
Tagalog difficult to imitate	.11	-.08	-.26**+	1	-.01	.07	.23*+	.07	.05	-.03	.20*	.20*	.20*	.13	.11	.06	.06	-.17	.08	-.09	.05
Tagalog familiar	-.09	-.13	.25*+	-.01	1	-.24*+	-.24*+	-.00	.07	.16	.05	.07	.02	.08	.10	.12	.18	.10	-.04	.24*+	.13
STMF	.39**+	.25*+	-.04	.07	-.24*+	1	.46**+	.14	.16	.13	.20*	.27**+	.24*+	.23*+	.33**+	.24*+	.09	.17	.30**+	.11	.24*+
STMB	.26**+	.23*+	-.04	.23*+	-.24*+	.46**+	1	.08	.20*	.08	.15	.23*+	.21*+	.19*	.29**+	.18	-.09	-.11	.05	-.03	.08
AMMAT	.35**+	.27**+	.21*	.07	-.00	.14	.08	1	.78**+	.40**+	.33**+	.26**+	.30**+	.39**+	.32**+	.36**+	.14	.12	.18*	.08	.23*+
AMMAR	.45**+	.34**+	.32**+	.05	.07	.16	.20*	.78**+	1	.50**+	.40**+	.33**+	.38**+	.47**+	.41**+	.45**+	.10	.04	.12	.16	.32**+
Number of musical instruments	.45**+	.19*	.27**+	-.03	.16	.13	.08	.40**+	.50**+	1	.54**+	.54**+	.54**+	.62**+	.59**+	.62**+	.27**+	.38**+	.29**+	.43**+	.42**+
New song melody	.52**+	.18	.15	.20*	.05	.20*	.15	.33**+	.40**+	.54**+	1	.93**+	.95**+	.75**+	.76**+	.73**+	.36**+	.47**+	.26**+	.52**+	.36**+
New song quality	.48**+	.14	.10	.20*	.07	.27**+	.23*+	.26**+	.33**+	.54**+	.93**+	1	.91**+	.78**+	.83**+	.75**+	.40**+	.48**+	.30**+	.56**+	.36**+
New song rhythm	.52**+	.17	.11	.20*	.02	.24*+	.21*+	.30**+	.38**+	.54**+	.95**+	.91**+	1	.76**+	.77**+	.74**+	.40**+	.51**+	.22*+	.50**+	.39**+
Happy Birthday melody	.50**+	.26**+	.04	.13	.08	.23*+	.19*	.39**+	.47**+	.62**+	.75**+	.78**+	.76**+	1	.88**+	.86**+	.32**+	.42**+	.29**+	.43**+	.36**+
Happy Birthday quality	.54**+	.20*	.01	.11	.10	.33**+	.29**+	.32**+	.41**+	.59**+	.76**+	.83**+	.77**+	.88**+	1	.89**+	.40**+	.44**+	.29**+	.53**+	.38**+
Happy Birthday rhythm	.54**+	.22*+	.06	.06	.12	.24*+	.18	.36**+	.45**+	.62**+	.73**+	.75**+	.74**+	.86**+	.89**+	1	.40**+	.46**+	.21*	.56**+	.33**+
Singing behaviour during childhood	.35**+	-.14	.03	.06	.18	.09	-.09	.14	.10	.27**+	.36**+	.40**+	.40**+	.32**+	.40**+	.40**+	1	.36**+	.07	.45**+	.19*
Self-estimation of singing capacity	.16	.11	.09	-.17	.10	.17	-.11	.12	.04	.38**+	.47**+	.48**+	.51**+	.42**+	.44**+	.46**+	.36**+	1	.11	.49**+	.19*
Highest level of education	.36**+	.15	.03	.08	-.04	.30**+	.05	.18*	.12	.29**+	.26**+	.30**+	.22*+	.29**+	.29**+	.21*	.07	.11	1	.40**+	.53**+
Performance anxiety	.37**+	.05	.14	-.09	.24*+	.11	-.03	.08	.16	.43**+	.52**+	.56**+	.50**+	.43**+	.53**+	.56**+	.45**+	.49**+	.40**+	1	.36**+
Number of foreign languages	.44**+	.29**+	.06	.05	.13	.24*+	.08	.23*+	.32**+	.42**+	.36**+	.36**+	.39**+	.36**+	.38**+	.33**+	.19*	.19*	.53**+	.36**+	1

All variables that correlated with Tagalog imitation, which is the dependent variable, were entered as independent ones. The results revealed that 39 percent of the variance can be explained by three predictors: *Happy Birthday melody*, *AMMAT* and *Performance anxiety*.

Table 40. Multiple regression *Tagalog imitation*

Multiple regression models explaining the variance in Tagalog imitation

	<i>R</i>	<i>R</i> ²	F Change	Sig. F Change	<i>B</i>	<i>SE B</i>	β	<i>p</i>
Model 1	.50	.25	25.05	0.000				
Constant					-0.40	0.59		
Happy Birthday melody					0.45	0.09	.50	< .01
Model 2	.58	.33	9.63	0.003				
Constant					-1.91	0.74		
Happy Birthday melody					0.34	0.09	.37	< .01
AMMAT					0.09	0.03	.32	< .01
Model 3	.62	.39	6.67	0.012				
Constant					-2.67	0.77		
Happy Birthday melody					0.22	0.10	.25	< .05
AMMAT					0.10	0.03	.35	< .01
Performance anxiety					0.03	0.01	.26	< .05

Dependent variable: Tagalog imitation

Figure 9. Variance in *Tagalog imitation* explained

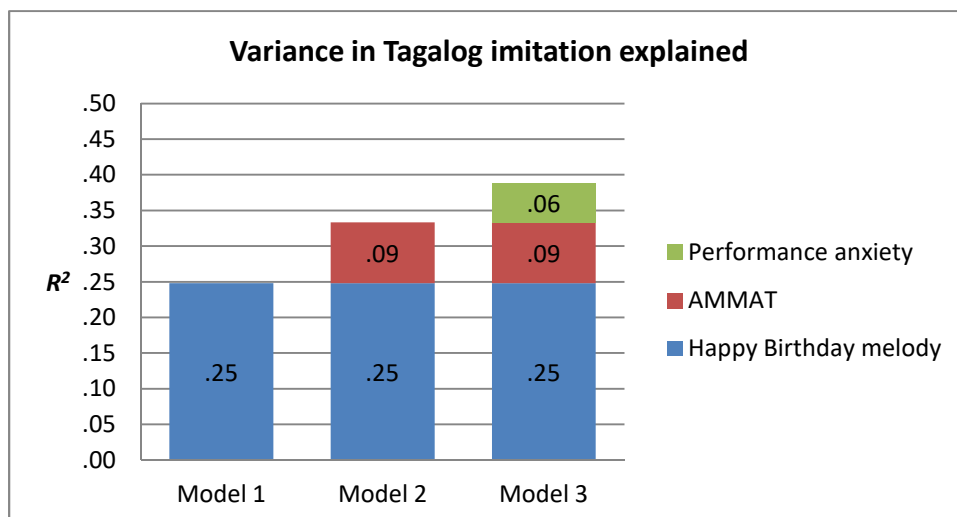


Table 41. Correlations for *Japanese imitation*

	Japanese imitation	Japanese perception	Japanese memorable	Japanese difficult to imitate	Japanese familiar	STMF	STMB	AMMAT	AMMAR	Number of musical instruments	New song melody	New song quality	New song rhythm	Happy Birthday melody	Happy Birthday quality	Happy Birthday rhythm	Singing behaviour during childhood	Self-estimation of singing capacity	Highest level of education	Performance anxiety	Number of foreign languages
Japanese imitation	1	.22*+	.33***	-.01	.08	.38***	.41***	.38***	.28***	.28***	.37***	.36***	.35***	.36***	.40***	.34***	.18	.08	.30***	.23*+	.29***
Japanese perception	.22*+	1	.29***	-.15	.10	.12	-.03	.13	.09	.13	.10	.12	.06	.01	.02	-.03	.14	.06	.23*+	.15	.25*+
Japanese memorable	.33***	.29***	1	-.49***	.24*+	.12	.07	.13	.08	.09	.04	.05	.05	-.04	-.04	-.01	.04	-.01	.19*	.19*	.16
Japanese difficult to imitate	-.01	-.15	-.49***	1	-.07	.12	-.01	.11	.20*	-.05	.04	.03	.04	.12	.16	.16	.00	-.11	.04	-.05	.10
Japanese familiar	.08	.10	.24*+	-.07	1	-.05	-.11	.19*	-.03	-.01	.11	.08	.12	-.03	-.05	.02	.08	-.02	.13	.16	.04
STMF	.38***	.12	.12	.12	-.05	1	.46***	.12	.15	.14	.20*	.27***	.24*+	.23*+	.33***	.24*+	.09	.17	.30***	.11	.24*+
STMB	.41***	-.03	.07	-.01	-.11	.46***	1	.12	.26***	.09	.15	.23*+	.21*	.19*	.29***	.18	-.09	-.11	.05	-.03	.08
AMMAT	.38***	.13	.13	.11	.19*	.12	.12	1	.75***	.38***	.28***	.22*+	.25*+	.36***	.29***	.34***	.11	.07	.16	.04	.18
AMMAR	.28***	.09	.08	.20*	-.03	.15	.26***	.75***	1	.48***	.37***	.31***	.36***	.45***	.39***	.43***	.07	-.00	.09	.14	.28***
Number of musical instruments	.28***	.13	.09	-.05	-.01	.14	.09	.38***	.48***	1	.53***	.54***	.53***	.61***	.60***	.62***	.27***	.38***	.29***	.43***	.42***
New song melody	.37***	.10	.04	.04	.11	.20*	.15	.28***	.37***	.53***	1	.93***	.95***	.75***	.76***	.73***	.36***	.47***	.26***	.52***	.36***
New song quality	.36***	.12	.05	.03	.08	.27***	.23*+	.22*+	.31***	.54***	.93***	1	.91***	.78***	.83***	.75***	.40***	.48***	.30***	.56***	.36***
New song rhythm	.35***	.06	.05	.04	.12	.24*+	.21*	.25*+	.36***	.53***	.95***	.91***	1	.76***	.77***	.74***	.40***	.51***	.22*+	.50***	.39***
Happy Birthday melody	.36***	.01	-.04	.12	-.03	.23*+	.19*	.36***	.45***	.61***	.75***	.78***	.76***	1	.88***	.86***	.32***	.42***	.29***	.43***	.36***
Happy Birthday quality	.40***	.02	-.04	.16	-.05	.33***	.29***	.29***	.39***	.60***	.76***	.83***	.77***	.88***	1	.89***	.40***	.44***	.29***	.53***	.38***
Happy Birthday rhythm	.34***	-.03	-.01	.16	.02	.24*+	.18	.34***	.43***	.62***	.73***	.75***	.74***	.86***	.89***	1	.40***	.46***	.21*	.56***	.33***
Singing behaviour during childhood	.18	.14	.04	.00	.08	.09	-.09	.11	.07	.27***	.36***	.40***	.40***	.32***	.40***	.40***	1	.36***	.07	.45***	.19*
Self-estimation of singing capacity	.08	.06	-.01	-.11	-.02	.17	-.11	.07	-.00	.38***	.47***	.48***	.51***	.42***	.44***	.46***	.36***	1	.11	.49***	.19*
Highest level of education	.30***	.23*+	.19*	.04	.13	.30***	.05	.16	.09	.29***	.26***	.30***	.22*+	.29***	.29***	.21*	.07	.11	1	.40***	.53***
Performance anxiety	.23*+	.15	.19*	-.05	.16	.11	-.03	.04	.14	.43***	.52***	.56***	.50***	.43***	.53***	.56***	.45***	.49***	.40***	1	.36***
Number of foreign languages	.29***	.25*+	.16	.10	.04	.24*+	.08	.18	.28***	.42***	.36***	.36***	.39***	.36***	.38***	.33***	.19*	.19*	.53***	.36***	1

All variables that correlate with Japanese imitation, the dependent variable, were entered as independent ones. The results revealed that 43 percent of the variance can be explained by four predictors: *AMMAT*, *Japanese memorable*, *New song melody* and *STMB*.

Table 42. Multiple regression *Japanese imitation*

Multiple regression models explaining the variance in Japanese imitation

	<i>R</i>	<i>R</i> ²	F Change	Sig. F Change	<i>B</i>	<i>SE B</i>	β	<i>p</i>
Model 1	.41	.17	15.14	0.000				
Constant					2.01	0.75		
AMMAT					0.11	0.03	.41	< .01
Model 2	.54	.29	12.87	0.001				
Constant					1.06	0.74		
AMMAT					0.10	0.03	.38	< .01
Japanese memorable					0.22	0.06	.35	< .01
Model 3	.61	.38	9.64	0.003				
Constant					0.32	0.74		
AMMAT					0.07	0.03	.27	< .01
Japanese memorable					0.22	0.06	.36	< .01
New song melody					0.25	0.08	.31	< .01
Model 4	.65	.43	6.29	0.014				
Constant					-0.24	0.75		
AMMAT					0.07	0.03	.26	< .01
Japanese memorable					0.20	0.06	.32	< .01
New song melody					0.21	0.08	.26	< .05
STMB					0.13	0.05	.24	< .05

Dependent variable: Japanese imitation

Figure 10. Variance in *Japanese imitation* explained

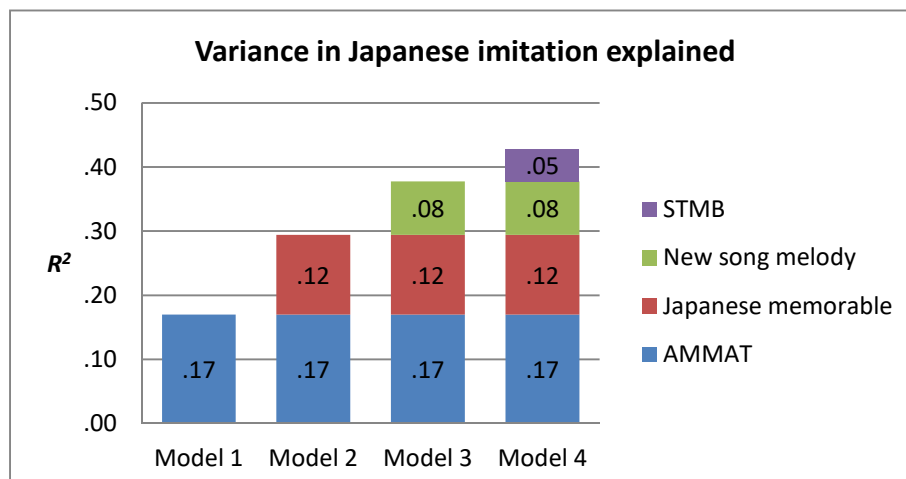


Table 43. Correlations for *Tamil imitation*

	Tamil imitation	Tamil memorable	Tamil difficult to imitate	Tamil familiar	STMF	STMB	AMMAT	AMMAR	Number of musical instruments	New song melody	New song quality	New song rhythm	Happy Birthday melody	Happy Birthday quality	Happy Birthday rhythm	Singing behaviour during childhood	Self-estimation of singing capacity	Highest level of education	Performance anxiety	Number of foreign languages
Tamil imitation	1	.19*	.02	.13	.36**+	.33**+	.17	.21*+	.23*+	.37**+	.40**+	.38**+	.26**+	.35**+	.25**+	.23*+	-.01	.30**+	.23*+	.38**+
Tamil memorable	.19*	1	.33**+	.45**+	.14	.01	-.10	-.11	-.20*	-.13	-.08	-.16	-.03	-.01	.01	.01	-.11	-.11	.07	-.12
Tamil difficult to imitate	.02	.33**+	1	-.20*	-.00	-.11	.19*	.19*	.15	.08	.04	.06	.01	.07	.06	.10	.03	.11	-.02	.14
Tamil familiar	.13	.45**+	-.20*	1	.09	-.07	-.02	.00	.02	-.02	.04	-.02	.02	.04	.02	.09	-.01	-.19*	.09	.07
STMF	.36**+	.14	-.00	.09	1	.46**+	.12	.15	.14	.20*	.27**+	.24*+	.23*+	.33**+	.24*+	.09	.17	.30**+	.11	.24*+
STMB	.33**+	.01	-.11	-.07	.46**+	1	.12	.26**+	.09	.15	.23*+	.21*+	.19*	.29**+	.18	-.09	-.11	.05	-.03	.08
AMMAT	.17	-.10	.19*	-.02	.12	.12	1	.75**+	.38**+	.28**+	.22*+	.25*+	.36**+	.29**+	.34**+	.11	.07	.16	.04	.18
AMMAR	.21*+	-.11	.19*	.00	.15	.26**+	.75**+	1	.48**+	.37**+	.31**+	.36**+	.45**+	.39**+	.43**+	.07	-.00	.09	.14	.28**+
Number of musical instruments	.23*+	-.20*	.15	.02	.14	.09	.38**+	.48**+	1	.53**+	.54**+	.53**+	.61**+	.60**+	.62**+	.27**+	.38**+	.29**+	.43**+	.42**+
New song melody	.37**+	-.13	.08	-.02	.20*	.15	.28**+	.37**+	.53**+	1	.93**+	.95**+	.75**+	.76**+	.73**+	.36**+	.47**+	.26**+	.52**+	.36**+
New song quality	.40**+	-.08	.04	.04	.27**+	.23*+	.22*+	.31**+	.54**+	.93**+	1	.91**+	.78**+	.83**+	.75**+	.40**+	.48**+	.30**+	.56**+	.36**+
New song rhythm	.38**+	-.16	.06	-.02	.24*+	.21*+	.25*+	.36**+	.53**+	.95**+	.91**+	1	.76**+	.77**+	.74**+	.40**+	.51**+	.22*+	.50**+	.39**+
Happy Birthday melody	.26**+	-.03	.01	.02	.23*+	.19*	.36**+	.45**+	.61**+	.75**+	.78**+	.76**+	1	.88**+	.86**+	.32**+	.42**+	.29**+	.43**+	.36**+
Happy Birthday quality	.35**+	-.01	.07	.04	.33**+	.29**+	.29**+	.39**+	.60**+	.76**+	.83**+	.77**+	.88**+	1	.89**+	.40**+	.44**+	.29**+	.53**+	.38**+
Happy Birthday rhythm	.25**+	.01	.06	.02	.24*+	.18	.34**+	.43**+	.62**+	.73**+	.75**+	.74**+	.86**+	.89**+	1	.40**+	.46**+	.21*	.56**+	.33**+
Singing behaviour during childhood	.23*+	.01	.10	.09	.09	-.09	.11	.07	.27**+	.36**+	.40**+	.40**+	.32**+	.40**+	.40**+	1	.36**+	.07	.45**+	.19*
Self-estimation of singing capacity	-.01	-.11	.03	-.01	.17	-.11	.07	-.00	.38**+	.47**+	.48**+	.51**+	.42**+	.44**+	.46**+	.36**+	1	.11	.49**+	.19*
Highest level of education	.30**+	-.11	.11	-.19*	.30**+	.05	.16	.09	.29**+	.26**+	.30**+	.22*+	.29**+	.29**+	.21*	.07	.11	1	.40**+	.53**+
Performance anxiety	.23*+	.07	-.02	.09	.11	-.03	.04	.14	.43**+	.52**+	.56**+	.50**+	.43**+	.53**+	.56**+	.45**+	.49**+	.40**+	1	.36**+
Number of foreign languages	.38**+	-.12	.14	.07	.24*+	.08	.18	.28**+	.42**+	.36**+	.36**+	.39**+	.36**+	.38**+	.33**+	.19*	.19*	.53**+	.36**+	1

All variables that correlate with Tamil imitation, the dependent variable, were entered as independent ones. The results revealed that 41 percent of the variance can be explained by two predictors: *New song rhythm* and *STMF*.

Table 44. Multiple regression *Tamil imitation*

Multiple regression models explaining the variance in Tamil imitation

	<i>R</i>	<i>R</i> ²	F Change	Sig. F Change	<i>B</i>	<i>SE B</i>	β	<i>p</i>
Model 1	.44	.19	17.52	0.000				
Constant					0.11	0.70		
New song rhythm					0.42	0.10	.44	< .01
Model 2	.51	.26	6.95	0.010				
Constant					-0.77	0.75		
New song rhythm					0.36	0.10	.37	< .01
STMF					0.17	0.07	.27	< .05

Dependent variable: Tamil imitation

Figure 11. Variance in *Tamil imitation* explained

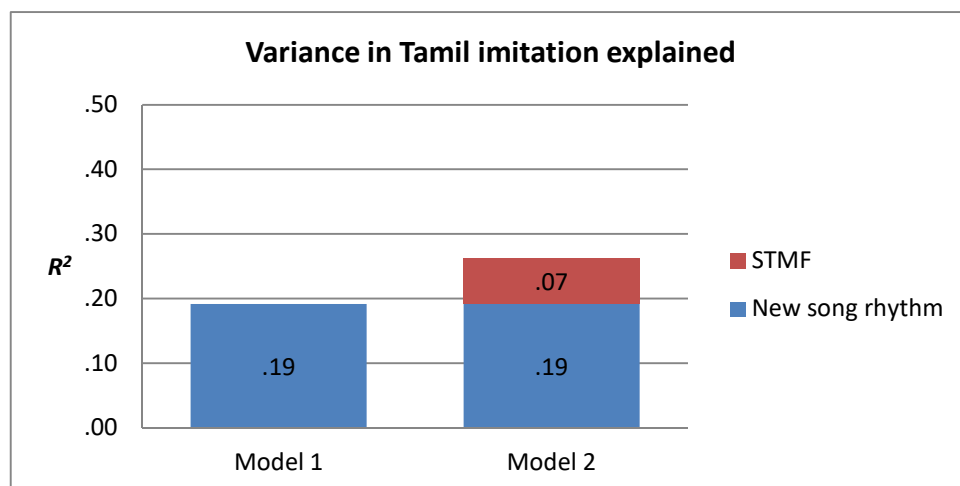


Table 45. Correlations for *Mandarin imitation*

	Mandarin imitation	Mandarin perception	Mandarin memorable	Mandarin difficult to imitate	Mandarin familiar	STMF	STMB	AMMAT	AMMAR	Number of musical instruments	New song melody	New song quality	New song rhythm	Happy Birthday melody	Happy Birthday quality	Happy Birthday rhythm	Singing behaviour during childhood	Self-estimation of singing capacity	Highest level of education	Performance anxiety	Number of foreign languages
Mandarin imitation	1	.01	.03	.20*	.20*	.31**+	.32**+	.27**+	.16	.28**+	.34**+	.35**+	.34**+	.27**+	.33**+	.27**+	.26**+	-.00	.25*+	.28**+	.38**+
Mandarin perception	.01	1	.01	-.12	.01	.23*+	.18	.00	.01	-.03	-.04	.02	-.05	.09	.10	.08	-.08	.05	.01	-.00	.06
Mandarin memorable	.03	.01	1	-.05	.33**+	-.13	-.12	.09	.07	-.09	-.16	-.24*+	-.18	-.28**+	-.28**+	-.22*+	.04	-.14	-.07	.04	.02
Mandarin difficult to imitate	.20*	-.12	-.05	1	.04	.22*+	.28**+	.11	.15	.06	.30**+	.33**+	.30**+	.27**+	.22*+	.24*+	.07	-.16	.35**+	.01	.21*
Mandarin familiar	.20*	.01	.33**+	.04	1	-.02	-.12	.06	.04	.04	.11	.04	.11	-.14	-.11	-.09	.01	-.03	.21*	.12	.29**+
STMF	.31**+	.23*+	-.13	.22*+	-.02	1	.46**+	.12	.15	.14	.20*	.27**+	.24*+	.23*+	.33**+	.24*+	.09	.17	.30**+	.11	.24*+
STMB	.32**+	.18	-.12	.28**+	-.12	.46**+	1	.12	.26**+	.09	.15	.23*+	.21*+	.19*	.29**+	.18	-.09	-.11	.05	-.03	.08
AMMAT	.27**+	.00	.09	.11	.06	.12	.12	1	.75**+	.38**+	.28**+	.22*+	.25*+	.36**+	.29**+	.34**+	.11	.07	.16	.04	.18
AMMAR	.16	.01	.07	.15	.04	.15	.26**+	.75**+	1	.48**+	.37**+	.31**+	.36**+	.45**+	.39**+	.43**+	.07	-.00	.09	.14	.28**+
Number of musical instruments	.28**+	-.03	-.09	.06	.04	.14	.09	.38**+	.48**+	1	.53**+	.54**+	.53**+	.61**+	.60**+	.62**+	.27**+	.38**+	.29**+	.43**+	.42**+
New song melody	.34**+	-.04	-.16	.30**+	.11	.20*	.15	.28**+	.37**+	.53**+	1	.93**+	.95**+	.75**+	.76**+	.73**+	.36**+	.47**+	.26**+	.52**+	.36**+
New song quality	.35**+	.02	-.24*+	.33**+	.04	.27**+	.23*+	.22*+	.31**+	.54**+	.93**+	1	.91**+	.78**+	.83**+	.75**+	.40**+	.48**+	.30**+	.56**+	.36**+
New song rhythm	.34**+	-.05	-.18	.30**+	.11	.24*+	.21*+	.25*+	.36**+	.53**+	.95**+	.91**+	1	.76**+	.77**+	.74**+	.40**+	.51**+	.22*+	.50**+	.39**+
Happy Birthday melody	.27**+	.09	-.28**+	.27**+	-.14	.23*+	.19*	.36**+	.45**+	.61**+	.75**+	.78**+	.76**+	1	.88**+	.86**+	.32**+	.42**+	.29**+	.43**+	.36**+
Happy Birthday quality	.33**+	.10	-.28**+	.22*+	-.11	.33**+	.29**+	.29**+	.39**+	.60**+	.76**+	.83**+	.77**+	.88**+	1	.89**+	.40**+	.44**+	.29**+	.53**+	.38**+
Happy Birthday rhythm	.27**+	.08	-.22*+	.24*+	-.09	.24*+	.18	.34**+	.43**+	.62**+	.73**+	.75**+	.74**+	.86**+	.89**+	1	.40**+	.46**+	.21*	.56**+	.33**+
Singing behaviour during childhood	.26**+	-.08	.04	.07	.01	.09	-.09	.11	.07	.27**+	.36**+	.40**+	.40**+	.32**+	.40**+	.40**+	1	.36**+	.07	.45**+	.19*
Self-estimation of singing capacity	-.00	.05	-.14	-.16	-.03	.17	-.11	.07	-.00	.38**+	.47**+	.48**+	.51**+	.42**+	.44**+	.46**+	.36**+	1	.11	.49**+	.19*
Highest level of education	.25*+	.01	-.07	.35**+	.21*	.30**+	.05	.16	.09	.29**+	.26**+	.30**+	.22*+	.29**+	.29**+	.21*	.07	.11	1	.40**+	.53**+
Performance anxiety	.28**+	-.00	.04	.01	.12	.11	-.03	.04	.14	.43**+	.52**+	.56**+	.50**+	.43**+	.53**+	.56**+	.45**+	.49**+	.40**+	1	.36**+
Number of foreign languages	.38**+	.06	.02	.21*	.29**+	.24*+	.08	.18	.28**+	.42**+	.36**+	.36**+	.39**+	.36**+	.38**+	.33**+	.19*	.19*	.53**+	.36**+	1

All variables that correlated with Mandarin imitation, which is the dependent variable, were entered as independent ones. The results revealed that 26 percent of the variance can be explained by three predictors: *New song rhythm*, *STMB* and *Mandarin familiar*.

Table 46. Multiple regression *Mandarin imitation*

Multiple regression models explaining the variance in Mandarin imitation

	<i>R</i>	<i>R</i> ²	F Change	Sig. F Change	<i>B</i>	<i>SE B</i>	β	<i>p</i>
Model 1	.39	.15	13.22	0.001				
Constant					0.89	0.40		
New song rhythm					0.21	0.06	.39	< .01
Model 2	.46	.21	5.21	0.025				
Constant					0.59	0.41		
New song rhythm					0.17	0.06	.32	< .01
STMB					0.07	0.03	.25	< .05
Model 3	.51	.26	4.84	0.031				
Constant					0.32	0.41		
New song rhythm					0.15	0.06	.28	< .01
STMB					0.08	0.03	.29	< .01
Mandarin familiar					0.05	0.02	.23	< .05

Dependent variable: Mandarin imitation

Figure 12. Variance in *Mandarin imitation* explained

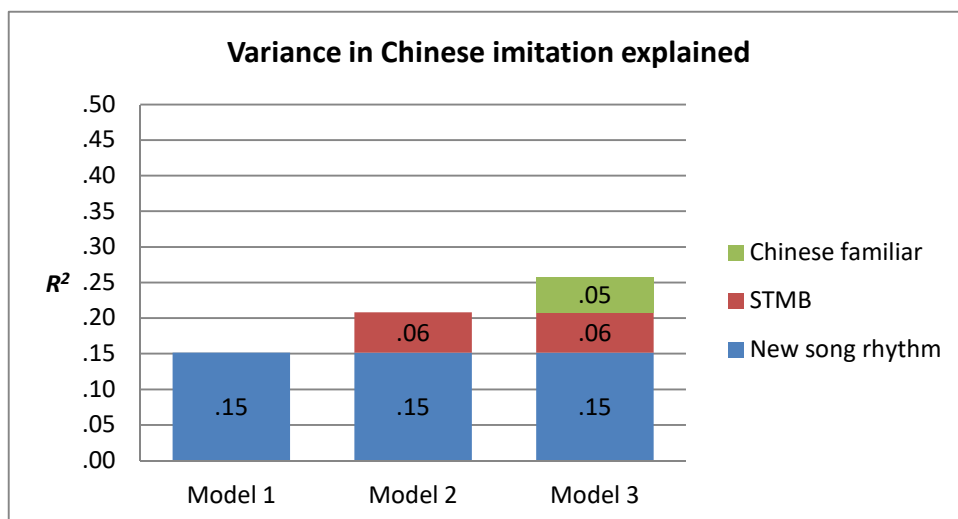


Table 47. Correlations for *Thai imitation*

	Thai imitation	Thai memorable	Thai difficult to imitate	Thai familiar	STMF	STMB	AMMAT	AMMAR	Number of musical instruments	New song melody	New song quality	New song rhythm	Happy Birthday melody	Happy Birthday quality	Happy Birthday rhythm	Singing behaviour during childhood	Self-estimation of singing capacity	Highest level of education	Performance anxiety	Number of foreign languages
Thai imitation	1	-.04	.37**+	-.05	.42**+	.39**+	.26**+	.27**+	.26**+	.46**+	.46**+	.46**+	.35**+	.40**+	.35**+	.21*+	.03	.22*+	.24*+	.30**+
Thai memorable	-.04	1	-.23**+	.31**+	-.17	-.14	.20*	.22*+	-.00	-.07	-.13	-.06	-.04	-.02	.00	.10	-.08	-.15	.04	-.10
Thai difficult to imitate	.37**+	-.23**+	1	-.11	.26**+	.22*+	.17	.10	.21*+	.42**+	.38**+	.41**+	.28**+	.27**+	.29**+	.05	.05	.39**+	.26**+	.21*+
Thai familiar	-.05	.31**+	-.11	1	-.05	-.19*	-.08	-.09	-.06	-.03	-.02	-.05	-.04	-.00	.01	.15	.01	.03	.13	.09
STMF	.42**+	-.17	.26**+	-.05	1	.46**+	.12	.15	.14	.20*+	.27**+	.24*+	.23*+	.33**+	.24*+	.09	.17	.30**+	.11	.24*+
STMB	.39**+	-.14	.22*+	-.19*	.46**+	1	.12	.26**+	.09	.15	.23*+	.21*+	.19*	.29**+	.18	-.09	-.11	.05	-.03	.08
AMMAT	.26**+	.20*	.17	-.08	.12	.12	1	.75**+	.38**+	.28**+	.22*+	.25*+	.36**+	.29**+	.34**+	.11	.07	.16	.04	.18
AMMAR	.27**+	.22*+	.10	-.09	.15	.26**+	.75**+	1	.48**+	.37**+	.31**+	.36**+	.45**+	.39**+	.43**+	.07	-.00	.09	.14	.28**+
Number of musical instruments	.26**+	-.00	.21*+	-.06	.14	.09	.38**+	.48**+	1	.53**+	.54**+	.53**+	.61**+	.60**+	.62**+	.27**+	.38**+	.29**+	.43**+	.42**+
New song melody	.46**+	-.07	.42**+	-.03	.20*+	.15	.28**+	.37**+	.53**+	1	.93**+	.95**+	.75**+	.76**+	.73**+	.36**+	.47**+	.26**+	.52**+	.36**+
New song quality	.46**+	-.13	.38**+	-.02	.27**+	.23*+	.22*+	.31**+	.54**+	.93**+	1	.91**+	.78**+	.83**+	.75**+	.40**+	.48**+	.30**+	.56**+	.36**+
New song rhythm	.46**+	-.06	.41**+	-.05	.24*+	.21*+	.25*+	.36**+	.53**+	.95**+	.91**+	1	.76**+	.77**+	.74**+	.40**+	.51**+	.22*+	.50**+	.39**+
Happy Birthday melody	.35**+	-.04	.28**+	-.04	.23*+	.19*	.36**+	.45**+	.61**+	.75**+	.78**+	.76**+	1	.88**+	.86**+	.32**+	.42**+	.29**+	.43**+	.36**+
Happy Birthday quality	.40**+	-.02	.27**+	-.00	.33**+	.29**+	.29**+	.39**+	.60**+	.76**+	.83**+	.77**+	.88**+	1	.89**+	.40**+	.44**+	.29**+	.53**+	.38**+
Happy Birthday rhythm	.35**+	.00	.29**+	.01	.24*+	.18	.34**+	.43**+	.62**+	.73**+	.75**+	.74**+	.86**+	.89**+	1	.40**+	.46**+	.21*+	.56**+	.33**+
Singing behaviour during childhood	.21*+	.10	.05	.15	.09	-.09	.11	.07	.27**+	.36**+	.40**+	.40**+	.32**+	.40**+	.40**+	1	.36**+	.07	.45**+	.19*
Self-estimation of singing capacity	.03	-.08	.05	.01	.17	-.11	.07	-.00	.38**+	.47**+	.48**+	.51**+	.42**+	.44**+	.46**+	.36**+	1	.11	.49**+	.19*
Highest level of education	.22*+	-.15	.39**+	.03	.30**+	.05	.16	.09	.29**+	.26**+	.30**+	.22*+	.29**+	.29**+	.21*+	.07	.11	1	.40**+	.53**+
Performance anxiety	.24*+	.04	.26**+	.13	.11	-.03	.04	.14	.43**+	.52**+	.56**+	.50**+	.43**+	.53**+	.56**+	.45**+	.49**+	.40**+	1	.36**+
Number of foreign languages	.30**+	-.10	.21*+	.09	.24*+	.08	.18	.28**+	.42**+	.36**+	.36**+	.39**+	.36**+	.38**+	.33**+	.19*	.19*	.53**+	.36**+	1

All variables that correlated with Tamil imitation, which is the dependent variable, were entered as independent ones. The results revealed that 29 percent of the variance can be explained by two predictors: *New song rhythm* and *STMF*.

Table 48. Multiple regression *Thai imitation*

Multiple regression models explaining the variance in Thai imitation

	<i>R</i>	<i>R</i> ²	F Change	Sig. F Change	<i>B</i>	<i>SE B</i>	β	<i>p</i>
Model 1	.45	.20	18.42	0.000				
Constant					0.10	0.38		
New song rhythm					0.23	0.05	.45	< .01
Model 2	.54	.29	9.63	0.003				
Constant					-0.45	0.40		
New song rhythm					0.20	0.05	.37	< .01
STMF					0.11	0.04	.31	< .01

Dependent variable: Thai imitation

Figure 13. Variance in *Thai imitation* explained

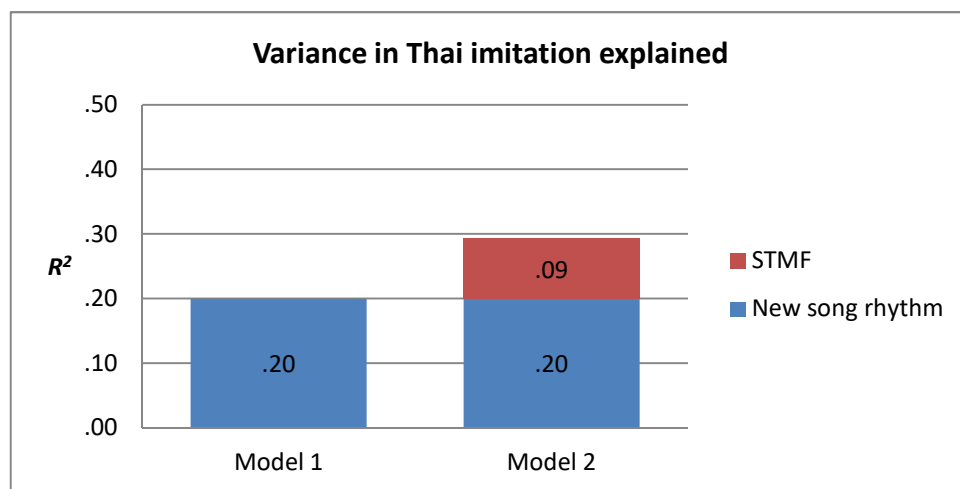


Table 49 demonstrates the predictors which explain some percentages of the variance of the individual language imitation tasks. The y-axis shows the individual languages, while the x-axis illustrates the 10 predictors which were found. The final line shows the count which demonstrates how often the 10 predictors were detected.

Table 49. Predictors

	Predictor 1 STMF	Predictor 2 New song rhythm	Predictor 3 STMB	Predictor 4 AMMAT	Predictor 5 New song melody	Predictor 6 Familiar (relates to each individual language)	Predictor 7 Memorable (relates to each individual language)	Predictor 8 Difficult to imitate (relates to each individual language)	Predictor 9 Happy Birthday melody	Predictor 10 Performance anxiety
Language										
Russian	x				x	x				
Farsi	x			x						
Hindi		x	x				x	x		
Tagalog				x					x	x
Japanese			x	x	x		x			
Tamil	x	x								
Mandarin		x	x			x				
Thai	x	x								
Count	4	4	3	3	2	2	2	1	1	1

11.4. Phonetic aptitude as measured by imitation tasks

Separate multiple regressions were performed for all eight languages to reveal predictors which are involved in explaining the variances of the performances. This aimed at identifying the main predictors of this investigation which could also be compared to the findings of previous research where the same stimuli for Tagalog, Hindi and Mandarin had already been used (Christiner 2013; Christiner & Reiterer 2013; Christiner et al. 2018). For illustration, a summary table about the main predictors, which were found to be involved in explaining some percentages of the variances of the individual imitation tasks for each of the languages is provided below (see table 49).

The regression models have shown that 10 predictors were involved in explaining the variances of the individual languages. Certain predictors occur more often, such as *STMF*, *New song rhythm*, *STMB* and *AMMAT*. Thus it can be concluded that the phonological STM, singing ability (rhythm and melody) and the ability to discriminate tonal differences (*AMMAT*) are the main predictors of the individual language imitation tasks. This represents the main findings which have also been found in past research (Christiner & Reiterer 2013; Christiner et al. 2018).

11.4.1. Verbal short-term memory

The *STMF* or the *STMB* contribute to the variances of the performances in seven out of eight languages. One of them is always included as a predictor, apart from *Tagalog imitation*. As the two measure the phonological STM capacity, it seems to be one of the best predictors for speech imitation in this investigation. This has already been promoted and shown in multiple investigations (Robinson 2002; Cain et al. 2004; Robinson 2005; Dörnyei 2005; Hu & Schuele 2005; Majerus et al. 2006; Hummel 2009; Wen & Skehan 2011; Vulchanova et al. 2014; Dörnyei & Ryan 2015; Wen et al. 2019) and will be discussed further in section 11.5.1.

11.4.2. Singing predictors

The individual singing parameter *New song rhythm* contributes to the variances of the performances in the languages Hindi, Tamil, Mandarin and Thai. Thus, this parameter is also a crucial predictor. To recap, the singing imitation task is based on the idea of learning a new song rather fast by listening to it three times where the participants had to sing the song purely from their memory. Sustaining the rhythmic structure of the original song was

therefore a challenging task which required the participant to keep the temporal features. The participants, who were good at sustaining the rhythmic structure of the new song they learned, were also better at imitating the four languages. This suggests that a sense for rhythm in song learning facilitates detecting the rhythmic structures of unintelligible languages. Research has shown that the segmentation of speech patterns is most difficult for learners of a foreign language in the early stages. Individuals often lack the ability to understand where a word begins or ends (Patel 2007). Even though the rhythm in language and music is quite diverse as language usually has no regular metrical structure, being able to sustain the rhythm of newly learnt songs improve the ability to recognise rhythmic cues and to detect and to remember segments of unintelligible speech. This has also been reported and found in previous investigations (Christiner 2013). Further evidence has also been provided in another study on children. Their foreign language capacity improved as a results of phonological training which included repeatedly singing of rhythmic melodies (Fonseca-Mora et al. 2015).

Melody in music has been shown to boost memory of verbal recall (Purnell-Webb & Speelman 2008). This suggests that melody “[...] seems to act as a path or a cue to evoke [...]” information (Fonseca-Mora 2000: 150). The singing parameters which contribute to the individual language imitation performances are *New song melody* for Russian and Japanese, and *Happy Birthday melody* for Tagalog. In this context, one difference has to be considered. The research design did not include a learning condition in which the same imitation tasks were sung and spoken, which is the common approach of most other research in this field. The two melody singing criteria in this investigation measure ability. Thus, a sense for melody facilitates the ability to treat unintelligible utterances in a more melodic way in general. Another interpretation is that some languages could be perceived as more melodic than others, which assists recall of utterances especially by those who sing and memorise melody well. This would also explain why those who have a sense for sustaining and generating melody possess enhanced phonetic imitation abilities.

11.4.3. Musical perception ability (tonal aptitude)

The musical aptitude measurement *AMMAT*, which provides information about the ability to discriminate tonal changes, contributes to the variance of the performances in three languages, namely Farsi, Tagalog and Japanese. Research has shown that tone transitions in melodies are strongly associated with prosodic variations of oral languages (Oechslin et al. 2010) with a dominant right hemispheric specialisation for linguistic information that is poor in content (Perkins et al. 1996). The imitation tasks were unintelligible to the participants,

which not only suggests that the processing of musical signals and meaningless linguistic material may overlap (Christiner et al. 2018), but also their generational processes to a certain extent.

The *AMMAT* is the only perceptual predictor that contributes to the variance of three of the individual language imitation tasks. This shows that tonal perception ability is crucial for both music and speech. In the factor analysis (table 32), it was shown that singing ability and tonal perception are less dependent. Both singing and music perception load onto separate factors. As was already discussed previously, some researchers have found a relationship between pitch perception and production, while others have not (see sections 4.6.3 and 11.2.1). Comparing tonal perception ability and its relationship to speech production, similar observations can be made. While for some language imitation tasks tonal discrimination ability is crucial, in others it is not. Although the *AMMAR* is correlated with all individual language imitation tasks apart from *Mandarin imitation*, it does not contribute to any of the variance of the imitation tasks (see tables 33, 35, 37, 39, 41, 43, 45, 47).

11.4.4. Performance anxiety, self-estimation concepts and speech perception

The concept *Performance anxiety* contributes to explaining ID in the ability to imitate Tagalog, which illustrates that unpleasant feelings can significantly influence the performances of oral speech. The influence of performance anxiety on foreign language learning has especially been related to oral performances. In the language learning environment, a low anxiety condition is very important and has been related to language success (Horwitz et al. 1986: 127). Imitation tasks should simulate a learning condition in which new language material needs to be repeated. This can cause considerable distress for the learners as well as for individuals who participate in studies. It is important to note that *Performance anxiety* only contributes to explaining the variance in Tagalog imitation. Both, however, also load onto the same factor as singing (see table 32) which indicates that Performance anxiety, Tagalog imitation and singing ability are related in a special way. In this respect evidence is mounting that Tagalog may be perceived to be more song-like. In consideration of the results in the follow-up analysis, Tagalog will be discussed further in the later section 13.2.

The participants of this investigation also had to indicate how familiar and memorable the language seemed to them. Furthermore, they were instructed to indicate how difficult they found the individual imitation tasks for each language. Even though the participants did not speak any of the languages they repeated, they were asked how familiar

the language seemed to them for each of the eight languages. Interestingly, two times the predictor *Familiar* contributed to the variances in the speech imitation performances for Russian and Mandarin. This indicates that despite the fact that participants did not speak either language, being more familiar with a language can contribute to the performance. Indeed, it is more likely that most participants had been exposed to Russian and Mandarin than to most of the other languages. This may suggest that a little exposure time can improve oral performances as well.

In addition, how memorable the language was to the participants predicted the variances in *Japanese* and *Hindi imitation*. The latter was also highly related to how difficult the *Hindi imitation* tasks were to the participants. These two parameters are quite interesting as well, since they show that participants seem to be aware of their own abilities and can estimate the quality of their performances quite precisely.

With the exception of the *AMMAT*, the other perceptual measurements failed to explain any variance of the eight language imitation tasks. The *AMMAR*, which correlated seven times with languages, failed as a predictor (see section 11.3). The same is true for all five language perception tasks (Farsi, Japanese, Mandarin, Tagalog and Hindi), although the languages, Hindi and Japanese, at least showed correlations between imitation and perception, while the others did not. This suggests, as is being found repeatedly, that production and perception of speech can largely be dissociated.

11.4.5. Comparisons of predictors with previous research

The discussion above shows that the individual predictors for the regression models are quite similar, with some occurring more often than others. Considering the results of two previous investigations where *Hindi*, *Tagalog* and *Mandarin imitation* performances were tested, distinct and similar predictors were found.

Some differences concerning study design of the previous studies have to be mentioned below. The slight differences in the design could also be responsible for the distinct findings. The previous research on Hindi was conducted by recruiting vocalists who received singing lessons (Christiner & Reiterer 2013), while most of the participants in this research are non-professional singers. The further former study which has focused on *Mandarin* and *Tagalog imitation* was conducted on children at the age of 10, while in the empirical part only adults were tested (Christiner et al. 2018). In addition, new predictors were included in this investigation as the test design was constantly improved. These may be

other reasons why former predictors could possibly become obsolete, replaced and consequently less predictive. The predictors of the studies are shown in table 50.

Table 50. Predictors of pervious and this investigation

<i>Dependent Variable</i>	<i>Predictors: previous investigations</i>	<i>Predictors: this investigation</i>
Hindi	STM total Singing Rhythm Singing Voice quality (Christiner & Reiterer 2013)	New song rhythm STMB Hindi memorable Hindi difficult to imitate
Mandarin	STMF IMMA tonal Singing performance overall (Christiner et al. 2018)	New song rhythm STMB Mandarin familiar
Tagalog	STMF IMMA rhythmic discrimination (Christiner et al. 2018)	Happy Birthday melody AMMAT Performance anxiety

11.4.6. Hindi, Tagalog, and Mandarin

The predictors for *Hindi imitation* are quite similar to those of previous investigations. Both regression models contain measurements of the phonological STM. While in previous investigations the *STM total* (forward and backward together) predicted the variances of the performances, in this investigation it was the *STMB* subtest. As already discussed, forward and backward spans load onto the same factors (Engle et al. 1999) and test the same concept.

Singing rhythm in previous investigations represents a total score of the rhythmic ability of both songs (familiar song singing and new song learning tasks), while in this investigation only the single score *New song rhythm* has been used. In both studies the participants' sense of rhythm while singing was rated by singing experts and is predictive to the variance in Hindi imitation. This shows that the singing parameter rhythm facilitates phonetic imitation ability in singers and non-singers. Therefore, this predictor should be integrated into singing research for professionals and non-professionals.

In marked contrast, the third predictor, *voice quality*, does not play any role in this investigation. The reason why *voice quality* contributed in previous research, however, is related to the choice of participants. Quality of voice is the marker that allows a distinction to be made between a professional and a non-professional voice (Sundberg 1988) and an ability or training-related effect that also improves speech imitation ability. This indicates that

predictors can also vary remarkably depending on the choice of participants. The predictors *Hindi memorable* and *Hindi difficult to imitate* also contribute to the model. These two predictors, however, were not part of previous investigations and will be discussed in section 11.5.2.

Mandarin performances should be treated with some caution for two reasons. First, previous research was conducted on children and second, singing measurements are only available as composite score (melody, rhythm and voice quality). For the Mandarin imitations, both regression models of past and this research contain measurements of the phonological *STM* which contribute to the variance of the Mandarin performances. In the same vein, for both adults and children singing also seems to positively influence *Mandarin imitation*. While for children global singing ability (tonal, rhythm and quality of voice together) contributed to the model, in this investigation, the sub-component, *New song rhythm* turned out to be a predictor. Differences, however, concern tonal ability, which has contributed to the variance in the performances of the children, although this could not be replicated in this study. This is surprising. The parameter *Mandarin familiar* which was predictive in this study is a new parameter that has been included the first time and cannot be compared to previous research.

As for the limitations for Mandarin mentioned above, they also apply to *Tagalog imitation*. Unexpectedly, children and adults do not share a single predictor that could explain the variance in the performances. While *STMF* and *IMMA* rhythm are predictors for Tagalog imitation of children, the predictors for adults were *Happy Birthday melody*, *AMMAT* and *Performance anxiety*. The different findings of the two studies could be a consequence of age differences. In chapter 9, group comparisons showed that before the age of thirteen there are significant differences in the ability to perceive unintelligible speech. Age-dependent differences between speech imitation tasks have not been investigated for particular ages. Research on pre-school children has shown that speech imitation tasks which consist of four to six syllables are predictive. However, in this study regression models did not reproduce meaningful results (Christiner & Reiterer 2018). In the latter investigation, *Tagalog imitation*, correlated with tonal and rhythmic musical perception ability and was also highly related to the singing ability of the pre-school children. (Christiner & Reiterer 2018).

Taken together, the following can be suggested. Tone languages may be predicted by singing ability as both Thai and Mandarin share the predictor *New song rhythm*, and singing ability has also been involved in explaining individual differences in performing Mandarin in previous investigations on children at the age of ten. In addition, research on pre-school

children support the close relationship between Tagalog imitation and singing. Tagalog was the only language which was highly related to singing ability in pre-school children (Christiner & Reiterer 2018).

11.4.7. Stress-timed, syllable-timed and mora-timed languages

As stated in the opening of this thesis, languages vary on their foot rhythm and are classified as stress-timed, syllable-timed and mora-timed languages accordingly.

Russian and *Farsi imitation*: These languages are considered to be stress-timed languages which both show only one overlapping predictor, *STMF*, which is a common predictor that is obviously crucial in all languages and cannot be considered language-specific. *Hindi* and *Tagalog imitation*: These languages which are classified as syllable-timed languages, do not share a single predictor. The same is true for Japanese and Tamil which are classified as mora-timed languages. Both also do not share a predictor. Thus it can be suggested that either a more precise test design has to be developed or more language material for each of the individual languages has to be provided in order to produce meaningful results.

11.5. Speech imitation total

For the analysis of the *Speech imitation total* score (dependent variable) regression analysis was applied. One reason for this is implied in the idea of phonetic aptitude. Precisely speaking, polyglots, who speak multiple languages possession high language aptitude. Polyglots are fast learners of new languages, which relates to their high capacity of the phonological memory (Papagno & Vallar 1995). Therefore, a test that includes multiple languages is required and is the main reason for creating a composite score which is comprised of all eight languages. Secondly, previous investigations included four to eight language samples, which includes that results could also have occurred by chance. A language imitation task that consists of eight languages with 32 samples and is conducted on an appropriate number of participants is hence more robust compared to previous investigations.

This is similar for the perception tasks. As already shown, the individual speech perception tasks did not predict any of the variances of the speech imitation tasks. If this failure is related to the fact that only five samples for each perception task of the individual languages have been included, the composite scores *PL perception* (25 items), *ML perception* (12 items) and *Overall Aptitude perception* that consists of all perception items could fix this issue. Essentially, both the speech imitation and perception tasks are unintelligible for the participants. This means that the basic condition is the same for all samples for the speech

perception and production tasks. Additionally, the inclusion of multiple languages aimed at being able to generalise the findings, which would allow some suggestions to be made which may be transferable to most languages. This knowledge could also be used to develop educational programmes.

The choice of the statistical analysis included running a factor analysis first to understand in how far the variables are related and might create meaningful composite scores, which has been discussed in section 11.2. This has shown that apart from the language perception tasks, all other scores are highly interrelated and composite scores indicate a very strong test design. Despite the lower connectivity of the perception tasks, a composite perception score has been created as discussed in section 11.2.4.

The total variables which were used are the *Speech imitation total*, the *Overall Aptitude perception*, the *PL perception* and the *ML perception* which intended to be robust test designs. The latter two sub-scores of the *Overall Aptitude perception* score have been included, as first, the *PL perception* resembles the speech imitation counterpart most, and second, the *ML perception* score is still a measurement which has to be inspected closely.

By the other total variables the following is meant: The *AMMATOTAL* (*AMMAT* + *AMMAR*) represents the perception measurement for musical aptitude, *STMT* (*STMF* + *STMB*) is the phonological *STM* capacity measurement, *New song total* (tonal, rhythm and quality of voice) and *Happy Birthday total* (tonal, rhythm and quality of voice) are the singing tasks. The *Speech imitation total* score is the dependent variable. Further criteria which were included are as follows: *All Languages memorable*, *All Languages difficult to imitate*, *All Languages familiar*, *Number of musical instruments*, *Singing behaviour during childhood*, *Singing regularly hours per week*, *Self-estimation of singing capacity*, *Performance anxiety*, *Number of foreign languages* and *Highest level of education*. All the independent variables were entered into the multiple linear regression models only if a probability of F-change < 0.05 was given (see table 51).

Table 51. Correlations for *Speech imitation total*

	Speech imitation total	Overall Aptitude perception	PL perception	ML perception	All Languages memorable	All Languages difficult to imitate	All Languages familiar	STMT	AMMATOTAL	Number of musical instruments	New song total	Happy Birthday total	Singing behaviour during childhood	Singing regularly hours per week	Self-estimation of singing capacity	Performance anxiety	Number of foreign languages	Highest level of education
Speech imitation total	1	.08	.16	-.17	.22*	.32**+	.09	.53**+	.35**+	.34**+	.52**+	.44**+	.27*+	-.03	.08	.31**+	.49**+	.39**+
Overall Aptitude perception	.08	1	.82**+	.56**+	-.08	.03	.05	.15	.19	.14	.11	.04	-.17	-.21	.08	-.00	.18	.06
PL perception	.16	.82**+	1	.08	.03	-.04	-.02	.16	.28*+	.18	.08	.07	-.10	-.16	.04	.03	.13	.10
ML perception	-.17	.56**+	.08	1	.02	.03	.20	-.01	-.11	-.02	-.03	-.08	-.17	-.09	.06	-.06	.02	-.11
All Languages memorable	.22*	-.08	.03	.02	1	-.06	.34**+	.04	.20	.12	-.02	-.03	.10	.14	-.00	.15	.07	-.00
All Languages difficult to imitate	.32**+	.03	-.04	.03	-.06	1	.23*	.17	.13	.03	.29**+	.19	.05	-.10	-.12	.05	.23*	.24*
All Languages familiar	.09	.05	-.02	.20	.34**+	.23*	1	-.11	.00	-.01	.09	-.03	.14	.03	.00	.12	.24*	.03
STMT	.53**+	.15	.16	-.01	.04	.17	-.11	1	.19	.14	.26*	.29**+	.00	-.04	.04	.06	.18	.17
AMMATOTAL	.35**+	.19	.28*+	-.11	.20	.13	.00	.19	1	.47**+	.35**+	.45**+	.11	.02	.10	.12	.29**+	.16
Number of musical instruments	.34**+	.14	.18	-.02	.12	.03	-.01	.14	.47**+	1	.54**+	.64**+	.27*+	.19	.38**+	.43**+	.42**+	.29**+
New song total	.52**+	.11	.08	-.03	-.02	.29**+	.09	.26*	.35**+	.54**+	1	.80**+	.40**+	.31**+	.50**+	.55**+	.38**+	.27*+
Happy Birthday total	.44**+	.04	.07	-.08	-.03	.19	-.03	.29**+	.45**+	.64**+	.80**+	1	.38**+	.34**+	.45**+	.51**+	.38**+	.28*+
Singing behaviour during childhood	.27*+	-.17	-.10	-.17	.10	.05	.14	.00	.11	.27*+	.40**+	.38**+	1	.48**+	.36**+	.45**+	.19	.07
Singing regularly hours per week	-.03	-.21	-.16	-.09	.14	-.10	.03	-.04	.02	.19	.31**+	.34**+	.48**+	1	.46**+	.40**+	.05	-.02
Self-estimation of singing capacity	.08	.08	.04	.06	-.00	-.12	.00	.04	.10	.38**+	.50**+	.45**+	.36**+	.46**+	1	.49**+	.19	.11
Performance anxiety	.31**+	-.00	.03	-.06	.15	.05	.12	.06	.12	.43**+	.55**+	.51**+	.45**+	.40**+	.49**+	1	.36**+	.40**+
Number of foreign languages	.49**+	.18	.13	.02	.07	.23*	.24*	.18	.29**+	.42**+	.38**+	.38**+	.19	.05	.19	.36**+	1	.53**+
Highest level of education	.39**+	.06	.10	-.11	-.00	.24*	.03	.17	.16	.29**+	.27*+	.28*+	.07	-.02	.11	.40**+	.53**+	1

All variables that correlated with the *Speech imitation total* score, which is the dependent variable, were entered as independent ones. The results revealed that 60 percent of the variance can be explained by four predictors: The *Number of foreign languages*, *STMT*, *New song total* and *All Languages memorable*.

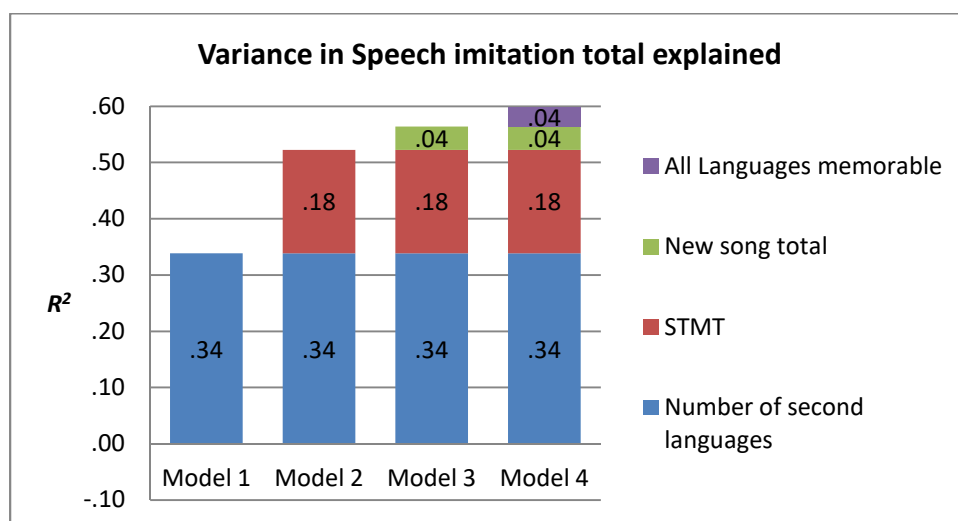
Table 52. Multiple regression *Speech imitation total*

Multiple regression models explaining the variance in Speech imitation total

	<i>R</i>	<i>R</i> ²	F Change	Sig. F Change	<i>B</i>	<i>SE B</i>	β	<i>p</i>
Model 1	.58	.34	38.88	0.000				
Constant					2.17	0.15		
Number of foreign languages					0.51	0.08	.58	< .01
Model 2	.72	.52	28.94	0.000				
Constant					0.73	0.30		
Number of foreign languages					0.44	0.07	.50	< .01
STMT					0.10	0.02	.44	< .01
Model 3	.75	.56	7.02	0.010				
Constant					0.16	0.36		
Number of foreign languages					0.36	0.07	.41	< .01
STMT					0.09	0.02	.39	< .01
New song total					0.14	0.05	.23	< .01
Model 4	.77	.60	6.46	0.013				
Constant					-0.49	0.43		
Number of foreign languages					0.34	0.07	.38	< .01
STMT					0.09	0.02	.37	< .01
New song total					0.16	0.05	.26	< .01
All Languages memorable					0.13	0.05	.19	< .05

Dependent variable: Speech imitation total

Figure 14. Variance in *Speech imitation total* explained



11.5.1. Phonetic aptitude, and the phonological short-term memory and foreign languages

The impact of the number of foreign languages spoken on the learning of new languages has been investigated in various ways. Individuals who learn a second or third language after their first language has been established such as successive and late bilinguals, as well as polyglots have been found to benefit from knowing larger phonological structures. Researchers distinguish between early and late bilinguals and it has generally been accepted that early and late bilingualism or multilingualism can improve the ability to acquire novel phonological forms (Kaushanskaya & Marian 2009; Hell & Mahn 1997; Papagno & Vallar 1995). In this investigation, 10 participants grew up bilingually and 1 trilingually, which indicates that the majority of the participants acquired foreign languages in a classroom environment or later. 48 of them spoke two or more languages (see section 10.2.1). Consequently, most of the participants are late bilinguals or late multilinguals. Comparable studies with a similar choice of participants have revealed that new word learning is related to the number of foreign languages spoken (Papagno & Vallar 1995; Hell & Mahn 1997). Particular emphasis has been given to phonological coding ability and new word learning (Cheung 1996; Hulme et al. 1991; Papagno et al. 1991). As a result, it has been suggested that “[m]ultilinguals probably also dispose of more refined (long-term) knowledge of phonological structures” (Hell & Mahn 1997: 536) which may lead to a higher adaptability and larger variety of possible phonological varieties in the mind. Interestingly, the predictor *Number of foreign languages* did not predict any of the eight individual language imitation performances, but the entire speech imitation measurement comprised of all eight languages.

This supports the claim that the more phonological structures are available the better the performances on various language imitation performances can be observed. This combined with improved phonological STM has been said to be responsible for improved foreign language performances (Hell & Mahn 1997), which is the second important predictor.

Language aptitude and its relationship to WM and STM has been discussed in multiple investigations (Robinson 2002; Cain et al. 2004; Robinson 2005; Dörnyei 2005; Hu & Schuele 2005; Majerus et al. 2006; Hummel 2009; Wen & Skehan 2011; Vulchanova et al. 2014; Dörnyei & Ryan 2015; Wen et al. 2019). One of the most investigated subcomponents of WM in the context of language aptitude is the phonological STM.

In this investigation the phonological STM was measured by using a digit span (forward and backward). It turned out to be one of the best predictors that contribute to the ability to imitate unintelligible speech. This is in concordance with previous investigations where the same findings have been reported (Christiner & Reiterer 2013; Christiner 2013; Christiner & Reiterer 2018; Christiner & Reiterer 2019). Individuals that retain and repeat phonetic elements faster can incorporate new sounds in a language learning situation more easily (Hummel 2009) and are considered to be exceptional learners. This is what imitation of unintelligible speech should simulate. Similarly, non-word repetition tasks have also been related to the learning of new phonological elements of languages, but not to the learning of the semantic interpretations of words (Gathercole & Baddeley 1990).

In this early stage of (foreign) language learning, the phonological loop capacity is a crucial predictor which can be seen as a potential that is associated with foreign language success. Evidence of the role of the phonological STM comes from multiple directions. For instance, clinical impairment studies show that patients who suffer from impairments of the phonological loop show poor abilities in foreign language learning (Baddeley 2010; Baddeley et al. 1998), which particularly affects phonological acquisition processes (Baddeley et al. 1998). In contrast, polyglots who possess high aptitude show an improved phonological STM capacity that has been related to the acquisition of non-words (Papagno & Vallar 1995). The phonological loop is also seen as a language acquisition device (Biedroń & Pawlak 2016). Thus the relationship of the STM and the imitation tasks has been expected.

11.5.2. Phonetic aptitude, singing and memorable

The phonetic relationship between language and music has been shown in multiple investigations (Chobert et al. 2014; François et al. 2013; Besson et al. 2011; Wong & Perrachione 2007; Wong et al. 2007), but only a few included singing measurements. Studies

have shown that singing facilitates the learning of new languages (Ludke et al. 2014). Singing in the foreign language classroom has often been employed as a learning tool (Anton 1990). For instance, foreign words are presented and learnt together with a melody. The technique is also said to serve “[...] as a mnemonic of storage of words in the long-term memory” (Gordon et al. 2010: 9889).

Comparing previous investigations and the findings of this research, it has been demonstrated that singing ability and speech imitation are not only related as a consequence of specific skills which are available for experts such as vocalists (Christiner 2013; Christiner & Reiterer 2013). It seems to be plausible that this relationship holds true for children (Christiner et al. 2018) and also for non-professional musicians or singers as shown in this investigation. Additional evidence has been provided by research on non-professional singers in which it has been suggested that accurate singing seems to be rather normally distributed and only 10 to 15 percent of individuals sing out of tune (Dalla Bella & Berkowska 2009; Sloboda 2005). Singing ability may therefore be an interesting tool for improving oral (foreign) language skills in general.

The predictor *Memorable* is an interesting criterion that indicates that individuals are highly aware of their own abilities, which directly affects their performances. This shows that self-estimation concepts may be useful tools which could more likely be used in educational context to facilitate foreign language choices. Strictly speaking, this is not new since self-estimation is a very common and accepted measurement. However, in the context of phonetic imitation aptitude it has largely been omitted so far. This indicates that future directions should include self-estimation measurements.

11.5.3. Discussion

As has been shown, 10 main predictors were able to explain the variances of the performances of the eight language imitation tasks. Certain predictors such as *STMF*, *New song rhythm*, *STMB* and *AMMAT* have occurred more often than others. It can be suggested that the phonological STM capacity, singing ability (rhythm and melody) and the ability to discriminate tonal differences (*AMMAT*) are the main predictors which could explain the variances of the performances of the individual language imitation tasks. Those four predictors were also found in previous investigations. This indicates that they seem to generally support speech imitation of unfamiliar languages.

The reason for the variation of predictors in the individual speech imitation tasks could also lie outside language typology and, for instance, rely on the choice of participants,

the stimuli, the test design and environmental aspects. This is especially evident if we look at the predictors of the entire *Speech imitation total* test. The *Number of foreign languages* contributes to the entire variances of the performances of all languages only in the composite condition. This is a clear indicator that the test design has an influence on the predictors. The number of foreign languages spoken is related to knowing larger amounts of phonological structures, which suggests that participants who speak multiple foreign languages are also better at imitating new languages.

This also has an effect on the idea of investigating whether typologically different languages are differently related to musical abilities. Some predictors which could explain a certain percentage of the variances of the performances occur in more languages, while others occur twice or once. There is some regularity and for all language imitation tasks, apart from Tagalog, STM capacity plays a significant role. However, this is a general predictor.

New song rhythm has also been isolated as a crucial predictor for four languages, namely Hindi, Tamil, Thai and Mandarin. Research has shown that singing rhythmic melodies improves foreign language capacity (Fonseca-Mora et al. 2015) and previously the criterion singing rhythm was also involved in explaining the variance of Hindi imitation performance in vocalists (Christiner & Reiterer 2013). Therefore, the singing criterion rhythm should be used for professional and non-professional singers if tested for their singing ability (see also section 11.4.6) and additional singing criterion rhythm also seems to be a more general predictor. Another important finding is that Mandarin and Thai imitation are both related to singing ability which has also been found previously in children (Christiner et al. 2018). This suggests that singing ability probably improve tone language imitation ability.

Some limitations may also concern the measurements. The singing criteria melody, rhythm and quality of voice show high loads onto the same factor (see table 32), which illustrates two aspects. One is very positive as it shows a very compact and well-designed singing measurement. Despite this, it could also be interpreted differently. The high interrelation of the singing criteria indicates that a global score could reveal the same results or the omission of one criterion could lead to a replacement by another one.

Other possible predictors such as vocal range have not been included and may influence results as well, since research is always incomplete. This also shows that future directions have to give a multidimensional approach which includes a set of measurements for all main factors: singing, speech imitation, speech and music perception. Partly this issue can be addressed by using composite scores that may be more robust and give general information as has been addressed in the analysis of the entire speech imitation test.

The *Speech imitation total* score has been used as a composite score which correlates with all other main variables, followed by a regression analysis. This aimed at giving a global idea about the predictors which may be relevant for all languages chosen within this investigation. Results revealed that the number of foreign languages, the phonological STM capacity, singing (*New song total*), and how memorable the language occurs to individuals could explain 60 percent of the variances of the performances within this sample. This shows that the number of foreign languages, the phonological STM capacity, singing ability, and how memorable the language appears to the participants seem to be most relevant for the imitation of new languages. In addition, how memorable a language seems to individuals also contributes to the variance of the performances, which indicates that self-awareness of one's own abilities directly affects performance.

Importantly, speech perception was not able to predict any of the imitation tasks, which reflect findings and suggestions of previous research. As has been discussed, comparisons between speech perception and production were limited by the number of items in each condition. To overcome this issue, the composite scores were used which revealed the same results. This suggests that the limitations of the individual conditions cannot account for the results since the same finding could be replicated in the complex condition as well.

11.6. Perception versus production

11.6.1. Group differences

Group comparisons between singers, instrumentalists, amateurs (participants who played a musical instrument as described in section 10.3) and non-musicians have been carried out to understand whether in some particular groups abilities are differently distributed. In previous research, for instance, it has been demonstrated that vocalists performed significantly better than instrumentalists in speech imitation ability. The instrumentalists also performed significantly better than the non-musicians did (Christiner & Reiterer 2015). In chapter 9, it was also demonstrated that participants who played a musical instrument were also better in speech perception ability. In this analysis, it will be shown whether these findings could be replicated. Additionally, it also aims at understanding in how far production and perception variables differentiate across groups.

Participants were divided into four groups of musical classification, singers (N = 15), instrumentalists (N = 15), amateurs (N = 21), and non-musicians (N = 35). Each participant was assigned to one group exclusively, according to the information the participant had given. The distribution of every language imitation task, language perception task, music perception

task and the singing tasks were tested for each group with the aim to determine whether there was a significant difference among them. In addition, tests to show differences between pairs of groups were also performed. The same was applied to the multi-item concepts *Singing behaviour during childhood* and *Performance anxiety*, which were added to the production variables. As the data was nonparametric, for overall distribution the Kruskal-Wallis test was chosen, and for pairs of groups the Mann-Whitney test was performed.

11.6.2. Results production

Table 53 shows the results of the group comparisons for all language imitation tasks, table 54 the results of the group comparisons for singing including *Singing behaviour during childhood* and *Performance anxiety*, while table 55 shows the descriptives. The latter also contains information about the median and mean for each group and task. This data shows that the higher the median and the mean the better the performance was. The only variable which must be read differently is *Performance anxiety*. This means that the higher the median and mean, the less anxious participants are. Figure 15 illustrates the mean differences for all language imitation tasks, singing tasks including *Singing behaviour during childhood*, and *Performance anxiety*.

For readability of the results, the main findings and group differences are described starting at section 11.6.2.1.

Table 53. Group comparisons language imitation

	Russian imitation		Japanese imitation		Hindi imitation		Tamil imitation		Mandarin imitation		Tagalog imitation		Thai imitation		Farsi imitation	
	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>
Kruskal-Wallis Test for 4 groups	0.00	sig	0.00	sig	0.01	sig	0.01	sig	0.00	sig	0.00	sig	0.00	sig	0.00	sig
Group comparison (Mann-Whitney)																
singers - instrumentalists	0.29	not sig*	0.76	not sig*	0.97	not sig*	0.29	not sig*	0.25	not sig*	0.56	not sig*	0.55	not sig*	0.62	not sig*
singers - amateurs	0.15	not sig*	0.01	sig*	0.05	not sig*	0.29	not sig*	0.32	not sig*	0.02	sig*	0.01	sig*	0.07	not sig*
singers – non-musicians	0.00	sig*	0.00	sig*	0.01	sig*	0.05	not sig*	0.01	sig*	0.00	sig*	0.00	sig*	0.00	sig*
instrumentalists - amateurs	0.01	sig*	0.02	sig*	0.04	not sig*	0.02	not sig*	0.06	not sig*	0.02	sig*	0.22	not sig*	0.03	not sig*
instrumentalists -non-musicians	0.00	sig*	0.00	sig*	0.01	sig*	0.00	sig*	0.00	sig*	0.00	sig*	0.03	not sig*	0.00	sig*
amateurs -non-musicians	0.07	not sig*	0.48	not sig*	0.29	not sig*	0.26	not sig*	0.15	not sig*	0.03	sig*	0.12	not sig*	0.23	not sig*

* Significance after Benjamini-Hochberg correction ($p < 0.05$)

Table 54. Group comparisons for singing including *Singing behaviour during childhood* and *Performance anxiety*

	Happy Birthday total		New song total		Singing behaviour during childhood		Performance anxiety	
	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>
Kruskal-Wallis Test for 4 groups	0.00	sig	0.00	sig	0.00	sig	0.00	sig
Group comparison (Mann-Whitney)								
singers - instrumentalists	0.00	sig*	0.00	sig*	0.33	not sig*	0.46	not sig*
singers - amateurs	0.00	sig*	0.00	sig*	0.01	sig*	0.01	sig*
singers -non-musicians	0.00	sig*	0.00	sig*	0.01	sig*	0.00	sig*
instrumentalists - amateurs	0.00	sig*	0.01	sig*	0.02	sig*	0.09	not sig*
instrumentalists -non-musicians	0.00	sig*	0.00	sig*	0.03	sig*	0.00	sig*
amateurs -non-musicians	0.03	sig*	0.17	not sig*	0.27	not sig*	0.18	not sig*

* Significance after Benjamini-Hochberg correction ($p < 0.05$)

Table 55. Descriptives

		Russian imitation	Japanese imitation	Hindi imitation	Tamil imitation	Mandarin imitation	Tagalog imitation	Thai imitation	Farsi imitation	Happy Birthday total	New song total	Singing behaviour during childhood	Perfor- mance anxiety
<i>N</i>	singers	15	15	15	15	15	15	15	15	15	15	15	15
<i>M</i>		4.04	5.49	3.77	3.32	2.64	3.56	2.03	2.97	8.21	8.11	6.88	6.67
<i>SE</i>		0.33	0.36	0.35	0.37	0.23	0.55	0.14	0.36	0.20	0.26	0.68	0.40
<i>Median</i>		3.75	5.54	3.60	2.70	2.40	3.19	2.10	2.69	8.17	8.38	8.00	7.00
<i>min</i>		2.06	1.25	1.85	1.40	1.50	0.81	1.00	0.75	6.92	6.54	2.25	3.00
<i>max</i>		6.31	6.83	7.85	5.70	5.00	7.81	2.85	6.50	9.50	9.17	9.75	8.67
<i>N</i>	instrumentalists	15	15	15	15	15	15	15	15	15	15	14	14
<i>M</i>		4.53	5.59	3.68	3.95	2.77	2.97	1.89	3.09	6.87	6.72	6.13	6.14
<i>SE</i>		0.34	0.32	0.22	0.33	0.17	0.29	0.19	0.29	0.26	0.32	0.51	0.53
<i>Median</i>		4.38	5.63	3.50	4.45	2.85	3.00	1.85	2.88	7.00	7.04	5.63	5.67
<i>min</i>		2.63	3.38	2.20	1.25	1.15	1.31	0.85	1.00	4.50	4.67	4.00	2.83
<i>max</i>		7.13	7.13	5.15	5.40	3.60	5.94	3.00	4.81	8.75	8.92	10.00	9.17
<i>N</i>	amateurs	21	21	21	21	21	21	21	21	21	21	21	21
<i>M</i>		3.49	4.55	3.05	2.90	2.40	2.19	1.55	2.18	5.75	5.70	4.27	4.89
<i>SE</i>		0.26	0.25	0.21	0.30	0.21	0.17	0.12	0.27	0.13	0.15	0.51	0.41
<i>Median</i>		3.50	4.33	2.95	2.50	2.35	2.19	1.40	2.13	5.83	5.67	3.75	4.67
<i>min</i>		1.75	2.88	1.45	1.45	0.90	0.94	0.30	0.19	4.83	4.58	0.25	2.00
<i>max</i>		7.00	7.04	4.90	7.40	5.20	3.69	3.00	4.88	6.67	7.13	9.00	8.00
<i>N</i>	non-musicians	35	35	35	35	35	35	35	35	35	35	35	35
<i>M</i>		3.16	4.37	2.77	2.55	2.02	1.78	1.42	1.81	5.29	5.15	4.76	4.15
<i>SE</i>		0.27	0.23	0.17	0.22	0.11	0.14	0.14	0.19	0.13	0.18	0.29	0.34
<i>Median</i>		2.81	4.21	2.80	2.20	1.95	1.69	1.25	1.56	5.33	5.42	4.25	4.00
<i>min</i>		1.19	2.21	0.80	0.85	0.90	0.50	0.10	0.31	3.67	2.75	1.50	0.33
<i>max</i>		7.63	7.42	4.75	5.55	3.50	4.06	3.95	5.44	6.83	6.96	8.75	8.33

11.6.2.1.Singers versus instrumentalists

Figure 15 shows the group comparisons of the individual language imitation tasks, singing ability, *Singing behaviour during childhood* and *Performance anxiety*. The results reveal that there was no significant difference between instrumentalists and vocalists for all language imitation tasks, *Singing behaviour during childhood* and *Performance anxiety*. Vocalists performed significantly better than instrumentalists in both singing tasks as expected (see section 11.6).

11.6.2.2.Singers versus amateurs

Singers' language performances in Japanese, Tagalog and Thai were significantly better than those of the amateurs. Singers' singing performances were also superior to the amateurs. As expected, singers were also less anxious (*Performance anxiety*) and sang more frequently during childhood.

11.6.2.3.Singers versus the non-musicians

Singers' language performances in all language imitation tasks were superior to the non-musicians, except for Tamil. As expected, singers were also less anxious (*Performance anxiety*) and sang more frequently during childhood than the non-musicians.

11.6.2.4.Instrumentalists versus amateurs

Instrumentalists' language performances in Russian, Japanese and Tagalog were significantly better than the amateurs. Instrumentalists were better at both singing tasks and sang more frequently during childhood than the amateurs. There was no significant difference between the two groups in terms of anxiety.

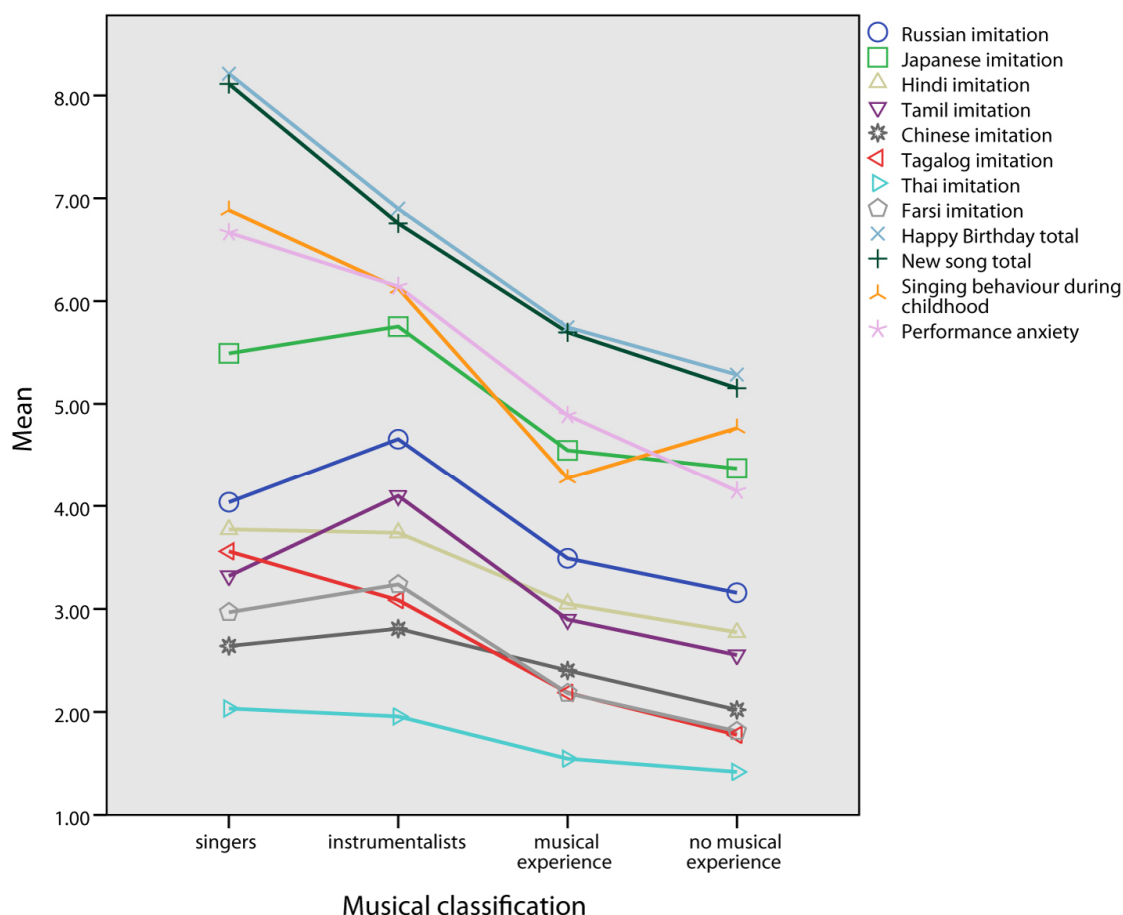
11.6.2.5.Instrumentalists versus non-musicians

Instrumentalists' language performances in all language imitation tasks were superior to the non-musicians, except for Thai. Instrumentalists were also less anxious (*Performance anxiety*) and sang more frequently during childhood than the non-musicians.

11.6.2.6.Amateurs versus non-musicians

Comparisons between both groups reveal that that there are only two parameters that are significantly different. These are *Tagalog imitation* and *Happy Birthday* singing, which are both performed better by the amateurs.

Figure 15. Mean differences



11.6.3. Results Perception

Table 56 shows the results of the group comparisons for all language perception tasks, table 57 the results of the group comparisons for the music perception variables *AMMATOTAL*, *AMMAT* and *AMMAR*, and table 58 contains the descriptives. Figure 16 illustrates the mean differences for all language and music perception tasks. For readability of the results, the main findings and group differences are described starting at section 11.6.3.1.

Table 56. Group comparisons for all language perception tasks

	Overall Aptitude perception		PL perception		ML perception		Tagalog perception		Japanese perception		Mandarin perception		Farsi perception		Hindi perception	
	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>
Kruskal-Wallis Test for 4 groups	0.39	not sig	0.07	not sig	0.60	not sig	0.05	not sig	0.32	not sig	0.96	not sig	0.14	not sig	0.37	not sig
Group comparison (Mann-Whitney)																
singers - instrumentalists	0.88	not sig*	0.19	not sig*	0.25	not sig*	0.61	not sig*	0.27	not sig*	0.71	not sig*	0.07	not sig*	0.13	not sig*
singers - amateurs	0.84	not sig*	0.63	not sig*	0.71	not sig*	0.61	not sig*	0.86	not sig*	0.79	not sig*	0.40	not sig*	0.27	not sig*
singers -non-musicians	0.19	not sig*	0.26	not sig*	0.58	not sig*	0.06	not sig*	0.52	not sig*	0.57	not sig*	0.70	not sig*	0.11	not sig*
instrumentalists - amateurs	0.84	not sig*	0.31	not sig*	0.27	not sig*	0.28	not sig*	0.15	not sig*	0.90	not sig*	0.26	not sig*	0.61	not sig*
instrumentalists -non-musicians	0.14	not sig*	0.01	not sig*	0.34	not sig*	0.01	not sig*	0.08	not sig*	0.92	not sig*	0.04	not sig*	0.98	not sig*
amateurs -non-musicians	0.31	not sig*	0.13	not sig*	0.71	not sig*	0.15	not sig*	0.59	not sig*	0.80	not sig*	0.26	not sig*	0.59	not sig*

* Significance after Benjamini-Hochberg correction ($p < 0.05$)

Table 57. Group comparisons for the music perception variables *AMMATOTAL*, *AMMAT* and *AMMAR*

	AMMATOTAL		AMMAT		AMMAR	
	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>	<i>p</i>	<i>sig</i>
Kruskal-Wallis Test for 4 groups	0.00	sig	0.00	sig	0.60	not sig
Group comparison (Mann-Whitney)						
singers - instrumentalists	0.02	sig*	0.05	not sig*	0.00	sig*
singers - amateurs	0.10	not sig*	0.17	not sig*	0.17	not sig*
singers -non-musicians	0.00	sig*	0.01	sig*	0.00	sig*
instrumentalists - amateurs	0.00	sig*	0.00	sig*	0.00	sig*
instrumentalists -non-musicians	0.00	sig*	0.00	sig*	0.00	sig*
amateurs -non-musicians	0.00	sig*	0.01	sig*	0.02	sig*

* Significance after Benjamini-Hochberg correction ($p < 0.05$)

Table 58. Descriptives

		Overall Aptitude perception	PL perception	ML perception	Tagalog perception	Japanese perception	Mandarin perception	Farsi perception	Hindi perception	AMMA- TOTAL	AMMAT	AMMAR
<i>N</i>	singers	15	15	15	15	15	15	15	14	15	15	15
<i>M</i>		0.59	0.63	0.49	0.67	0.74	0.62	0.54	0.63	0.72	0.69	0.75
<i>SE</i>		0.03	0.03	0.04	0.06	0.05	0.06	0.05	0.06	0.03	0.04	0.02
<i>Median</i>		0.62	0.68	0.50	0.60	0.80	0.60	0.40	0.70	0.75	0.75	0.75
<i>min</i>		0.41	0.45	0.17	0.40	0.40	0.20	0.40	0.20	0.53	0.48	0.58
<i>max</i>		0.70	0.82	0.75	1.00	1.00	1.00	1.00	0.80	0.90	0.90	0.90
<i>N</i>	instrumentalists	15	15	15	15	15	15	15	14	15	15	15
<i>M</i>		0.59	0.69	0.44	0.71	0.80	0.66	0.65	0.54	0.82	0.80	0.84
<i>SE</i>		0.02	0.03	0.03	0.06	0.06	0.04	0.06	0.04	0.02	0.03	0.02
<i>Median</i>		0.62	0.68	0.42	0.80	0.83	0.60	0.60	0.60	0.83	0.80	0.85
<i>min</i>		0.49	0.50	0.17	0.20	0.20	0.40	0.20	0.20	0.63	0.53	0.73
<i>max</i>		0.70	0.91	0.64	1.00	1.00	1.00	1.00	0.80	1.00	1.00	1.00
<i>N</i>	amateurs	20	20	20	20	20	20	20	19	21	21	21
<i>M</i>		0.59	0.64	0.49	0.63	0.73	0.65	0.58	0.57	0.67	0.63	0.71
<i>SE</i>		0.02	0.03	0.03	0.05	0.04	0.04	0.05	0.04	0.01	0.02	0.02
<i>Median</i>		0.57	0.64	0.50	0.60	0.80	0.60	0.60	0.60	0.68	0.63	0.73
<i>min</i>		0.41	0.45	0.17	0.20	0.40	0.40	0.20	0.20	0.58	0.50	0.58
<i>max</i>		0.76	0.82	0.83	1.00	1.00	1.00	1.00	0.80	0.79	0.80	0.85
<i>N</i>	non-musicians	35	35	35	35	35	35	35	34	35	35	35
<i>M</i>		0.55	0.59	0.48	0.53	0.70	0.67	0.50	0.54	0.61	0.57	0.65
<i>SE</i>		0.02	0.02	0.02	0.03	0.04	0.03	0.04	0.03	0.01	0.01	0.02
<i>Median</i>		0.54	0.59	0.50	0.60	0.60	0.60	0.40	0.60	0.61	0.55	0.68
<i>min</i>		0.35	0.36	0.17	0.20	0.20	0.40	0.00	0.20	0.44	0.45	0.43
<i>max</i>		0.76	0.82	0.83	1.00	1.00	1.00	1.00	0.80	0.81	0.80	0.83

11.6.3.1.Singers versus instrumentalists

The group comparisons show that no significant difference in speech perception ability across all groups could be detected. Singers, however, were outperformed by the instrumentalists on two of the musicality perception measurements, i.e. *AMMAR* and *AMMATOTAL* (see section 11.7).

11.6.3.2.Singers versus amateurs

Group comparisons between the singers and amateurs show that there is no significant difference for all perception tasks in language and music.

11.6.3.3.Singers versus non-musicians

Group comparisons on singers and non-musicians have shown that there is no significant difference for all speech perception tasks. However, singers outperformed the non-musicians on all music perception tasks, i.e. *AMMATOTAL*, *AMMAT*, and *AMMAR*.

11.6.3.4.Instrumentalists versus amateurs

Group comparisons between instrumentalists and amateurs show that there is no significant difference for all speech perception tasks. However, instrumentalists outperformed the amateurs on all music perception tasks, i.e. *AMMATOTAL*, *AMMAT*, and *AMMAR*.

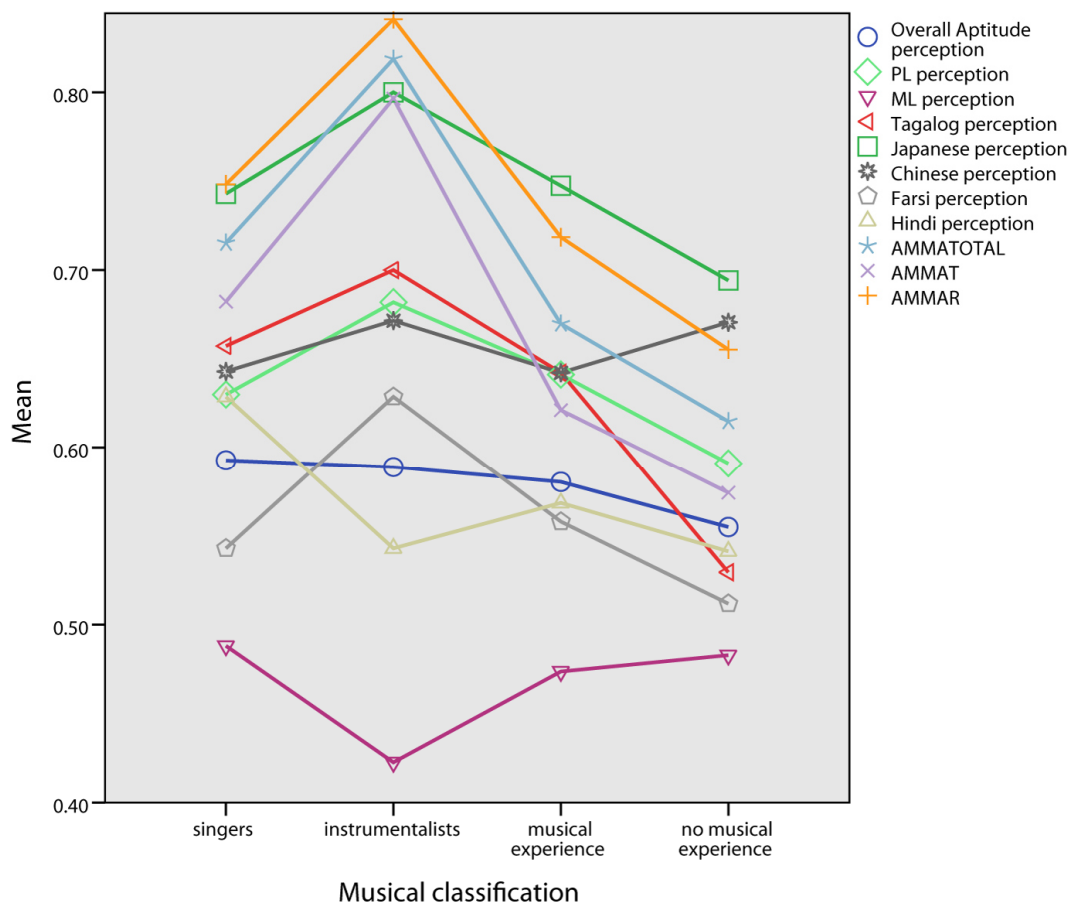
11.6.3.5.Instrumentalists versus the non-musicians

Group comparisons between instrumentalists and amateurs show that there is no significant difference for all speech perception tasks. However, instrumentalists also outperformed the non-musicians on all music perception tasks, i.e., *AMMATOTAL*, *AMMAT*, and *AMMAR*.

11.6.3.6.Amateurs versus non-musicians

Group comparisons revealed again no significant difference between amateurs and non-musicians on the speech perception tasks. However, amateurs also outperformed non-musicians on all music perception tasks, i.e. *AMMATOTAL*, *AMMAT*, and *AMMAR*.

Figure 16. Mean differences



11.7. Instrumentalists and vocalists

In previous research, it has continually been noted that singing ability is a better predictor for speech imitation ability than musical perception measurements (Christiner & Reiterer 2013; Christiner 2013). Consequently, it could be assumed that vocalists and instrumentalists who train different abilities could be differentiated based on their imitation ability as well. As singing requires vocal motor training, it could be suggested that they outperform instrumentalists who train other abilities. Indeed, this was found in previous research where vocalists outperformed instrumentalists and non-musicians on speech imitation tasks (Christiner & Reiterer 2015). Interestingly, singers and instrumentalists could not be differentiated based on their musical perception ability as measured by the AMMA test. However, instrumentalists also outperformed non-musicians on the imitation tasks. Consequently, it has been concluded that adaptive plasticity for speech imitation and phonetic aptitude rely on both improved perception and production ability (Christiner & Reiterer 2015).

In this investigation similar and distinct findings have been made. The difference between singers and instrumentalists could not be replicated. There was no significant difference between the two groups across all eight language imitation tasks. This could also be the case since there were fewer singers and instrumentalists included than in previous investigations, in which the number of participants was considerably higher and the analysis more robust. Despite this, the basic results and conclusions are not entirely different. Both singers and instrumentalists outperformed non-musicians on seven of eight language imitation tasks. This suggests that, on average, trained musicians, both singers and instrumentalists, are better at imitating unfamiliar language material than non-musicians.

In previous investigations, amateurs have not been included as a fourth category. In this investigation vocalists and instrumentalists performed better in three language imitation tasks compared to amateurs. The vocalists were better at imitating Japanese, Tagalog, and Thai and the instrumentalists were better at imitating Russian, Japanese, and Tagalog, compared to amateurs in this investigation. The latter, however, only performed significantly better in one task than the non-musicians did. Importantly, a trend could be observed for all musical categories, namely that singers and instrumentalists performed best in the language imitation tasks, followed by the amateurs and the non-musicians. For illustration, table 55 shows the median and mean of the group performances. The findings suggest that musical aptitude, singing ability and musical training all improve phonetic aptitude as measured by imitation tasks.

In marked contrast, for the speech perception abilities no significant differences were observed among all groups, while the opposite was found for the musical aptitude measurements: *AMMATOTAL*, *AMMAT* and *AMMAR*. Results show that instrumentalists performed significantly better than the singers did on two of the *AMMA* measurements (*AMMATOTAL* and *AMMAR*). No significant difference was observed among singers and the amateurs. The vocalists, instrumentalists and amateurs all performed significantly better than the non-musicians in all *AMMA* measurements. For illustration, table 55 shows the median and mean of the group performances. This suggests that instrumentalists were most proficient on the musical aptitude measurements, followed by singers, amateurs and non-musicians.

Taken together, the findings suggest that music perception ability as well as singing ability seem to be indicators that can improve speech imitation abilities. Vocalists and instrumentalists are trained musicians whose training and abilities are quite diverse. Instrumentalists train a new ability, while singers have to improve their vocal motor apparatus. Both groups, however, have to train musical perception. The results suggest that

phonetic aptitude is facilitated by both enhanced music perception and vocal motor ability. Thus, it is worth examining probable reasons as to why vocalists and instrumentalists seem to be better endowed for speech imitation abilities.

11.7.1. Vocalists and phonetic aptitude

Vocalists have shown to be better than amateurs and no-musicians in phonetic aptitude as measured by language imitation tasks. This has been observed and replicated several times (Christiner & Reiterer 2013; Christiner & Reiterer 2019; Christiner et al. 2018). A refined oromotor system such as that of singers may also positively influence language functions, i.e. the imitation of unfamiliar languages. Evidence for this has been provided from multiple angles such as singing experts, neuroscience and psychology.

Voice experts stress the relationships between singing and speaking and claim that both have much in common, as they are generated by the same vocal apparatus (García-López & Gavilán Bouzas 2010), which is also a reason why many singing experts started comparing singing and speaking processes (Sundberg 1977, 1999; Roberts & Davies 1975; Rutkowski 1996; Rinta & Welch 2009; Rinta & Welch 2008; McClellan 2011; García-López & Gavilán Bouzas 2010; Christiner & Reiterer 2013, 2015; Christiner et al. 2018; Christiner & Reiterer 2018). In this respect, contrary to language, production processes are in the foreground in singing research. Voice experts analyse the impact of vocal motor commands, i.e. laryngeal positions (McClellan 2011; Sundberg 1999; Ekholm et al. 1998). For instance, vocal timbre and vowel quality can be altered by the lowering and raising of the larynx (Ekholm et al. 1998; Sundberg 1999). Vocal techniques for singers include multiple other generational instructions such as glottis closure, tongue coordination, breathing techniques, body posture and intrinsic as well as extrinsic muscle control (García-López & Gavilán Bouzas 2010). Singing training focuses on refining the interplay between the individual components of the vocal organ, the respiratory system, the vocal folds and the vocal tract (Sundberg 1999: 172) with which vocalisation is realised.

Overlaps between singing and speaking have also been detected in neuroscientific investigations. Research has shown that the neural network of song and speech production is largely shared. Bilateral activation in both speech and song production has been observed in the superior temporal gyrus, the superior temporal sulcus and in the inferior precentral and postcentral gyrus (Ozdemir et al. 2006). Neuroscientific research on classical singing training has also concluded that especially kinaesthetic ability improves (Kleber et al. 2010) and intensive vocal motor training seems to leave a deep imprint in the brain structure in general.

Vocal motor training leads to “an increase in volume and microstructural complexity of specific white-matter tracts connecting regions that are fundamental to sound perception, production, and its feedforward and feedback control [..]” (Halwani et al. 2011: 1). Interestingly, this effect has not been reported for general musical abilities, which suggests that they are specific to vocal motor training. The three studies share that the process of production seems to be mainly involved to explain these findings. Vocal expertise and the ability to alter and modify vocalisation may be responsible for a higher ability to imitate unfamiliar speech material.

11.7.2. Instrumentalists and phonetic aptitude

Instrumentalists, like singers, were better at imitating unfamiliar speech material. Tendencies have even shown that they were, on average, slightly better than the vocalists. The instrumentalists, however, were also good singers, even though they performed less well than the vocalists did. Instrumentalists also performed significantly better in all singing tasks compared to the amateurs and the non-musicians. This shows that previous research is in concordance with this investigation (Christiner & Reiterer 2015). Singing ability turned out to be one of the most important predictors for several language imitation tasks, which has been represented in the regression models (section 11.3).

Like singers, instrumentalists possess high imitation abilities which could also be related to enhanced music perception ability, which facilitate language production processes as well. The instrumentalists were also significantly better in the *AMMATOTAL* and its rhythmic subtest compared to the singers and to all other groups. This suggests that the instrumentalists of this investigation were also in the possession of enhanced perceptual musical aptitude.

Interdisciplinary research on music and speech found that the processing of both shows large overlaps (Peretz et al. 2015; Wong et al. 2007; Rogalsky et al. 2011; Patel 2011; Ding et al. 2017; Asaridou & McQueen 2013; Perrachione et al. 2013) that result in positive transfer from music to language (Christiner & Reiterer 2013, 2015; Christiner et al. 2018; Hu et al. 2013; Reiterer et al. 2011; Milovanov 2009; Milovanov & Tervaniemi 2011). More specifically, enhanced musical perception has also been related to the perception of speech (Chobert et al. 2014; François et al. 2013; Besson et al. 2011; Moreno et al. 2009; Wong & Perrachione 2007) and the capacity of verbal memory (Ho et al. 2003; Moreno et al. 2011). Additionally, musical perception aptitude has also an effect on generational processes such as

speech imitation. The key role may be found in the tonal perception parameter which also contributed to three of the individual speech imitation tasks as shown in section 11.3.

Interestingly, no group differences could be detected for the speech perception ability, contrary to the music perception. This is an unexpected finding since at least it has been suggested that musicians perform better than the less experienced musicians in the language perception tasks. In chapter 9, the opposite was found. Participants who played a musical instrument outperformed non-musicians in all main speech perception tasks: the *Overall Aptitude perception*, the *PL perception* and the *ML perception* (see 9.10.3). However, a trend can be observed for the main variables, the *Overall Aptitude perception* and the *PL perception* conditions, in which singers and instrumentalists perform slightly better than the amateurs and the non-musicians – though this difference is non-significant. As the groups are considerably smaller, the results could differ on larger samples.

In addition to the above shown results, further relationships between perception ability of speech and music were presented in previous chapters. The *AMMATOTAL* and the *PL conditions* show correlations as illustrated in table 51. However, further evidence for a relationship between *Farsi* and *Tagalog perception*, the two sub-scores of the *PL conditions*, and *AMMAT* and *AMMAR* has been provided in the factorial design (see table 32), in which they all load onto the same factor. This indicates that even though the relationship between music and speech perception is not as strong as for the production variables speech imitation and singing, they are interrelated on several dimensions.

11.7.3. Performance anxiety

Performance anxiety is associated with a very unpleasant feeling that can occur in many situations, especially if performances are evaluated (Cheng et al. 2009). Even though the majority of studies focus on negative effects of performance anxiety, it has also been shown that anxiety is potentially an adaptive system which can influence performances positively (Cheng et al. 2009). The latter, however, is not part of this investigation as the difference of anxiety levels between musicians and non-musicians should be related to the language and music performances.

Studies on musical abilities that include performances, such as singing tasks or the playing of short musical pieces, astonishingly seldom include anxiety measurements. This is also true if the test design includes group comparisons between musicians and non-musicians. Research has noted that anxiety affects individual differences in music and language performances. Additionally, studies have revealed that more advanced musicians will also

perform remarkably better under conditions of increased anxiety (Hamann & Sobaje 1983; Kokotsaki & Davidson 2003). This suggests that individual musical performances are influenced by musical expertise as was discussed in chapter 6. In this investigation a questionnaire has been developed that aims to reveal how anxious the participants are generally while performing as well as within this investigation, see section 7.7.2.

Several statistical analyses were performed which aimed at determining whether *Performance anxiety* influences the results of the language and music performances or not. First of all, *Performance anxiety* correlates with all language imitation tasks apart from Hindi. This is also the reason why *Performance anxiety* has been included in the regression analysis for all other language imitation tasks apart from Hindi (see section 11.3). Results showed that *Performance anxiety* appeared only once as a predictor that could explain the variance of Tagalog performances in the regression model, as illustrated in section 11.3.

Secondly, group comparisons were undertaken to understand whether musical expertise or proficiency has an effect on *Performance anxiety* in this sample. As expected, the instrumentalists and vocalists were less anxious, in comparisons to all other groups. No significant difference between instrumentalists and singers was observed. Compared to the vocalists, amateurs and non-musicians were significantly more anxious, which demonstrates that both groups more likely face unpleasant feelings while performing language or music tasks.

Group comparisons between instrumentalists, amateurs and non-musicians show a significant difference between the instrumentalist group and the non-musicians only. Generally speaking, the group comparisons revealed that singers and instrumentalists are those who are less anxious compared to the other groups (for clarification, the higher the score, the less anxious the participants). This observation, however, does not allow general conclusions to be made about the influence of performance anxiety to other variables. Presumably, as *Performance anxiety* only predicted the variance in the *Tagalog imitation* task, its influence seems to be rather low for the other languages.

In chapter 6, it has been argued that the impact of anxiety may also depend on the measurements or tasks participants have to fulfil. This means that in some instances integrating concepts that measure anxiety may be more or less important. Being more precise, whereas direct oral performances such as the imitation of new utterances or singing may be more influenced by anxiety, perceptual measurements, which are often executed with headphones alone, may be less affected by unpleasant feelings. For instance, the relationship

between speech imitation and singing ability could be very strong because singers are less anxious in both vocal behaviours.

Partial correlations make it possible to observe as to whether a linear relationship still exists if it is controlled by another variable which could influence their relationship. The relationships of speech imitation to the *AMMA* test, the two different singing tasks and the *STM* have been investigated.

The original Spearman correlations for the most prominent variables are as follows: The *Speech imitation ability* is significantly correlated with the *New song total* $r_s = 0.52$, p (two-tailed) < 0.01 , the *Happy Birthday total*, $r_s = 0.44$, p (two-tailed) < 0.01 , the *AMMATOTAL* $r_s = 0.35$, p (two-tailed) < 0.01 and the *STMT* $r_s = 0.53$, p (two-tailed) < 0.01 , as shown in table 51. The partial correlations of the same variables are described below and reveal that when *Performance anxiety* on the relationship between *Speech imitation total* and *New song total* is controlled $r = 0.39$, p (two-tailed) < 0.01 . When *Performance anxiety* on the relationship between *Speech imitation total* and *Happy Birthday total* is controlled $r = 0.34$, p (two-tailed) < 0.01 . This shows that their relationship diminishes but is still significant.

In marked contrast, when *Performance anxiety* on the relationship between *Speech imitation total* and the *AMMATOTAL* is controlled $r = 0.37$, p (two-tailed) < 0.01 . When *Performance anxiety* on the relationship between *Speech imitation total* and the *STMT* is controlled $r = 0.53$, p (two-tailed) < 0.01 . This shows that their relationship stays the same.

Consequently, *Performance anxiety* has an effect on the relationship between singing ability and speech imitation. The correlation coefficients of both singing variables are reduced and their relationship to the speech imitation ability slightly diminishes. Their relationship, however, is still highly significant. This is a crucial finding as it suggests that the relationship between singing ability and speech imitation is only slightly influenced by anxiety. In marked contrast, the same effect has not been observed for the relationship between *Speech imitation total* and the *AMMATOTAL* measurement, as well as for the phonological *STM*

As has been suggested, oral performances such as the imitation of new utterances or singing are influenced by anxiety, while perceptual measurements may be less affected by unpleasant feelings. The findings of this investigation support this notion. However, even though the results have revealed that *Performance anxiety* influences the relationship between singing ability and speech imitation, their linear relationship still remains significant. This additionally supports

11.7.4. Singing behaviour during childhood

Research on singing acquisition processes are underrepresented in the literature (Tsang et al. 2011), and how they relate to other cognitive domains has largely been neglected so far. Researchers have claimed that shared and distinct processes of language and singing acquisition, as well as the question as to whether there may be sensitive periods in which singing ability develops in a better and more elaborate way, can reveal interesting information which may also give insights to understand other domains (Cohen 2008). For instance, the acquisition of music, language and singing all seem to have sensitive periods in which acquisition develops faster.

Taking languages into account, it has been concluded that the earlier a foreign language is acquired the better the proficiency. Additionally, it has largely been accepted that acquiring the ability to play a musical instrument early in life is also related to achievement. It is largely unknown whether this is also true for singing or whether singing during childhood influences other cognitive domains.

For analysing the singing behaviour of the participants during childhood, a multi-item scale concept was developed that aims to reveal how frequently the participants sang during childhood (see section 7.7.1). Singing behaviour during childhood correlated with six of the eight different language imitation tasks with both singing tasks but not with any of the *AMMA* measurements (see tables 31, 33, 35, 37, 39, 41, 43, 45). *Singing behaviour during childhood*, however, did not predict any variance of the speech imitation tasks in regression models (see section 11.3). This shows that the variances of speech imitation cannot be explained by the *Singing behaviour during childhood*.

Group comparisons of *Singing behaviour during childhood* have revealed crucial aspects about singers and instrumentalists. There is no difference in singing behaviour among singers and instrumentalists. However, the professional musicians sang significantly more often during childhood than both amateurs and the non-musicians in this investigation. This is of course an interesting finding and could imply that becoming a musician may be related to someone's singing behaviour during childhood. Interestingly, the absence of any relationship to the *AMMA* measurements as indicated by the correlations in table 51 is surprising. Their difference could be seen as follows: While the *AMMA* test is a musical aptitude measurement, *Singing behaviour during childhood* maybe a more training related musical measurement. Both may be interesting ways of understanding training effects versus aptitude and should be investigated more intensively in the future.

Furthermore, *Singing behaviour during childhood* also correlated with most of the language imitation tasks. This shows that, even though singing in childhood was not able to predict language imitation tasks, it should be considered for further research. This investigation is one of the first which includes this parameter in phonetic aptitude research. Some limitations still need to be mentioned. Although the multi-item concept showed high reliability, a single concept may not be enough to reveal the high potential which this research field may offer. More complex measurements for *Singing behaviour during childhood* or longitudinal studies are required to give a multidimensional picture.

11.8. Discussion

As has been illustrated in the group comparisons, several findings have to be discussed. First of all, both instrumentalists and vocalists largely outperformed non-musicians on language imitation tasks. This has also been detected in previous research (Christiner & Reiterer 2015). One dissimilarity concerns the musician group: there was no significant difference among the vocalists and instrumentalists in the language imitation tasks. Despite this, the findings are similar to former investigations and suggest that both vocal motor training as targeted at singers, as well as musical perception ability, enhance speech imitation ability.

Contrary to this, speech perception ability showed completely different results as no significant difference was found among any of the groups. This result is different to that in chapter 9, where the participants who played a musical instrument outperformed the non-musicians. However, it has to be considered that the number of participants in each group was considerably lower in the empirical investigation. In addition, a trend could be observed which showed that both singers and instrumentalists were better in the *Overall Aptitude perception* and the *PL perception* conditions, even though it was statistically non-significant.

Group comparisons between *Performance anxiety* and *Singing behaviour during childhood* have also been carried out. They produce two interesting results. First, the professional musicians (vocalists and instrumentalists) were less anxious than the other groups. Second, this suggests that anxiety levels have to be measured especially if oral performances are included in the research design. Even though the effect of anxiety is quite low in this investigation, it has to be inspected more closely in future research.

Singing behaviour during childhood also showed crucial results. Even though it was less predictive for speech imitation ability, vocalists and instrumentalists sang more frequently than amateurs and non-musicians during their childhood. In addition, the musical

aptitude measurements *AMMAR* and *AMMAT* do not correlate with this concept (see section 11.3), which reveals a demand for further investigations that demonstrate whether both are interrelated or not.

12. Final discussion

12.1. Aptitude tests

Language aptitude components are rather diverse and include various measurements that focus on memory, new and artificial language learning, applying grammar rules, language perception and production tasks, learning processes, visual cues and sound relationships as discussed in chapter 2.3.

According to this research, two different aptitude components should be added to existing language aptitude batteries, phonetic imitation tasks which measure imitation ability and musical measurements which contribute to language aptitude. Regarding the latter with the exception of very few language aptitude batteries which have included pitch discrimination (Pimsleur et al. 1962) and pitch perception ability (Nakamura 1988), musical measurement as a concept of language aptitude has largely been neglected. This is surprising since research has increasingly shown that language and musical aptitude are highly intertwined. The second parameter concerns imitation ability; this tests vocal motor skills, which are less dependent on typology.

12.1.1. Speech imitation as a language aptitude component

The speech imitation tasks of the eight individual languages were highly interrelated since they show high loads onto the same factor (see table 32). This suggests that language imitation tasks measure a phonetic imitation ability which is less dependent on language typology. Thus, it shows that the original Hindi test developed by Dogil and Reiterer (2009), is a very robust measurement which reveals the potential to imitate unfamiliar phonetic language material.

It has also been shown that the imitation tasks all load onto the same factor as the *STMF* and *STMB* measurements (see table 32), which suggests that imitation tasks also represent STM measurements that test verbal spans and capacities in general. Concerning aptitude tests, it shows that unintelligible speech imitation tasks should be integrated into language aptitude testing. However, it may be better to refer to this concept as phonetic

imitation aptitude, which is a measurement that also includes the testing of vocal motor abilities. In addition, speech imitation tasks could also be used as verbal STM measurements.

12.1.2. Singing and music perception ability as a component of language aptitude batteries

Musical measurements should also be components of language aptitude batteries. Foreign language capacity is associated with oral proficiency, which is one of the most important aspects that facilitates adequate communication (Dalton-Puffer et al. 1997). Native-like pronunciation skills, therefore, are crucial aspects that are associated with language aptitude. As could be demonstrated, singing ability and music perception ability contribute to imitating unfamiliar new foreign speech material and presumably to the ability to pronounce foreign languages. In this respect the main analysis has revealed that mainly singing, the two singing sub-components rhythm and melody, and the perceptual tonal discrimination ability as measured by the *AMMA* test have been related to speech imitation ability (see section 11.4).

Aptitude for music perception is also related to aptitude in speech perception. In the factorial design (see table 32), music and unfamiliar speech perception ability (*Farsi* and *Tagalog perception*) were components of one factor. Additionally, the miscellaneous results between the relationship of tonal perception ability and some of the individual speech imitation tasks (see section 11.4) more likely suggest that two different musical aptitude factors should be included, one for singing (music production), and a second one for music perception. The two musical aptitude factors could especially be relevant for revealing phonetic aptitude.

The findings of this investigation demand the demonstration of possible musical aptitude components which should be integrated in language aptitude batteries. As not all singing criteria illustrated in section 4.7 have been included, it is difficult to provide a complete concept. For singing ability, evaluation criteria (melody, rhythm and voice quality) of this investigation could be used, including measurements of vocal range as well as further singing tasks with and without lyrics. Apart from those, singing behaviour during childhood (an improved more complex version as used in this investigation), singing background and concepts that provides information about negative and positive effects of performance anxiety of individuals should be included.

The perceptual musical aptitude component could be quite diverse. In this investigation, only rhythmic and tonal musical aptitude tests have been used, but in the future, research could include measurements of consonance, dissonance and tonal memory. However,

measurements of elementary auditory skills such as loudness, tone length, timbre, overtone and fundamental tone preferences could also be included, to name but a few.

12.2. Predictors of phonetic imitation aptitude

This investigation has shown that 10 different predictors were able to explain a certain percentage of the variance of language imitation ability in eight different languages (see section 11.4). These are: *STMF*, *New song rhythm*, *STMB*, *AMMAT*, *New song melody*, *Familiar*, *Memorable*, *Difficult to imitate*, *Happy Birthday melody* and *Performance anxiety*. Certain predictors occur more often, such as *STMF*, *New song rhythm*, *STMB* and *AMMAT*.

Thus it can be concluded that based on the results, the phonological STM both forward and backward, singing ability (rhythm and melody) and the ability to discriminate tonal differences (*AMMAT*) are the main predictors of the individual eight language imitation tasks. Those essentially represent the main predictors which were also found in past research (Christiner & Reiterer 2013; Christiner et al. 2018). There is just little evidence that predictors vary depending on language typology for tone languages.

One of the phonological STM predictors, *STMF* or *STMB* is always included in regression models, with the exception of *Tagalog imitation* and is a general predictor. STM predictors and their relationship to languages have been discussed in multiple investigations (Sagarra & Herschensohn 2010; Robinson 2005; Gathercole et al. 1997; Gathercole 2006; Speciale et al. 2004; Mackey & Sachs 2012; Wen & Skehan 2011; Wen et al. 2019; Korecky-Kröll et al. 2019). In addition, there is the *New song rhythm* criterion which predicts both tone languages. Singing as a composite score (consisting of tonal, rhythm and quality of voice measurements) has also been related to *Mandarin imitation* in children in former investigations (Christiner et al. 2018). This result shows that future research should continue examining the relationship of singing ability and tone languages. However, it should also be noted that singing predictors also explained the variances of the performances in a number of other non-tone languages (see section 11.4). Therefore, this singing criterion seems to be a more general predictor as well.

Further evidence that phonetic imitation tasks are less dependent on typology comes from the factor analysis (see table 32), which has revealed that it may be more plausible that the imitation tasks of all languages measure a very similar concept. Thus it can be suggested that the language imitation tasks as used in this investigation represent a general phonetic imitation ability. Therefore, it is more likely that the main reoccurring predictors of the eight languages are all relevant for speech imitation ability (see table 49). The differences in the

individual languages could additionally be influenced by factors such as participants selection, individual differences or environmental influences.

This is also shown in the results of the entire *Speech imitation total* measurement. The Number of foreign languages contributes to the variances of the performances of all languages only in the composite condition. This finding is not surprising. Research has shown that the *Number of foreign languages* contributes to the learning of new foreign languages in general. The difference between the results of the *Speech imitation total* and the individual eight languages can thus be found in the test design which favours the *Number of foreign languages* instead of predictors which occur in single language imitation tasks. Mastering multiple foreign languages goes alongside knowing larger numbers of phonological structures, which facilitates the ability to imitate multiple languages. This highlights the necessity of understanding individual differences as a multidimensional construction.

To recap, all eight languages comprised in the *Speech imitation total* performance have revealed four predictors, the *Number of foreign languages*, *STMT*, *New song total* and *All Languages memorable*, which account for 60 percent of the variances in the performances in this sample.

The first predictor, the *Number of foreign languages*, has already been discussed and indicates that the larger the repertoire of phonological structures available, the better the performances on various different language imitation performance tasks. This combined with improved phonological STM seems to improve foreign language performance (Hell & Mahn 1997). Referring to the entire WM complex in which the STM is considered to be a subcomponent, it has been suggested that the WM complex could replace the entire language aptitude construction (see section 5.3). In this respect it is very difficult to comment on this for two reasons. First, this investigation includes STM measurements only and not complex WM measurements. Second, as WM and STM terminology are used interchangeably in the literature (see chapter 5), even though they measure different cognitive processes, it is difficult to compare research results in general. As has been replicated in this investigation, the phonological STM is one of the most important predictors for phonetic imitation ability.

Singing ability has also turned out to be an important criterion which is related to phonetic imitation aptitude. This may be the case as both depend on motor commands, and the better the singing ability, the better the imitation ability.

In addition, how memorable a language occurs to individuals also contributes to the variance of their performance, which shows that self-awareness of one's own abilities directly affects performance. Individuals tend to be aware of their own imitation ability, which could

support L2 choices and may be a promising self-estimation measurement and tool that could be used as an easily and reliably executable measurement and concept in educational context.

The speech perception measurements all failed to explain any variance of the speech imitation tasks (see section 11.4). Considering the limitations of the research, the speech perception tasks of the individual languages consisted of five items and those of the speech imitation tasks of four items for each of the five languages (Tagalog, Hindi, Japanese, Mandarin and Farsi) which have been compared directly. Therefore, it could be suggested that the low number of items of each task could have led to this dissociation.

As a control function, the *Speech imitation total* (comprised of eight languages) and the three speech perception scores, *Overall Aptitude perception*, *ML perception*, and *PL perception* were created and compared to all main other variables of this investigation (see tables 51 and 52). Apart from a low correlation between *Speech imitation total* and *PL perception* (see table 51) no further relationship between speech perception and production were detected. This indicates that perception and production can largely be dissociated on the basis of the measurements used in this research.

12.3. Dissociation between perception and production

Results of previous investigations in which vocalists performed significantly better than instrumentalists in the imitation of unintelligible speech material could not be replicated (Christiner & Reiterer 2015). This means that no significant differences were found between vocalists and instrumentalists in their ability to imitate speech in eight languages. Both groups, however, performed significantly better on seven of eight language imitation tasks compared to the non-musicians, and on three of eight compared to amateurs. This suggests that instrument playing and singing both facilitate phonetic imitation ability (see section 11.6.2).

In chapter 9, participants who play musical instruments outperformed non-musicians in speech perception ability. In section 11.6, no significant difference was observed in group comparisons between instrumentalists, vocalists, amateurs and non-musicians, even though a trend has been identified for vocalists and instrumentalists in this investigation. In both groups the *Overall Aptitude perception* and the *PL perception* median was higher (see table 56). In general, the findings suggest that musical training facilitates speech perception processes.

Evidence for a relationship between unintelligible speech perception and music perception aptitude may also be shown in the factorial analysis (see table 32) and the correlational analysis (see table 51). In the former, tonal and rhythmic musical perception

loaded onto the same factor as *Farsi* and *Tagalog perception*. In the latter, the *PL perception* score correlated with the *AMMATOTAL* score. However, it should be noted that this of course shows a limitation of the research which requires investigation in future research.

As expected, group differences among professionals, amateurs and non-musicians in the musicality measurements have been found (see tables 57 and 58). *AMMAT*, *AMMAR* and *AMMATOTAL* performances were significantly better for singers and instrumentalists compared to the musical experience group and the non-musicians. Instrumentalists, however, were also significantly better than the singers in the *AMMAR* and *AMMATOTAL*. This indicates that the professional music group were better at discriminating incongruities in musical statements.

The investigation facilitates the idea that perception and production for both music and speech can be dissociated. Speech perception and imitation of unintelligible speech seem to be largely independent (see tables 32, 51, and 52). This resembles findings of other researchers. Foreign speech perception and production can be dissociated on a behavioural basis as well as partially on an anatomical one (Golestani & Pallier 2007). Consequently, the authors of the study concluded that individuals who perceive foreign speech sounds faster do not necessarily pronounce them better. This finding has been replicated in the present study, which demonstrates that the imitation ability is not related to that of the perceptual counterpart.

Concerning music perception and production (singing), the findings are similar. The factorial design has shown that music perception ability (tonal and rhythm) loads onto a different factor than singing. This more likely supports the idea that music perception and production can also be dissociated. Following previous research, investigations on singing ability and pitch perception produced contradictory results (Estis et al. 2009). While poor and inaccurate singing ability has often been related to pitch perception deficits (Sloboda et al. 2005; Dalla Bella & Berkowska 2009; Hutchins et al. 2014; Tremblay-Champoux et al. 2010), evidence has also been provided that pitch production such as in singing and pitch perception can largely be dissociated (Loui et al. 2009). This has been replicated in this investigation. The dissociation between perception and production also requires that common notions about learning and acquisition processes in music and language have to be reframed and rethought.

In general, it is largely accepted that perception precedes the production of vocal output, which suggests that there must be a relationship between both which often cannot be replicated. The results of this investigation, however, would best fit in what has been

described as a dual route model. The *Linked Dual Representation Model* proposes “vocal-motor representation can be mediated by the symbolic representation” (Hutchins & Moreno 2013: 8). In other words, vocal processing can be processed in two different ways. First, it can be encoded as a kind of symbolic representation which is associated with conscious perception, and second, it is also possible to encode motoric representations directly (Hutchins & Moreno 2013: 7). Additionally, it has been suggested that “the degree to which a tone is initially encoded symbolically or motorically is influenced by the intention of the listener” (Hutchins & Moreno 2013: 7). This would also explain why vocalists outperformed instrumentalists in speech imitation ability in previous investigations as it is very plausible that singers prefer vocal-motor coding while imitating new languages. This would entail that specific groups may prefer one of the two options.

As the predictors for the individual languages of speech imitation tasks are rather similar, despite the great differences between these languages, only ten individual predictors have been found for eight languages (see section 11.4). Furthermore, the factor analysis that the language imitation tasks are highly interrelated (see section 11.2.2). They represent a more general phonetic imitation ability. Sometimes the *AMMAT* or singing parameters predicted the variances of the performances (see 11.4). This possibly depends on how the languages are perceived and indicate that the motor or the conscious perceptive approach is flexible and depends on the participants’ choices. This would also indicate that both ways are possible but depend on multiple factors.

The developers of the Hi-LAB aptitude battery also claim that high level attainment in second and foreign language proficiency largely relies on cognitive and perceptual abilities (Linck et al. 2013). The rational idea of this investigation supports this notion as especially the perceptual measurements of the individual languages were less connected. This assumes that it is more likely that speech perception varies more than the imitation tasks, which indeed is the case (see section 11.2.3). However, this is also a paradox. Language aptitude is associated with oral language performance, the production part, which is based on the fact that perceptual benefits of gifted learners are transferred to speech production. If perception and production of speech can be dissociated, it is essential to understand under which conditions a transfer between both takes place. This should be addressed in future research.

12.4. Performance anxiety and singing behaviour during childhood

The two predictors *Singing behaviour during childhood* and *Performance anxiety* were also included as measurements. Even though the former could not explain any variances

of the speech performances (see section 11.4), group comparisons revealed that musicians tend to sing more often from early childhood on (see section 11.7.4). *Performance anxiety*, on the other hand, is reduced in singers and instrumentalists, which shows that similar concepts that provide information about anxiety levels should be integrated into the research design if group comparisons are undertaken between musicians and non-musicians. Even though the effect of anxiety is not large in this investigation, nonetheless it seems to play a role. If the relationship between *Speech imitation total* and both singing tasks is controlled for *Performance anxiety*, the relationship diminishes slightly (see 11.7.3). This effect has not been found for the perceptual musicality tests, which suggests that those are less affected by anxiety conditions. Furthermore, *Performance anxiety* also turned out to be a predictor which was involved in explaining the variances of *Tagalog imitation*, which at least suggests that it has played a role for one of the eight individual language imitation tasks.

Consequently, both concepts should be improved and integrated in future aptitude research in music and language. For *Singing behaviour during childhood* this thesis should be a starting point for further investigations that work on singing acquisition processes and their relationship to language capacities.

13. Outlook

13.1. Early-acquired effects, training related influence and innate predispositions

In chapters 2 and 3 the origin of aptitude was discussed including various assumptions about nature and nurture related individual differences. Research postulates that in this respect two relevant distinctions should be regarded. First of all, understanding early-acquired capacities which are differentiated from innate predispositions are essential. This distinction would be beneficial to aptitude research and would reveal abilities which are free of training related effects. In this respect, however, a severe problem is that early-acquired capacities and training related effects cannot be distinguished precisely and have to be discussed separately as well.

Currently, it can be suggested that individual differences in capacities cannot be explained by environmental influence alone, but evidence from brain research and genetic studies show that innate predispositions have an effect on capacities on multiple domains. For instance, this has been found for elaborate speech perception and production (Golestani et al. 2011; Golestani & Zatorre 2009), as well as for excellent musical abilities (Pulli et al. 2008; Mosing et al. 2014; Butkovic et al. 2015). Therefore, it can be assumed that a composition of

innate, early acquired and training related effects can explain achievement or mastery in performances. Future directions should find methods in which innate, early acquired and training related effects can be distinguished.

Additionally, it has to be identified when acquisition processes start, or whether a non-acquisition situation even exists. Musical and linguistic features are also acquired prenatally (McMullen & Saffran 2004), which indicates that it is very difficult to draw a line between what has been acquired and what has been given by nature. In this respect, single disciplines alone cannot provide answers which enable a distinction to be made between innate capacities and early-acquired ones. As shown, there is a serious demand for interdisciplinary research, which should address the aforementioned issues in more detail.

13.2. Language aptitude perception and production measurements

The results of this investigation also show that the two language aptitude measurements (perception and production) should be improved. First of all, the language perception measurements included pure and mixed strings of languages (see 9.3). The *ML perception* conditions were less informative in revealing individual differences and also showed fewer correlations (tables 27 and 51) than the *PL perception* to other variables, which suggests that this condition could be removed. Instead, further pure language conditions should be added.

The speech imitation tasks of the eight languages could also be improved in several ways. *Tagalog imitation* has unexpectedly shown to be very specific compared to the other seven language imitation tasks. Evidence is mounting that Tagalog is perceived to be more musical or song-like than any of the other languages. This assumption is based on the three main analyses run in chapter 11. First, *Tagalog imitation* was the only speech imitation task which also loads onto the singing factor (see section 11.2.1). This could mean that Tagalog measures similar aspects to the ones that singing tasks do. Second, the variance in Tagalog performances could partly be predicted by *Performance anxiety* (see table 40). The latter also loads onto the singing factor (see section 11.2.1). This illustrates that *Performance anxiety* has an effect on *Tagalog imitation* and singing. Third, group comparisons between professional musicians (vocalists and instrumentalists), amateurs and non-musicians revealed that performance anxiety is significantly higher in non-musicians and amateurs (see tables 54 and 55).

Taken together, it may be the case that *Tagalog imitation* has more likely been associated with song-like parameters. This could be the reason why Tagalog deviates from all

other languages included in this investigation. Even though this assumption cannot be proved within this study, this finding is very crucial for future research. Speech imitation tasks should also be rated for how melodic they occur to the participants. This would provide information about as to which languages are perceived as being more musical. In a wider sense, this could be a novel way to reveal ID in phonetic aptitude and its relationship to musical abilities.

As has been discussed in the sections 11.3 and 11.5, even though the results of this investigation suggest that there may be a common imitation ability which is less dependent on language typology, future research should retest this notion by selecting more specific language features. This could be achieved by developing imitation tasks which contain non-native language features, which are additionally difficult to generate.

To recap, the development of the language imitation tasks in this investigation was based on three main criteria which were all found to be relevant in previous research. They should be unintelligible to the participants, have a length between nine and eleven syllables and should be typologically different. Future research should integrate a selection of specific, for the participants, non-native language features. For instance, retroflex sounds are rather common in Hindi but not in German. Retroflex consonants require high level of vocal ability as they are generated with the tongue curled back (Christiner & Reiterer 2015).

Another option would be to develop measurements that are based on testing phonetic contrasts which are unfamiliar to the participants. For instance, retroflex versus dental stop distinctions as in Hindi are less often part of European languages (Werker & Tees 2005). In this respect, non-native contrasts would be suitable for both production and perception tests. This could reveal whether the ability to imitate typologically different languages could be derived more easily from this approach.

For tone languages, another option would be to imitate pairs of phrases which are lexically equal but include a sound change. One example of this is the Mandarin pair *qùnnálǐ* and *qùnnàlǐ*. The syllable *ná* in the former word indicates a rising tone, while in the latter *nà* represents a falling tone. This could also be used as a production and perception test.

13.3. Education and aptitude

The results of this investigation also have implications on education. In general, aptitude in the educational context has often been rejected for several reasons. Aptitude is considered to be a rather stable trait, described as a kind of initial state of readiness (Carroll 1981), which is less likely influenced by the environment (Sparks et al. 2011). In addition, the relationship between foreign language aptitude and foreign language achievement is also

unclear (Wen et al. 2017). Taken together, it is quite obvious that the idea of including aptitude included in educational programmes, which aim at improving performances through instruction, is not very attractive.

A thorough analysis, however, reveals the opposite. Aptitude is said to turn into talent by training, which implies that the educational setting is the best place in which aptitudes can be supported and transformed. Therefore, it shows that the role of aptitude in educational environment has to be promoted among teachers. Currently, they are often forced to reject the idea of integrating aptitude in their educational programmes since proper, applicable tools that enable them to understand the aptitude of individuals do not exist or are hardly ever available. This suggests that new measurements and useful tools that provide information about aptitude and ID of individuals should be developed and made accessible to teachers as well as to people in general.

13.4. Foreign language and music education

This investigation also demonstrates that singing and musical activities represent general tools which have to be re-installed in educational settings more frequently. Musical capacity improves foreign language pronunciation skills and multiple other cognitive domains which are useful in the context of learning. Music also has uniting characteristics which facilitate social bonding, in addition to its being entertaining. Recent trends in schools include singing activities in foreign language classes in primary school settings. Developing new educational programmes which focus on music and language programmes would be an important step towards introducing new ways of teaching. For instance, the emergence of a *Song, Music and Language Integrated Learning* concept in the school environment as an equivalent to *Content and Language Integrated Learning* (CLIL), in which other subjects are taught in foreign languages, would be an interesting option. *Song, Music and Language Integrated Learning* could be a concept in which foreign languages and musical abilities are taught simultaneously or within the same framework. Singing, for instance, would facilitate pronunciation skills and create awareness of the phonological structures and features of foreign languages. At the same time this educational programme could also focus on generational processes for which there is not much time in most foreign language classes.

13.5. Ethical discussion

It is important to note that aptitude research also has to hold an ethical discussion on the findings and results revealed. Possessing aptitude is a valued and desired good which has

a very high social standing. However, it is also clear that individuals possess different aptitudes which can vary tremendously. Science has the responsibility to show that the varying degree of aptitudes of individuals should be positively connotated. This requires an ethical discussion as research on aptitude is growing enormously at the present moment.

A further issue may also be the term *aptitude* itself. Individuals without impairment can learn languages or acquire the music which is typical for their culture. This shows that in all individuals a certain kind of latent potential is recognisable. It may be better to speak of “a degree of aptitude” instead of *aptitude* alone. This term would be an inclusive term, which suggests that everyone has potential and the degree is variable.

We are on the threshold to understanding positive transfer from one domain to another, which is based on individual differences that show that our abilities vary considerably. This requires foresight and careful handling of the knowledge generated.

14. Conclusion

This investigation has shown that 10 different predictors were able to explain the variance of language imitation ability in eight different languages (see section 11.4). These are: *STMF*, *New song rhythm*, *STMB*, *AMMAT*, *New song melody*, *Familiar*, *Memorable*, *Difficult to imitate*, *Happy Birthday melody* and *Performance anxiety*. Certain predictors occur more often such as *STMF*, *New song rhythm*, *STMB* and *AMMAT*. Regression analysis of the entire *Speech imitation score* which is comprised of all eight languages reveals four predictors. These are: *Number of foreign languages*, *STMT*, *New song total* and *All Languages memorable*, and account for 60 percent of the variances of the speech imitation performances (see section 11.5).

As has been demonstrated, imitation tasks may be less suitable to explain typological differences as their relationship resembles more a single concept which may best be defined as phonetic imitation aptitude. Nonetheless, tone language imitation can be predicted by singing ability as both Thai and Mandarin share the predictor *New song rhythm*, and singing ability has also been involved in explaining individual differences in performing Mandarin in previous investigations. However, it is also likely that the most frequently occurring predictors facilitate the acquisition of most languages and the reason why one predictor occurs in some languages, but not in others may lie outside the imitation tasks.

This investigation has revealed that speech imitation and speech perception of unintelligible language material can largely be dissociated. In marked contrast, singing ability

and the imitation of new speech material are highly related. Unintelligible speech perception, however, seems to be related to music perception aptitude (see section 11.2.3). Tonal ability as measured by perception tests is the only predictor for phonetic imitation aptitude which has also been involved in explaining the variances of speech imitation tasks of three out of eight particular languages (see section 11.4.1).

Finally, language aptitude measurements are diverse, but two new components should be added to language aptitude batteries: language imitation tasks which measure phonetic imitation ability, and a musical component that consists of a music production (singing) and perception part. As the journey of this thesis has shown, even though music and language can be recognised as two separate faculties, they simply cannot be understood separately.

15. References

- Alloway, Tracy P. 2007. Automated working memory assessment. London: Pearson Assessment.
- Altenmüller, Eckart. 2008. Neurology of musical performance. *Clinical medicine* 8(4). 410–413.
- Amir, Ofer, Noam Amir & Liat Kishon-Rabin. 2003. The effect of superior auditory skills on vocal accuracy. *Journal of the Acoustical Society of America* 113(2). 1102–1108.
- Anton, Ronald J. 1990. Combining singing and psychology. *Hispania* 73(4). 1166.
- Apfelstadt, Hilary. 1984. Effects of melodic perception instruction on pitch discrimination and vocal accuracy of kindergarten children. *Journal of Research in Music Education* 32(1). 15–24.
- Arbib, Michael A. 2005. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and brain sciences* 28(2). 105-24; discussion 125-67.
- Asaridou, Salomi S. & James M. McQueen. 2013. Speech and music shape the listening brain: Evidence for shared domain-general mechanisms. *Frontiers in Psychology* 4. 321.
- Ayotte, Julie, Isabelle Peretz & Krista Hyde. 2002. Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain* 125(Pt 2). 238–251.
- Baddeley, Alan D. 1992. Working memory. *Science* 255(5044). 556–559.
- Baddeley, Alan D. 2000. The episodic buffer: A new component of working memory? *Trends in cognitive sciences* 4(11). 417–423.
- Baddeley, Alan D. 2003a. Working memory and language: An overview. *Journal of Communication Disorders* 36(3). 189–208.
- Baddeley, Alan D. 2003b. Working memory: Looking back and looking forward. *Nature reviews. Neuroscience* 4(10). 829–839.
- Baddeley, Alan D. 2010. Working memory. *Current biology* 20(4). R136-40.
- Baddeley, Alan D. 2012. Working memory: Theories, models, and controversies. *Annual review of psychology* 63. 1–29.
- Baddeley, Alan D., Susan Gathercole & Costanza Papagno. 1998. The phonological loop as a language learning device. *Psychological review* 105(1). 158–173.
- Baddeley, Alan D., Graham J. Hitch & Richard J. Allen. 2009. Working memory and binding in sentence recall. *Journal of Memory and Language* 61(3). 438–456.

- Baddeley, Alan D., Vivien Lewis & Giuseppe Vallar. 1984. Exploring the Articulatory Loop. *Quarterly Journal of Experimental Psychology Section A* 36(2). 233–252.
- Balkwill, Laura-Lee & William F. Thompson. 1999. A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception* 17(1). 43–64.
- Banai, Karen & Merav Ahissar. 2013. Musical experience, auditory perception and reading-related skills in children. *PloS one* 8(9). e75876.
- Bangert, Marc, Thomas Peschel, Gottfried Schlaug, Michael Rotte, Dieter Drescher, Hermann Hinrichs, Hans-Jochen Heinze & Eckart Altenmüller. 2006. Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage* 30(3). 917–926.
- Baran-Łucarz, Małgorzata. 2016. The relationship between language anxiety and the actual and perceived levels of foreign language pronunciation. *Studies in Second Language Learning and Teaching* 1(4). 491.
- Barbo, Matjaž. 2019. Die Bedeutung in der Musik und die Musik in der Bedeutung. In Oswald Panagl (ed.), *Kratylos*, vol. 64, 169-176. Wiesbaden: Ludwig Reichert.
- Baumann, Simon, Susan Koenke, Martin Meyer, Kai Lutz & Lutz Jäncke. 2005. A network for sensory-motor integration: What happens in the auditory cortex during piano playing without acoustic feedback? *Annals of the New York Academy of Sciences* 1060. 186–188.
- Benner, Jan, Martina Wengenroth, Julia Reinhardt, Christoph Stippich, Peter Schneider & Maria Blatow. 2017. Prevalence and function of Heschl's gyrus morphotypes in musicians. *Brain structure & function* 222(8). 3587–3603.
- Benner, Uta. 2005. *Syllables in speech production: A study of the mental syllabary*. University of Stuttgart: Stuttgart.
- Bentley, Arnold. 1966. Measures of musical abilities. Windsor, England: NFER-Nelson.
- Bentley, Arnold. 1969. Some research interests and findings. *Journal of Research in Music Education* 17(1). 41.
- Berke, Gerald S. & Jennifer L. Long. 2010. Functions of the larynx and production of sounds. In *Handbook of mammalian vocalization: An integrative neuroscience approach* (Handbook of Behavioral Neuroscience 19), vol. 19, 419–426. London: Academic.
- Berkowska, Magdalena & Simone Dalla Bella. 2009. Acquired and congenital disorders of sung performance: A review. *Advances in cognitive psychology* 5. 69–83.

- Berkowska, Magdalena & Simone Dalla Bella. 2013. Uncovering phenotypes of poor-pitch singing: The Sung Performance Battery (SPB). *Frontiers in Psychology* 4. 714.
- Besson, Mireille, Julie Chobert & Céline Marie. 2011. Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology* 2. 94.
- Best, Catherine. 1995. A direct realist perspective on cross-language speech perception. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research*, 171–204. Baltimore, MD: York Press.
- Biedroń, Adriana & Mirosław Pawlak. 2016. New conceptualizations of linguistic giftedness. *Language Teaching* 49(2). 151–185.
- Billroth, Theodor. 1895. *Wer ist musikalisch?* Berlin: Gebrüder Paetel.
- Birdsong, David & Jan Vanhove. 2016. Age of second language acquisition: Critical periods and social concerns. In Elena Nicoladis & Simona Montanari (eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (Language and the human lifespan series), 163–181. Washington, DC & Berlin: American Psychological Association; Walter de Gruyter.
- Bongaerts, Theo. 1999. Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners, in second language acquisition and the critical period hypothesis. In David Birdsong (ed.), *Second language acquisition and the critical period hypothesis* (Second language acquisition research), 133–159. Mahwah, NJ: Erlbaum.
- Bradburn, Norman M. & Carrie Miles. 1979. Vague quantifiers. *Public Opinion Quarterly* 43(1). 92.
- Brown, Rachel M., Joyce L. Chen, Avrum Hollinger, Virginia B. Penhune, Caroline Palmer & Robert J. Zatorre. 2013. Repetition suppression in auditory-motor regions to pitch and temporal structure in music. *Journal of cognitive neuroscience* 25(2). 313–328.
- Brown, Steven & Joseph Jordania. 2013. Universals in the world's musics. *Psychology of Music* 41(2). 229–248.
- Burgess, Neil & Graham J. Hitch. 2006. A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language* 55(4). 627–652.
- Butkovic, Ana, Fredrik Ullén & Miriam A. Mosing. 2015. Personality related traits as predictors of music practice: Underlying environmental and genetic influences. *Personality and Individual Differences* 74. 133–138.

- Cain, Kate, Jane Oakhill & Peter Bryant. 2004. Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology* 96(1). 31–42.
- Calvo, Manuel G. & Michael W. Eysenck. 1996. Phonological working memory and reading in test anxiety. *Memory (Hove, England)* 4(3). 289–305.
- Cao, Chuan, Ming Li, Jian Liu & Yonghong Yan. 2008. A study on singing performance evaluation criteria for untrained singers. *IEEE Staff*. 1475–1478.
- Carroll, John B. 1958. A factor analysis of two foreign language aptitude batteries. *Journal of general psychology* 59(1). 3–19.
- Carroll, John B. 1981. Twenty-five years of research on foreign language aptitude. In Karl C. Diller (ed.), *Individual differences and universals in language learning aptitude*, 83–118. Rowley, MA: Newbury House.
- Carroll, John B. 1989. The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher* 18(1). 26–31.
- Carroll, John B. 1990. Cognitive abilities in foreign language aptitude: Then and now. In Thomas Parry & Charles W. Stansfield (eds.), *Language aptitude reconsidered*, 11–29. Englewood Cliffs, NJ: Prentice Hall.
- Carroll, John B. 1993. *Human cognitive abilities*. Cambridge: Cambridge University Press.
- Carroll, John B. & Stanley Sapon. 1959. *Modern Language Aptitude Test (MLAT)*. New York, NY: The Psychological Corporation.
- Case, Robbie, D. Midian Kurland & Jill Goldberg. 1982. Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology* 33(3). 386–404.
- Chen, Joyce L., Virginia B. Penhune & Robert J. Zatorre. 2008. Moving on time: Brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *Journal of cognitive neuroscience* 20(2). 226–239.
- Cheng, Wen-Nuan K., Lew Hardy & David Markland. 2009. Toward a three-dimensional conceptualization of performance anxiety: Rationale and initial measurement development. *Psychology of Sport and Exercise* 10(2). 271–278.
- Cheung, Him. 1996. Nonword span as a unique predictor of second-language vocabulary language. *Developmental Psychology* 32(5). 867–873.
- Chobert, Julie, Clément François, Jean-Luc Velay & Mireille Besson. 2014. Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive

- processing of syllabic duration and voice onset time. *Cerebral cortex* 24(4). 956–967.
- Christiner, Markus. 2013. *Singing performance and language aptitude: Behavioural study on singing performance and its relation to the pronunciation of a second language*. University of Vienna: Vienna Master Thesis.
- Christiner, Markus. 2018. Let the music speak: Examining the relationship between music and language aptitude in pre-school children. In Susanne M. Reiterer (ed.), *Exploring language aptitude: Views from psychology, the language sciences, and cognitive neuroscience* (English Language Education 16), vol. 16, 149–166. Cham: Springer Nature.
- Christiner, Markus & Susanne Reiterer. 2019. Music, song and speech. In John Truscott & Michael Sharwood Smith (eds.), *The Internal Context of Bilingual Processing* (Bilingual Processing and Acquisition 8), vol. 3, 131–156. Amsterdam: John Benjamins.
- Christiner, Markus & Susanne M. Reiterer. 2013. Song and speech: Examining the link between singing talent and speech imitation ability. *Frontiers in Psychology* 4. 874.
- Christiner, Markus & Susanne M. Reiterer. 2015. A Mozart is not a Pavarotti: Singers outperform instrumentalists on foreign accent imitation. *Frontiers in Human Neuroscience* 9. 482.
- Christiner, Markus & Susanne M. Reiterer. 2018. Early influence of musical abilities and working memory on speech imitation abilities: Study with pre-school children. *Brain Sciences* 8(9). 169.
- Christiner, Markus, Stefanie Rüdigger & Susanne M. Reiterer. 2018. Sing Chinese and tap Tagalog? Predicting individual differences in musical and phonetic aptitude using language families differing by sound-typology. *International Journal of Multilingualism* 15(4). 455–471.
- Cohen, Annabel J. 2008. Advancing interdisciplinary research in singing through a shared digital repository. *Journal of the Acoustical Society of America* 123(5). 3380.
- Cohen, Jacob. 1983. The cost of dichotomization. *Applied Psychological Measurement* 7(3). 249–253.
- Colom, Roberto, Carmen Flores-Mendoza, Ma Á. Quiroga & Jesús Privado. 2005. Working memory and general intelligence: The role of short-term storage. *Personality and Individual Differences* 39(5). 1005–1014.

- Connor Desai, Saoirse & Stian Reimers. 2019. Comparing the use of open and closed questions for Web-based measures of the continued-influence effect. *Behavior research methods* 51(3). 1426–1440.
- Conway, Andrew R. A., Christopher Jarrold, Michael J. Kane, Akira Miyake & John N. Towse. 2008. Variation in working memory: An introduction. In Andrew R. A. Conway (ed.), *Variation in working memory*, 3–18. Oxford: Oxford Univ. Press.
- Conway, Andrew R. A., Michael J. Kane, Michael F. Bunting, D. Z. Hambrick, Oliver Wilhelm & Randall W. Engle. 2005. Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review* 12(5). 769–786.
- Conway, Andrew R. A., Michael J. Kane & Randall W. Engle. 2003. Working memory capacity and its relation to general intelligence. *Trends in cognitive sciences* 7(12). 547–552.
- Conway, Andrew R.A., Nelson Cowan, Michael F. Bunting, David J. Theriault & Scott R.B. Minkoff. 2002. A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence* 30(2). 163–183.
- Cooper, Nancy A. 1995. Children's singing accuracy as a function of grade level, gender, and individual versus unison singing. *Journal of Research in Music Education* 43(3). 222–231.
- Costa-Giomi, Eugenia. 2003. Young children's harmonic perception. *Annals of the New York Academy of Sciences* 999. 477–484.
- Coumel, Marion, Markus Christiner & Susanne M. Reiterer. 2019. Second language accent faking ability depends on musical abilities, not on working memory. *Frontiers in Psychology* 10. 257.
- Cowan, Nelson. 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological bulletin* 104(2). 163–191.
- Cowan, Nelson. 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences* 24(1). 87-114; discussion 114-85.
- Cowan, Nelson. 2008. Chapter 20 What are the differences between long-term, short-term, and working memory? In Wayne S. Sossin (ed.), *Progress in brain research: Essence of memory* (Progress in Brain Research 169), vol. 169, 323–338. Amsterdam: Elsevier.

- Cross, Ian. 2001. Music, cognition, culture, and evolution. *Annals of the New York Academy of Sciences* 930. 28–42.
- Crowder, Robert G., Mary L. Serafine & Bruno Repp. 1990. Physical interaction and association by contiguity in memory for the words and melodies of songs. *Memory & cognition* 18, <https://doi.org/10.3758/BF03198480>.
- Dalla Bella, Simone. 2015. Defining poor-pitch singing. *Music Perception* 32(3). 272–282.
- Dalla Bella, Simone & Magdalena Berkowska. 2009. Singing proficiency in the majority: Normality and "phenotypes" of poor singing. *Annals of the New York Academy of Sciences* 1169. 99–107.
- Dalla Bella, Simone, Jean-François Giguère & Isabelle Peretz. 2007. Singing proficiency in the general population. *Journal of the Acoustical Society of America* 121(2). 1182–1189.
- Dalla Bella, Simone, Jean-François Giguère & Isabelle Peretz. 2009. Singing in congenital amusia. *Journal of the Acoustical Society of America* 126(1). 414–424.
- Dalla Bella, Simone, Alexandra Tremblay-Champoux, Magdalena Berkowska & Isabelle Peretz. 2012. Memory disorders and vocal performance. *Annals of the New York Academy of Sciences* 1252. 338–344.
- Dalton-Puffer, Christiane, Gunther Kaltenboeck & Ute Smit. 1997. Learner attitudes and L2 pronunciation in Austria. *World Englishes* 16(1). 115–128.
- Daneman, Meredyth & Patricia A. Carpenter. 1980. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19(4). 450–466.
- Darwin, Charles. 1871. *The descent of man, and selection in relation to sex* (2). London: John Murray.
- Davies, Ann D. M. & Emlyn Roberts. 1975. Poor pitch singing: A survey of its incidence in school children. *Psychology of Music* 3(2). 24–36.
- DeKeyser, Robert M., Iris Alfi-Shabtay & Dorit Ravid. 2010. Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics* 31(3). 413–438.
- Demorest, Steven M. & Peter Q. Pfordresher. 2015. Singing accuracy development from K-Adult. *Music Perception* 32(3). 293–302.
- Demorest, Steven M., Peter Q. Pfordresher, Simone Dalla Bella, Sean Hutchins, Psyche Loui, Joanne Rutkowski & Graham F. Welch. 2015. Methodological perspectives on singing accuracy. *Music Perception* 32(3). 266–271.

- DeVellis, Robert F. 2012. *Scale development: Theory and applications*, 3rd edn. (Applied social research methods series 26). Thousand Oaks, CA: Sage.
- Ding, Nai, Aniruddh D. Patel, Lin Chen, Henry Butler, Cheng Luo & David Poeppel. 2017. Temporal modulations in speech and music. *Neuroscience and biobehavioral reviews* 81(Pt B). 181–187.
- Dogil, Grzegorz & Susanne M. Reiterer (eds.). 2009. *Language talent and brain activity*. Berlin: Mouton de Gruyter.
- Dollmann, Jörg, Irena Kogan & Markus Weißmann. 2019. Speaking accent-free in L2 beyond the critical period: The compensatory role of individual abilities and opportunity structures. *Applied Linguistics*. amz029.
- Dörnyei, Zoltán. 2005. *The psychology of the language learner: Individual differences in second language acquisition* (Second language acquisition research). Mahwah, N.J: L. Erlbaum.
- Dörnyei, Zoltán. 2006. Themes in SLA research. *AILA Review* 19. 42–68.
- Dörnyei, Zoltán. 2014. Researching complex dynamic systems: ‘Retrodictive qualitative modelling’ in the language classroom. *Language Teaching* 47(1). 80–91.
- Dörnyei, Zoltán, Peter D. MacIntyre & Alastair Henry (eds.). 2015. *Motivational dynamics in language learning* (Second language acquisition 81). Bristol: Multilingual Matters.
- Dörnyei, Zoltán & Stephen Ryan. 2015. *The psychology of the language learner revisited* (Second language acquisition research series). New York, NY & London: Routledge.
- Dörnyei, Zoltán & Peter Skehan. 2003. Individual differences in second language learning. In Catherine Doughty & Michael H. Long (eds.), *The handbook of second language acquisition* (Blackwell handbooks in linguistics), 589–630. Malden, MA: Blackwell.
- Dörnyei, Zoltán & Tatsuya Taguchi. 2010. *Questionnaires in second language research: Construction, administration, and processing* (Second language acquisition research series). New York, NY: Routledge.
- Doughty, Catherine, Susan G. Campbell, Meredith A. Mislevy, Michael F. Bunting, Anita R. Bowles & Joel T. Koeth. 2008. Predicting near-native ability: The factor structure and reliability of the Hi-LAB. In T. Prior, Y. Watanabe & S-K. Lee (eds.), *Selected proceedings of the 2008 Second Language Research Forum*, 10–31. Somerville, MA.
- Doughty, Catherine & Michael H. Long (eds.). 2003. *The handbook of second language acquisition* (Blackwell handbooks in linguistics). Malden, MA: Blackwell.
- Dowling, W. Jay. 1999. The development of music perception and cognition. *Academic Press series in cognition and perception: A series of monographs and treatises*. 603–625.

- Dudycha, Arthur L. & James B. Carpenter. 1973. Effects of item format on item discrimination and difficulty. *Journal of Applied Psychology* 58(1). 116–121.
- Ehrman, Madeline E. & Rebecca L. Oxford. 1995. Cognition plus: Correlates of language learning success. *Modern Language Journal* 79(1). 67.
- Ekholm, Elizabeth, Georgios C. Papagiannis & Françoise P. Chagnon. 1998. Relating objective measurements to expert evaluation of voice quality in western classical singing: Critical perceptual parameters. *Journal of Voice* 12(2). 182–196.
- Engel de Abreu, Pascale M. J. & Susan E. Gathercole. 2012. Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology* 104(4). 974–986.
- Engle, Randall W., Stephen W. Tuholski, James E. Laughlin & Andrew R. A. Conway. 1999. Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of experimental psychology* 128(3). 309–331.
- Estis, Julie M., Joana K. Coblentz & Robert E. Moore. 2009. Effects of increasing time delays on pitch-matching accuracy in trained singers and untrained individuals. *Journal of Voice* 23(4). 439–445.
- Fehm, Lydia & Katja Schmidt. 2006. Performance anxiety in gifted adolescent musicians. *Journal of anxiety disorders* 20(1). 98–109.
- Field, Andy. 2009. *Discovering statistics using SPSS*, 3rd edn. Los Angeles, CA: Sage.
- Figueira, Jessica S. B., Leticia Oliveira, Mirtes G. Pereira, Luiza B. Pacheco, Isabela Lobo, Gabriel C. Motta-Ribeiro & Isabel A. David. 2017. An unpleasant emotional state reduces working memory capacity: Electrophysiological evidence. *Social cognitive and affective neuroscience* 12(6). 984–992.
- Fitch, W. Tecumseh. 2006. On the biology and evolution of music. *Music Perception* 24(1). 85–88.
- Foddy, William H. 1993. *Constructing questions for interviews and questionnaires: Theory and practice in social research*. Cambridge & New York, NY: Cambridge University Press.
- Fonseca-Mora, Carmen. 2000. Foreign language acquisition and melody singing. *ELT Journal* 54(2). 146–152.
- Fonseca-Mora, Carmen, Pilar Jara-Jiménez & María Gómez-Domínguez. 2015. Musical plus phonological input for young foreign language readers. *Frontiers in Psychology* 6. 286.

- François, Clément, Julie Chobert, Mireille Besson & Daniele Schön. 2013. Music training for the development of speech segmentation. *Cerebral cortex* 23(9). 2038–2043.
- Fritz, Thomas, Sebastian Jentschke, Nathalie Gosselin, Daniela Sammler, Isabelle Peretz, Robert Turner, Angela D. Friederici & Stefan Koelsch. 2009. Universal recognition of three basic emotions in music. *Current biology* 19(7). 573–576.
- Fukuda, Keisuke, Edward Vogel, Ulrich Mayr & Edward Awh. 2010. Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic bulletin & review* 17(5). 673–679.
- Gagné, François. 2004. Transforming gifts into talents: The DMGT as a developmental theory. *High Ability Studies* 15(2). 119–147.
- Gagné, François. 2008. Building gifts into talents: Brief overview of the DMGT 2.0. *Gifted* 152. 5–9.
- García-López, Isabel & Javier Gavilán Bouzas. 2010. La voz cantada. *Acta otorrinolaringologica espanola* 61(6). 441–451.
- Gardner Robert. 1990. Attitudes, motivation, and personality as predictors of success in foreign language learning. In Thomas Parry & Charles W. Stansfield (eds.), *Language aptitude reconsidered*, 179–221. Englewood Cliffs, NJ: Prentice Hall.
- Gaser, Christian & Gottfried Schlaug. 2003. Brain structures differ between musicians and non-musicians. *Journal of Neuroscience* 23(27). 9240–9245.
- Gathercole, Susan E. 2006. Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics* 27(4). 513–543.
- Gathercole, Susan E. & Alan D. Baddeley. 1990. The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology* 81(4). 439–454.
- Gathercole, Susan E. & Alan D. Baddeley. 1993. *Working memory and language* (Essays in cognitive psychology). Hove: Erlbaum.
- Gathercole, Susan E., Graham J. Hitch, Elisabet Service & Amanda J. Martin. 1997. Phonological short-term memory and new word learning in children. *Developmental Psychology* 33(6). 966–979.
- Gathercole, Susan E., Emily Lamont & Tracy P. Alloway. 2007. Working Memory in the Classroom. In Susan J. Pickering (ed.), *Working memory and education*, 219–240. Amsterdam: Elsevier.

- Geringer, John M. 1983. The Relationship of Pitch-Matching and Pitch-Discrimination Abilities of Preschool and Fourth-Grade Students. *Journal of Research in Music Education* 31(2). 93–99.
- Gerstman, Hubert L. 1964. A case of aphasia. *Journal of speech and hearing disorders* 29. 89–91.
- Geschwind, Norman. 1971. Current concepts: Aphasia. *New England journal of medicine* 284(12). 654–656.
- Goetze, Mary. 1989. A comparison of the pitch accuracy of group and individual singing in young children. *Bulletin of the Council for Research in Music Education*, www.jstor.org/stable/40318325.
- Goetze, Mary, Nancy Cooper & Carol J. Brown. 1990. Recent research on singing in the general music classroom. *Bulletin of the Council for Research in Music Education* (104). 16–37.
- Golestani, Narly & Christophe Pallier. 2007. Anatomical correlates of foreign speech sound production. *Cerebral cortex* 17(4). 929–934.
- Golestani, Narly, Cathy J. Price & Sophie K. Scott. 2011. Born with an ear for dialects? Structural plasticity in the expert phonetician brain. *Journal of Neuroscience* 31(11). 4213–4220.
- Golestani, Narly & Robert J. Zatorre. 2009. Individual differences in the acquisition of second language phonology. *Brain and language* 109(2-3). 55–67.
- Gordon, Edwin. 1965. Musical aptitude profile manual. Boston, MA: Houghton Mifflin.
- Gordon, Edwin. 1967a. *A three-year study of the musical aptitude profile*. Chicago, IL: GIA.
- Gordon, Edwin. 1967b. The musical aptitude profile. *Music Educators Journal* 53(6). 52–54.
- Gordon, Edwin. 1979. *Primary measures of music audiation*. Chicago, IL: GIA.
- Gordon, Edwin. 1981. *The manifestation of developmental music aptitude in the audiation of "same" and "different" as sound in music*. Chicago, IL: GIA.
- Gordon, Edwin. 1982. *Intermediate measures of music audiation*. Chicago, IL: GIA.
- Gordon, Edwin. 1989a. *Advanced measures of music audiation*. Chicago, IL: GIA.
- Gordon, Edwin. 1989b. *Audie: A game for understanding and analyzing your child's musical potential*. Chicago, IL: GIA.
- Gordon, Edwin. 1997. *Learning sequences in music: Skill, content, and patterns*, 1997th edn. Chicago, IL: GIA.
- Gordon, Edwin E. 2013. *Music learning theory for newborn and young children*. Chicago, IL: GIA.

- Gordon, Reyna L., Daniele Schön, Cyrille Magne, Corine Astésano & Mireille Besson. 2010. Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PloS one* 5(3). e9889.
- Gould, A. Oren. 1969. Developing Specialized Programs for Singing in the Elementary School. *Bulletin of the Council for Research in Music Education*, www.jstor.org/stable/40317056.
- Gouzouasis, Peter, Martin Guhn & Nand Kishor. 2007. The predictive relationship between achievement and participation in music and achievement in core Grade 12 academic subjects. *Music Education Research* 9(1). 81–92.
- Grashel, John. 2008. The measurement of musical aptitude in 20th century United States: A brief history. *Bulletin of the Council for Research in Music Education* (176). 45–49.
- Green, Georgia A. 1990. The effect of vocal modeling on pitch-matching accuracy of elementary schoolchildren. *Journal of Research in Music Education* 38(3). 225.
- Grigornko, Elena L., Robert J. Sternberg & Madeline E. Ehrman. 2000. A theory-based approach to the measurement of foreign language learning ability: The Canal-F theory and test. *Modern Language Journal* 84(3). 390–405.
- Groome, Lynn J., Donna M. Mooney, Scherri B. Holland, Lisa A. Smith, Jana L. Atterbury & Roscoe A. Dykman. 1999. Behavioral state affects heart rate response to low-intensity sound in human fetuses. *Early Human Development* 54(1). 39–54.
- Gupta, Chitrallekha, Haizhou Li & Ye Wang. 2018. A technical framework for automatic perceptual evaluation of singing quality. *APSIPA Transactions on Signal and Information Processing* 7. 1.
- Halwani, Gus F., Psyche Loui, Theodor Rüber & Gottfried Schlaug. 2011. Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists, and non-musicians. *Frontiers in Psychology* 2. 156.
- Hamann, Donald L. & Martha Sobaje. 1983. Anxiety and the college musician: a study of performance conditions and subject variables. *Psychology of Music* 11(1). 37–50.
- Hannon, Erin E. & Laurel J. Trainor. 2007. Music acquisition: Effects of enculturation and formal training on development. *Trends in cognitive sciences* 11(11). 466–472.
- Hansen, Mads, Mikkel Wallentin & Peter Vuust. 2013. Working memory and musical competence of musicians and non-musicians. *Psychology of Music* 41(6). 779–793.
- Hébert, Sylvie, Amélie Racette, Lise Gagnon & Isabelle Peretz. 2003. Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain* 126(Pt 8). 1838–1850.

- Hell, Janet G. van & Andrea C. Mahn. 1997. Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning* 47(3). 507–546.
- Hepper, Peter G. 1991. An examination of fetal learning before and after birth. *Irish Journal of Psychology* 12(2). 95–107.
- Hernandez, Arturo E. & Ping Li. 2007. Age of acquisition: Its neural and computational mechanisms. *Psychological bulletin* 133(4). 638–650.
- Ho, Yim-Chi, Mei-Chun Cheung & Agnes S. Chan. 2003. Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology* 17(3). 439–450.
- Hockett, Charles F., Robert Ascher, George A. Agogino, Ray L. Birdwhistell, Alan L. Bryan, J. D. Clark, Carleton S. Coon, Earl W. Count, Robert Cresswell, A. R. Diebold, Theodosius Dobzhansky, R. D. Givens, Gordon W. Hewes, Ilse Lehisté, Margaret Mead, Ashley Montagu, Hans G. Mukarovsky, John Pfeiffer, Bernard Pottier, Adolph H. Schultz, H. L. Smith, James L. Swauger, George L. Trager, Eugene Verstraelen & Roger W. Wescott. 1964. The human revolution [and comments and reply]. *Current Anthropology* 5(3). 135–168.
- Hollien, Harry. 1993. That golden voice - talent or training? *Journal of Voice* 7(3). 195–205.
- Honing, Henkjan. 2011. *Musical Cognition: A science of listening*. Somerset: Taylor and Francis.
- Hornbach, Christina M. & Cynthia C. Taggart. 2005. The relationship between developmental tonal aptitude and singing achievement among kindergarten, first-, second-, and third-grade students. *Journal of Research in Music Education* 53(4). 322–331.
- Horwitz, Elaine K. 1986. Preliminary evidence for the reliability and validity of a foreign language anxiety scale. *TESOL Quarterly* 20(3). 559–562.
- Horwitz, Elaine K., Michael B. Horwitz & Joann Cope. 1986. Foreign Language Classroom Anxiety. *Modern Language Journal* 70(2). 125–132.
- Hu, Chieh-Fang & C. Melanie Schuele. 2005. Learning nonnative names: The effect of poor native phonological awareness. *Applied Psycholinguistics* 26(3). 343–362.
- Hu, Xiaochen, Hermann Ackermann, Jason A. Martin, Michael Erb, Susanne Winkler & Susanne M. Reiterer. 2013. Language aptitude for pronunciation in advanced second language (L2) learners: Behavioural predictors and neural substrates. *Brain and language* 127(3). 366–376.

- Hulme, Charles, Sarah Maughan & Gordon D.A. Brown. 1991. Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language* 30(6). 685–701.
- Hummel, Kirsten, M. 2009. Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics* 30(2). 225–249.
- Hunt, T., F. Wallace, S. Doran, K. Buynitzky & R. Schwarz. 1929. Language aptitude test. Washington, DC: George Washington University.
- Huron, David. 2001. Is music an evolutionary adaptation? *Annals of the New York Academy of Sciences* 930. 43–61.
- Hutchins, Sean, Pauline Larrouy-Maestri & Isabelle Peretz. 2014. Singing ability is rooted in vocal-motor control of pitch. *Attention, perception & psychophysics* 76(8). 2522–2530.
- Hutchins, Sean & Sylvain Moreno. 2013. The linked dual representation model of vocal perception and production. *Frontiers in Psychology* 4. 825.
- Hutchins, Sean, Jean M. Zarate, Robert J. Zatorre & Isabelle Peretz. 2010. An acoustical study of vocal pitch matching in congenital amusia. *Journal of the Acoustical Society of America* 127(1). 504–512.
- Hutchins, Sean M. & Isabelle Peretz. 2012. A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of experimental psychology* 141(1). 76–97.
- Hyde, Krista L. & Isabelle Peretz. 2004. Brains that are out of tune but in time. *Psychological science* 15(5). 356–360.
- Ibrahim, Karim M., David Grunberg, Kat Agres, Chitrlekha Gupta & Ye Wang. 2017. Intelligibility of sung lyrics: a pilot study. Suzhou: Proceedings of the 18th International Society for Music Information Retrieval Conference, ISMIR. 686–693.
- Imfeld, Adrian, Mathias S. Oechslin, Martin Meyer, Thomas Loenneker & Lutz Jancke. 2009. White matter plasticity in the corticospinal tract of musicians: A diffusion tensor imaging study. *NeuroImage* 46(3). 600–607.
- Iverson, Jana M. 2010. Developing language in a developing body: The relationship between motor development and language development. *Journal of child language* 37(2). 229–261.
- Jackendoff, Ray & Fred Lerdahl. 2006. The capacity for music: What is it, and what's special about it? *Cognition* 100(1). 33–72.

- Johnson, J. S. & E. L. Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology* 21(1). 60–99.
- Jørgensen, H. 2008. What is a talent, how is it identified, and how is it developed? Oslo: Presented at the conference “Precollege education and talent development”, European Association of Conservatoires (AEC).
- Juffs, Alan & Michael Harrington. 2011. Aspects of working memory in L2 learning. *Language Teaching* 44(2). 137–166.
- Kaushanskaya, Margarita & Viorica Marian. 2009. The bilingual advantage in novel word learning. *Psychonomic bulletin & review* 16(4). 705–710.
- Kawase, Satoshi. 2016. Associations among music majors’ personality traits, empathy, and aptitude for ensemble performance. *Psychology of Music* 44(2). 293–302.
- Keith, Robert L. & Arnold E. Aronson. 1975. Singing as therapy for apraxia of speech and aphasia: Report of a case. *Brain and language* 2. 483–488.
- Kershenbaum, Ayelet, M. L. Nicholas, Eileen Hunsaker & Lauryn Zipse. 2019. Speak along without the song: What promotes fluency in people with aphasia? *Aphasiology* 33(4). 405–428.
- Kisilevsky, Barbara S., Sylvia M. J. Hains, Kang Lee, Xing Xie, Hefeng Huang, Hai H. Ye, Ke Zhang & Zengping Wang. 2003. Effects of experience on fetal voice recognition. *Psychological science* 14(3). 220–224.
- Kleber, B., R. Veit, N. Birbaumer, J. Gruzelier & M. Lotze. 2010. The brain of opera singers: Experience-dependent changes in functional activation. *Cerebral cortex* 20(5). 1144–1152.
- Klingberg, Torkel, Hans Forssberg & Helena Westerberg. 2002. Training of working memory in children with ADHD. *Journal of clinical and experimental neuropsychology* 24(6). 781–791.
- Koelsch, Stefan, Thomas Fritz, Katrin Schulze, David Alsop & Gottfried Schlaug. 2005a. Adults and children processing music: An fMRI study. *NeuroImage* 25(4). 1068–1076.
- Koelsch, Stefan, Thomas C. Gunter, Matthias Wittfoth & Daniela Sammler. 2005b. Interaction between syntax processing in language and in music: An ERP Study. *Journal of cognitive neuroscience* 17(10). 1565–1577.

- Koelsch, Stefan, Katrin Schulze, Daniela Sammler, Thomas Fritz, Karsten Müller & Oliver Gruber. 2009. Functional architecture of verbal and tonal working memory: An fMRI study. *Human Brain Mapping* 30(3). 859–873.
- Kokotsaki, Dimitra & Jane W. Davidson. 2003. Investigating Musical Performance Anxiety among Music College Singing Students: A quantitative analysis. *Music Education Research* 5(1). 45–59.
- Korecky-Kröll, Katharina, Neriman Dobek, Verena Blaschitz, Sabine Sommer-Lolei, Monika Boniecki, Kumru Uzunkaya-Sharma & Wolfgang U. Dressler. 2019. Vocabulary as a central link between phonological working memory and narrative competence: evidence from monolingual and bilingual four-year-olds from different socioeconomic backgrounds. *Language and speech* 62(3). 546–569.
- Kormos, Judit & Anna Safar. 2008. Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition* 11(2). 261–271.
- Krashen, Stephen D. 1985. *The input hypothesis: Issues and implications*. London: Longman.
- Kraus, Nina & Bharath Chandrasekaran. 2010. Music training for the development of auditory skills. *Nature reviews. Neuroscience* 11(8). 599–605.
- Krumhansl, Carol L. & Frank C. Keil. 1982. Acquisition of the hierarchy of tonal functions in music. *Memory & cognition* 10(3). 243–251.
- Kuhl, Patricia K., Jean E. Andruski, Inna A. Chistovich, Ludmilla A. Chistovich, Elena V. Kozhevnikova, Viktoria L. Ryskina, Elvira I. Stolyarova, Ulla Sundberg & Francisco Lacerda. 1997. Cross-language analysis of phonetic units in language addressed to infants. *Science* 277(5326). 684–686.
- Larrouy-Maestri, Pauline, Yohana Lévêque, Daniele Schön, Antoine Giovanni & Dominique Morsomme. 2013. The evaluation of singing voice accuracy: A comparison between subjective and objective methods. *Journal of Voice* 27(2). 259.e1-259.e5.
- Law, Lily N. C. 2012. *Assessing and understanding individual differences in music perception abilities*. York: University of York PhD thesis.
- Law, Lily N. C. & Marcel Zentner. 2012. Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PloS one* 7(12). e52508.
- Lecanuet, J. P., C. Granier-Deferre, A. J. DeCasper, R. Maugeais, A. J. Andrieu & M. C. Busnel. 1987. Perception et discrimination foetales de stimuli langagiers; mise en

- évidence à partir de la réactivité cardiaque; résultats préliminaires. *Comptes rendus de l'Académie des sciences. Serie III, Sciences de la vie* 305(5). 161–164.
- Leighton, Geraldine & Alexandra Lamont. 2006. Exploring children's singing development: Do experiences in early schooling help or hinder? *Music Education Research* 8(3). 311–330.
- Lenneberg, Eric H. 1967. The biological foundations of language. *Hospital Practice* 2(12). 59–67.
- Leonard, C. M., C. Puranik, J. M. Kuldau & L. J. Lombardino. 1998. Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: Where is it? *Cerebral cortex* 8(5). 397–406.
- Levitin, Daniel J. 2012. What does it mean to be musical? *Neuron* 73(4). 633–637.
- Lieberman, Alvin M. & Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21(1). 1–36.
- Lietz, Petra. 2010. Research into questionnaire design: A summary of the literature. *International Journal of Market Research* 52(2). 249–272.
- Limb, Charles J. 2006. Structural and functional neural correlates of music perception. *The anatomical record. Part A, Discoveries in molecular, cellular, and evolutionary biology* 288(4). 435–446.
- Linck, Jared A., Meredith M. Hughes, Susan G. Campbell, Noah H. Silbert, Medha Tare, Scott R. Jackson, Benjamin K. Smith, Michael F. Bunting & Catherine J. Doughty. 2013. Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning* 63(3). 530–566.
- Linck, Jared A., Peter Osthus, Joel T. Koeth & Michael F. Bunting. 2014. Working memory and second language comprehension and production: A meta-analysis. *Psychonomic bulletin & review* 21(4). 861–883.
- Livingstone, Frank B. 1973. Did the Australopithecines Sing? *Current Anthropology* (14). 25–29.
- Locke, John L., Kaaren E. Bekken, Laura McMinn-Larson & Debra Wein. 1995. Emergent control of manual and vocal-motor activity in relation to the development of speech. *Brain and language* 51(3). 498–508.
- Loosli, Sandra V., Martin Buschkuehl, Walter J. Perrig & Susanne M. Jaeggi. 2012. Working memory training improves reading processes in typically developing children. *Child neuropsychology* 18(1). 62–78.

- Loui, Psyche, David Alsop & Gottfried Schlaug. 2009. Tone deafness: a new disconnection syndrome? *Journal of Neuroscience* 29(33). 10215–10220.
- Ludke, Karen M., Fernanda Ferreira & Katie Overy. 2014. Singing can facilitate foreign language learning. *Memory & cognition* 42(1). 41–52.
- Lynn, Richard, R. Graham Wilson & Adrienne Gault. 1989. Simple musical tests as measures of Spearman's g. *Personality and Individual Differences* 10(1). 25–28.
- Mackey, Alison & Rebecca Sachs. 2012. Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning* 62(3). 704–740.
- Macnamara, Brooke N., David Z. Hambrick & Frederick L. Oswald. 2014. Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological science* 25(8). 1608–1618.
- Majerus, Steve, Martine Poncelet, Christelle Greffe & Martial van der Linden. 2006. Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology* 93(2). 95–119.
- Manzano, Örjan de & Fredrik Ullén. 2018. Same genes, different brains: Neuroanatomical differences between monozygotic twins discordant for musical training. *Cerebral cortex* 28(1). 387–394.
- Masataka, Nobuo. 2009. The origins of language and the evolution of music: A comparative perspective. *Physics of life reviews* 6(1). 11–22.
- McClellan, Josef W. 2011. *Comparative analysis of speech level singing and traditional vocal training in the United States*. Memphis, TN: University of Memphis Dissertation.
- McClelland, James L. & Jeffrey L. Elman. 1986. The TRACE model of speech perception. *Cognitive psychology* 18(1). 1–86.
- McGrath, Robert E. & Glenn D. Walters. 2012. Taxometric analysis as a general strategy for distinguishing categorical from dimensional latent structure. *Psychological methods* 17(2). 284–293.
- McMullen, Erin & Jenny R. Saffran. 2004. Music and language: A developmental comparison. *Music Perception* 21(3). 289–311.
- Meara, Paul. 2005. LLAMA language aptitude tests. Swansea: Lognostics.
- Michaelis, Christian F. 1805. Über die Prüfung musikalischer Fähigkeiten. In J. F. Reichard (ed.), *Berlinische Musikalische Zeitung*, vol. 56, 222–230. Berlin.

- Milovanov, Riia. 2009. Musical aptitude and foreign language learning skills: Neural and behavioural evidence about their connections.
- Milovanov, Riia & Mari Tervaniemi. 2011. The Interplay between musical and linguistic aptitudes: A review. *Frontiers in Psychology* 2. 321.
- Mithen, Steven. 2007. *The singing Neanderthals: The origins of music, language, mind and body*. Cambridge, MA: Harvard Univ. Press.
- Mogle, Jacqueline A., Benjamin J. Lovett, Robert S. Stawski & Martin J. Sliwinski. 2008. What's so special about working memory? An examination of the relationships among working memory, secondary memory, and fluid intelligence. *Psychological science* 19(11). 1071–1077.
- Moog, Helmut. 1976. The development of musical experience in children of pre-school age. *Psychology of Music* 4(2). 38–45.
- Moore, Robert E., Julie Estis, Susan Gordon-Hickey & Christopher Watts. 2008. Pitch discrimination and pitch matching abilities with vocal and nonvocal stimuli. *Journal of Voice* 22(4). 399–407.
- Moreno, Sylvain. 2009. Can music influence language and cognition? *Contemporary Music Review* 28(3). 329–345.
- Moreno, Sylvain, Ellen Bialystok, Raluca Barac, E. Glenn Schellenberg, Nicholas J. Cepeda & Tom Chau. 2011. Short-term music training enhances verbal intelligence and executive function. *Psychological science* 22(11). 1425–1433.
- Moreno, Sylvain, Carlos Marques, Andreia Santos, Manuela Santos, São L. Castro & Mireille Besson. 2009. Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral cortex* 19(3). 712–723.
- Morosan, Patricia, Joerg Rademacher, Axel Schleicher, Katrin Amunts, Thorsten Schormann & Karl Zilles. 2001. Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage* 13(4). 684–701.
- Mosing, Miriam A., Guy Madison, Nancy L. Pedersen, Ralf Kuja-Halkola & Fredrik Ullén. 2014. Practice does not make perfect: No causal effect of music practice on music ability. *Psychological science* 25(9). 1795–1803.
- Moyer, Alene. 2014. What's age got to do with it? Accounting for individual factors in second language accent. *Studies in Second Language Learning and Teaching* 3. 443–464.
- Müllensiefen, Daniel, Bruno Gingras, Jason Musil & Goldsmiths, Lauren, Stewart. 2012. A new tool for measuring musical sophistication: The Goldsmiths Musical

- Sophistication Index. Thessaloniki: 12th International Conference on Music Perception and Cognition (ICMPC).
- Nakamura, Yoshihiro. 1988. Foreign language aptitude in foreign language learning. *Bulletin of the Faculty of Education, Kagoshima University* 40. 163–202.
- Nardo, Davide & Susanne M. Reiterer. 2009. Musicality and phonetic language aptitude. In Grzegorz Dogil & Susanne M. Reiterer (eds.), *Language talent and brain activity*, 213–256. Berlin: Mouton de Gruyter.
- Nasir, Sazzad M. & David J. Ostry. 2009. Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of the United States of America* 106(48). 20470–20475.
- Newman, Ernest. 1905. *Musical studies*. London: John Lane.
- Nielsen, Jens B. & Leonardo G. Cohen. 2008. The Olympic brain. Does corticospinal plasticity play a role in acquisition of skills required for high-performance sports? *Journal of physiology* 586(1). 65–70.
- Nimmons, Grace L., Robert S. Kang, Ward R. Drennan, Jeff Longnion, Chad Ruffin, Tina Worman, Bevan Yueh & Jay T. Rubinstein. 2008. Clinical assessment of music perception in cochlear implant listeners. *Otology & neurotology* 29(2). 149–155.
- Norris, Dennis. 2017. Short-term memory and long-term memory are still different. *Psychological bulletin* 143(9). 992–1009.
- Norton, Andrea, Lauryn Zipse, Sarah Marchina & Gottfried Schlaug. 2009. Melodic intonation therapy: Shared insights on how it is done and why it might help. *Annals of the New York Academy of Sciences* 1169. 431–436.
- Oechslin, Mathias S., Martin Meyer & Lutz Jäncke. 2010. Absolute pitch-functional evidence of speech-relevant auditory acuity. *Cerebral cortex* 20(2). 447–455.
- Oikkonen, J., Y. Huang, P. Onkamo, L. Ukkola-Vuoti, P. Raijas, K. Karma, V. J. Vieland & I. Järvelä. 2015. A genome-wide linkage and association study of musical aptitude identifies loci containing genes related to inner ear development and neurocognitive functions. *Molecular psychiatry* 20(2). 275–282.
- O'Malley, Michael J. & Anna U. Chamot. 1990. *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- Omori, Koichi, Ashutosh Kacker, Linda M. Carroll, William D. Riley & Stanley M. Blaugrund. 1996. Singing power ratio: Quantitative evaluation of singing voice quality. *Journal of Voice* 10(3). 228–235.

- Oz, Huseyin. 2014. Big five personality traits and willingness to communicate among foreign language learners in Turkey. *Social Behavior and Personality* 42(9). 1473–1482.
- Ozdemir, Elif, Andrea Norton & Gottfried Schlaug. 2006. Shared and distinct neural correlates of singing and speaking. *NeuroImage* 33(2). 628–635.
- Papagno, Costanza, Tim Valentine & Alan D. Baddeley. 1991. Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language* 30(3). 331–347.
- Papagno, Costanza & G. Vallar. 1995. Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology Section A* 48(1). 98–107.
- Park, Hansoo, Seungbok Lee, Hyun-Jin Kim, Young S. Ju, Jong-Yeon Shin, Dongwan Hong, Marcin von Grotthuss, Dong-Sung Lee, Changho Park, Jennifer H. Kim, Boram Kim, Yun J. Yoo, Sung-Il Cho, Joohon Sung, Charles Lee, Jong-Il Kim & Jeong-Sun Seo. 2012. Comprehensive genomic analyses associate UGT8 variants with musical ability in a Mongolian population. *Journal of medical genetics* 49(12). 747–752.
- Parry, Thomas & James Child. 1990. Preliminary investigation of the relationship between VORD, MLAT and language proficiency. In Thomas Parry & Charles W. Stansfield (eds.), *Language aptitude reconsidered*, 30–66. Englewood Cliffs, NJ: Prentice Hall.
- Patel, Aniruddh D. 2003. Language, music, syntax and the brain. *Nature neuroscience* 6(7). 674–681.
- Patel, Aniruddh D. 2007. *Music, language, and the brain*. Oxford: Oxford University Press.
- Patel, Aniruddh D. 2011. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology* 2. 142.
- Patel, Aniruddh D. & Joseph R. Daniele. 2003. An empirical comparison of rhythm in language and music. *Cognition* 87(1). B35-B45.
- Payne, J. Scott & Paul J. Whitney. 2002. Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *Calico* 20, www.jstor.org/stable/24149607.
- Peper, Jiska S., Rachel M. Brouwer, Dorret I. Boomsma, René S. Kahn & Hilleke E. Hulshoff Pol. 2007. Genetic influences on human brain structure: A review of brain imaging studies in twins. *Human Brain Mapping* 28(6). 464–473.
- Peretz, Isabelle. 2008. Musical disorders. *Current Directions in Psychological Science* 17(5). 329–333.
- Peretz, Isabelle, Lise Gagnon, Sylvie Hebert & Joel Macoir. 2004. Singing in the brain: Insights from cognitive neuropsychology. *Music Perception* 21(3). 373–390.

- Peretz, Isabelle, Dominique Vuvan, Marie-Élaine Lagrois & Jorge L. Armony. 2015. Neural overlap in processing music and speech. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 370(1664). 20140090.
- Perkins, Judy M., Jane A. Baran & Jack Gandour. 1996. Hemispheric specialization in processing intonation contours. *Aphasiology* 10(4). 343–362.
- Perrachione, Tyler K., Evelina G. Fedorenko, Louis Vinke, Edward Gibson & Laura C. Dilley. 2013. Evidence for shared cognitive processing of pitch in music and language. *PloS one* 8(8). e73372.
- Petersen, Calvin R. & Antoine R. Al-Haik. 1976. The development of the defense language aptitude battery (DLAB). *Educational and Psychological Measurement* 36(2). 369–380.
- Pfordresher, Peter Q. & Steven Brown. 2007. Poor-pitch singing in the absence of "tone deafness". *Music Perception* 25(2). 95–115.
- Pfordresher, Peter Q., Andrea R. Halpern & Emma B. Greenspon. 2015. A mechanism for sensorimotor translation in singing. *Music Perception* 32(3). 242–253.
- Pfordresher, Peter Q. & James T. Mantell. 2014. Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive psychology* 70. 31–57.
- Phillips, Elaine M. 1991. Anxiety and oral competence: Classroom dilemma. *The French Review* 65, www.jstor.org/stable/394560.
- Phillips, Elaine M. 1992. The effects of language anxiety on students' oral test performance and attitudes. *Modern Language Journal* 76(1). 14.
- Phillips, Kenneth H. & Randall E. Aitchison. 1997a. Effects of psychomotor instruction on elementary general music students' singing performance. *Journal of Research in Music Education* 45(2). 185–196.
- Phillips, Kenneth H. & Randall E. Aitchison. 1997b. The relationship of singing accuracy to pitch discrimination and tonal aptitude among third-grade students. *Contributions to Music Education* 24(1). 7–22.
- Phillips, Kenneth H., Randall E. Aitchison & Yolisa P. Nompula. 2002. The relationship of music aptitude to singing achievement among fifth-grade students. *Contributions to Music Education* 29(1). 47–58.
- Pimsleur, Paul. 1963. Predicting success in high school foreign language courses. *Educational and Psychological Measurement* 23(2). 349–357.
- Pimsleur, Paul. 1966. Pimsleur language aptitude battery. New York, NY: Harcourt Brace Javanovich.

- Pimsleur, Paul, Ludwig Mosberg & Andrew L. Morrison. 1962. Student factors in foreign language learning. *Modern Language Journal* 46(4). 160.
- Platz, Friedrich, Reinhard Kopiez, Andreas C. Lehmann & Anna Wolf. 2014. The influence of deliberate practice on musical achievement: A meta-analysis. *Frontiers in Psychology* 5. 646.
- Price, Harry E., Cornelia Yarbrough, Marilyn Jones & Randall S. Moore. 1994. Effects of male timbre, falsetto, and sine-wave models on interval matching by inaccurate singers. *Journal of Research in Music Education* 42(4). 269–284.
- Pulli, K., K. Karma, R. Norio, P. Sistonen, H. H. H. Göring & I. Järvelä. 2008. Genome-wide linkage scan for loci of musical aptitude in Finnish families: Evidence for a major locus at 4q22. *Journal of medical genetics* 45(7). 451–456.
- Purnell-Webb, Patricia & Craig P. Speelman. 2008. Effects of music on memory for text. *Perceptual and motor skills* 106(3). 927–957.
- Rademacher, Joerg, Patricia Morosan, Thorsten Schormann, Axel Schleicher, Christian Werner, Hans J. Freund & Karl Zilles. 2001. Probabilistic mapping and volume measurement of human primary auditory cortex. *NeuroImage* 13(4). 669–683.
- Rapee, Ronald M. 1993. The utilisation of working memory by worry. *Behaviour Research and Therapy* 31(6). 617–620.
- Reiterer, Susanne M., Xiaochen Hu, Michael Erb, Giuseppina Rota, Davide Nardo, Wolfgang Grodd, Susanne Winkler & Hermann Ackermann. 2011. Individual differences in audio-vocal speech imitation aptitude in late bilinguals: Functional neuro-imaging and brain morphology. *Frontiers in Psychology* 2. 271.
- Renzulli, Joseph S. 1978. What makes giftedness? Reexamining a definition. *Phi Delta Kappan* (60). 180–184.
- Renzulli, Joseph S. 2016. The three-ring conception of giftedness: A developmental model for promoting creativve productivity. In S. M. Reis (ed.), *Reflections on gifted education*, 55–86. Waco, TX: Prufrock Press.
- Révész, Andrea. 2012. Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning* 62(1). 93–132.
- Rinta, Tiija & Graham F. Welch. 2008. Should singing activities be included in speech and voice therapy for prepubertal children? *Journal of Voice* 22(1). 100–112.
- Rinta, Tiija E. & Graham F. Welch. 2009. Perceptual connections between prepubertal children's voices in their speaking behavior and their singing behavior. *Journal of Voice* 23(6). 677–686.

- Roberts, Emlyn & Ann D. M. Davies. 1975. Poor pitch singing: response of monotone singers to a program of remedial training. *Journal of Research in Music Education* 23(4). 227–239.
- Robinson, Peter (ed.). 2002. *Individual differences and instructed language learning* (2). Amsterdam: John Benjamins.
- Robinson, Peter. 2005. Aptitude and second language acquisition. *Annual Review of Applied Linguistics* 25. 46–73.
- Robinson, Peter. 2019a. Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA. In Robert M. DeKeyser & Goretti Prieto Botana (eds.), *Doing SLA research with implications for the classroom* (Language Learning & Language Teaching 52), vol. 2, 211–266. Amsterdam: John Benjamins.
- Robinson, Peter. 2019b. Learning conditions, aptitude complexes, and SLA. In Robert M. DeKeyser & Goretti Prieto Botana (eds.), *Doing SLA research with implications for the classroom* (Language Learning & Language Teaching 52), vol. 2, 113–133. Amsterdam: John Benjamins.
- Rogalsky, Corianne, Feng Rong, Kourosh Saberi & Gregory Hickok. 2011. Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. *Journal of Neuroscience* 31(10). 3843–3852.
- Rogers, Maria, Heungsun Hwang, Maggie Toplak, Margaret Weiss & Rosemary Tannock. 2011. Inattention, working memory, and academic achievement in adolescents referred for attention deficit/hyperactivity disorder (ADHD). *Child neuropsychology* 17(5). 444–458.
- Rogers, Vivienne, Paul Meara, Thomas Barnett-Legh, Clare Curry & Emma Davie. 2017. Examining the LLAMA aptitude tests. *Journal of the European Second Language Association* 1(1). 49–60.
- Rolstad, Sindre, John Adler & Anna Rydén. 2011. Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in health* 14(8). 1101–1108.
- Roman, Adrienne S., David B. Pisoni & William G. Kronenberger. 2014. Assessment of working memory capacity in preschool children using the missing scan task. *Infant and child development* 23(6). 575–587.
- Rutkowski, Joanne. 1996. The effectiveness of individual/small-group singing activities on kindergartners' use of singing voice and developmental music aptitude. *Journal of Research in Music Education* 44(4). 353–368.

- Rutkowski, Joanne. 2013. The nature of children's singing voices: characteristics and assessment. *The Phenomenon of Singing* 1. 201–209.
- Rutkowski, Joanne & Martha Snell Miller. 2002. A longitudinal study of elementary children's acquisition of their singing voices. *Update: Applications of Research in Music Education* 22(1). 5–14.
- Sagarra, Nuria & Julia Herschensohn. 2010. The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua* 120(8). 2022–2039.
- Salamé, Pierre & Alan D. Baddeley. 1989. Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology Section A* 41(1). 107–122.
- Salvador, Karen. 2010. How can elementary teachers measure singing voice achievement? A critical review of assessments, 1994-2009. *Update: Applications of Research in Music Education* 29(1). 40–47.
- Sasaki, Miyuki. 1999. *Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analyses* (Theoretical studies in second language acquisition 6). New York, NY: Lang.
- Sasaki, Miyuki. 2012. The modern language aptitude test (paper-and-pencil version). *ELT Journal* 29(2). 315–321.
- Schellenberg, E. Glenn, Emmanuel Bigand, Benedicte Poulin-Charronnat, Cécilia Garnier & Catherine Stevens. 2005. Children's implicit knowledge of harmony in Western music. *Developmental science* 8(6). 551–566.
- Schellenberg, Glenn E. 2005. Music and cognitive abilities. *Current Directions in Psychological Science* 14(6). 317–320.
- Schellenberg, Glenn E. 2011. Music lessons, emotional intelligence, and IQ. *Music Perception* 29(2). 185–194.
- Schlaug, Gottfried, Andrea Norton, Sarah Marchina, Lauryn Zipse & Catherine Y. Wan. 2010. From singing to speaking: Facilitating recovery from nonfluent aphasia. *Future neurology* 5(5). 657–665.
- Schmader, Toni & Michael Johns. 2003. Converging evidence that stereotype threat reduces working memory capacity. *Journal of personality and social psychology* 85(3). 440–452.

- Schneider, Peter, Michael Scherg, H. Günter Dosch, Hans J. Specht, Alexander Gutschalk & André Rupp. 2002. Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature neuroscience* 5(7). 688–694.
- Schneider, Peter, Vanessa Sluming, Neil Roberts, Stefan Bleeck & André Rupp. 2005. Structural, functional, and perceptual differences in Heschl's gyrus and musical instrument preference. *Annals of the New York Academy of Sciences* 1060. 387–394.
- Schneiderman, E. I. & C. Desmarais. 1988. The talented language learner: Some preliminary findings. *Interlanguage studies bulletin (Utrecht)* 4(2). 91–109.
- Schön, Daniele, Cyrille Magne & Mireille Besson. 2004. The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology* 41(3). 341–349.
- Schulze, Katrin & Stefan Koelsch. 2012. Working memory for speech and music. *Annals of the New York Academy of Sciences* 1252. 229–236.
- Schuman, Howard & Stanley Presser. 1981. *Questions and answers in attitude surveys: Experiments on question form, wording, and context* (Quantitative studies in social relations). New York, NY: Acad. Press.
- Schwarz, Norbert & Daphna Oyserman. 2001. Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation* 22(2). 127–160.
- Seashore, Carl E. 1919a. *Manual of instructions and interpretations for measures of musical talent*. New York, NY: Columbia Graphophone.
- Seashore, Carl E. 1919b. *The psychology of musical talent*. Boston, MA: Silver, Burdett and Company.
- Seashore, Carl E. & M. Metfessel. 1925. Deviation from the regular as an art principle. *Proceedings of the National Academy of Sciences* 11(9). 538–542.
- Seither-Preisler, Annemarie, Richard Parncutt & Peter Schneider. 2014. Size and synchronization of auditory cortex promotes musical, literacy, and attentional skills in children. *Journal of Neuroscience* 34(33). 10937–10949.
- Seither-Preisler, Annemarie & Peter Schneider. 2015. Positive Effekte des Musizierens auf Wahrnehmung und Kognition aus neurowissenschaftlicher Perspektive. In Günther Bernatzky & Gunter Kreutz (eds.), *Musik und Medizin: Chancen für Therapie, Prävention und Bildung*, vol. 8, 375–393. Vienna: Springer Vienna.
- Seither-Preisler, Annemarie, Sabrina Turker, Susanne M. Reiterer, Markus Christiner & Peter Schneider. 2018. Neurobiologische Erkenntnisse zum Wechselspiel von

- musikalischer Begabung und Expertise sowie biografischen Einflüssen im Entwicklungsverlauf. *Journal für Begabtenförderung* 2. 15–24.
- Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics in Language Teaching* 10(1-4).
- Semal, Catherine, Laurent Demany, Kazuo Ueda & Pierre-André Hallé. 1996. Speech versus nonspeech in pitch memory. *Journal of the Acoustical Society of America* 100(2 Pt 1). 1132–1140.
- Sergeant, Desmond & Graham F. Welch. 2008. Age-related changes in long-term average spectra of children's voices. *Journal of Voice* 22(6). 658–670.
- Serrallach, Bettina, Christine Groß, Valdis Bernhofs, Dorte Engelmann, Jan Benner, Nadine Gündert, Maria Blatow, Martina Wengenroth, Angelika Seitz, Monika Brunner, Stefan Seither, Richard Parncutt, Peter Schneider & Annemarie Seither-Preisler. 2016. Neural biomarkers for dyslexia, ADHD, and ADD in the auditory cortex of children. *Frontiers in neuroscience* 10. 324.
- Sichivitsa, Veronica O. 2007. The influences of parents, teachers, peers and other factors on students' motivation in music. *Research Studies in Music Education* 29(1). 55–68.
- Skehan, Peter. 1991. Individual differences in second language learning. *Studies in Second Language Acquisition* 13(2). 275–298.
- Skehan, Peter. 2019. Theorising and updating aptitude. In Robert M. DeKeyser & Goretti Prieto Botana (eds.), *Doing SLA research with implications for the classroom* (Language Learning & Language Teaching 52), vol. 2, 69–93. Amsterdam: John Benjamins.
- Sloboda, John. 2005. *Exploring the musical mind: Cognition, emotion, ability, function*. Oxford: Oxford University Press.
- Sloboda, John. 2008. The ear of the beholder. *Nature* 454. 32–33.
- Sloboda, John A., Karen J. Wise & Isabelle Peretz. 2005. Quantifying tone deafness in the general population. *Annals of the New York Academy of Sciences* 1060. 255–261.
- Sparks, Richard & Leonore Ganschow. 2001. Aptitude for learning a foreign language. *Annual Review of Applied Linguistics* 21. 90–111.
- Sparks, Richard, Nancy Humbach, J. O. N. Patton & Leonore Ganschow. 2011. Subcomponents of second-language aptitude and second-language proficiency. *Modern Language Journal* 95(2). 253–273.

- Speciale, Giovanna, Nick C. Ellis & Tracey Bywater. 2004. Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics* 25(2). 293–321.
- Spielberger, Charles, Richard Gorsuch, Robert Lushene, P. R. Vagg & Gehard Jacobs. 1983. Manual for the state-trait anxiety inventory. Palo Alto, CA: Consulting Psychologists Press.
- Spolsky, Bernard. 1995. Prognostication and language aptitude testing, 1925-62. *ELT Journal* 12(3). 321–340.
- Stahl, Benjamin, Sonja A. Kotz, Ilona Henseler, Robert Turner & Stefan Geyer. 2011. Rhythm in disguise: Why singing may not hold the key to recovery from aphasia. *Brain* 134(Pt 10). 3083–3093.
- Stamou, Lelouda, Charles P. Schmidt & Jere T. Humphreys. 2010. Standardization of the Gordon primary measures of music audiation in Greece. *Journal of Research in Music Education* 58(1). 75–89.
- Stansfield, Charles W. 1988. Pimsleur language aptitude battery. In Daniel Keyser & Richard Sweetland (eds.), *Test Critiques*, 7th edn., 438–445. Kansas City, MO: Test Corporation of America.
- Stansfield, Charles W. & Daniel J. Reed. 2019. The MLAT at 60 years. In Zhisheng Wen, Peter Skehan, Adriana Biederon, Shaofeng Li & Richard L. Sparks (eds.), *Language aptitude: Advancing theory, testing, research and practice*, 15–32. New York, NY: Routledge.
- Stauffer, S. 1986. *An investigation of the effects of melodic and harmonic context on the development of singing ability in primary grade children*. Michigan, MI: University of Michigan Dissertation.
- Steinke, Willi R., Lola L. Cuddy & Lorna S. Jakobson. 2001. Dissociations among functional subsystems governing melody recognition after right-hemisphere damage. *Cognitive neuropsychology* 18(5). 411–437.
- Stocké, Volker & Christian Hunkler. 2007. Measures of desirability beliefs and their validity as indicators for socially desirable responding. *Field Methods* 19(3). 313–336.
- Strait, Dana L., Jane Hornickel & Nina Kraus. 2011. Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behavioral and brain functions* 7. 44.
- Sundberg, Johan. 1977. Vibrato and vowel identification. *Archives of Acoustics* (2). 257–266.

- Sundberg, Johan. 1988. *The science of the singing voice*. Dekalb, IL: Northern Illinois Univ.
- Sundberg, Johan. 1999. The perception of singing. In Diana Deutsch (ed.), *The psychology of music*, 2nd edn. (Academic Press series in cognition and perception), 171–212. San Diego, CA: Acad. Press.
- Swanson, H. Lee & Tracy P. Alloway. 2012. Working memory, learning, and academic achievement. In Karen R. Harris, Steve Graham & Timothy C. Urdan (eds.), *APA educational psychology handbook* (APA handbooks in psychology), 327–366. Washington, DC: American Psychological Association.
- Tafari, Johannella & Donatella Villa. 2002. Musical elements in the vocalisations of infants aged 2–8 months. *British Journal of Music Education* 19(01).
- Thompson, William F., E. Glenn Schellenberg & Gabriela Husain. 2004. Decoding speech prosody: Do music lessons help? *Emotion (Washington, D.C.)* 4(1). 46–64.
- Thorn, Annabel S. C. & Susan E. Gathercole. 2001. Language differences in verbal short-term memory do not exclusively originate in the process of subvocal rehearsal. *Psychonomic bulletin & review* 8(2). 357–364.
- Tiedemann, Joachim. 1989. Measures of cognitive styles: A critical review. *Educational Psychologist* 24(3). 261–275.
- Tillmann, Barbara, Katrin Schulze & Jessica M. Foxton. 2009. Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and cognition* 71(3). 259–264.
- Tourangeau, Roger & Tom W. Smith. 1996. Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly* 60(2). 275.
- Trehub, Sandra E. 2001. Musical predispositions in infancy. *Annals of the New York Academy of Sciences* 930. 1–16.
- Trehub, Sandra E., Judith Becker & Iain Morley. 2015. Cross-cultural perspectives on music and musicality. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 370(1664). 20140096.
- Trehub, Sandra E., David S. Hill & Stuart B. Kamenetsky. 1997. Parents' sung performances for infants. *Canadian journal of experimental psychology* 51(4). 385–396.
- Trehub, Sandra E. & Laurel J. Trainor. 1998. Singing to infants: Lullabies and playsongs. *Advances in Infancy Research* (12). 43–77.

- Tremblay-Champoux, Alexandra, Simone Dalla Bella, Jessica Phillips-Silver, Marie-Andrée Lebrun & Isabelle Peretz. 2010. Singing proficiency in congenital amusia: Imitation helps. *Cognitive neuropsychology* 27(6). 463–476.
- Tsang, Christine D., Rayna H. Friendly & Laurel J. Trainor. 2011. Singing development as a sensorimotor interaction problem. *Psychomusicology* 21(1-2). 31–44.
- Turker, Sabrina, Sabine Sommer-Lolei & Markus Christner. 2018. Sprachtalent und Musikgenie – zwei Seiten einer Münze? Zusammenspiel musikalischer und sprachlicher Fähigkeiten durch umsetzungsnahe Ideen im schulischen und familiären Bereich. *Journal für Begabtenförderung* 2. 53–60.
- Turner, Marilyn L. & Randall W. Engle. 1989. Is working memory capacity task dependent? *Journal of Memory and Language* 28(2). 127–154.
- Vulchanova, Mila, Camilla H. Foyn, Randi A. Nilsen & Hermundur Sigmundsson. 2014. Links between phonological memory, first language competence and second language competence in 10-year-old children. *Learning and Individual Differences* 35. 87–95.
- Wallentin, Mikkel, Andreas H. Nielsen, Morten Friis-Olivarius, Christian Vuust & Peter Vuust. 2010. The musical ear test, a new reliable test for measuring musical competence. *Learning and Individual Differences* 20(3). 188–196.
- Wapnick, Joel & Elizabeth Ekholm. 1997. Expert consensus in solo voice performance evaluation. *Journal of Voice* 11(4). 429–436.
- Watkins, Kate E., Antonio P. Strafella & Tomas Paus. 2003. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41(8). 989–994.
- Watts, Christopher, Kathryn Barnes-Burroughs, Mary Andrianopoulos & Megan Carr. 2003. Potential factors related to untrained singing talent: A survey of singing pedagogues. *Journal of Voice* 17(3). 298–307.
- Wechsler, David. 1939. *The measurement of adult intelligence*. Baltimore, MD: Williams & Wilkins.
- Welch, Graham F. 1986. A developmental view of children's singing. *British Journal of Music Education* 3(3). 295–303.
- Welch, Graham F. 1994. The assessment of singing. *Psychology of Music* 22(1). 3–19.
- Welch, Graham F. 1998. Early childhood musical development. *Research Studies in Music Education* 11(1). 27–41.

- Welch, Graham F. 2006. Singing and vocal development. In Gary McPherson (ed.), *The child as musician: A handbook of musical development*, 311–330. Oxford & New York, NY: Oxford University Press.
- Welch, Graham F. 2007. Singing as communication. In Dorothy Miell (ed.), *Musical communication*, 239–260. Oxford: Oxford University Press.
- Welch, Graham F., E. Himonides, J. Saunders, I. Papageorgi, T. Rinta, C. Preti, C. Stewart, J. Lani & J. Hill. 2011. Researching the first year of the national singing programme sing up in england: An initial impact evaluation. *Psychomusicology* 21(1-2). 83–97.
- Wen, Zhisheng. 2012. Foreign language aptitude. *ELT Journal* 66(2). 233–235.
- Wen, Zhisheng, Adriana Biedroń & Peter Skehan. 2017. Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching* 50(1). 1–31.
- Wen, Zhisheng & Peter Skehan. 2011. A new perspective on foreign language aptitude research: Building and supporting a case for "working memory as language aptitude". *Ilha do Desterro* 60. 15–44.
- Wen, Zhisheng, Peter Skehan, Adriana Biederon, Shaofeng Li & Richard L. Sparks (eds.). 2019. *Language aptitude: Advancing theory, testing, research and practice*. New York, NY: Routledge.
- Wengenroth, Martina, Maria Blatow, Armin Heinecke, Julia Reinhardt, Christoph Stippich, Elke Hofmann & Peter Schneider. 2014. Increased volume and function of right auditory cortex as a marker for absolute pitch. *Cerebral cortex* 24(5). 1127–1137.
- Werker, Janet F. & Richard C. Tees. 2005. Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental psychobiology* 46(3). 233–251.
- Williamson, Victoria J., Alan D. Baddeley & Graham J. Hitch. 2010a. Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Memory & cognition* 38(2). 163–175.
- Williamson, Victoria J., Claire McDonald, Diana Deutsch, Timothy D. Griffiths & Lauren Stewart. 2010b. Faster decline of pitch memory over time in congenital amusia. *Advances in cognitive psychology* 6. 15–22.
- Wing, Herbert D. 1948. *Tests of musical ability and appreciation: An investigation into the measurement, distribution, and development of musical capacity*. London: Cambridge University Press.
- Wing, Herbert D. 1962. A revision of the Wing musical aptitude test. *Journal of Research in Music Education* 10(1). 39–46.

- Wong, Patrick C. M. & Tyler K. Perrachione. 2007. Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics* 28(4). 565–585.
- Wong, Patrick C. M., Erika Skoe, Nicole M. Russo, Tasha Dees & Nina Kraus. 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience* 10(4). 420–422.
- Woodrow, Lindy. 2006. Anxiety and speaking English as a second language. *RELC Journal* 37(3). 308–328.
- Yoshiura, Takashi, Shuichi Higano, Ana Rubio, David A. Shrier, Wingchi E. Kwok, Soji Iwanaga & Yuji Numaguchi. 2000. Heschl and superior temporal gyri: Low signal intensity of the cortex on T2-weighted MR images of the normal brain. *Radiology* 214(1). 217–221.
- Young, William T. 1973. The Bentley “measures of musical abilities”: A congruent validity report. *Journal of Research in Music Education* 21(1). 74–79.
- Zafar, Shahila & K. Meenakshi. 2012. Individual learner differences and second language acquisition: A review. *Journal of Language Teaching and Research* 3(4).
- Zatorre, Robert J., Joyce L. Chen & Virginia B. Penhune. 2007. When the brain plays music: Auditory-motor interactions in music perception and production. *Nature reviews. Neuroscience* 8(7). 547–558.
- Zuk, Jennifer, Christopher Benjamin, Arnold Kenyon & Nadine Gaab. 2014. Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PloS one* 9(6). e99868.

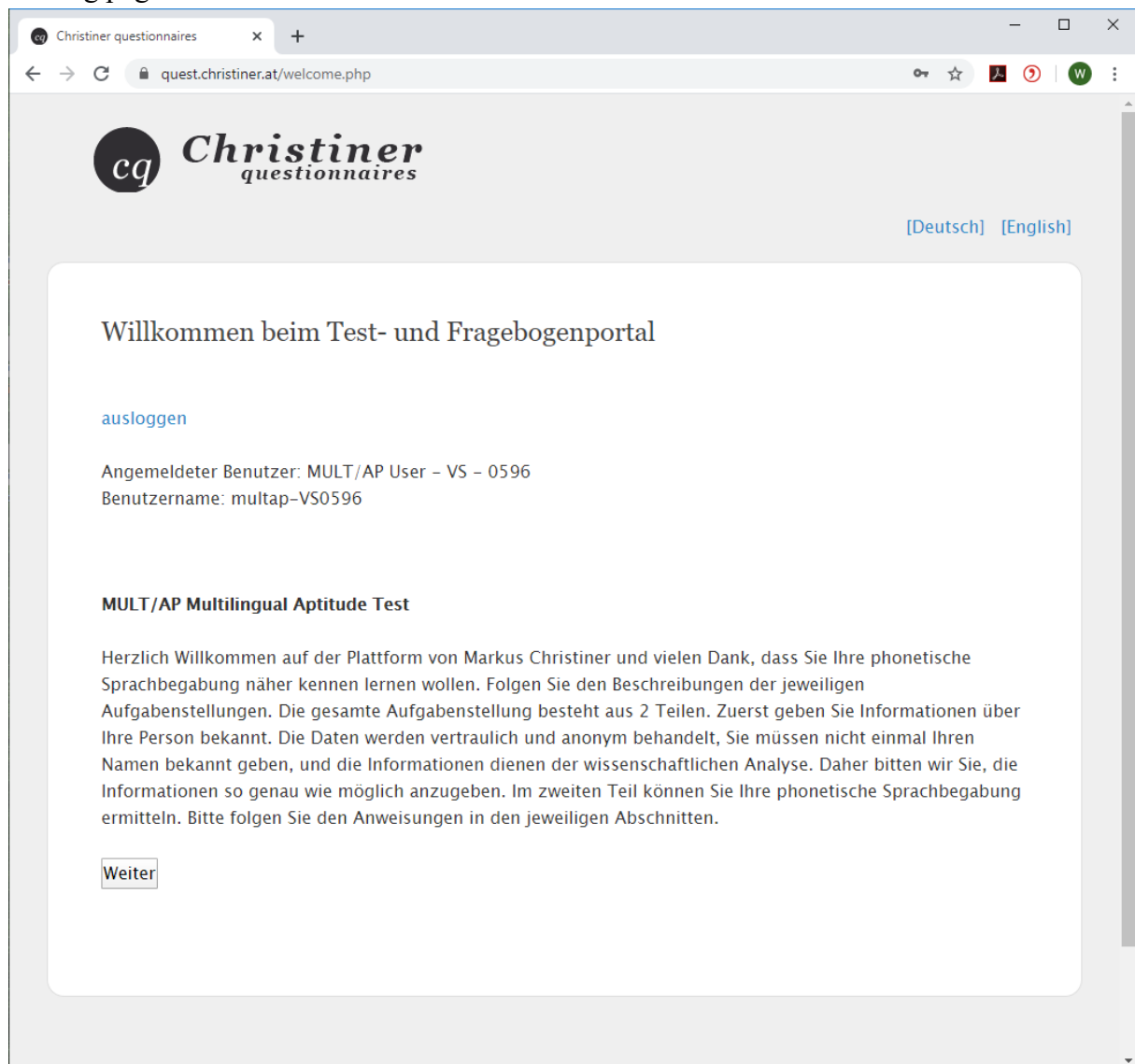
Appendix

Appendix 1. Questionnaire development (Screenshots)

A1.1. Participants

A1.1.1. Factual questions (selected screenshots)

Starting page



Overview

The screenshot shows a web browser window with the address bar displaying "quest.christiner.at/schedule.php". The page header features the "cq Christiner questionnaires" logo on the left and language links "[Deutsch]" and "[English]" on the right. The main content area is titled "MULT/AP Multilingual Aptitude Test". Below this, there is a section "Persönliche Informationen" with a text prompt "In diesem Teil geben Sie uns bitte ein paar Informationen über sich bekannt." followed by a button labeled "Jetzt eingeben". The next section is "MULT/AP Test ABC-VS", which contains a detailed paragraph about the test format: "Sie hören im folgenden Abschnitt eine Kette von aneinander gereihten Wörtern in Fremdsprachen (Multilinguale Sprachkette), die Sie möglicherweise noch nie gehört haben. Danach hören Sie entweder eine, zwei oder drei aufeinanderfolgende Vergleichsphrasen/Wörter (Vergleichsphrasen), die entweder in der zuvor gehörten Kette inkludiert waren oder nicht. Sie sollen danach sagen, ob diese Phrase bzw. diese Phrasen in der zuvor gehörten Kette tatsächlich inkludiert waren oder nicht und wie sicher Sie sich sind. Wenn die Phrase bzw. die Phrasen inkludiert waren, wählen Sie „ja“ wenn nicht, dann „nein“. Waren nicht alle Phrasen oder keine Phrase enthalten, dann ist mit „nein“ zu antworten. Danach wählen Sie, wie sicher Sie sich sind. Wenn Sie sich sehr sicher sind, wählen Sie „sehr sicher“ aus, wenn Sie ziemlich sicher sind, wählen Sie „ziemlich sicher“ aus, und wenn Sie unsicher sind, wählen Sie „weiß nicht“ aus. Sie können jedes Vergleichspaar nur einmal hören. Nachdem Sie Ihre Ergebnisse eingetragen haben, geht es weiter in die nächste Runde. Es sind 37 Runden bis diese Aufgabenstellung erledigt ist." At the bottom of this section are two buttons: "Ausprobieren" and "Jetzt eingeben". A "Startseite" link is located at the bottom left of the page.

Christiner questionnaires

quest.christiner.at/schedule.php

cq Christiner questionnaires

[\[Deutsch\]](#) [\[English\]](#)

MULT/AP Multilingual Aptitude Test

Persönliche Informationen

In diesem Teil geben Sie uns bitte ein paar Informationen über sich bekannt. [Jetzt eingeben](#)

MULT/AP Test ABC-VS

Sie hören im folgenden Abschnitt eine Kette von aneinander gereihten Wörtern in Fremdsprachen (**Multilinguale Sprachkette**), die Sie möglicherweise noch nie gehört haben. Danach hören Sie entweder eine, zwei oder drei aufeinanderfolgende Vergleichsphrasen/Wörter (**Vergleichsphrasen**), die entweder in der zuvor gehörten Kette inkludiert waren oder nicht. Sie sollen danach sagen, ob diese Phrase bzw. diese Phrasen in der zuvor gehörten Kette tatsächlich inkludiert waren oder nicht und wie sicher Sie sich sind. Wenn die Phrase bzw. die Phrasen inkludiert waren, wählen Sie „ja“ wenn nicht, dann „nein“. Waren nicht alle Phrasen oder keine Phrase enthalten, dann ist mit „nein“ zu antworten. Danach wählen Sie, wie sicher Sie sich sind. Wenn Sie sich sehr sicher sind, wählen Sie „sehr sicher“ aus, wenn Sie ziemlich sicher sind, wählen Sie „ziemlich sicher“ aus, und wenn Sie unsicher sind, wählen Sie „weiß nicht“ aus. Sie können jedes Vergleichspaar nur einmal hören. Nachdem Sie Ihre Ergebnisse eingetragen haben, geht es weiter in die nächste Runde. Es sind 37 Runden bis diese Aufgabenstellung erledigt ist.

[Ausprobieren](#) [Jetzt eingeben](#)

[Startseite](#)

Step 1 / gender

The screenshot shows a web browser window with the address bar displaying "quest.christiner.at/EInfo_eingeben.php". The page header includes the "cq Christiner questionnaires" logo and language links "[Deutsch]" and "[English]". The main content area is titled "Eingabe 1 von 22" and contains the instruction "Bitte geben Sie Ihr Geschlecht an." (Please specify your gender). Below this, there are two radio button options: "weiblich" (female) and "männlich" (male). At the bottom of the form, there is a "Weiter" (Next) button.

Step 2 / age

The screenshot shows the same web browser window as Step 1, but the page is titled "Eingabe 2 von 22". The instruction "Bitte geben Sie Ihr Alter in Jahren an." (Please specify your age in years) is displayed. Below this, a red error message "Bitte eine ganze Zahl eingeben." (Please enter a whole number) is shown above an empty text input field. A "Weiter" (Next) button is located at the bottom of the form.

Step 5 / education

Christiner questionnaires

quest.christiner.at/ElInfo_eingeben.php

Eingabe 5 von 22

Bitte geben sie Ihren höchsten Bildungsabschluss an.

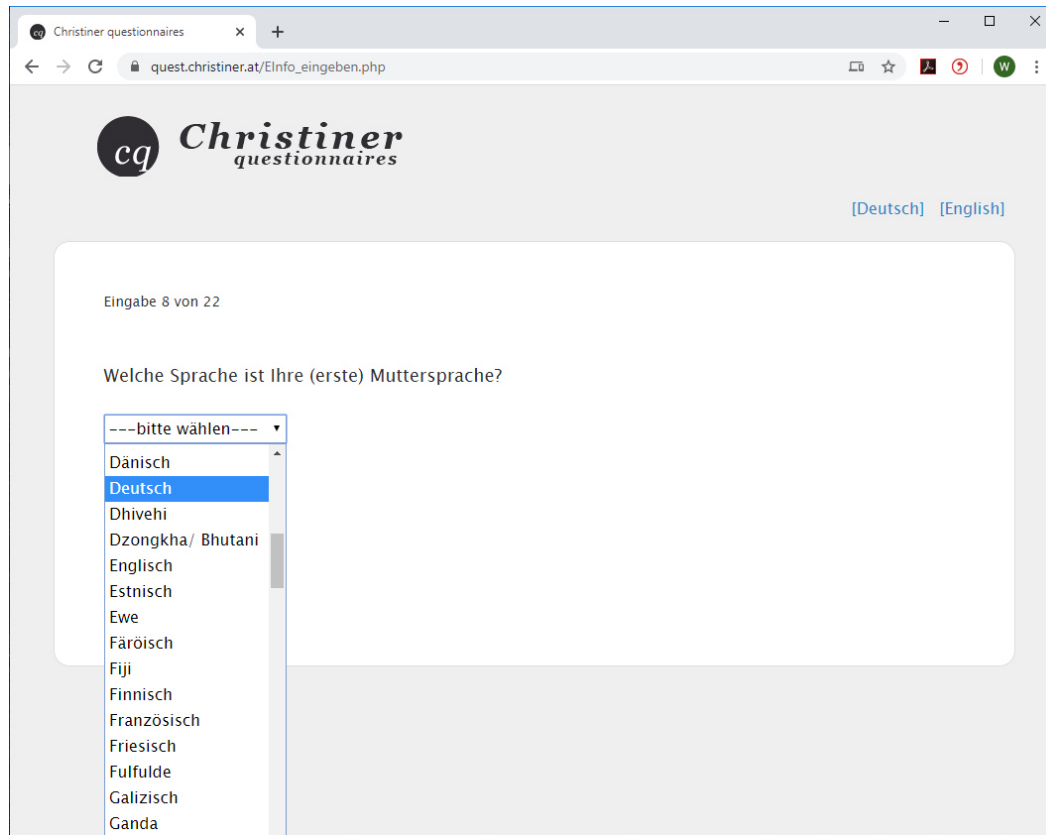
Trifft keine Bezeichnung exakt zu, wählen Sie bitte den am ehesten passenden Bildungsabschluss.

- ☐ keine Schulbildung
- ☐ **D (Deutschland):** Grundschule
Ö (Österreich): Volksschule
- ☐ **D:** Hauptschul-/Realschulabschluss, Gymnasium/Gesamtschule (ohne Abitur)
Ö: Hauptschule, AHS-Unterstufe, Neue Mittelschule
- ☐ **D:** Berufsschule, Berufsfachschule, Lehrabschluss
Ö: Lehrabschlussprüfung, berufsbildende mittlere Schulen (BMS), z.B. HAS
- ☐ **D:** Abitur (Gymnasium, Gesamtschule), Fachoberschule/Berufsfachschule mit Fachhochschulreife/Hochschulreife
Ö: AHS-Oberstufe mit Matura, Matura Abendschule, Höhere berufsbildende Schulen mit Reife- und Diplomprüfung, Berufsreifeprüfung
- ☐ **D:** Kollegs, 2- und 3-jährige Programme an Ausbildungsstätten / Schulen für Gesundheits- und Sozialberufe, Berufsfachschule mit Berufsabschluss
Ö: Akademien (PädAk, SozAk), Kollegs, Lehramtsprüfung für allgemeinbildende Pflichtschulen, Gesundheits- und Krankenpflegeschule
- ☐ **D:** Meisterausbildung
Ö: Meister- und Werkmeisterausbildung, Bauhandwerkerschulen (Meisterprüfung)
- ☐ **D:** Bachelor (Universität, Fachhochschule, Diplomstudiengang)
Ö: Fachhochschule, Universität (Bachelor)
- ☐ **D:** Fachhochschule, Universität (Master, Diplom, Staatsprüfung)
Ö: Fachhochschule, Universität (Magister, Diplomingenieur, Master, MSc, MBA,...)
- ☐ **D:** Promotion
Ö: Doktorat

Weiter

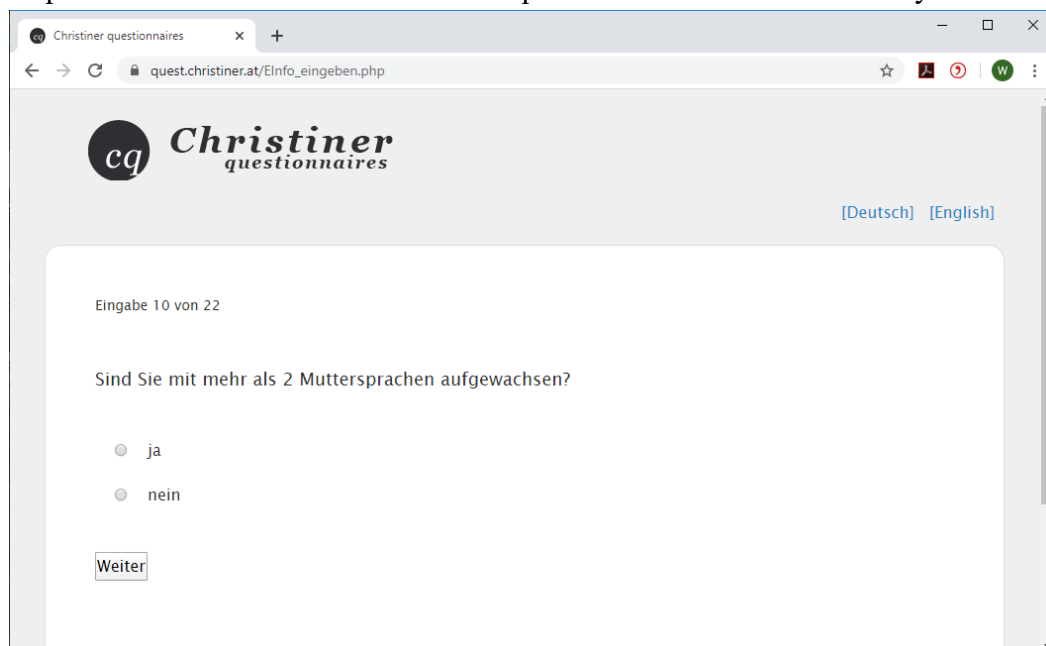
Step 8 / mother language

All spoken languages according to ISO-639-1



The screenshot shows a web browser window with the URL `quest.christiner.at/Einfo_eingeben.php`. The page header includes the logo "cq Christiner questionnaires" and language links "[Deutsch]" and "[English]". The main content area displays "Eingabe 8 von 22" and the question "Welche Sprache ist Ihre (erste) Muttersprache?". Below the question is a dropdown menu with the placeholder text "---bitte wählen---". The menu is open, showing a list of languages: Dänisch, Deutsch (highlighted), Dhivehi, Dzongkha/ Bhutani, Englisch, Estnisch, Ewe, Färöisch, Fiji, Finnisch, Französisch, Friesisch, Fulfulde, Galizisch, and Ganda.

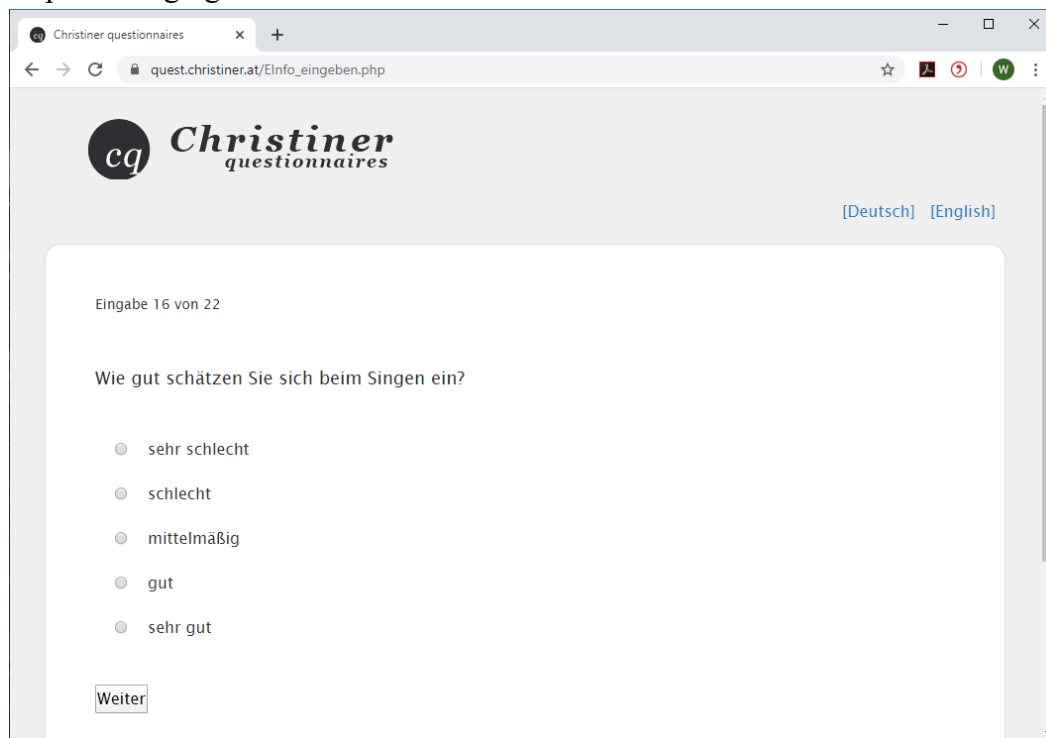
Step 10 / Control variable if the former question was answered correctly



The screenshot shows the same web browser window as before, but at a different point in the questionnaire. The page header is identical. The main content area displays "Eingabe 10 von 22" and the question "Sind Sie mit mehr als 2 Muttersprachen aufgewachsen?". Below the question are two radio button options: "ja" and "nein". At the bottom left of the form is a button labeled "Weiter".

A1.1.2. Singing self-estimation

Step 16 / singing self-estimation



Christiner questionnaires

quest.christiner.at/Einfo_eingeben.php

cq Christiner questionnaires

[Deutsch] [English]

Eingabe 16 von 22

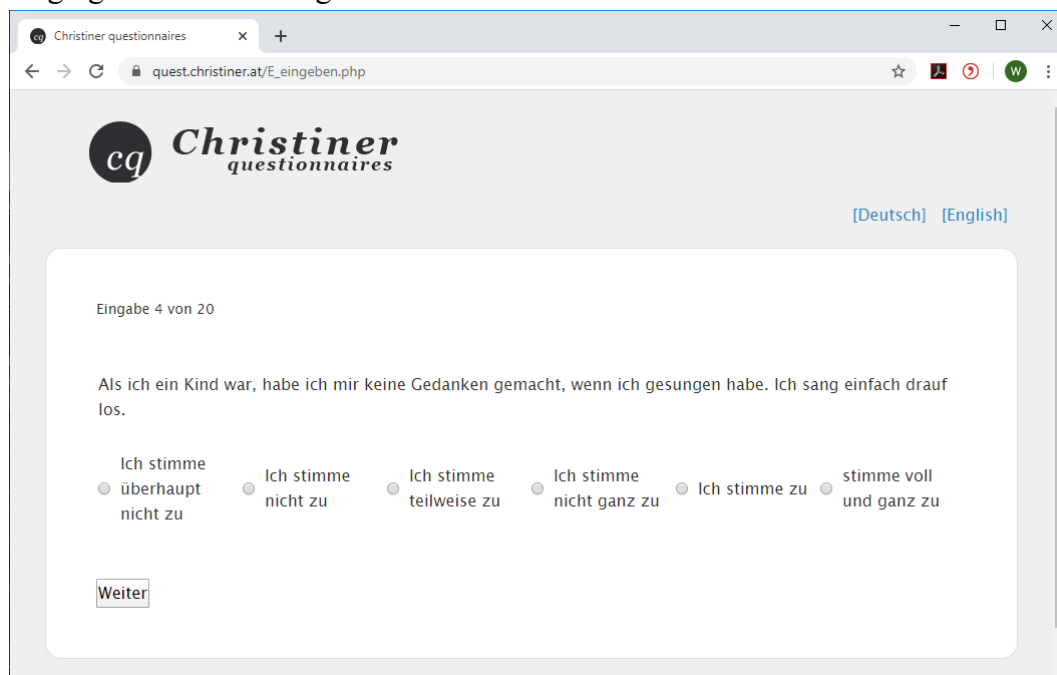
Wie gut schätzen Sie sich beim Singen ein?

- ☐ sehr schlecht
- ☐ schlecht
- ☐ mittelmäßig
- ☐ gut
- ☐ sehr gut

Weiter

A1.1.3. Singing behaviour during childhood and performance anxiety

Singing behaviour during childhood



Christiner questionnaires

quest.christiner.at/E_eingeben.php

cq Christiner questionnaires

[Deutsch] [English]

Eingabe 4 von 20

Als ich ein Kind war, habe ich mir keine Gedanken gemacht, wenn ich gesungen habe. Ich sang einfach drauf los.

- ☐ Ich stimme überhaupt nicht zu
- ☐ Ich stimme nicht zu
- ☐ Ich stimme teilweise zu
- ☐ Ich stimme nicht ganz zu
- ☐ Ich stimme zu
- ☐ stimme voll und ganz zu

Weiter

Performance anxiety

The screenshot shows a web browser window with the address bar displaying 'quest.christiner.at/E_eingeben.php'. The page header includes the 'cq Christiner questionnaires' logo and language links '[Deutsch]' and '[English]'. The main content area is titled 'Eingabe 2 von 20'. It contains a statement: 'Mir ist es lieber, wenn ich zuhause und alleine singe, wo mich niemand hört.' Below this statement is a horizontal rating scale with six radio button options: 'Ich stimme überhaupt nicht zu', 'Ich stimme nicht zu', 'Ich stimme teilweise zu', 'Ich stimme nicht ganz zu', 'Ich stimme zu', and 'stimme voll und ganz zu'. A 'Weiter' button is located at the bottom left of the question area.

A1.2. Ratings speech

Rating Russian / starting page

The screenshot shows a web browser window with the address bar displaying 'quest.christiner.at/welcome.php'. The page header includes the 'cq Christiner questionnaires' logo and language links '[Deutsch]' and '[English]'. The main content area is titled 'Welcome to the online survey'. It contains a 'log out' link, the text 'Registered user: Rating Russian – rater 06' and 'Username: PD-rater-ru-06', and a section titled 'Rating Russian'. Below this section is a welcome message: 'Welcome on the platform of Markus Christiner. Please press the continue button. You will receive further information and instructions to the following tasks.' A 'Continue' button is located at the bottom left of the main content area.

Overview rating Russian

Christiner questionnaires

quest.christiner.at/schedule.php

☆

W

 **Christiner**
questionnaires

[\[Deutsch\]](#) [\[English\]](#)

Ratings for foreign speech imitation

Dear participant,

I want to thank you for your valuable help and your contribution to rate speech files. Please follow the entire instructions. In the following sections you should evaluate how well adults imitated new speech material they have never heard before by indicating it with a number. 0 is the worst score someone can get, while 10 is the highest score for the best performance. Most of the speakers have not spoken or have been thought in this language before. Thus it may be that utterances may not be understood clearly. The original sentences have a certain length which may be one reason why some sentences may not be meaningful to you as native speakers. Your task is to rate how well the participants were able to imitate the original sound files (which is always the first one on the top) and whether the content of the sentence spoken by the participants is intelligible. Thank you for your professional work.

To consider is that

- participants listened 3 times the original file before they repeated what they were listening to. The participants were instructed to repeat the original file one time only.
- You should rate one section within a single sitting. Each section is comprised of around 80 files.

If you have questions or difficulties please contact the experimenter, Markus Christiner :
markus.christiner@univie.ac.at

Rating Russian 1

Enter now

Rating an item

Christiner questionnaires

quest.christiner.at/RE_II_eingeben.php

cq

Christiner
questionnaires

[\[Deutsch\]](#) [\[English\]](#)

Step 1 of 87

▶ 0:00 / 0:04

This is the original soundfile the participant was listening to:

▶ 0:00 / 0:02

This is the soundfile of the participant:

You should rate the soundfile of the participant now!

The scale is ranging between 0 and 10. 10 is the highest score someone can receive. 0 is the lowest score someone can receive.

Score I: The overall imitation score

lowest

0

1

2

3

4

5

6

7

8

9

10

highest

Score II: Intelligibility

Have you understood the content?

not at all

0

1

2

3

4

partly

5

6

7

8

completely

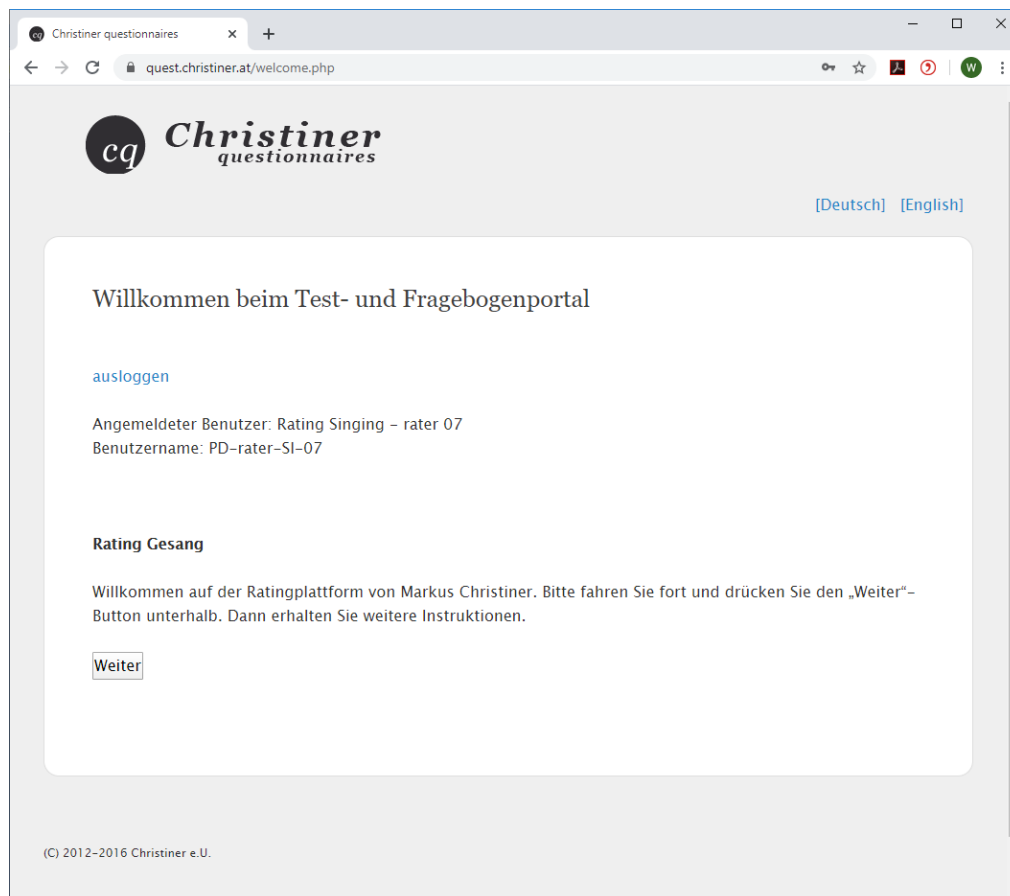
9

10

Continue

A1.3. Rating singing

Rating singing / starting page



Overview rating singing

Christiner questionnaires x +

quest.christiner.at/schedule.php

Liebe Rater!

Bevor Sie die Gesangsaufnahmen, die Sie in Kürze hören werden, bewerten, sollten Sie diese Anweisungen durchlesen.

Möglicherweise ist eine Frage von Ihnen: „Was soll ich eigentlich bewerten?“

Zunächst ist eine kritische Beurteilung notwendig.

Generell sollten Sie Ihren eigenen Eindruck von der gehörten Aufnahme bewerten. Sie können dafür zwischen 0 und 10 Punkte vergeben, wobei 10 die höchste Punktezahl ist, die man für das jeweilige Kriterium erhalten kann. 0 hingegen steht für eine sehr schlechte Bewertung. Wählen Sie eine angemessene Punktezahl für jedes Kriterium pro Aufnahme aus. Sie müssen für alle 4 Kriterien Punkte vergeben, um den nächsten Teilnehmer bewerten zu können. **Bitte bewerten Sie ganz schnell und ohne lange zu zögern.**

Die Teilnehmer mussten einen Teil eines ihnen unbekannten Liedes nach dreimaligem Hören wiedergeben. Dabei hatten sie keinen Text zur Verfügung und mussten rein aus ihrer Erinnerung diese Aufgabe erledigen. Das Lied durfte in einer anderen Tonhöhe gesungen werden. (Anmerkung: Der ursprüngliche Ausschnitt ist in den folgenden Schritten immer derselbe.)

Melodie:
Wie gut wurde die Melodie nachempfunden?

Stimmqualität:
Vibrato, Color, Resonanz etc.
Hat der/die SängerIn eine angenehme, aussagekräftige Stimme etc.?

Rhythmus:
Wie gut wurde die Rhythmik wiedergegeben. Beziehungsweise wie rhythmisch sind die nachgesungenen Phrasen in sich.

Texttreue:
Wie gut wurde der Text wiedergegeben?

Danke für Ihre Teilnahme!

Rating an item

Eingabe 1 von 87

Das ist die originale Sounddatei, die angehört wurde:

Das ist die Aufnahme des Teilnehmers:

Liedtext: Whenever I miss, whenever I miss, I miss your smiling.

Bitte bewerten Sie jetzt die Aufnahme des Teilnehmers!
Bitte bewerten Sie auf einer Skala von 0 bis 10. Der schlechteste Wert ist 0 und der beste ist 10.

Bitte bewerten Sie nach den untenstehenden Kriterien.

Melodie
sehr schlecht sehr gut
0 1 2 3 4 5 6 7 8 9 10

Stimmqualität
sehr schlecht sehr gut
0 1 2 3 4 5 6 7 8 9 10

Rhythmus
sehr schlecht sehr gut
0 1 2 3 4 5 6 7 8 9 10

Texttreue
sehr schlecht sehr gut
0 1 2 3 4 5 6 7 8 9 10

Weiter

Appendix 2. MULT AP development

A2.1. Overview Stimlines

11 Pages database output

Multap - VS

Stimline 2

S-Id	SCS	SCNR	Pos	Language	Syll	Stim-No	Identifier	f_c	SL-Pos
SL02									
	0	0	1	Russian	3	3	RU-03-03		1
	0	0	2	Persian	3	2	FA-03-02		2
	0	0	3	Tagalog	3	3	TL-03-03		3
	0	0	4	Mandarin	3	3	ZH-03-03		4
	0	0	5	Hindi	3	3	HI-03-03		5
	0	0	6	Persian	3	3	FA-03-03		6
	0	0	7	Russian	3	2	RU-03-02		7
	0	0	8	Tagalog	3	2	TL-03-02		8
	0	0	9	Thai	3	3	TH-03-03		9
	0	0	10	Tagalog	3	1	TL-03-01		10
SC02-A2									
	f								
	A	2	1	Persian	3	1	FA-03-01	f	
SC02-B2									
	f								
	B	2	1	Russian	3	1	RU-03-01	f	
	B	2	2	Tagalog	3	1	TL-03-01	f	10

Multap - VS

Stimline 3

S-Id	SCS	SCNR	Pos	Language	Syll	Stim-No	Identifier	f_c	SL-Pos
SL03									
	0	0	1	Thai	3	3	TH-03-03		1
	0	0	2	Thai	3	2	TH-03-02		2
	0	0	3	Japanese	3	3	JA-03-03		3
	0	0	4	Mandarin	3	3	ZH-03-03		4
	0	0	5	Hindi	3	1	HI-03-01		5
	0	0	6	Hindi	3	2	HI-03-02		6
	0	0	7	Persian	3	3	FA-03-03		7
	0	0	8	Tagalog	3	3	TL-03-03		8
	0	0	9	Tagalog	3	2	TL-03-02		9
	0	0	10	Persian	3	1	FA-03-01		10
	0	0	11	Mandarin	3	2	ZH-03-02		11
	0	0	12	Hindi	3	3	HI-03-03		12
SC03-B1									
	<u>c</u>								
	B	1	1	Mandarin	3	3	ZH-03-03	c	4
	B	1	2	Hindi	3	1	HI-03-01	c	5

Multap - VS

Stimline 5

S-Id	SCS	SCNR	Pos	Language	Syll	Stim-No	Identifier	f_c	SL-Pos
SL05									
	0	0	1	Hindi	4	2	HI-04-02		1
	0	0	2	Hindi	4	3	HI-04-03		2
	0	0	3	Japanese	4	3	JA-04-03		3
	0	0	4	Thai	4	3	TH-04-03		4
	0	0	5	Thai	4	2	TH-04-02		5
	0	0	6	Tagalog	4	3	TL-04-03		6
	0	0	7	Russian	4	2	RU-04-02		7
	0	0	8	Japanese	4	2	JA-04-02		8
	0	0	9	Hindi	4	1	HI-04-01		9
	0	0	10	Mandarin	4	3	ZH-04-03		10
SC05-A1									
	ε								
	A	1	1	Russian	4	2	RU-04-02	c	7
SC05-D1									
	ε								
	D	1	1	Japanese	4	3	JA-04-03	c	3
	D	1	2	Russian	4	2	RU-04-02	c	7
	D	1	3	Japanese	4	2	JA-04-02	c	8

Multap - VS

Stimline 6

S-Id	SCS	SCNR	Pos	Language	Syll	Stim-No	Identifier	f_c	SL-Pos
SL06	0	0	1	Mandarin	4	3	ZH-04-03		1
	0	0	2	Mandarin	4	2	ZH-04-02		2
	0	0	3	Tagalog	4	3	TL-04-03		3
	0	0	4	Thai	4	1	TH-04-01		4
	0	0	5	Thai	4	2	TH-04-02		5
	0	0	6	Hindi	4	1	HI-04-01		6
	0	0	7	Hindi	4	2	HI-04-02		7
	0	0	8	Japanese	4	2	JA-04-02		8
	0	0	9	Persian	4	1	FA-04-01		9
	0	0	10	Persian	4	2	FA-04-02		10
	0	0	11	Russian	4	2	RU-04-02		11
	0	0	12	Tagalog	4	2	TL-04-02		12
SC06-B2	f	B	2	1	Mandarin	4	1	ZH-04-01	f
		B	2	2	Hindi	4	2	HI-04-02	f 7

Multap - VS

Stimline 8

S-Id	SCS	SCNR	Pos	Language	Syll	Stim-No	Identifier	f_c	SL-Pos
SL08	0	0	1	Russian	5	3	RU-05-03		1
	0	0	2	Persian	5	2	FA-05-02		2
	0	0	3	Persian	5	1	FA-05-01		3
	0	0	4	Tagalog	5	2	TL-05-02		4
	0	0	5	Thai	5	2	TH-05-02		5
	0	0	6	Mandarin	5	1	ZH-05-01		6
	0	0	7	Hindi	5	1	HI-05-01		7
	0	0	8	Japanese	5	3	JA-05-03		8
	0	0	9	Persian	5	3	FA-05-03		9
	0	0	10	Russian	5	2	RU-05-02		10
SC08-A1	c	A	1	1	Tagalog	5	2	TL-05-02	c 4

Multap - VS

Stimline 11

S-Id	SCS	SCNR	Pos	Language	Syll	Stim-No	Identifier	f_c	SL-Pos
SL11									
	0	0	1	Mandarin	6	2	ZH-06-02		1
	0	0	2	Mandarin	5	3	ZH-05-03		2
	0	0	3	Thai	5	2	TH-05-02		3
	0	0	4	Thai	7	2	TH-07-02		4
	0	0	5	Thai	3	3	TH-03-03		5
	0	0	6	Thai	9	2	TH-09-02		6
	0	0	7	Mandarin	7	1	ZH-07-01		7
	0	0	8	Thai	5	1	TH-05-01		8
	0	0	9	Thai	9	1	TH-09-01		9
	0	0	10	Mandarin	3	2	ZH-03-02		10
SC11-A1 <u>ç</u>									
	A	1	1	Mandarin	5	3	ZH-05-03	c	2
SC11-A2 <u>f</u>									
	A	2	1	Mandarin	4	1	ZH-04-01	f	

Multap - VS

Stimline 13

S-Id	SCS	SCNR	Pos	Language	Syll	Stim-No	Identifier	f_c	SL-Pos
SL13									
	0	0	1	Tagalog	5	1	TL-05-01		1
	0	0	2	Mandarin	3	3	ZH-03-03		2
	0	0	3	Hindi	3	5	HI-03-05		3
	0	0	4	Japanese	4	3	JA-04-03		4
	0	0	5	Persian	1	2	FA-01-02		5
	0	0	6	Russian	2	2	RU-02-02		6
	0	0	7	Tagalog	2	2	TL-02-02		7
	0	0	8	Tagalog	5	3	TL-05-03		8
	0	0	9	Thai	3	3	TH-03-03		9
	0	0	10	Tagalog	9	2	TL-09-02		10
	0	0	11	Tamil	11	2	TA-11-02		11
	0	0	12	Mandarin	5	1	ZH-05-01		12
SC13-A1									
	c								
	A	1	1	Tagalog	5	3	TL-05-03	c	8
SC13-B2									
	f								
	B	2	1	Mandarin	4	1	ZH-04-01	f	
	B	2	2	Mandarin	5	1	ZH-05-01	f	12
SC13-D1									
	c								
	D	1	1	Hindi	3	5	HI-03-05	c	3
	D	1	2	Persian	1	2	FA-01-02	c	5
	D	1	3	Tagalog	5	3	TL-05-03	c	8

Multap - VS

Stimline 14

S-Id	SCS	SCNR	Pos	Language	Syll	Stim-No	Identifier	f_c	SL-Pos
SL14									
	0	0	1	Mandarin	3	2	ZH-03-02		1
	0	0	2	Mandarin	7	2	ZH-07-02		2
	0	0	3	Mandarin	9	3	ZH-09-03		3
	0	0	4	Mandarin	9	2	ZH-09-02		4
	0	0	5	Mandarin	6	2	ZH-06-02		5
	0	0	6	Mandarin	5	3	ZH-05-03		6
	0	0	7	Mandarin	4	3	ZH-04-03		7
	0	0	8	Mandarin	3	3	ZH-03-03		8
SC14-A2									
	f								
	A	2	1	Mandarin	4	2	ZH-04-02	f	
SC14-B1									
	c								
	B	1	1	Mandarin	3	3	ZH-03-03	c	8
	B	1	2	Mandarin	6	2	ZH-06-02	c	5
SC14-B2									
	f								
	B	2	1	Mandarin	4	1	ZH-04-01	f	
	B	2	2	Mandarin	5	3	ZH-05-03	f	6
SC14-D1									
	c								
	D	1	1	Mandarin	7	2	ZH-07-02	c	2
	D	1	2	Mandarin	9	2	ZH-09-02	c	4
	D	1	3	Mandarin	5	3	ZH-05-03	c	6
SC14-D2									
	c								
	D	2	1	Mandarin	6	2	ZH-06-02	c	5
	D	2	2	Mandarin	9	2	ZH-09-02	c	4
	D	2	3	Mandarin	4	3	ZH-04-03	c	7

Multap - VS

Stimline 15

S-Id	SCS	SCNR	Pos	Language	Syll	Stim-No	Identifier	f_c	SL-Pos
SL15									
	0	0	1	Tagalog	3	1	TL-03-01		1
	0	0	2	Tagalog	5	1	TL-05-01		2
	0	0	3	Tagalog	7	1	TL-07-01		3
	0	0	4	Tagalog	5	2	TL-05-02		4
	0	0	5	Tagalog	9	2	TL-09-02		5
	0	0	6	Tagalog	7	3	TL-07-03		6
	0	0	7	Tagalog	4	3	TL-04-03		7
	0	0	8	Tagalog	7	2	TL-07-02		8
SC15-A2									
	<u>f</u>								
	A	2	1	Tagalog	5	3	TL-05-03	f	
SC15-A3									
	<u>f</u>								
	A	2	1	Tagalog	3	2	TL-03-02	f	
SC15-B2									
	<u>f</u>								
	B	2	1	Tagalog	3	2	TL-03-02	f	
	B	2	2	Tagalog	4	3	TL-04-03	f	7
SC15-D1									
	<u>c</u>								
	D	1	1	Tagalog	5	1	TL-05-01	c	2
	D	1	2	Tagalog	5	2	TL-05-02	c	4
	D	1	3	Tagalog	7	3	TL-07-03	c	6
SC15-D2									
	<u>f</u>								
	D	2	1	Tagalog	5	2	TL-05-02	f	4
	D	2	2	Tagalog	7	3	TL-07-03	f	6
	D	2	3	Tagalog	9	1	TL-09-01	f	

Multap - VS

Stimline 16

S-Id	SCS	SCNR	Pos	Language	Syll	Stim-No	Identifier	f_c	SL-Pos
SL16									
	0	0	1	Japanese	5	2	JA-05-02		1
	0	0	2	Japanese	7	2	JA-07-02		2
	0	0	3	Japanese	7	1	JA-07-01		3
	0	0	4	Japanese	4	2	JA-04-02		4
	0	0	5	Japanese	9	1	JA-09-01		5
	0	0	6	Japanese	7	3	JA-07-03		6
	0	0	7	Japanese	4	3	JA-04-03		7
	0	0	8	Japanese	5	3	JA-05-03		8
SC16-A1									
	<u>c</u>								
	A	1	1	Japanese	5	2	JA-05-02	c	1
SC16-A2									
	<u>f</u>								
	A	2	1	Japanese	5	1	JA-05-01	f	
SC16-B3									
	<u>c</u>								
	B	3	1	Japanese	5	2	JA-05-02	c	1
	B	3	2	Japanese	7	3	JA-07-03	c	6
SC16-D3									
	<u>c</u>								
	D	3	1	Japanese	4	2	JA-04-02	c	4
	D	3	2	Japanese	7	3	JA-07-03	c	6
	D	3	3	Japanese	5	2	JA-05-02	c	1
SC16-D4									
	<u>f</u>								
	D	4	1	Japanese	4	3	JA-04-03	f	7
	D	4	2	Japanese	5	1	JA-05-01	f	
	D	4	3	Japanese	7	1	JA-07-01	f	3

Multap - VS

Stimline 17

S-Id	SCS	SCNR	Pos	Language	Syll	Stim-No	Identifier	f_c	SL-Pos
SL17									
	0	0	1	Persian	4	1	FA-04-01		1
	0	0	2	Persian	5	2	FA-05-02		2
	0	0	3	Persian	4	3	FA-04-03		3
	0	0	4	Persian	7	2	FA-07-02		4
	0	0	5	Persian	9	1	FA-09-01		5
	0	0	6	Persian	7	1	FA-07-01		6
	0	0	7	Persian	5	3	FA-05-03		7
	0	0	8	Persian	5	1	FA-05-01		8
SC17-A1 <u>ع</u>									
	A	1	1	Persian	4	3	FA-04-03	c	3
SC17-A2 <u>ف</u>									
	A	2	1	Persian	4	2	FA-04-02	f	
SC17-B1 <u>ع</u>									
	B	1	1	Persian	5	2	FA-05-02	c	2
	B	1	2	Persian	7	2	FA-07-02	c	4
SC17-D1 <u>ع</u>									
	D	1	1	Persian	4	1	FA-04-01	c	1
	D	1	2	Persian	5	2	FA-05-02	c	2
	D	1	3	Persian	7	1	FA-07-01	c	6
SC17-D2 <u>ع</u>									
	D	2	1	Persian	7	2	FA-07-02	c	4
	D	2	2	Persian	4	3	FA-04-03	c	3
	D	2	3	Persian	7	1	FA-07-01	c	6

Multap - VS

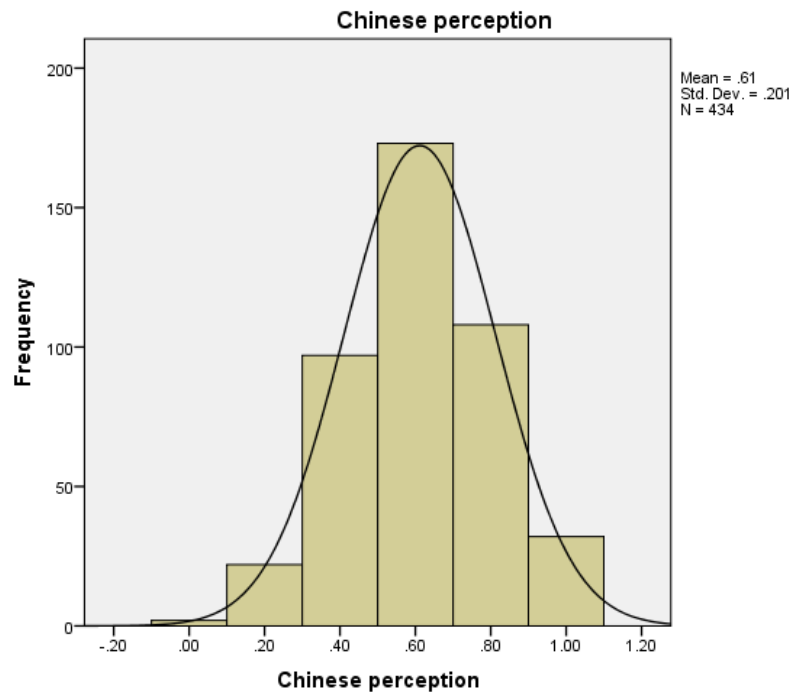
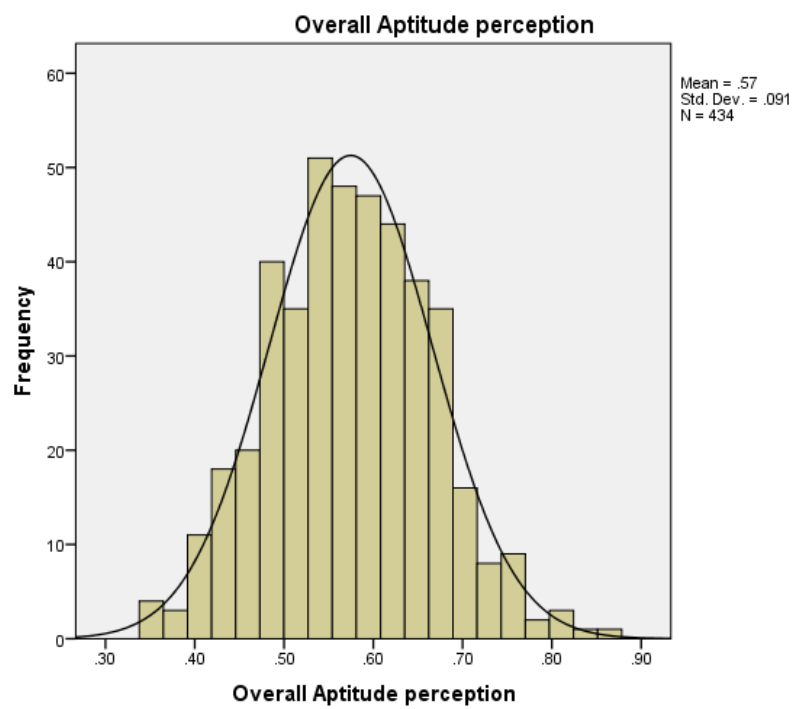
Stimline 18

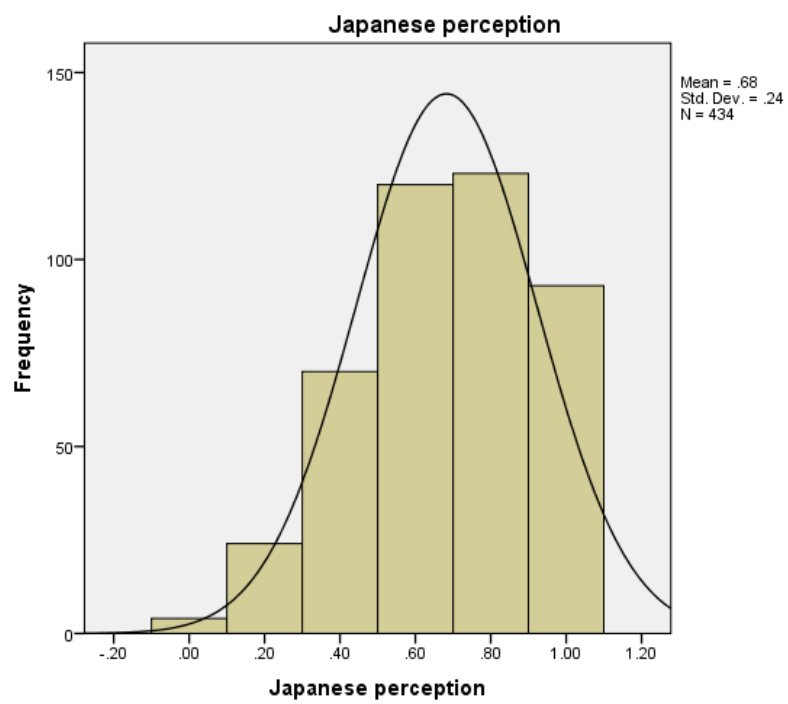
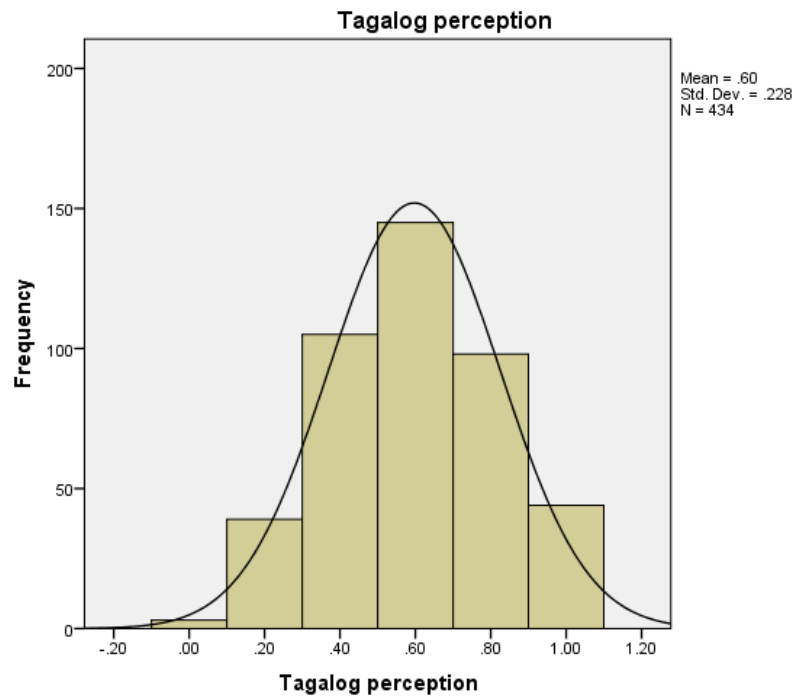
S-Id	SCS	SCNR	Pos	Language	Syll	Stim-No	Identifier	f_c	SL-Pos
SL18									
	0	0	1	Hindi	6	1	HI-06-01		1
	0	0	2	Hindi	3	2	HI-03-02		2
	0	0	3	Hindi	4	2	HI-04-02		3
	0	0	4	Hindi	3	1	HI-03-01		4
	0	0	5	Hindi	9	1	HI-09-01		5
	0	0	6	Hindi	6	2	HI-06-02		6
	0	0	7	Hindi	3	3	HI-03-03		7
	0	0	8	Hindi	3	6	HI-03-06		8
SC18-A1 <u>ॐ</u>									
	A	1	1	Hindi	3	2	HI-03-02	c	2
SC18-A2 <u>f</u>									
	A	1	1	Hindi	5	1	HI-05-01	f	
SC18-B1 <u>f</u>									
	B	2	1	Hindi	4	2	HI-04-02	f	3
	B	2	2	Hindi	5	1	HI-05-01	f	
SC18-D1 <u>ॐ</u>									
	D	4	1	Hindi	3	1	HI-03-01	c	4
	D	4	2	Hindi	6	1	HI-06-01	c	1
	D	4	3	Hindi	4	2	HI-04-02	c	3
SC18-D2 <u>f</u>									
	D	4	1	Hindi	4	3	HI-04-03	f	
	D	4	2	Hindi	3	6	HI-03-06	f	8
	D	4	3	Hindi	4	2	HI-04-02	f	3

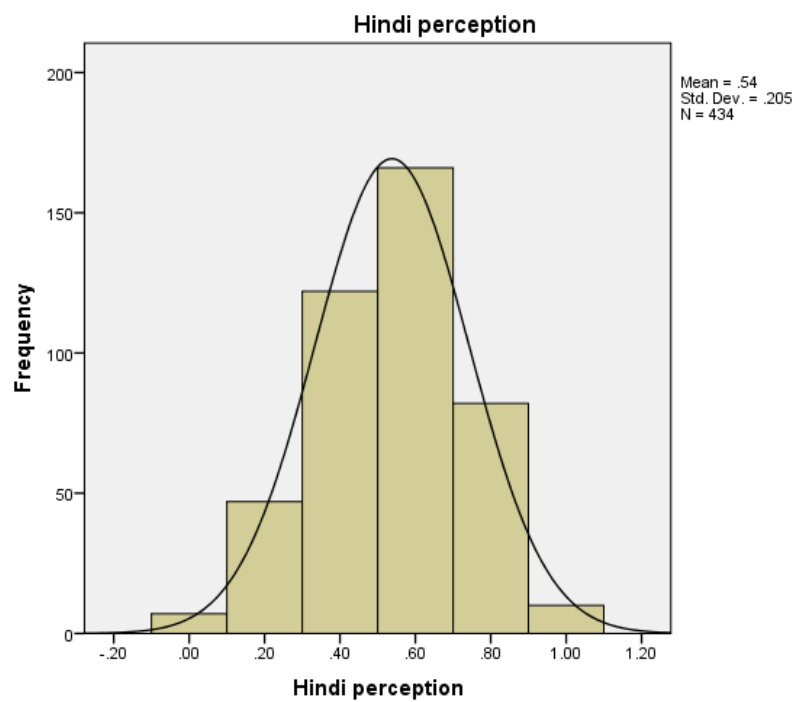
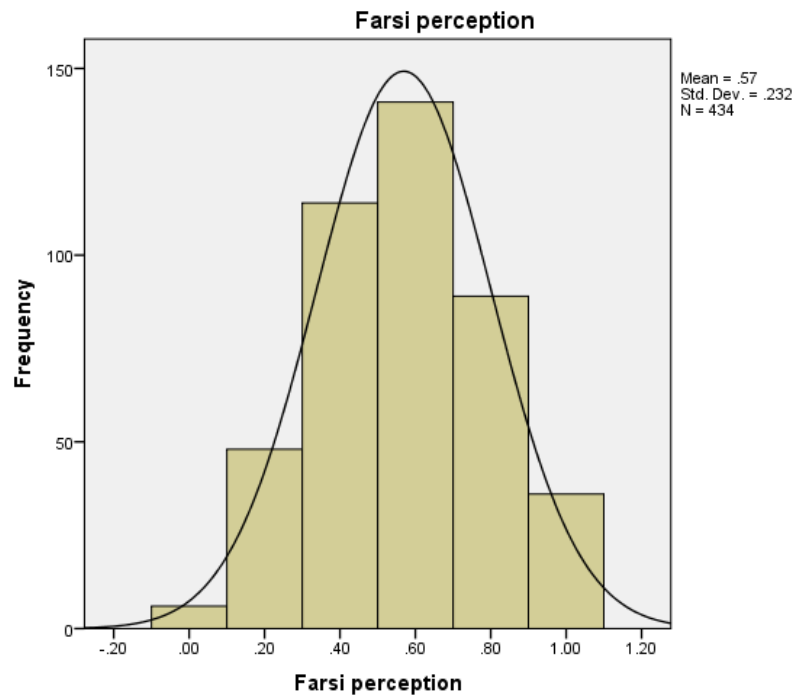
A2.2. Overview Multap Variables

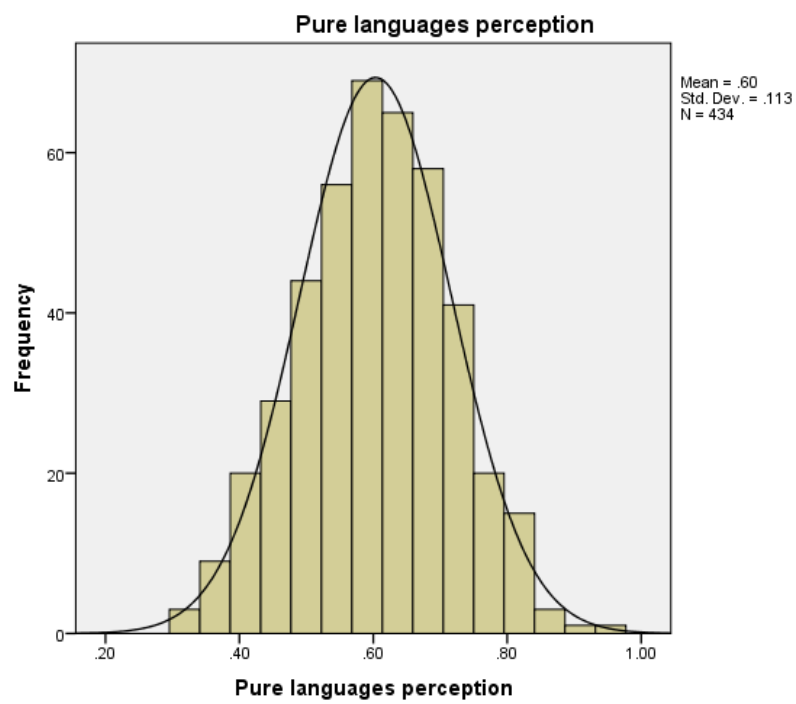
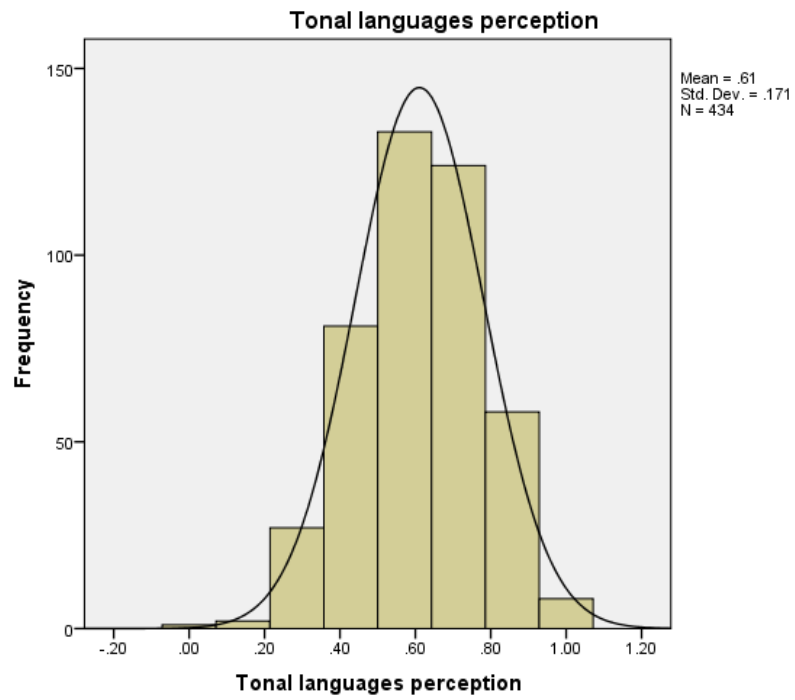
Statistics									Test for Normality	
<i>N</i> = 434									Shapiro-Wilk <i>df</i> = 434	
	<i>M</i>	<i>SE</i>	<i>SD</i>	<i>Variance</i>	<i>min</i>	<i>max</i>	<i>Skewness</i>	<i>Kurtosis</i>	<i>Statistic</i>	<i>Sig.</i>
Overall Aptitude perception	.57	.004	.09	.01	.35	.86	.10	-.11	0.99	< 0.01
Mandarin perception	.61	.010	.20	.04	.00	1.00	-.09	-.20	0.98	< 0.01
Tagalog perception	.60	.011	.23	.05	.00	1.00	-.04	-.56	0.77	< 0.01
Japanese perception	.68	.012	.24	.06	.00	1.00	-.41	-.49	0.91	< 0.01
Farsi perception	.57	.011	.23	.05	.00	1.00	-.02	-.51	0.93	< 0.01
Hindi perception	.54	.010	.20	.04	.00	1.00	-.19	-.23	0.91	< 0.01
Tonal languages perception	.61	.008	.17	.03	.00	1.00	-.17	-.16	0.93	< 0.01
Pure languages perception	.60	.005	.11	.01	.32	.95	-.02	-.25	0.92	< 0.01
Mixed languages perception	.52	.007	.14	.02	.17	.92	-.03	-.18	0.94	< 0.01

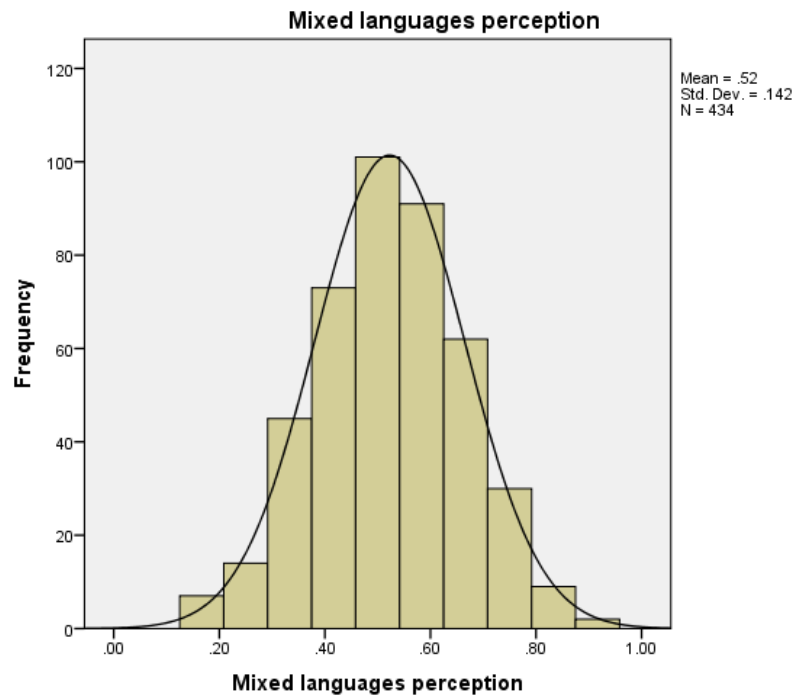
A2.3. Histograms



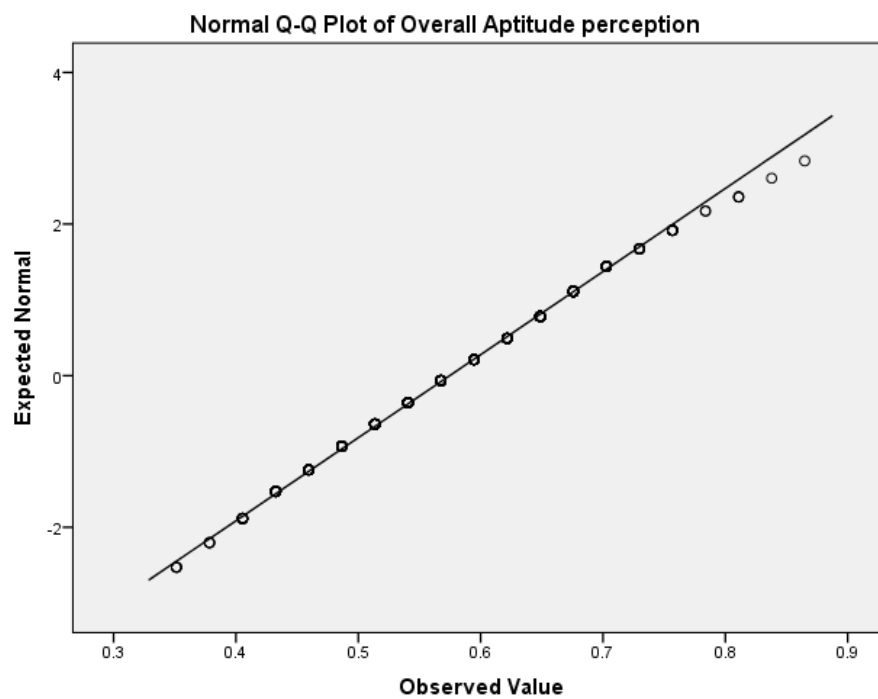


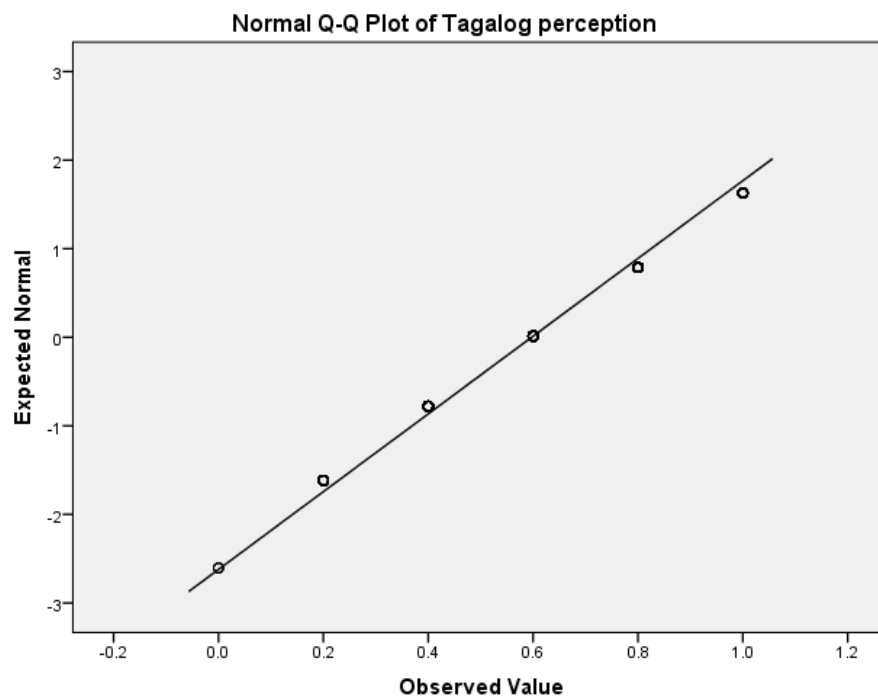
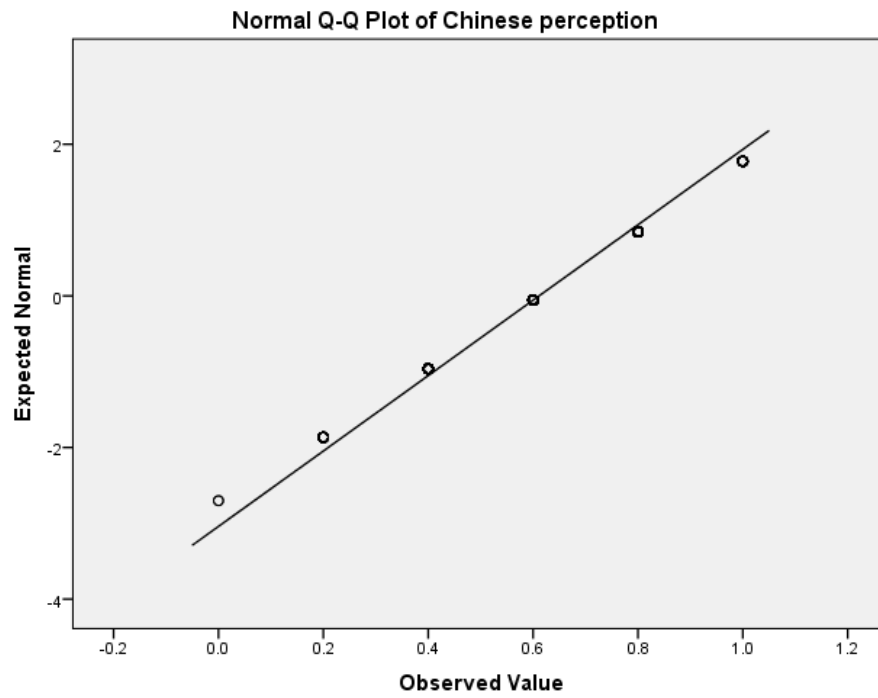


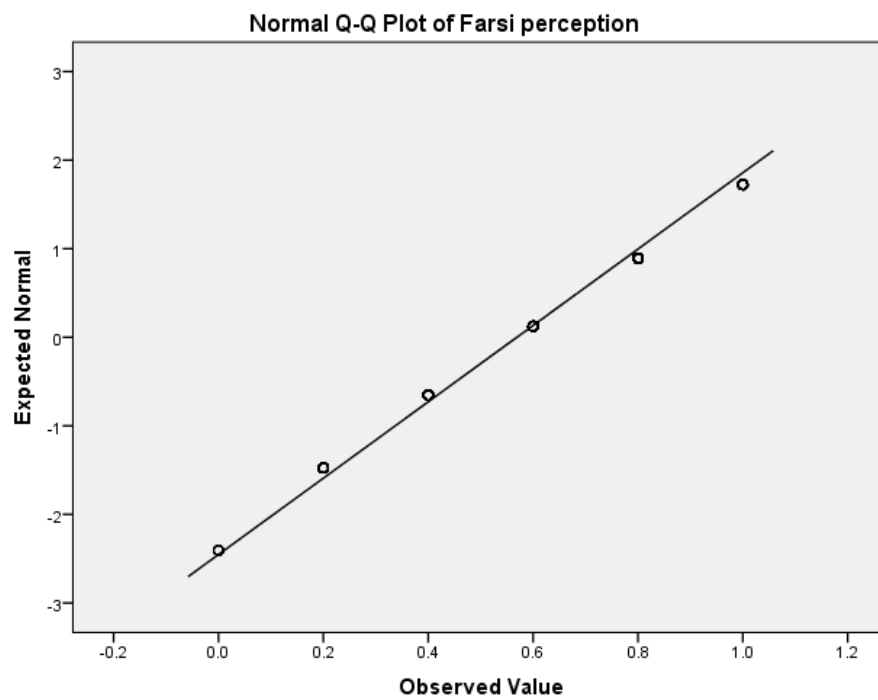
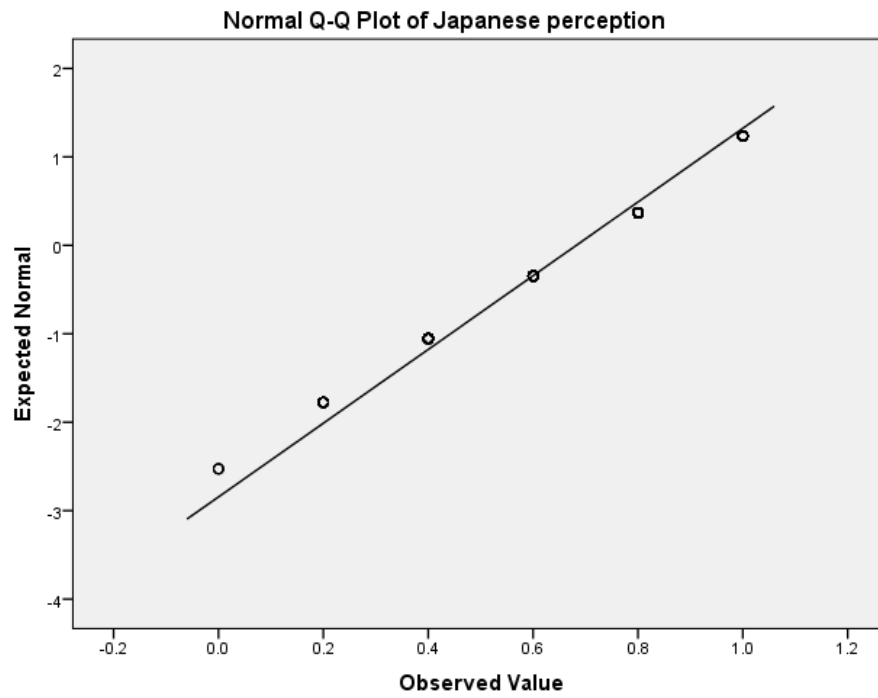


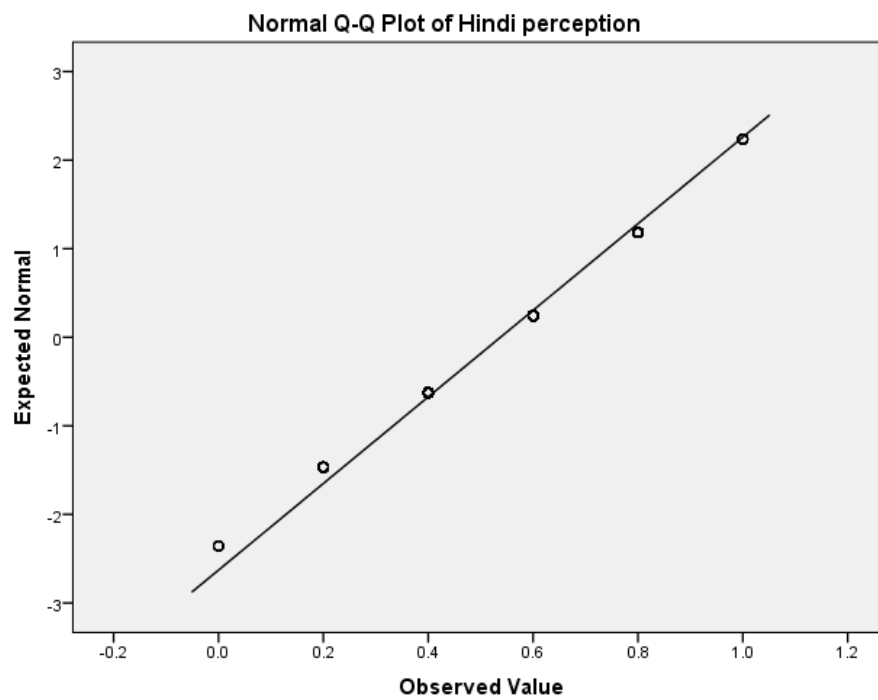
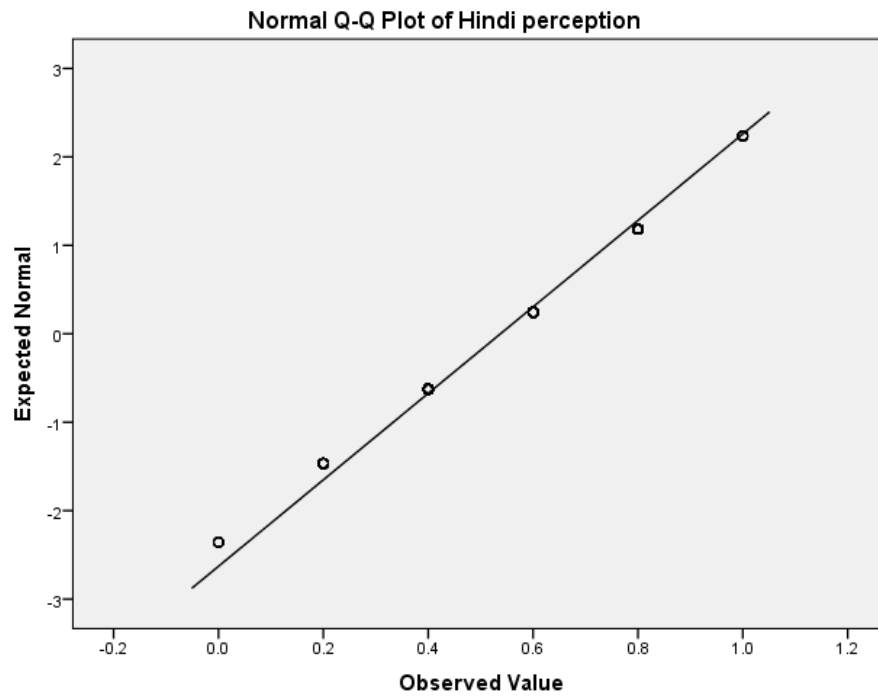


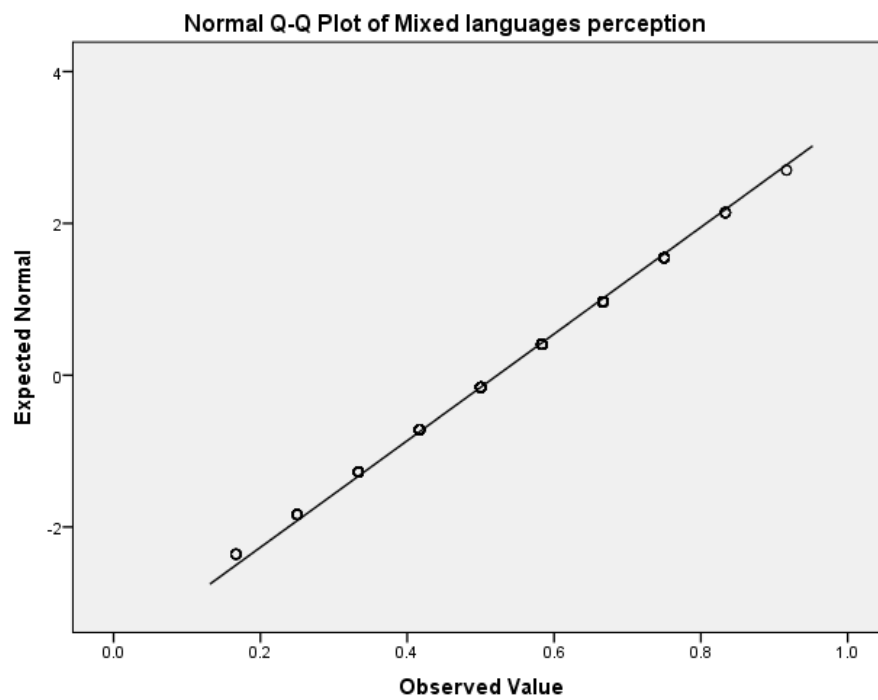
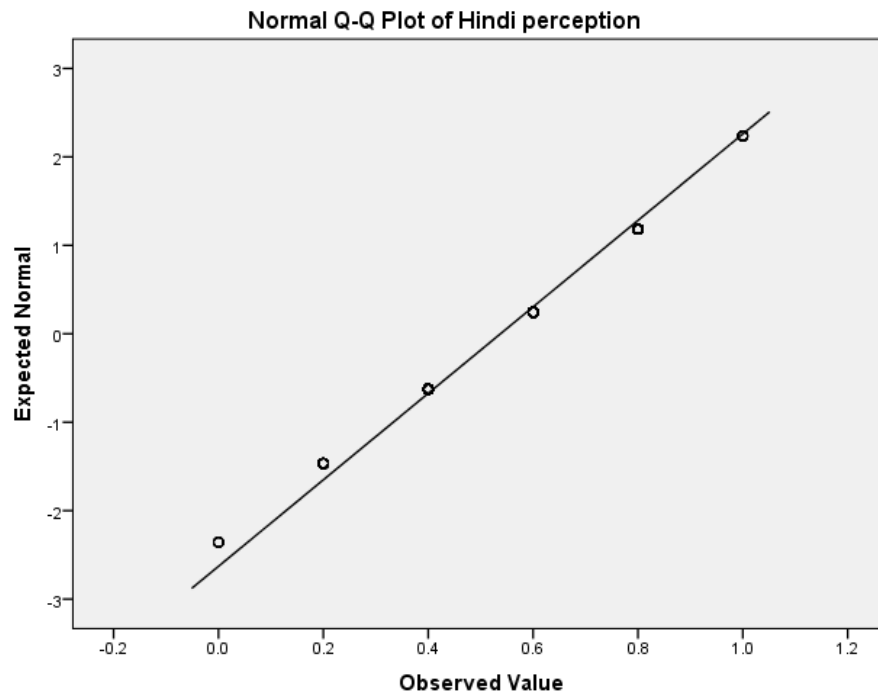
A2.4. Q-Q Plots





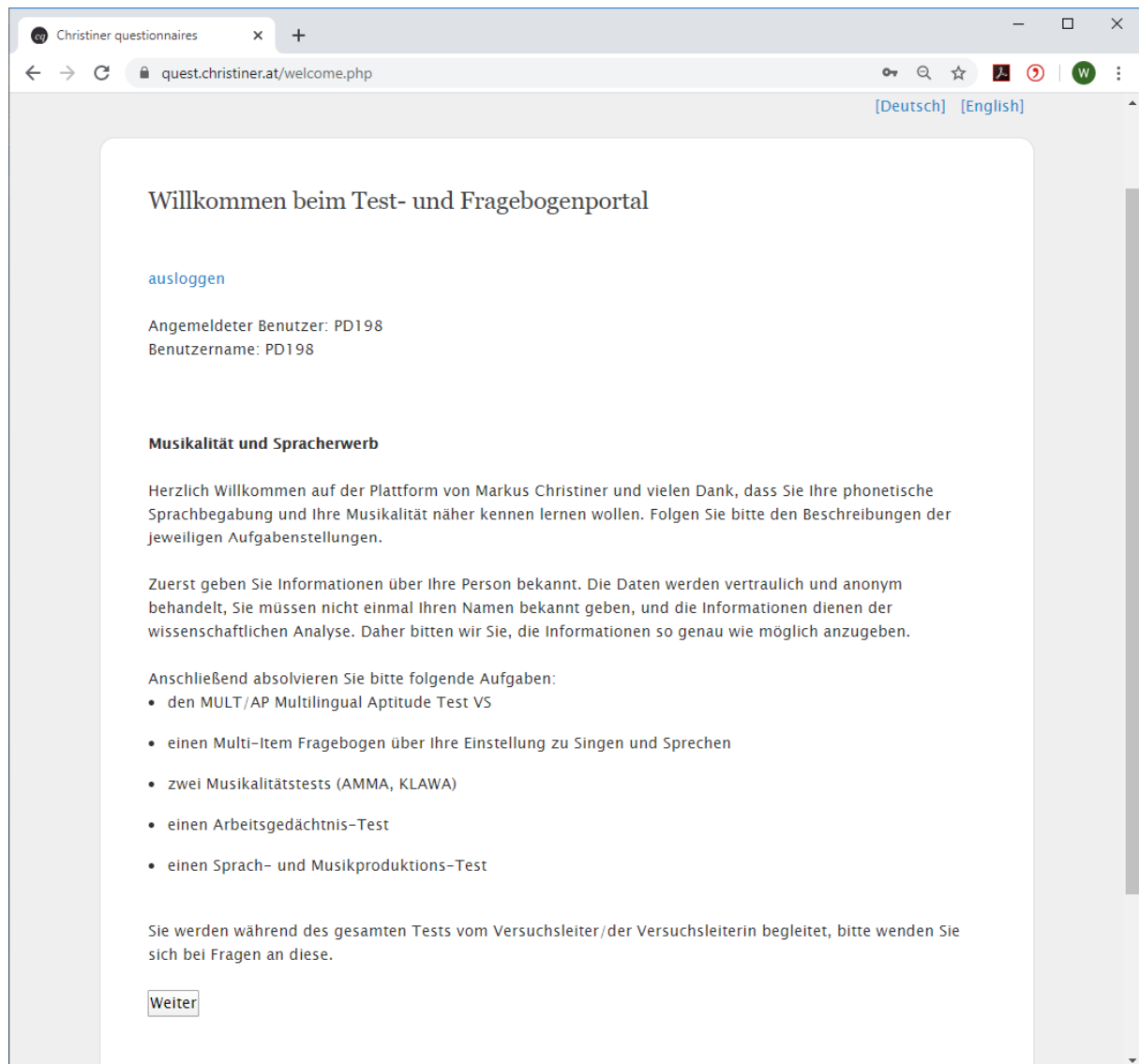






Appendix 3. Empirical investigation

Starting page



Appendix 4. Key variables

A4.1. Perception

Perception	Statistics					Test for Normality	
	<i>M</i>	<i>SE</i>	<i>MDN</i>	<i>min</i>	<i>max</i>	Shapiro-Wilk <i>Statistic</i>	<i>Sig.</i>
Overall Aptitude perception	.58	.01	.57	.35	.76	0.98	0.15
PL perception	.63	.01	.64	.36	.91	0.98	0.20
ML perception	.48	.02	.50	.17	.83	0.96	0.01
Mandarin perception	.65	.02	.60	.20	1.00	0.90	0.00
Tagalog perception	.61	.02	.60	.20	1.00	0.91	0.00
Japanese perception	.73	.02	.80	.20	1.00	0.89	0.00
Farsi perception	.55	.03	.60	.00	1.00	0.91	0.00
Hindi perception	.56	.02	.60	.20	.80	0.86	0.00
AMMATOTAL	54.56	.95	53.00	35.00	80.00	0.99	0.52
AMMAT	25.86	.55	25.00	18.00	40.00	0.96	0.01
AMMAR	28.70	.46	29.00	17.00	40.00	0.99	0.55

A4.2. Production

Production	Statistics					Test for Normality	
	<i>M</i>	<i>SE</i>	<i>MDN</i>	<i>min</i>	<i>max</i>	Shapiro-Wilk <i>Statistic</i>	<i>Sig.</i>
Russian imitation	3.63	.16	3.44	1.19	7.63	0.91	0.00
Japanese imitation	4.82	.15	4.92	1.25	7.42	0.98	0.17
Hindi imitation	3.17	.12	3.00	.80	7.85	0.96	0.01
Tamil imitation	3.01	.15	2.65	.85	7.40	0.94	0.00
Mandarin imitation	2.35	.09	2.28	.90	5.20	0.95	0.00
Tagalog imitation	2.40	.15	2.06	.50	7.81	0.83	0.00
Thai imitation	1.64	.08	1.50	.10	3.95	0.96	0.01
Farsi imitation	2.33	.14	2.19	.19	6.50	0.96	0.01
Speech imitation total	2.95	.10	2.75	1.18	5.83	0.97	0.05
Singing total	6.13	.14	5.93	3.50	9.19	0.96	0.02
New song total	6.08	.16	5.92	2.75	9.17	0.97	0.02
Happy Birthday total	6.19	.14	5.96	3.67	9.50	0.97	0.02
New song melody	5.73	.18	5.56	2.00	9.25	0.97	0.08
New song quality	5.71	.16	5.38	2.75	9.25	0.95	0.00
New song rhythm	6.78	.15	6.75	3.38	9.63	0.98	0.16
Happy Birthday melody	6.22	.17	6.25	3.00	9.50	0.97	0.06
Happy Birthday quality	5.58	.15	5.25	2.50	9.50	0.95	0.00
Happy Birthday rhythm	6.76	.12	6.50	4.25	9.50	0.97	0.05
Singing behaviour during childhood	28.96	.99	28.00	9.00	48.00	0.97	0.09
Singing regularly hours per week	2.04	.48	.50	.00	28.00	0.50	0.00
Self-estimation of singing capacity	2.85	.12	3.00	1.00	5.00	0.91	0.00
Performance anxiety	42.62	1.38	41.00	14.00	67.00	0.97	0.03

A4.3. Phonological STM and further criteria

Phonological STM and further criteria	Statistics					Test for Normality	
	<i>M</i>	<i>SE</i>	<i>MDN</i>	<i>min</i>	<i>max</i>	Shapiro-Wilk <i>Statistic</i>	<i>Sig.</i>
STMF	7.44	.22	7.00	3.00	13.00	0.97	0.04
STMB	7.79	.26	8.00	3.00	13.00	0.94	0.00
STM	15.23	.41	15.00	8.00	25.00	0.96	0.01
All Languages memorable	4.88	.14	5.13	2.25	7.38	0.97	0.04
All Languages difficult to imitate	6.81	.17	7.13	2.25	9.50	0.95	0.00
All Languages familiar	4.39	.18	4.38	1.13	8.38	0.99	0.50
Highest level of education	5.58	.26	5.00	1.00	9.00	0.88	0.00

Appendix 5. Behavioural concepts

A5.1. Concept singing during childhood

Concept singing during childhood	
	Cronbach's α if Item Deleted
Wenn ich als Kind die Möglichkeit hatte, habe ich in der Kirche laut mitgesungen.	.79
Als Kind habe ich mir nie Gedanken gemacht, ob das was ich singe, richtig oder falsch klingt.	.84
Als Kind habe ich gerne in einem Chor, mit Freunden oder zu festlichen Anlässen wie Weihnachten gesungen.	.77
Mein Kindheitstraum war es, ein berühmter Sänger zu werden.	.81
Als ich ein Kind war, habe ich mir keine Gedanken gemacht, wenn ich gesungen habe. Ich sang einfach drauf los.	.81
Ich habe in meiner Kindheit, wann immer ich konnte, gesungen.	.77
Als Kind war ich im Schulchor.	.80
Als Kind habe ich öfter gesungen als meine Freunde.	.77

A5.2. Concept Performance anxiety

Concept Performance anxiety	
	Cronbach's α if Item Deleted
Wenn ich Sprachaufnahmen höre, welche ich nachahmen soll, fühle ich mich unwohl.	.88
Sprachen zu imitieren gefällt mir sehr.	.88
Ich bin nicht nervös, wenn ich etwas spontan nachahmen soll.	.87
Etwas nachzusingen macht, mir Spaß.	.86
Ich habe kein Problem damit, etwas vor anderen Leuten zu singen.	.85
Singen und Musizieren, auch vor anderen Leuten, macht mir Spaß.	.85
Mir ist es lieber, wenn ich zuhause und alleine singe, wo mich niemand hört.	.86
Vor anderen Menschen zu sprechen, ist für mich kein Problem.	.86
Ich halte gerne Reden, auch wenn viele Leute anwesend sind.	.85
Mir macht es Spaß, wenn mich andere hören.	.85
Vor Präsentationen bin ich meist sehr nervös.	.87
Auf einer Bühne zu stehen, gefällt mir besonders gut.	.85

Appendix 6. Reliability analysis – Raters

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Russian Item 1 - Rater 1	51.56	469.38	0.53	0.90
Russian Item 1 - Rater 2	54.38	454.66	0.59	0.90
Russian Item 1 - Rater 3	54.91	457.97	0.64	0.89
Russian Item 1 - Rater 4	51.53	516.16	0.55	0.90
Russian Item 2 - Rater 1	54.33	490.22	0.56	0.90
Russian Item 2 - Rater 2	56.59	480.83	0.64	0.89
Russian Item 2 - Rater 3	56.56	483.73	0.62	0.89
Russian Item 2 - Rater 4	52.76	519.06	0.53	0.90
Russian Item 3 - Rater 1	54.52	462.28	0.61	0.89
Russian Item 3 - Rater 2	55.27	475.37	0.62	0.89
Russian Item 3 - Rater 3	57.02	491.41	0.63	0.89
Russian Item 3 - Rater 4	52.94	511.89	0.56	0.90
Russian Item 4 - Rater 1	55.28	493.19	0.49	0.90
Russian Item 4 - Rater 2	54.81	488.86	0.65	0.89
Russian Item 4 - Rater 3	56.76	474.21	0.68	0.89
Russian Item 4 - Rater 4	52.35	507.95	0.59	0.90

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Japanese Item 1 - Rater 1	111.34	979.68	0.63	0.92
Japanese Item 1 - Rater 2	109.28	991.66	0.61	0.92
Japanese Item 1 - Rater 3	110.95	1,011.79	0.49	0.93
Japanese Item 1 - Rater 4	111.78	996.60	0.58	0.92
Japanese Item 1 - Rater 5	112.62	982.59	0.54	0.93
Japanese Item 1 - Rater 6	108.97	967.94	0.72	0.92
Japanese Item 2 - Rater 1	109.02	1,003.25	0.63	0.92
Japanese Item 2 - Rater 2	109.30	1,018.64	0.61	0.92
Japanese Item 2 - Rater 3	109.76	1,031.46	0.43	0.93
Japanese Item 2 - Rater 4	110.21	1,025.70	0.41	0.93
Japanese Item 2 - Rater 5	112.36	1,020.30	0.50	0.93
Japanese Item 2 - Rater 6	108.23	1,037.50	0.45	0.93
Japanese Item 3 - Rater 1	112.19	1,027.19	0.49	0.93
Japanese Item 3 - Rater 2	110.83	1,009.32	0.58	0.92
Japanese Item 3 - Rater 3	111.22	997.44	0.61	0.92
Japanese Item 3 - Rater 4	113.59	1,004.76	0.52	0.93
Japanese Item 3 - Rater 5	114.49	1,041.31	0.41	0.93
Japanese Item 3 - Rater 6	109.60	1,004.67	0.60	0.92
Japanese Item 4 - Rater 1	111.41	975.94	0.76	0.92
Japanese Item 4 - Rater 2	110.58	1,004.55	0.65	0.92
Japanese Item 4 - Rater 3	110.41	978.97	0.64	0.92
Japanese Item 4 - Rater 4	110.36	975.15	0.62	0.92
Japanese Item 4 - Rater 5	113.94	1,020.62	0.57	0.92
Japanese Item 4 - Rater 6	109.69	983.02	0.60	0.92

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Tamil Item 1 - Rater 1	55.31	706.31	0.61	0.92
Tamil Item 1 - Rater 2	54.21	718.59	0.66	0.92
Tamil Item 1 - Rater 3	57.87	708.68	0.62	0.92
Tamil Item 1 - Rater 4	55.95	731.10	0.56	0.92
Tamil Item 1 - Rater 5	57.56	694.98	0.65	0.92
Tamil Item 2 - Rater 1	55.92	693.82	0.64	0.92
Tamil Item 2 - Rater 2	56.23	670.20	0.72	0.92
Tamil Item 2 - Rater 3	57.57	713.21	0.70	0.92
Tamil Item 2 - Rater 4	56.02	722.73	0.67	0.92
Tamil Item 2 - Rater 5	57.24	677.17	0.74	0.92
Tamil Item 3 - Rater 1	58.08	747.61	0.46	0.92
Tamil Item 3 - Rater 2	58.79	751.32	0.44	0.92
Tamil Item 3 - Rater 3	58.81	756.69	0.52	0.92
Tamil Item 3 - Rater 4	57.36	764.33	0.44	0.92
Tamil Item 3 - Rater 5	59.45	756.60	0.46	0.92
Tamil Item 4 - Rater 1	57.24	709.86	0.58	0.92
Tamil Item 4 - Rater 2	57.24	699.36	0.65	0.92
Tamil Item 4 - Rater 3	58.47	717.80	0.67	0.92
Tamil Item 4 - Rater 4	56.81	744.84	0.57	0.92
Tamil Item 4 - Rater 5	59.14	734.43	0.58	0.92

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Hindi Item 1 - Rater 1	60.58	451.12	0.58	0.87
Hindi Item 1 - Rater 2	58.42	439.80	0.62	0.87
Hindi Item 1 - Rater 3	58.80	437.71	0.47	0.87
Hindi Item 1 - Rater 4	57.53	433.66	0.57	0.87
Hindi Item 1 - Rater 5	57.29	434.42	0.57	0.87
Hindi Item 2 - Rater 1	62.30	457.72	0.63	0.87
Hindi Item 2 - Rater 2	60.27	432.03	0.55	0.87
Hindi Item 2 - Rater 3	61.57	444.20	0.60	0.87
Hindi Item 2 - Rater 4	60.08	430.17	0.64	0.87
Hindi Item 2 - Rater 5	61.33	439.99	0.70	0.87
Hindi Item 3 - Rater 1	61.36	468.02	0.47	0.87
Hindi Item 3 - Rater 2	59.90	492.71	0.05	0.88
Hindi Item 3 - Rater 3	61.15	468.15	0.41	0.87
Hindi Item 3 - Rater 4	59.26	462.87	0.45	0.87
Hindi Item 3 - Rater 5	59.52	437.97	0.39	0.88
Hindi Item 4 - Rater 1	62.02	464.63	0.58	0.87
Hindi Item 4 - Rater 2	58.70	493.98	0.08	0.88
Hindi Item 4 - Rater 3	62.10	477.20	0.46	0.87
Hindi Item 4 - Rater 4	61.53	448.98	0.53	0.87
Hindi Item 4 - Rater 5	62.12	450.10	0.47	0.87

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Mandarin Item 1 - Rater 1	42.99	229.52	0.60	0.85
Mandarin Item 1 - Rater 2	42.16	229.93	0.53	0.85
Mandarin Item 1 - Rater 3	43.63	239.51	0.47	0.86
Mandarin Item 1 - Rater 4	44.67	236.83	0.59	0.85
Mandarin Item 1 - Rater 5	45.06	223.16	0.69	0.85
Mandarin Item 2 - Rater 1	44.43	247.42	0.48	0.86
Mandarin Item 2 - Rater 2	43.31	237.63	0.46	0.86
Mandarin Item 2 - Rater 3	44.67	259.54	0.31	0.86
Mandarin Item 2 - Rater 4	45.65	255.55	0.43	0.86
Mandarin Item 2 - Rater 5	45.17	240.12	0.50	0.85
Mandarin Item 3 - Rater 1	45.26	246.97	0.49	0.86
Mandarin Item 3 - Rater 2	43.70	238.85	0.55	0.85
Mandarin Item 3 - Rater 3	46.13	266.96	0.09	0.87
Mandarin Item 3 - Rater 4	46.52	259.19	0.46	0.86
Mandarin Item 3 - Rater 5	45.19	241.49	0.56	0.85
Mandarin Item 4 - Rater 1	45.44	253.57	0.26	0.86
Mandarin Item 4 - Rater 2	43.17	241.11	0.46	0.86
Mandarin Item 4 - Rater 3	45.48	255.08	0.35	0.86
Mandarin Item 4 - Rater 4	46.60	260.95	0.49	0.86
Mandarin Item 4 - Rater 5	44.86	248.47	0.41	0.86

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Tagalog Item 1 - Rater 1	34.24	404.66	0.59	0.88
Tagalog Item 1 - Rater 2	32.05	427.66	0.55	0.88
Tagalog Item 1 - Rater 3	34.13	417.35	0.53	0.88
Tagalog Item 1 - Rater 4	36.89	421.91	0.48	0.89
Tagalog Item 2 - Rater 1	35.05	404.07	0.49	0.89
Tagalog Item 2 - Rater 2	34.94	394.44	0.60	0.88
Tagalog Item 2 - Rater 3	35.04	408.01	0.54	0.88
Tagalog Item 2 - Rater 4	37.28	407.63	0.63	0.88
Tagalog Item 3 - Rater 1	36.64	416.02	0.52	0.88
Tagalog Item 3 - Rater 2	36.51	420.44	0.51	0.89
Tagalog Item 3 - Rater 3	35.82	424.39	0.49	0.89
Tagalog Item 3 - Rater 4	37.67	432.34	0.59	0.88
Tagalog Item 4 - Rater 1	37.36	424.00	0.59	0.88
Tagalog Item 4 - Rater 2	37.82	435.58	0.57	0.88
Tagalog Item 4 - Rater 3	37.05	416.76	0.63	0.88
Tagalog Item 4 - Rater 4	37.87	423.61	0.71	0.88

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Thai Item 1 - Rater 1	29.03	173.61	0.61	0.85
Thai Item 1 - Rater 2	31.19	190.39	0.50	0.85
Thai Item 1 - Rater 3	28.93	182.40	0.49	0.85
Thai Item 1 - Rater 4	32.35	201.12	0.53	0.86
Thai Item 1 - Rater 5	32.06	183.40	0.56	0.85
Thai Item 2 - Rater 1	29.45	182.89	0.48	0.85
Thai Item 2 - Rater 2	31.83	197.79	0.44	0.86
Thai Item 2 - Rater 3	28.70	184.40	0.54	0.85
Thai Item 2 - Rater 4	32.03	202.81	0.41	0.86
Thai Item 2 - Rater 5	31.83	193.06	0.48	0.85
Thai Item 3 - Rater 1	30.67	184.65	0.54	0.85
Thai Item 3 - Rater 2	32.38	205.23	0.28	0.86
Thai Item 3 - Rater 3	30.10	185.27	0.48	0.85
Thai Item 3 - Rater 4	32.33	206.55	0.27	0.86
Thai Item 3 - Rater 5	32.21	196.83	0.38	0.86
Thai Item 4 - Rater 1	31.20	194.80	0.47	0.85
Thai Item 4 - Rater 2	32.19	198.29	0.58	0.85
Thai Item 4 - Rater 3	29.81	185.19	0.49	0.85
Thai Item 4 - Rater 4	32.26	206.12	0.34	0.86
Thai Item 4 - Rater 5	32.03	191.12	0.44	0.86

Appendix 7. List of abbreviations

Note: The variable names used in this investigation are printed in italics, while general abbreviations, which refer to internationally used terminologies, measurements and common abbreviations are not in italics. For the test descriptions in chapter 2, 3 and 4, the sub-tests if referred to in the text have been indicated by italics as well.

List of abbreviations		
Category	Internal abbreviation	Description
Music perception	<i>AMMATOTAL</i>	The total score of the AMMA test
	<i>AMMAT</i>	The tonal sub-score of the AMMA test
	<i>AMMAR</i>	The rhythmic sub-score of the AMMA test
Singing ability	<i>Singing total</i>	The total score of both singing tasks
	<i>New song total</i>	The total score of the song learning condition
	<i>Happy Birthday total</i>	The total score of the familiar song singing task
	<i>New song melody</i>	The singing criteria melody of the song learning condition (indicates how well the original melody was sustained - key did not play any role)
	<i>New song quality</i>	The singing criteria quality of voice of the song learning condition
	<i>New song rhythm</i>	The singing criteria rhythm of the song learning condition (indicates how well the original rhyhtm was sustained)
	<i>Happy Birthday melody</i>	The singing criteria melody of the familiar song singing task "Happy Birthday"
	<i>Happy Birthday quality</i>	The singing quality of voice of the familiar song singing task "Happy Birthday"
	<i>Happy Birthday rhythm</i>	The singing criteria rhyhtm of the familiar song singing task "Happy Birthday"
	<i>Self-estimation of singing capacity</i>	The self-estimation score of the singing ability as estimated by the participants
	<i>Singing behaviour during childhood</i>	The score of the questionnaire which focusses on asking for how frequently the participants sang during their childhood
	<i>Singing regularly hours per week</i>	The hours the participants sang in a week

Category	Internal abbreviation	Description
Speech perception	MULT AP	Multilingual aptitude test
	<i>Overall Aptitude perception</i>	The total score of all language perception tasks (37 items)
	<i>ML perception</i>	The total score of the mixed language perception tasks (12 items)
	<i>PL perception</i>	The total score of the pure language perception tasks (which is comprised of the following five speech perception tasks and consists of 25 items)
	<i>Mandarin perception</i>	Total score of the Mandarin perception tasks (5 items)
	<i>Tagalog perception</i>	Total score of the Tagalog perception tasks (5 items)
	<i>Japanese perception</i>	Total score of the Japanese perception tasks (5 items)
	<i>Farsi perception</i>	Total score of the Farsi perception tasks (5 items)
	<i>Hindi perception</i>	Total score of the Hindi perception tasks (5 items)
Speech imitation	<i>Speech imitation total</i>	The total score of all language production tasks (32 items)
	<i>Russian imitation</i>	The total score of the Russian production tasks (4 items)
	<i>Japanese imitation</i>	The total score of the Japanese production tasks (4 items)
	<i>Hindi imitation</i>	The total score of the Hindi production tasks (4 items)
	<i>Tamil imitation</i>	The total score of the Tamil production tasks (4 items)
	<i>Mandarin imitation</i>	The total score of the Mandarin production tasks (4 items)
	<i>Tagalog imitation</i>	The total score of the Tagalog production tasks (4 items)
	<i>Thai imitation</i>	The total score of the Thai production tasks (4 items)
Education	<i>Farsi imitation</i>	The total score of the Farsi production tasks (4 items)
	<i>Highest level of education</i>	The score which describes the educational level of the participants
	<i>Highest education of mother</i>	The score which describes the educational level of the participants' mothers
	<i>Highest education of father</i>	The score which describes the educational level of the participants' fathers
	<i>Number of foreign languages</i>	The total number of the foreign languages spoken (the proficiency in each language has not been considered)
Performance anxiety	<i>Number of musical instruments</i>	The total number of musical instruments played
	<i>Performance anxiety</i>	The score of the questionnaire which focusses on asking revealing performance anxiety

Category	Internal abbreviation	Description
Self-estimation concepts	<i>Russian memorable</i>	The total score of how memorable Russian occurred to the participants
	<i>Russian difficult to imitate</i>	The total score of how difficult Russian was to the participants
	<i>Russian familiar</i>	The total score of how familiar Russian was to the participants
	<i>Japanese memorable</i>	The total score of how memorable Japanese occurred to the participants
	<i>Japanese difficult to imitate</i>	The total score of how difficult Japanese was to the participants
	<i>Japanese familiar</i>	The total score of how familiar Japanese was to the participants
	<i>Tamil memorable</i>	The total score of how memorable Tamil occurred to the participants
	<i>Tamil difficult to imitate</i>	The total score of how difficult Tamil was to the participants
	<i>Tamil familiar</i>	The total score of how familiar Tamil was to the participants
	<i>Hindi memorable</i>	The total score of how memorable Hindi occurred to the participants
	<i>Hindi difficult to imitate</i>	The total score of how difficult Hindi was to the participants
	<i>Hindi familiar</i>	The total score of how familiar Hindi was to the participants
	<i>Mandarin memorable</i>	The total score of how memorable Mandarin occurred to the participants
	<i>Mandarin difficult to imitate</i>	The total score of how difficult Mandarin was to the participants
	<i>Mandarin familiar</i>	The total score of how familiar Mandarin was to the participants
	<i>Tagalog memorable</i>	The total score of how memorable Tagalog occurred to the participants
	<i>Tagalog difficult to imitate</i>	The total score of how difficult Tagalog was to the participants
	<i>Tagalog familiar</i>	The total score of how familiar Tagalog was to the participants
	<i>Thai memorable</i>	The total score of how memorable Thai occurred to the participants
	<i>Thai difficult to imitate</i>	The total score of how difficult Thai was to the participants
	<i>Thai familiar</i>	The total score of how familiar Thai was to the participants
	<i>Farsi memorable</i>	The total score of how memorable Farsi occurred to the participants
	<i>Farsi difficult to imitate</i>	The total score of how difficult Farsi was to the participants
	<i>Farsi familiar</i>	The total score of how familiar Farsi was to the participants

Category	Internal abbreviation	Description
Self-estimation concepts (continued)	<i>All Languages memorable</i>	The total score of how memorable all eight language imitation tasks (Russian, Japanese, Tamil, Hindi, Mandarin, Tagalog, Thai and Farsi) occurred to the participants
	<i>All Languages difficult to imitate</i>	The total score of how difficult all eight language imitation tasks (Russian, Japanese, Tamil, Hindi, Mandarin, Tagalog, Thai and Farsi) were to the participants
	<i>All Languages familiar</i>	The total score of how familiar all eight language imitation tasks (Russian, Japanese, Tamil, Hindi, Mandarin, Tagalog, Thai and Farsi) were to the participants
Phonological short-term memory	<i>STMF</i>	Total score of the short-term memory forward
	<i>STMB</i>	Total score of the short-term memory backward
	<i>STMT</i>	Total score of the short-term memory forward and backward together

Category	International abbreviation	Description
Short-term memory	STM	Short-term memory (general abbreviations are not used in italic)
Working memory	WM	Working memory
Languages	L2	Second language
	L1	First language
Aptitude	DMGT	Model of Giftedness and Talent
	MLAT	The Modern Language Aptitude Test
	PLAB	The Pimsleur Language Aptitude Battery
	DLAB	The Defense Language Aptitude Battery
	VORD	A language aptitude test by Parry and Child
	CANAL-F test	Cognitive Ability for Novelty in Acquisition of Language - Foreign
	LLAMA	A language aptitude test developed by Paul Meara
	Hi-LAB	High-level Language Aptitude Battery
	DLI	Defense Language Institute
	HABLA	Horne's Assessment of Basic Linguistic Abilities
	AFLAAT	Al-Haik's Foreign Language Auditory Aptitude Test
	VORD	A language aptitude test based on an artificial language
	ID	individual differences
	IQ	intelligence quotient
	MAP	The Musical Aptitude Profile
	PMMA	Primary Measures of Musical Audiation
	IMMA	Intermediate Measures of Musical Audiation
	AMMA	Advanced Measures of Musical Audiation
	PROMS	Profile of Music Perception Skills
	Gold-MSI	Goldsmiths' Musical Sophistication Index
Singing	SVDM	Singing Voice Development Measure
	SSAP	Seattle Singing Accuracy Protocol

Appendix 8. Zusammenfassung

In dieser Dissertation wird der Zusammenhang zwischen phonetischer Begabung in Fremdsprachen und Singen, Musikperzeption, Kurzzeitgedächtnis, Leistungsnervosität, Gesangsverhalten in der Kindheit, Bildungsstand und weiteren Selbsteinschätzungsfragen untersucht. Die Ergebnisse zeigen, dass die Sprachproduktion von neuem Sprachmaterial, das für die Teilnehmer nicht zu verstehen ist, sowie deren Perzeption offensichtlich trennbar ist, da diese kaum zusammenhängen. Die Begabung zur Tondiskriminierung scheint dabei eine besondere Fähigkeit zu sein, die Abweichungen in der Fähigkeit, bestimmte Sprachen nachahmen zu können, erklären kann.

Die phonetische Begabung, die durch das Nachahmen von neuen, für die Teilnehmer, nicht verständlichen Sprachen gemessen worden ist, hat gezeigt, dass zehn Prädiktoren, die die Unterschiede in der Imitationsfähigkeit von acht Sprachen teilweise erklären können, gefunden worden sind. Die wichtigsten sind das phonologische Kurzzeitgedächtnis, Singen (Melodie und Rhythmus) und Tonperzeptionsfähigkeit. Andere Prädiktoren, die die Einprägsamkeit sowie den Schwierigkeitsgrad der Nachahmbarkeit einer Sprache festgestellt haben, und überdies den Bekanntheitsgrad der Sprachen und Leistungsangst, sind ebenfalls ermittelt worden.

Das Testverfahren ist so gewählt worden, dass acht Sprachen als gesamtes Testdesign verwendet worden sind, mit dem Ziel, zu verstehen, welche Fähigkeiten für das Imitieren von vielen Sprachen benötigt werden, wie es der Fall ist bei polyglotten Personen. Die Ergebnisse zeigen, dass die Prädiktoren in den einzelnen Sprachen abweichen. Dabei können 60 Prozent der Abweichungen der Sprachimitation mit vier Prädiktoren erklärt werden: der Anzahl der Fremdsprachen, das phonologische Kurzzeitgedächtnis, wie leicht die Teilnehmer ein neues Lied erlernen und wie einprägsam ihnen die Sprachen vorgekommen sind.

Imitationsaufgaben scheinen relativ sprachunabhängig zu sein und eignen sich daher weniger für eine Untersuchung typologischer Sprachunterschiede. Die einzigen Ausnahmen könnten bei Tonsprachen gefunden werden, welche immer Gesangsprädiktoren aufweisen. Generell ist aber davon auszugehen, dass wiederkehrende Prädiktoren Sprachimitation generell fördern.

Resultierend aus der Untersuchung kann davon ausgegangen werden, dass Singen und musikalisches Training generell unterstützende Funktionen aufweisen, die die Aussprachefähigkeit in Fremdsprachen positiv fördert. Diese scheinen vor allem im

Zusammenhang mit phonetischer Begabung zu stehen. Die Ergebnisse können auch auf Begabungstests übertragen werden, denn Sprachimitationsaufgaben scheinen sehr einfach durchführbare Messwerkzeuge zu sein. Im selben Zuge sollten musikalische Aspekte wie Singen und Musikperzeptionstests auch den Einzug in Sprachbegabungstests halten, da diese aussagekräftige Informationen zu individuellen Sprachfähigkeiten liefern.