universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„The effect of singing performance on vocabulary acquisition in foreign language learning "

verfasst von / submitted by

Liisi Sofia Törmäkangas, BA

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

Master of Arts, MA

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 867
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Allgemeine Linguistik: Grammatiktheorie und kognitive Linguistik
Betreut von / Supervisor:	Assoz. Prof. Mag. Dr. Susanne Maria Reiterer

Abstract

Die vorliegende Masterarbeit gibt einen Überblick über die Ergebnisse der empirischen Studie über die Einwirkung von Gesang auf den Vokabularerwerb beim Fremdsprachenlernen. Musikalität gilt als eine der diversen individuellen Faktoren, die den Prozess Fremdsprachlernen fördern. In den vergangenen Jahren wurden zahlreiche Studien durchgeführt, die einen positiven Zusammenhang zwischen musikalischer Kompetenz und verschiedenen Teilbereichen von Sprachlernen zeigen. In dieser Studie wurde untersucht, ob TeilnehmerInnen, die Lieder in einer Fremdsprache (Finnisch) singen, beim Vokabularerwerbe Vorteile haben im Vergleich zu den TeilnehmerInnen, die den Liedern nur zuhören. Zusätzlich wurde untersucht, ob TeilnehmerInnen mit Gesangserfahrung von ihrer musikalischen Erfahrung profitieren. Die Ergebnisse dieser Studie konnten nicht als significant nachgewiesen werden.

Schlüsselwörter: language aptitude, foreign language learning, vocabulary acquisition, musical aptitude, singing

Contents

Acknowledgments

I would like to thank my supervisor, Assoz. Prof. Mag. Dr. Susanne Maria Reiterer for her excellent guidance and encouragement through each stage of the process of my master thesis, inspiring my interest in the field of language aptitude and supporting my study design from the very beginning of this long journey.

My incredible colleagues, dear linguists, have defined the path of my research. Willing to sacrifice their time and patience for discussing the issues of my study design and volunteering to be my guinea pigs, they have given a huge contribution for the sake of my experiment.
Franz Schmid and Lukas Nemestothy, thank you for all the support during the years of my master's degree, I am extremely grateful for the discussions, I have learned so much from you.

Each and every one of the participants of my master's study - thank you for your contribution.

I want to thank my parents Ulla and Seppo for always being there for me, and for letting me grow up in an environment where music was always around and highly appreciated. It is the greatest thing to know that my happiness means the world to you.

Due to the love and care of Irmela Schödl, I have managed through my studies, learned two additional languages, and made it all work out while living abroad, things that I am extremely grateful for.

To all my wonderful friends in Vienna and Finland, all the things we have experienced together have formed me into the person I am nowadays. All that love I have received from you is the main resource of my strength and never-ending energy – you keep me going. Also, I would have never completed this thesis without the support of Riina Hiirikoski and Bartlomiej Cieczka, thank you for always taking such good care of me.

Special thanks I owe to Sebastian Schödl, for always hoping the best for me, standing on my side, seeing my strengths better than I do, and for carrying me through my weakest moments.

Finally, yet importantly, to my dear Lukas, your friendship is one of the most precious things I have ever achieved, vielen herzlichen Dank für die schönen Jahre an der Universität Wien.

THEORETICAL PART

1 Introduction

Language makes us human – it allows us to communicate with each other, share our thoughts, and express feelings. We learn it effortlessly, use it abundantly but only rarely think about it. Typically, the first language is acquired as an infant without being taught, whereas subsequent languages learned after puberty or in adulthood are studied to different degrees of competence under various conditions.

Learning to speak a foreign language can be indeed a challenging and time-consuming project, particularly when learning the language as an adult. When encountered with a foreign language, the first obstacle learners will face is the segmentation of foreign speech cascade into words, where at first the speech sounds like an uninterrupted stream of meaningless sounds. Learners tackle the challenging task of figuring out the word boundaries, which are not necessarily marked by consistent acoustical cues such as pauses or accents. Moreover, even if reliable acoustical cues for word boundaries are present, some minimal lexical knowledge would be a prerequisite for their efficient use (Jusczyk, 1995, Vinter & Perruchet, 2000).

The lack of a continuous target language auditory input is considered to be one of the biggest obstacles during the language learning process. Compared to the first language acquisition, where babies receive sonorous stimuli already in their mother's womb, in foreign language learning the auditory input is limited to occasional situations, for example in the classroom. Depending on which language is to be learned, the differences between the source language and the target language influence the degree of difficulty as well.

It is widely known that people learn things differently and at varied tempos and correspondingly this also applies to foreign language learning. The speed of internalization of the taught material depends on partly biological factors such as age, gender, the influence of the first language, and the genetic learning aptitude, or due to social factors like culture and socioeconomic status. Additionally, the association of psychological factors like motivation or personality to language learning has been extensively studied during the last decades (Herschensohn & Young-Scholten, 2012). Moreover, individual differences such as different learning styles and learning strategies are causing disparities between the language learners. Individual differences among language learners have indeed gained researchers's attention in a growing fashion in the last decades. In the 21st century, the broad investigation of language learning differences has already resulted in a shift towards more individualized teaching and learning methods in educational settings.

Yet another research interest in the domain of individual factors is the connection between musical aptitude and the talent to acquire novel languages. Many studies have shown that music and rhythm, and especially songs and singing, function as powerful aids to language learning. However, there is still considerable uncertainty regarding the reasons, which properties of music influence the language learning process, and to what degree musicality is affecting it. According to the most recent findings, different factors such as structural and motivational properties of instrumental music and songs seem to be one of the underlying sources of optimal language learning. Concurrently, the relation between language learners' language aptitude and musical aptitude is gaining more attention (Fonseca-Mora, Toscano-Fuentes, & Wermke, 2011).

Already the bare listening to music has been proven to affect learning processes. In 1993, the well-known theory of the Mozart Effect was introduced by Rauscher, Shaw, & Ky, as they claimed that after listening to Mozart's sonata for 10 minutes, normal subjects showed significantly better spatial reasoning skills than after periods of listening to relaxation instructions or pure silence. Even though newer studies have provided controversial results of the Mozart Effect, other studies have shown that background music is linked with improved performance on language learning tasks (Kang & Williamson, 2014). However, relevant factors to the results of learning success are the structure and volume of music, since the music that is high in complexity and loudness interferes with cognitive performance (Kämpfe, Sedlmeier, & Renkewitz, 2011; Schellenberg, 2012; Thompson, Schellenberg, & Letnic, 2012). On the contrary, the presence of low complex music, referring to non-verbal music, possesses stable tonality, and has minimal changes in tempo and amplitude, has been linked to improvements in task performance. Moreover, recent research also suggests that the presence of music may aid memory for new verbal materials (Wallace, 1994; Legg, 2009; Thiessen, Hill, & Saffran, 2005).

Another rather well-investigated domain where musical aptitude affects language learning is the connection of musical talent and speech imitation and pronunciation. Some scholars claim pronunciation and accent imitation to be the most important aspects of foreign language acquisition: to be able to get understood while speaking a foreign language, one must be capable of producing the right sounds. According to Seliger (1978), there are many critical periods for different aspects of language, and the ability to master a native accent in a foreign language, also referred to as phonetic ability, is the first one to be lost in puberty. Furthermore, in language learning, the phonetic ability is considered to be the only sub-ability that is completely subjected to a critical period (Moyer, 2014). Recent studies have shown a significant relation between speech imitation and musical expertise, where musicians outperformed non-musicians in language imitation tasks (Schön, Magne, & Besson, 2004, Milovanov, Huotilainen, Välimäki, Esquef, & Tervaniemi, 2008, Nardo & Reiterer, 2009, Reiterer, et al., 2011). Additionally, Christiner & Reiterer (2015) investigated if there are differences in results between different kinds of musical aptitudes, such as instrument playing versus singing. In their study, both instrumentalists and vocalists retained a higher ability to imitate unintelligible speech and foreign accents compared to non-musicians and non-singers. Interestingly, when compared between vocalists and instrumentalists, vocalists outperformed instrumentalists significantly. These findings indicate that different musical aptitudes contribute differently at least to the ability to imitate foreign speech. If vocalists truly benefit in this specific domain of language learning, the question arises, how can these benefits be implemented in other aspects of foreign language acquisition?

Although there is a growing number of studies dedicated to the link between music and language learning, very few of them aim their attention to the importance of how songs and singing affect foreign language learning and especially how singing might facilitate vocabulary acquisition in foreign language learning. The present thesis targets to contribute to the previous research and to enlighten the connection between foreign language aptitude and musical aptitude. Moreover, this thesis attempts to augment the body of research with the latest insights from an empirical study that was conducted by testing young German-speaking adults. This thesis aims to determine, whether a) singing songs improve L2 vocabulary acquisition more than merely listening to songs and b) if musical people with a singing background benefit from their musical experience.

2 Foreign language aptitude

First of all, it is important to delve into the definition of foreign language aptitude. Simply explained, foreign language aptitude refers to the specific talent that allows an individual to learn foreign languages more effectively than others (Carroll, 1962, 1981,1990; Dörnyei & Skehan, 2003). However, this definition cannot be successfully summarized in one sentence due to the multifaceted nature of the concept of language aptitude. During the last decades, various notions such as talent, skill, aptitude, language learning ability, or giftedness, have been under examination. Many authors agree with the Differentiated Model for Giftedness and Talent developed by Gagné & Deci (2005), according to which the term aptitude refers to the innate property that develops into a certain skill, which is then called a talent (Stern & Neubauer, 2013; Reiterer, 2018). To put it another way, aptitude is seen as an undeveloped, biologically inherited predisposition for acquiring a skill, something that people are born with. In comparison to that, talent is the outcome of what an individual develops over time out of aptitude, the proficiency that can be achieved.

When talking about (foreign) language aptitude (in the following abbreviated as FLA), it is necessary to keep in mind that this concept never applies to first language acquisition or multilingualism. Instead, language aptitude is always linked to the acquisition of novel languages. Already in the 1960s, the American psychologist J.B. Carroll, who is also considered as the father of FLA, defined language aptitude to be the state of readiness of individuals, which provides them with a certain capacity and facility to learn foreign languages if motivation and opportunity are provided. Learning a foreign language never takes place out of context and requires high cognitive abilities, such as working memory and metalinguistic awareness.

According to Carroll (1958), language aptitude consists of four components, which can be divided as follows:

1) Phonetic Coding Ability, refers to the ability to code unfamiliar phonetic material so that it can be retained and subsequently retrieved

2) Grammatical Sensitivity, is the ability to handle and identify grammatical functions that words fulfill in sentences

3) Inductive Language Learning Ability, capacity to infer syntactic and morphological forms, rules, and patterns from new linguistic content

4) Rote Learning Ability, which signifies the capacity to learn phonetic and grammatical associations

Even though Carroll's model of components dates back to the 1960s, it is still used as the basis for today's view of language aptitude. Theoretical advancement to Carroll's work proposed by Robinson (2005, 2012) introduces a model of language aptitude, where the concept of aptitude complexes includes not only the components of FLA but also focuses on the pedagogical side of language learning, such as learning circumstances. Additionally, it emphasizes the interaction between the different aptitude components. In Skehan's model (1986, 1991, 2002), Carrell's four categories have been merged into three, comprising the grammatical sensitivity and inductive language ability to one united category that goes by the name of language analytic ability. In this model, the three components of language aptitude are then combined with corresponding processes so that auditory processing refers to the phonetic coding ability, language processing to the language analytic ability, and memory to the associative memory.

The number of subcomponents associated with FLA denotes the complexity of this concept and provides an explanation for that language aptitude is something that 1) cannot be measured with one single test format and 2) cannot be expressed through one single score. Although FLA has formerly been seen as a relatively stable individual characteristic, current research provides evidence that FLA is a dynamic, multi-faceted composite of various cognitive skills that can be altered through practice under certain circumstances (Reiterer, 2018). Since language aptitude consists of several distinguished skills, the adapted test patterns for measuring language aptitude are often divided into different subsections. The next chapter provides an overview of the different language aptitude tests used in research.

2.1 Language aptitude tests

The most widely used test pattern for testing language aptitude is the Modern Language Aptitude Test (MLAT), published by Carroll and Sapon in 1959. As well as Carroll's work in language aptitude research, MLAT continues yet nowadays to serve as the standard basis for other similar approaches which aim to measure an individual's aptitude for learning a foreign language (Grigorenko, Sternberg, & Ehrman, 2000). In the MLAT five sections are testing phonemic coding ability, rote memory, and grammatical sensitivity. The fourth component underlying language aptitude according to Carroll, inductive reasoning ability, is not measured. The purpose of the MLAT is to predict the rate at which people will learn a foreign language in an instructed setting. Its primary use has been the placement of students in foreign language programs with curricular options as well as the selection of students for foreign language studies. Thus, some educators apply the test to determine the student's language strengths and weaknesses.

MLAT has received criticism regarding the availability, since the test pattern is only limited to a narrow group of operators, such as government agencies, missionary groups, or licensed clinical psychologists. Therefore, individual researchers, teachers, or students do not have access to the MLAT. Additionally, since the test format, "Paper-and-Pencil-Version", has not changed since its first release in 1959, it does not serve today's technological demands of research.

Resembling test batteries have been developed during the last decades to develop the language aptitude testing approach. Pimsleur's Language Aptitude Battery (PLAB) (Pimsleur, 1966) is similarly intended to predict an individual's success in foreign language learning. The main difference to MLAT is that the target group consists of a younger population. A similar type of test pattern developed by Petersen and Al-Haik (1976) is the Defense Language Aptitude Battery (DLAB), which like MLAT and PLAB, pursues to categorize students due to their proficiency.

An example of a more recent testing battery is the Cognitive Ability for Novelty in Acquisition of Language (Foreign) (CANAL-F) developed by Grigorenko, Sternberg, and Ehrmann (2000). CANAL-F is based on the model of language aptitude consisting of four essential aspects: 1) Knowledge Acquisition Processes, 2) Levels of Processing, 3) Modes of Input, and 4) Memory Processes (Rysiewicz, 2008). The difference to the previous approaches is that CANAL-F is described as dynamic, referring to the test method where individuals are asked to learn a new artificial language called Ursulu during the test.

Another recent language aptitude test is the LLAMA (Meara, 2005), which is similar to the aforementioned test batteries based on the core components of language aptitude defined by Carroll. The LLAMA test battery is computer-based and available online for free. This open-source principle imposes a big advantage compared to the other restrictive approaches. Even though LLAMA is based on MLAT, it operates language independent. This enables its use on a wider scale than MLAT, which is available only in a limited number of languages. It is precisely for the reason of the easy access and availability, that the LLAMA test battery was selected to be used as the measurement of language aptitude in this thesis.

3 Musical aptitude

In literature, the concept of musicality is constantly linked to language aptitude as an enhancing factor in language learning. Nevertheless, defining the meaning of musicality is challenging due to the different research angles of given authors. What makes an individual musical? Is musicality something innate, a stable element that is inherited, or can it be acquired and developed through practice? On one hand, some authors focus on the technical achievements of musicians, such as reading or remembering music, playing an instrument, whereas other authors associate musicality with the level of expression a musician can bring to work and the assorted sorts of receptive musical skills (Levitin, 2012).

Various definitions are used to refer to musical talents, such as musicality, musical ability, musical intelligence, and musical giftedness. These characterizations are broadly referring to a talent for music. The term talent, defined by many authors, is used to describe something being regarded as special, an exceptional capability in a given domain, and something is regarded as a potential, e.g., capable of development (Jorgensen, 2008). The term musical intelligence is used in the context where the cognitive content of musicality is underlined, as in the theory of multiple intelligences introduced by Gardner (1999). According to this theory, musical intelligence, as well as linguistic intelligence, is part of nine bits of intelligence that explain human cognition. Music is identified as something that requires intelligence, similar to a school subject such as language or math. This definition differs from musicality in a way that musicality is equated with talent and skill, something that can be nurtured and developed, whereas musical intelligence underlines the cognitive content of the music.

While Gardner’s idea of musical intelligence is understood as a single, general factor, Mark & Reimer (2003) portray musicality as a concept where an individual can be musically intelligent in various ways, for instance being intelligent in composing, performing, improvising, listening, in music theory, musicology or music teaching. Correspondingly, there are different levels of achievement that can be reached, i.e., individuals can show competence in some of these subdomains but not necessarily in all of them. As Reimer formulates it, the definition of musical intelligence should serve more as a tool for thought instead of a strict prescription, since the ability of an individual to practice music can be realized through proper education.

Gordon (1997) approaches the definition of musicality from the perspective that every human is born with some level of a capacity to develop musicality and prefers the term music aptitude. In Gordon’s point of view, there is a clear distinction between aptitude and achievement. Music aptitude refers to an inner possibility, the potential to learn or achieve in music, whereas music achievement expresses the outer reality, all entities an individual has learned in music. In a survey addressed to music teachers, Brandstrom (1999) discovered two views of musicality. *An absolute view* sees music as something that can be inherited and measured by music achievement, such as performing, composing, and improvising. According to this approach, only the minority of the population is musical and can demonstrate musicality. *A realistic view* recognizes being musical as something that all humans are capable of and many modern music teaching methods, such as Orff, Dalcroze, and Suzuki advocate and are based on this belief.

Nardo and Reiterer (2009) use the definition of musical talent concerning a predominantly innate tendency to understand and appreciate, perform, or create music outstandingly. According to them, a distinction between talent and musicality could be drawn to differentiate the nature vs. nurture approaches. In this context, talent or aptitude refers to the innate tendency to perform well in a given domain, the degree of which varies among different individuals, and which is rather independent of experience. Musicality or ability, on the other hand, is more likely associated with the resulting skill developed in interaction with the environment through training and practice.

Yet another way to define musicality is to examine the standards for arts education, as suggested by Jaffurs (2004). It is generally understood that musicality can be acquired either formally or informally. Formal music education practices are methods that are used by music teachers in classrooms and informal music education practices consist of methods that students develop on their own. The standards used in formal practices are provided by National Standards for Music Education (MENC,1994), developed by approximately 90 000 music educators in the United States. The nine content standards for art education include singing, performing on instruments, improvising, composing, arranging, reading, and notating music, listening, evaluating, curriculum integration as well as historical and cultural understanding of music. While these standards serve as guidelines for the formal music practices, the informal music practices are described to be natural, spontaneous responses to music, which receive no evaluation and follow no guidance. These informal practices value qualities and define musicality with concepts like playing with expressiveness or feeling, timbre sensitivity, repertoire readiness, imitating skills, and the ability to get along with other musicians, as stated in a survey with amateur musicians (Green, 2008).

The comparison between formal and informal music practices demonstrates the diverse perspectives that should be considered while defining the concept of musicality. As Jaffurs (2004) points out, musicality has a range of multiplicity in its definition, it is more than just a skill but a partially innate concept that is strengthened by a nurturing environment. Musicality does not have to be taught traditionally to be learned and there are countless ways to practice musicality through different music styles and instruments. Leaning on the comparison of both formal and informal music practices, Jaffurs proposes a new definition of musicality that includes different properties. Firstly, musicality is the competence to be technically proficient and expressive on an instrument or voice and to compose good music. Secondly, musicality is also the ability to remember musical experiences and listen globally, and to grasp all. Additionally, musicality is the desire to find out what is important and to anticipate what is needed, as well as to keep the music alive and to feel what counts. Thirdly, musicality is to know how and where to fit into the overall feel of the piece. Finally, musicality is to find delight and joy in the listening, performing, and composing of music and musicality requires a high degree of interpersonal skills.

Jaffurs' definition includes a wide range of diverse features and explains for its part the difficulty to create a clear-cut characterization of musicality. The next matter to be tackled is the question of how musicality can be measured, especially when the definition consists of such a variety of components. The first thing to be addressed is the question of whether talent is an exclusively innate phenomenon or if nurture plays a role. Most of the authors acknowledge musicality to be innate and hereditary, although the role of contingent learning factors and the social environment should not be underestimated. Especially in the domain of music, where the ability to play an instrument for instance, can be finely trained. The puzzle is to separate aptitude from achievement.

The next question that arises is whether talent is normally distributed in the population or is it rather an “all-or-none” phenomenon. The popular belief states that only some individuals are gifted whilst others are not, but the recent scientific literature confirms music aptitude, among other human characteristics, is normally distributed in the population (Jorgensen, 2008). In other words, everybody has a certain degree of potential to achieve musical skills, with relatively fewer very high- and low-talented persons, and the majority with an average aptitude. One widely used approach is to classify individuals as musicians and nonmusicians based on the outcome of interest. The main limitation of this kind of approach is that dichotomizing individuals into musicians and nonmusicians might turn out misleading, since being a “nonmusician” does not denote an absence of musical ability (Law & Zentner, 2012; Zentner & Strauß, 2017). Instead, the musical ability of nonmusicians might remain undiscovered, or the circumstances may have prevented its development. According to the terminology introduced by Zentner and Strauss (2017), individuals among the musically untrained, who are still able to reach a high level of musical proficiency if given the time and opportunity to do so, are referred to as musical sleepers, because of their existing but dormant musical skills. In contrast to that, the term sleeping musicians refers to individuals whose musical proficiency languishes despite multiple years of training.

In conclusion, musical aptitude is a very broad concept that everyone possesses to some degree. It consists of several various abilities, which are based on both innate predispositions and experience. Different circumstances and the amount of musical training contribute to the level and degree of musical aptitude.

The next chapter addresses the question, of how musical aptitude can be measured, i.e., what sub-components and fundamental abilities musicality consists of.

3.1. How to measure musicality

To be able to score musical skills, it needs to be defined, what is measured and how. What is agreed on by the majority of the authors regarding measuring musicality is the importance of perception and cognition of musical patterns and structures. Shuter-Dyson (1999) proposes a system of five groups of fundamental abilities that form the base for musicality tests: tonal, rhythmic, kinesthetic, aesthetic, and creative abilities. Each one of these elements can be broken down further into other sub-components and each element is contingent on improvement with age and exposure.

Tonal abilities consist of pitch perception, sense of tonality, and harmony-polyphony. Pitch is one of the primary auditory sensations and plays a defining role in music and speech. In the speech, rising and falling pitch contours provide a basis for defining prosody, and in tone languages pitch contours are helping in the comprehension of the meaning of the words, prominent examples being Mandarin and Cantonese. In music, pitch sequences define melody and simultaneous combinations of pitch define harmony (McDermott & Oxenham, 2008; Oxenham, 2012). In the context of musical aptitude, pitch perception comprises the ability to discriminate between different pitches. Sense of tonality refers to the development of a pitch system where the tone relations are specifically defined based on their inferred relation to the tonic (the first note or degree of the diatonic scale, major or minor). In other words, the sense of tonality tells us whether a piece of music sounds right or not. Harmony-polyphony sums up the ability to detect incorrectness or violations in chord sequences (sequences where several melodies are sounding at the same time). The rhythmic abilities include sub-components such as meter abstraction, perception of rhythmic structures, rhythmic anticipation, practorhythmic factor, and tempo-tapping. Meter abstraction, perception of rhythmic structures, rhythmic anticipation, and tempo-tapping all refer to abilities to perceive, comprehend and create rhythmic patterns, whereas practorhytmic factor means the coordination of limbs in rhythmic movements. The kinesthetic abilities refer to the motor components of music performance. These abilities have been investigated to influence the ability to improvise, the expressivity, as well as the auditory perception of a performer (McPherson, 1993/1994; Davidson, 1993).

Finally, both aesthetic and creative abilities are again complex concepts to define. It is hard to say, what is creativity? What factors make a music performance aesthetic? Broadly speaking, aesthetic abilities are linked to definitions such as expression, appreciation, and emotion. Creativity in the context of musicality is often associated with improvisation (Vaughan, 1977; Webster, 1983; Wang & Kageff, 1985) and personality traits such as imagination, curiosity, and anxiety (Swanner, 1985). Furthermore, all creative abilities are linked to musical extensiveness, flexibility, and originality (Nardo & Reiterer, 2009).

When contrasting all the previously introduced concepts of musicality, it is noticeable that different authors prefer their views to define, what musicality is about and what subcomponents belong to it. Therefore, it is self-evident that the various test patterns developed to measure musical aptitude vary in the same way. To provide an overview of the idea of how musicality is measured, the most popular approaches are shortly discussed in the next chapters.

3.2 Seashore´s Measures of Musical Talents

The oldest standardized test available to measure musical aptitude is Seashore’s Measures of Musical Talents, which was published in 1919, followed by a revision in 1939 (Seashore, Lewis, & Saetveit, 1960). The Seashore Musical aptitude test is founded on the presumption that musicality consists of six different domains, each of them forming a subtest of the musicality test. The test is based on a strong psychophysical approach focusing on the very basic sensory capabilities. Each category of the test presents subjects with a series of pairs of tones, with the assignment of discriminating a certain physical characteristic between them. The different subcategories of the Seashore Musical aptitude test are pitch, loudness, rhythm, time, timbre, and tonal memory. To give an example, each subtest presents two items, such as notes, or rhythmic patterns and the subject’s task is to indicate whether these items differ or not. In addition to that, in each subtest only one component is varied at the time, meanwhile, others are kept constant and as simple as possible. This and the fact that the test measures immediate sensory acts that are not improved with practice enables testing for both young and old subjects as well as for musicians and non-musicians (Nardo & Reiterer 2009).

Seashore’s test was the pioneering work in the field of musicality tests, and it contains historical importance being the first test pattern to be fully standardized. Notwithstanding, it has received heavy critique regarding its weakness in predicting musical ability. According to Wing (1970), Seashore’s test tests the most basic qualities of musical capacity in a very elementary way. For example, the pitch subtests are too simple for a musician, and the measures are too subtle to be discriminated against. The time and intensity tests are according to Wing the least satisfactory since they do not test time and intensity as they are used in music.

3.3 Wing’s standardized tests of musical intelligence

The main idea of Wing’s test, developed and revised between the late 1930s and 1970, was to create a test that combines the pragmatic experience of musicians and the experimental experience of psychologists. Wing’s test of music aptitude is independent of musical training and indicates the following items:

1) the mental processes implicated in music fruition

2) the distribution of musical aptitude in the population

3) the development of musical aptitude with age and learning

4) the influence of the environment and culture

According to Wing, the preceding test patterns (including Seashore's) suffered from deficiencies, such as ignoring qualities a musician regards as desirable, and they measured only one aspect of music and treated such a measure as general musical capacity. Hence, previous test patterns were validated on rather small groups and the test scores were not adequately standardized, plus the effect of musical training on the test scores was neglected. Furthermore, these tests were difficult to repeat due to the narrow age range used. Wing aimed to build a test pattern that would cover the weaknesses of the previous test patterns (Wing. 1970, Nardo & Reiterer, 2009).

Wing's test pattern consists of two sub-tests, where the first test focuses on perceptive aspects (ability subtests) and the second test covers cognitive components (appreciation subtests). The main aspects Wing holds for crucial when defining musical talent is the musical ability and musical appreciation. Musical ability, according to Wing, includes not only the ability to play an instrument, but also the speed in learning to play, the ability to perform an aural test, and the ability to carry out musical activities such as composing. Musical appreciation in turn is used to refer to the power to recognize and evaluate artistic merit in music, as well as to the aesthetic judgments of music.

The newer version of Wing's test dates to 1970 and consists of seven subtests listed below (Wing 1970):

1) chord analysis, detecting the number of notes played in a single chord

2) pitch change, detecting an alteration of a single note in a repeated chord

3) memory, detecting an alteration of a note in a short melody

4) rhythmic accent, choosing the better rhythmic accent in two performances

5) harmony, judging the more appropriate of two harmonizations

6) intensity, judging the more appropriate mode of varying loudness in two performances of the same melody

7) phrasing, judging the more appropriate phrasing in two performances – grouping of notes by pauses, legato and staccato playing, etc.

Compared to the preceding test formats, Wing's test benefits from the use of original material, higher reliability, and validity, as well as from easier and more uniform administration. The sub-tests require no special knowledge of musical technicalities, and the degree of difficulty enables the use of the test pattern among a wide age range. The time used to complete Wing's test pattern takes approximately one hour.

3.4 Gordon's series of musicality tests

Another popular series of musicality tests in the literature is the test pattern devised by Gordon in the 1960s. One reason for the recognition of Gordon's test format traces to the art of it, which allows testing subjects from various age groups as well as from different expertise levels at the same time. In addition to that, Gordon's tests have been recorded on CD which has served the purposes of the digital era. The various aspects of musicality tested in Gordon's tests focus on components such as pitch ability, rhythmic ability, performance, and expression preferences, as well as the ability to improvise, for instance.

Characteristic for Gordon's way to approach musicality is the clear distinction between aptitude and achievement. For Gordon, music aptitude refers to an inner possibility, the potential to learn or prosper in music. On the contrary, music achievement indicates an outer reality, all that somebody has learned in music. Following this distinction, Gordon claims that individuals with a higher level of music achievement have a correspondingly high level of music aptitude, which is not necessarily the case the other way around. According to Gordon, this means that individuals with low music aptitude can achieve more success in music achievement when receiving proper instruction compared to individuals with average music aptitude who only receive improper instruction.

Although the distinction between aptitude and achievement seems like a statement for music aptitude being a product of nurture, Gordon assumes that music aptitude consists of both nature and nurture. According to him, music aptitude is something innate, where heredity does influence it, yet not being the determinator of it. Gordon argues that the environment plays a decisive role when it comes to the facilitation of music aptitude – the innate potential accompanied by a rich music environment is fundamental for ideal music aptitude.

The musicality tests developed by Gordon can be split into two groups in respect of the different features which are tested. The developmental music aptitude tests focus on tonal and rhythmic patterns, where the responses consist of “same/different” or “same/not same” answers. These kinds of tests are targeted at children aged three to four, PMMA tests (Primary Measures of Music Audiation, 1979) aim for children in grades from kindergarten until third grade, and IMMA tests (Intermediate Measures of Music Audiation, 1982), include the target group with the age range of children from six to eleven years. The other group of Gordon’s tests, stabilized music aptitude tests, apply “same/different”, “like/different”, and “yes/no” responses, as in test formats like AMMA (Advanced Measures of Music Audiation, 1989a), or MAP (Musical Aptitude Profile, 1965). The AMMA test targets a group of high school/university students and consists of approximately 30 items, where a short musical statement is followed by a short musical answer with the same number of notes. The task of the participant is to answer whether the musical statement and answer are the same or different, and if the answer is different, the participant needs to tell if the difference is tonal or rhythmic. The MAP test functions in a similar way as AMMA, where the task of the participants is to listen to two melodies but instead of indicating whether the melodies are the same or different, participants need to decide which one sounds “better” (Gordon E. E., 1979, 1998, 2004).

3.5 PROMS

One of the newer types of musicality tests is the Profile of Music Perception Skills (PROMS; Law & Zentner, 2012), which was developed to answer the demands of criticism addressed to the previous test formats. It has been argued that the older approaches were not capable of identifying the musical talent in musically untrained individuals, and PROMS was created to counteract this limitation. PROMS measures perceptual musical skills objectively across multiple modalities, such as timbre, tuning, melody, pitch, rhythm, and tempo. The distinctive feature in comparison to the older musicality test formats is, according to the developers, the objectivity for the comprehensive assessment of music perception skills and unlike the other tests, PROMS is the first test battery to include timbre, tuning, and advanced rhythm discrimination tasks to the analysis. Through PROMS it is possible to test musical ability both in musicians as well as in non-musicians. Additionally, the high stability of across-browsers and OS provides PROMS the advantage to be perfectly suited for online data collection.

PROMS is available in different versions, with different durations, which is a major advantage for the test format allowing it to be more time-effective (Zentner & Strauss 2017). Full PROMS is the original test version containing all nine test components are included and the duration to complete the test is approximately 60 minutes. A shorter version of Full PROMS is PROMS-S with 8 subcategories and the completion time is about 35 minutes. Modular PROMS is an adaptation where the various subtests can be configured in a customized way. Finally, the shortest versions of PROMS are the Brief PROMS, which test melody, metric accent, tempo, and tuning with a duration of 30 minutes, and Mini PROMS, which is a shorter version of Brief PROMS with a duration of 15 minutes. In addition to these, the developers of PROMS are working on an even shorter version of PROMS, which is named Micro-PROMS and is planned to have a duration time of 10 minutes.

Furthermore, PROMS is available in various languages e.g., Arabic, Chinese, English, German, French, Spanish, Norwegian, Russian, and Japanese, which enables the test to be used in international settings. Multi culture-fair stimuli make PROMS suitable also for cross-cultural comparisons.

As previously mentioned, PROMS consists of nine subtests that test various modalities such as pitch, rhythm, and sound quality. In each part of the test battery, the task of the participant is to indicate whether the presented stimuli are the same or different.

Due to the easy accessibility (online access at the University of Innsbruck) and relatively short duration (15-20 minutes), Mini-PROMS was chosen for the musicality test for the empirical part of this thesis.

4 Music and language

Music is rhythm, a reflection of our physiological life. Furthermore, music is melody, a sequence of sounds that are closely connected with the emotional life of humans. And lastly, music is harmony and simultaneity of sounds that are analyzed and studied (Fonseca-Mora, Toscano-Fuentes, & Wermke, 2011). So far, the concepts of language aptitude and musical aptitude have been closely examined and the next matter to be addressed is the link between these two. The following chapter focuses on the core and essence of music and language and how these two systems are intertwined.

4.1 Evolutionary origins of music and language

To be able to understand, why and how music and particularly singing play a role in language learning, it is necessary to have a look at the common ground of music and language. As summarized by Winters & Griffin (2014), these two substantial human universals define identities, represent thoughts, symbolize feelings, and unite communities through social practices. Until this day, it is still highly debated, which came first, music or language.

The speculations of music's origin have a long history and several academics have debated about the possibility of music being an evolutionary adaption. Some authors believe that music has never possessed any survival value and merely exploits an existing pleasure channel, whereas other authors advocate the theory that music has at one time contributed directly to human survival but has subsequently shifted to something irrelevant, a piece of evolutionary litter (Huron, 2001, Masataka, 2009). Among others, Pinker (2013) stated that instead of having an adaptive role, music is more like an "auditory cheesecake", a byproduct of natural selection that just happened.

The most common proposals for the evolutionary theories of music identified by Huron (2001) encompass the communal effects of music. Music may have developed to create and maintain social cohesion while contributing to group solidarity and promoting altruism, and consequently increasing the effectiveness of collective actions, for example when defending against a predator or attacking a rival clan. Similarly, music might have contributed to the coordination of group work, such as pulling a heavy object, in favor of group effort. Additionally, the essence of music could have contributed originally to conflict reduction, suggesting music and singing being safer social activities than bare discussions that could lead to arguments.

One of the most popular theories to endorse the assumption, music is an evolutionary adaption, is the mate selection theory. According to this, music-making behaviors may have arisen as courtship behavior. Examples of dance and song playing an integral part in human courtship can be found all over the world when observing Salsa dancing in a Latin club, or the ritual singing and dancing of the Masai in Kenya, or night clubs offering a perfect setting for close-by singing and dancing (Binns, 2016). Charles Darwin (1872) suggested that the origin of music has arisen because of the sexual selection in mating calls. Equivalent to the peacock's tail which is a classic example of sexual selection, the preferences of women could create an escalating competition for even more elaborate and beautiful melodies. That would mean that the best musicians or dancers have more reproductive success than their less musical colleagues. Supporting this theory, Miller's (2000) hypothesis draws a line between the appearance of musical interests and sexual selection in adolescence. Furthermore, Miller reported that male musicians produced at least 10 times more music than females and were most productive around the age of 30, near the time of peak mating effort and activity. Although these views have gained lots of interest in the research field and for the first glimpse, they seem to offer an easy explanation for the origin of music, no evidence would indicate that one sex is more musical than the other. Even though women may be impressed by men who serenade them, in the same manner, women are capable of serenading men. Therefore, as the particular genetic preference is established by the opposite sex is a precondition for the presence of sexual selection, no known example meets the requirements of sexual dimorphism.

Other evolutionary theories of music listed by Huron include perceptual development and motor skill development theories, where music is hypothesized to provide a sort of "exercise" for hearing, in the sense of music teaching people to be more perceptive. Alternatively, singing could have provided possibilities to refine motor skills, where music and particularly singing could have been a precursor for the development of speech. Huron also mentions the idea that the origin of music stems from a way to usefully convey information. This theory of transgenerational communication would explain the ubiquity of folk ballads and epics, that successfully convey communication over long periods.

After considering all the proposed theories, the lack of sufficient evidence for music being an evolutionary adaption remains. Some basic evidence exists, such as the assumption that for a behavior to be adaptive, it must be very old. In this manner, music and music-making do indeed possess great antiquity. Additionally, behavioral specializations have been associated with specific anatomical or functional brain structures (Huron 2001). In this context, the neurological evidence is consistent with the possibility that there are specialized music-related brain structures. Furthermore, the universality of music should not be ignored in the discussions of the evolutionary theory of music. Music does exist all over the world in all cultures and an even more crucial indication for the evolutionary origin of music is the ability of human newborns to react to music stimuli. In other words, music seems to be hardwired into the brains of human babies.

Meanwhile the lack of sufficient evidence sides with the assumption that music serves no obvious adaptive purpose, some hypotheses indicate a shared evolutionary history between language and music due to their similarities (Masataka, 2009). If the evolution of music provides a huge puzzle for evolutionary biologists, correspondingly controversial is the debate on the origin of human language. Despite all the controversy, a certain consensus about the origin of the language system exists. Presumably, the language system has developed using natural selection, where more accurate communication helped early humans to survive and reproduce. A great adherence of scholars also supports the view of learning activity being genetically preprogrammed. Children, regardless of the language system being acquired, learn the subcomponents of the system one after another during acquisition, with the time course remaining largely consistent.

What comes to the academic debate over the relationship between music and language regarding their evolutionary aspect, an increasing amount of evidence exists favoring the idea that either language went through a musical phase during evolution, or that music is derived from language (Dunbar, 2004, Wilson, 2012). For instance, studies focusing on infant-directed Speech (IDS), a specialized style of speech that caregivers use when talking to infants, promote the fact that this speech register is based on musical rather than linguistic aspects when it comes to conveying the meaning. Also, acoustic analysis examinations show that there is a distinctive prosodic structure in IDS, vowel hyperarticulation, that might facilitate word learning. This structure is characteristic also in Foreigner-Directed Speech, where it has been suggested to act as a didactic device for the instruction of language learners (Papousek & Hwang, 1991, Kemler Nelson, Hirsh Pasek, & Jusczyk, 1989, Uther, Knoll, & Burnham, 2007). Another proof for a common evolutionary history between music and language, according to Binns (2016), is the prosody of traditional songs. The musical patterns of babies crying correspond to those of their mother’s language already long before the children can form words. Just like each language takes shape differently from music, traditional music evolves accordingly around the given language. According to Binns, the prosody found in traditional music is one of the easiest ways of illustrating a language’s rhythm, word emphasis, word intonation, and pronunciation.

After highlighting the various evolutionary theories of both music and language, in the scope of the next chapter, there are several similarities that music and language have in common, as well as the features that separate them from each other.

4.2 Similarities and differences of music and language

Language and music both share a vast number of characteristics and are strongly intertwined. They are universal and specific to human beings and greatly culture-bond, the essential elements being rhythm and melody. Both act in accordance with rule-governed, fixed structure, either by the way words are used in a language or by musical notes in music, in order to create a sentence or a melody (Fonseca-Mora et al. 2011). Besides, music and language resemble each other through the property of recursion in a way where a sentence can be lengthened indefinitely by the insertion of words or other sentences. Similarly, a musical piece can be altered by the insertion of phrases or other musical units (Trehub, 2003).

Additionally, language is broadly viewed as a medium for human communication, consisting of the use of words in a structured way either in spoken, written, or gestural expression. While declaring love, ordering a meal, or asking for directions, people do not speak just for being heard, but to be understood. But when having a closer look, the speech seems to be the symbolic part of language, while the sound is the bearer of its message. Accordingly, the natural environment for both music and language is auditory, particularly vocal. These two auditory phenomena consist of perceptually discrete elements that are organized into hierarchically structured sequences; where an individual note belongs to the larger constituent of a musical composition, similarly in language phonemes form together broader discourse units (Arleo, 2000; Sloboda, 1985; Patel, 2010). Finally, the fundamental characteristics that music and language have in common, are the processing of sounds, the conveyance of messages, the learning by exposure, and the sharing of intrinsic features like pitch, volume, prominence, stress, tone, rhythm, and pauses (Fonseca-Mora, 2000).

Apart from these similarities, there are also differences where language and music part company, summarized by Fonseca-Mora et al (2011). To be addressed first are the differences concerning meaning. While there is no need to translate music, a language normally needs to be translated into another language to be understood. This remark does not imply music is not meaningful, because both music and language do show the duality of patterning through the combination of discrete, meaningless elements which form meaningful structures. However, the resulting musical pieces are not similarly meaningful as the verbal utterances are since music lacks semanticity. Musical pieces are therefore those of a “non-referential” nature, whereas verbal utterances do possess referential features.

Similarly, regarding competence, differences between music and language arise. Most people are familiar with different music styles whereas being competent in many languages is exceptional. Secondly, whereas grammar provides meaning to language, musical rules have nothing to do with the meaning of music. The evolution of the range of musical styles is a rapid process compared to the evolution of grammar of a language. Thirdly, the numbers and variety of tones are similar in all cultures, whereas the numbers of phonemes connected to language vary.

4.3 Music in first language acquisition

In the first language acquisition, speech and song are spontaneously developed at the same time. Besides being able to understand what is being said, crucially important for language acquisition is to learn how something is being said. Intonation, one of the most essential prosodic features is the key to the perception of word stress and the recognition of sentence structures. In the first language acquisition (from now on L1 acquisition), intonation enables language to become accessible to a child. Prosody makes the segmentation of audible speech stream and recognition of meaningful structural elements possible and from the very beginning, the melody is the most noticeable prosodic element in the acquisition process (Fonseca-Mora et al. 2011).

The integration of music and language accompanies language development already from the prenatal phases of a baby. Several studies have demonstrated the ability of fetuses and newborns to distinguish between voices, for instance, the voice of the mother versus a stranger or male versus female voice (Kisilevsky, et al., 2003; Lecaneut & Granier-Deferre, 1993) The perception of music by the fetus in the utero is well investigated (Kisilevsky et al. 2003). Upon birth, infants show a preference for those melodies and rhythms to which they were exposed prenatally, and they will most frequently respond to their mother's voice (James, Spencer, & Stepsis, 2002; Damstra-Wijmenga, 2009). These findings show the extensive musical perceptual capabilities of human infants. Subsequently, before babies can pronounce complete words, the very first sounds babies produce are crying sounds, which demonstrate familiarity with the intonation patterns of their caregiver's language (Rukholm, 2015). Still well before their first birthday, babies can reproduce the intonation of the ambient language and they start to imitate rhythm and melodic contours which are later replaced by phonemes (Mampe, Friederici, Christophe, & Wermke, 2009).

Language acquisition is strongly based on interaction and for communication to become successful, affection is the mediating force. Caregivers speak to infants and small children in a highly melodic manner that can be distinguished from normal speech patterns by differing pitch levels, tempo, and rhythm as well as by exaggerated melodic expressions. This speech register is known as child-directed speech, (CDS) or as previously mentioned, infant-directed speech (IDS). This is the very first form of interaction and language input that a child receives (Fernald, 1989). More recent evidence also demonstrates the preference of infants for the CDS and exemplifies how messages are conveyed to the preverbal human infant using melody (Fernald A., 1992, 1993, Falk, 2011). Musical abilities are presumably very significant in the acquisition and processing of language. Features such as speech melody, meter, rhythm, and timbre function as prosodic cues of language, thus infants acquire much information about word and phrase boundaries through them (Jusczyk P. W., 1999).

CDS also shares many similar features with teacher talk, which has been described as a simplified code created to help the hearer to learn and understand language (Fonseca-Mora et al. 2011). Besides the melodic features, these speech registers also share features such as the frequent use of repetition, expansions, preference for simplified vocabulary, change in voice volume, and modification of intonational contours. These features are seen as species-specific learning guidance towards language (Wermke & Mende, 2009)

Exposure to the combination of musical and linguistic input accompanies children throughout their childhood – music is integrated into lullabies, television programs, video, and computer games and is providing background for activities. Melodies and music accompany the first language acquisition in general and are also present in the language teaching context as well.

5 Cognitive aspects of music and language

“*If we were able to explain music, we could possibly find the key to human thought-”*

-Claude Levi-Strauss

So far, the intertwined nature of music and language has been examined closely as well as the way music accompanies language acquisition during the different stages of human life. Also, the properties that link music and language to each other, and the features separating these two concepts have been presented. The next thing is to scrutinize the cognitive grounds of both music and language, which will be covered in the following sections.

Language is considered the most human of all cognitive functions because it permits us to communicate thoughts relevant to the current context, to explain stories that happened in the past, and to project into the future to plan upcoming events. People typically utter hundreds or even thousands of words per day, which requires a complex processing system subsuming attention, memory, and motor abilities. Several stages of processing are included in creating speech: morphology, phonology, semantics, syntax, and pragmatics (Besson, Chobert, & Marie, 2011).

Although music requires in a similar manner different levels of processing, such as rhythm, melody, and harmony, it does not function the same way as language. In contrast to language, music does not allow people to discuss present, past, and future events. What comes to the expertise and practice of music, all individuals do not master music to similar extents, whereas in language the level of capability is more or less the same over the population.

5.1 Language in the human brain

Language and communication activate almost every part of the human brain. Even though the use of language prevails, other cognitive functions such as attention, memory, emotion, and executive processes are simultaneously involved. For centuries, neuroscience has attempted to explain the responsible neural networks that explain how our brain “understands”, “speaks” and “writes”. Through observations from patients with language impairments in the past centuries, brain research has managed to locate several language-specific areas in the brain, such as Broca’s and Wernicke’s areas discovered in the late nineteenth century (Friederici, 2017). Located in the left hemisphere, Broca’s area is involved in the production of speech, and Wernicke’s area is linked to the comprehension of speech. Since the late twentieth century, as new neuroscientific methods and technical innovations developed, for instance, fMRI (functional magnetic resonance imaging) and PET (positron emission tomography), the relation between language and the brain can be observed in living persons meanwhile the language is processed. Other language-relevant brain regions are the motor cortex, which is associated with the articulation of speech, and the arcuate fasciculus, a bundle of nerve fibers responsible for the connections between Broca’s and Wernicke’s areas (Yule, 2014).

L1 processing, which refers to the brain activity when hearing, understanding, and saying a word, is arranged by a definite pattern. The initial stage of the language comprehension process is the auditory perception, implying the moment when the word is heard in the first place. Firstly, the processing of phonemes and the acoustic-phonological analysis takes place in the left middle portion of the superior temporal gyrus (STG), lateral to Heschl’s gyrus, where the primary auditory cortex (PAC) is located. The auditorily presented words are processed in a region anterior to Heschl’s gyrus in the left STG. The next step to occur is the processing of syntactic information of the perceived words, which takes place in the anterior superior temporal cortex. Studies conducted by fMRI have reported lexical-semantic processes to take place in the middle temporal gyrus (MTG). Compared to the word level, the semantic processes at the sentential level are more challenging to locate, but arguably they seem to be localized in the anterior temporal lobe, posterior temporal cortex, and angular gyrus (Schön, Gordon, Campagne, & Magne, 2010).

Once the perceived words have been processed and understood, the process for speech production begins. Speech production has been localized to generally take place in the Broca's area. Injuries to the left hemisphere are known to cause deficits in language production. Neurologists typically classify aphasias into two major categories based upon how spoken language abilities are affected. Fluent aphasia patients produce quite effortless speech while the content of the speech varies from good to incomprehensible. In non-fluent aphasia instead, the speech seems slow and effortful marked by reduced grammatical complexity and phrase length, but with a relative sparing of content word use (Callan, Callan, & Jones, 2014).

Several functional neuroimaging studies have explored single word production, which have identified the left hemisphere to be involved in word production. One major challenge for speech production studies is that naturalistic or spontaneous speech production is seldomly being explored due to the unpredictable head movements causing image artifacts, especially for fMRI (Gracco, Tremblay, & Pike, 2005). Therefore, many of the studies focus on imaging single-word production. For instance, Indefrey & Levelt (2004) identified 11 regions in the left hemisphere and 4 regions in the right hemisphere, which are linked to the core processes of word production. In addition, it was possible to differentiate specific tasks for certain regions, such as lexical selection linked to the left MTG (middle temporal gyrus), self-monitoring to the left STG (superior temporal gyrus), articulation to the sensorimotor cortex, and phonological code retrieval to the Wernicke's area.

However, language production seems to be a more complex process than assumed previously. Language-related processes that are linked to speech processes have been speculated to involve the right hemisphere as well. More detailed, Davis, et al. (2008) proposed, that the processing of intonation is located in the right hemisphere, involving the frontal operculum and regions in the superior temporal gyrus.

5.2 Music in the human brain

The neural mechanisms of the human brain linked to music, particularly producing and perceiving music, create a great challenge for cognitive neuroscience. Over the last decade, music processing has become an area of intense and systematic study. A major motivation for the research is that the study of brain organization offers a better view to understand the brain plasticity and the inner working of music processing. Additionally, analyzing brain anomalies provides information about to what extent and at what level music processing recruits neural networks that are distinct from those involved in other auditory-vocal functions such as language (Peretz & Zatorre, 2005).

Some brain regions that have widely been considered language-specific, such as Broca's and Wernicke's areas, are found to be activated also when processing music (Keenan, Freymann, Stroud, & Walter, 2001; Levitin & Menon, 2003). The neuropsychological literature presumes that melodic and temporal structures of music are processed independently in the brain. Studies show that brain damage can interfere with discrimination of pitch relations while sparing the accurate interpretation of time relations, and similarly, the rhythmic discrimination of musical events can be impaired while extraction of pitch content is spared (Ayotte, Peretz, Rousseau, Bard, & Bojanowski, 2000). The computation of pitch relations has been located to take place in the right temporal neocortex, as studies of patients with focal brain excisions demonstrate (Liégeois-Chauvel, Peretz, Babai, Laguitton, & Chauvel, 1998). In addition, damages to the right anterolateral part of Heschl´s gyrus have been found to cause deficits in specific aspects of pitch processing, such as the ability to determine the direction of pitch change and analysis of pitch information (Johnsrude, Owen, White, Zhao, & Bohbot, 2000; Tramo, Shah, & Braida, 2002).

For the processing of temporal organization of musical sequences, two different categories of time relations are relevant: the segmentation of an ongoing sequence into temporal groups of events based on their durational values and the extraction of an underlying temporal regularity or beat (Fraisse, 1982). Findings from studies with patients with lesions of the right temporal auditory cortex report that the right hemisphere is responsible for the meter whereas the grouping of sequences locates in the left hemisphere. The patients with such brain damage were not able to tap the beat or generate a steady pulse, whereas the discrimination and reproduction of irregular temporal sequences remained intact (Fries & Swilhart, 1990).

When analyzing the effect of musical expertise on brain anatomy, it is remarkable how the different types of musical abilities (i.e., playing different instruments) are imprinted in the human brain. Studies of Magnetic Resonance Imaging (MRI) and Magnetoencephalography (MEG) show the development of different motor abilities in the brain structure due to extensive practicing of an instrument (Besson, Chobert & Marie 2011). For example, pianists acquire extraordinary velocity for fingers of both hands, whereas violinists only use the left hand because the right-hand holds the bow. For violinists, the representation of left-hand fingers in the motor cortex is larger than in non-musicians, where there are no differences between the representation of hands (Elbert, Pantev, Wienbruch, Rockstsroh, & Taub, 1995). Because playing an instrument usually requires efficient hemispheric transfers, the corpus callosum of musicians is larger than the one of non-musicians (Schlaug, Jäncke, Huang, & Steinmetz, 1995; Schmithorst & Wilke, 2002). Studies by Amunts, et al. (1997) and Hutchinson, Whiteley, Smith, & Connors (2003) also demonstrate that the posterior part of the precentral gyrus is more symmetrical in pianists than in non-musicians.

Worth noting is the discovery, that several regions known to be important for language processing are correspondingly more developed in musicians than in non-musicians: Broca' areas (Levitin and Menon 2003; Maess, Koelsch, Gunter, & Friederici, 2001; Vuust, Roepstorff, Wallentin, Mouridsen, & Ostergaard, 2006), the left Planum Temporale (Keenan et al. 2001; Schlaug et al. 1995) and the inferior frontal gyrus and the inferior lateral part of the temporal lobe (Gaser & Schlaug, 2003; Luders, Toga, Lepore, & Gaser, 2009). Furthermore, several studies discovered that white matter architecture is shaped by musical expertise (Bengtsson, et al., 2005; Imfeld, Oechslin, Meyer, Loenneker, & Jancke, 2009). In their study, Imfeld et al. identified, that the water diffusion values in the cortico-spinal fiber tract of musicians were lower than those of non-musicians. This finding is interpreted to reflect the plasticity of the axonal membrane, which is linked to the production of sounds.

Whereas playing an instrument activates more parts of the brain controlling motor movements, singing, albeit being a form of musical expression, resembles most closely the spoken voice through to its basic acoustic profile. In the study of Schön et al. (2010), subjects were presented either with spoken or sung stimuli, and the results revealed both conditions activating similar clusters in the middle and superior temporal gyrus. Similar outcomes were found in the study of Callan, Tsytsarev, Hanakawa, & Callan, (2006), where six well-known songs were compared both in spoken and sung form. Compared to speech, singing was reported to cause greater activity in the right planum temporale.

Even though singing and speaking activate partly the same brain regions, several studies suggest that there are differences in the neural processes which underlie singing and speaking. For example, many aphasia patients that have severe deficits in their ability to speak as a result of left inferior frontal lobe damage, maintain the ability to sing words without much deficit (Hebert, Racette, Gagnon, & Peretz, 2003). This phenomenon has been explained due to the central role of the left hemisphere in propositional speech e.g., spontaneously produced utterances involving language processing in context. In comparison, automatic speech, as well as the singing familiar songs, are associated to activate the right hemisphere. Support for this hypothesis provides transcranial magnetic stimulation to the left inferior frontal cortex, where it causes speech arrest in right-handed individuals, while singing the same words, prevails intact (Epstein, Harris, Stanley, & Kanwisher, 1999; Stewart, Walsh, Frith, & Rothwell, 2001). Equivalently, patients with damages in the frontal right hemisphere have been reported to lose the ability in the musical domain (amusia) and singing, while speech remains unaffected (Peretz et al.1997).

Differences in neural activity for speech and singing have been reported by brain imaging studies as well. It has been shown that covertly singing a well-known non-lyrical tone involves the right sensorimotor cortex including the posterior inferior frontal gyrus (IFG), as well as the left cerebellum, when in comparison, covertly speaking a highly overlearned word string is activating the left sensorimotor cortex including posterior IFG and the right cerebellum (Ackermann & Riecker, 2004; Riecker, Ackermann, Wildgruber, & Dogil, 2000). An interesting finding is also the fact that while singing, brain regions that are involved with rewards, such as nucleus accumbens, posterior cingulate, orbital cortex, and parahippocampal gyrus, are activated (Callan et al. 2004). This speaks for a greater emotional component involved in the processing of singing compared to speaking.

6 Singing

„*Those who wish to sing, always find a song.*“

-Swedish Proverb

Singing is a universal human capacity that constitutes the most widespread mode of musical expression that can be performed at least to some extent by the majority of humans. The universality of this ability implies that singing could have arisen as an evolutionary adaption and that singing as well as other musical activities may have evolved as a mechanism of social bonding (Huron 2006). Support for this allegation can be drawn from the fact that the release of neuropeptides such as oxytocin and endorphin is known to be associated not only with social bonding behaviors such as in mother-infant bonds and romantic relationships, but they are also released during synchronic behaviors that involve some physical exertion as in the act of laughing or dancing (Tarr, Launay, & Dunbar, 2014). Additionally, the positive effects of singing and especially those of group singing have been investigated largely during the last decades. People describe singing to be a source of enjoyment, as well as a way to improve physical and psychological wellbeing. When compared to other activities that require long-lasting practice before the advantages can be seen, singing is argued to be rewarding to anyone who attempts it (Unwin, Kenny, & Davis, 2002). In other words, the satisfaction with singing does not appear to be dependent on the level of expertise, as many amateur choir members have limited music training backgrounds. Furthermore, various biological, social, and psychological benefits such as improved mood, quality of life, and well-being as well as reduced feelings of depression and isolation are associated with singing in groups (Abril, 2007).

Singing has been used as a part of the treatment of patients suffering from dementia. Administered as a music therapy intervention, many case studies confirm the improvements in the well-being and the quality of life of dementia patients, through the improved quality of breathing by encouraging deeper breaths and increasement in oxytocin uptake. In addition, participation in focused music therapy groups might help to reduce depressive symptoms in dementia patients (Ashida, 2000).

When examining the quality of life and the wellbeing of individuals and society, the first attributes to be connected to them are traditional factors such as health and socioeconomic position. However, factors such as several close friends, the amount of trust or distrust, and participation in associations also influence the well-being (Hyyppä, 2011). Choir singing, where the basis for participation is formed by making music together, has been mentioned as an example of such social activity. According to Hyyppä, singing together promotes happier and longer life with higher social capital.

As mentioned previously, it is commonly believed that musicality is an innate tendency. Several studies have shown that singing is a developing skill that can be learned and improved through practice. In the study of Numminen (2005), adults with poor pitch singing received sensing lessons for 1,5 years and by the end of the research, the students were able to sing a song in tune at least when accompanied by piano or teacher’s voice.

6.1 Singing and songs in foreign language learning

Singing and songs have been widely used in classroom situations as language teaching techniques to support encoding and retrieval of information for decades. The psychological effects of music solely have been studied at length. Music does not only increase relaxation, but also enhances energy levels and brings joy due to the increased amount of oxytocin produced in the brain (Grape, Sandgren, Hansson, Ericson, & Theorell, 2003; Huron, 2006). Listening to music generates neural activities which operate deeper in the areas of the auditory cortex, activating the frontal, temporal, parietal, and subcortical areas which are related among other things to attention (Schellenberg, Nakata, Hunter, & Tamoto, 2007), semantics, and syntactic processing, memory, and motor functions (Koelsch, et al., 2004), as well as the limbic and paralimbic systems which are associated with the processing of emotions and memory (Koelsch, Fritz, Cramon, Müller, & Friederici, 2006). One well-known method to utilize music in classroom situations is to use instrumental music without lyrics. In an educational context, instrumental music may provide a relaxed atmosphere and it helps the development of the creative process while isolating students from external noises (Fonseca-Mora 2002). Already during the 1990s and early 2000s, the "Mozart Effect" received lots of attention from disciplinary researchers. This phenomenon suggested that listening to Mozart's music might improve the performance of certain kinds of mental tasks, such as spatial reasoning (Ivanov & Geake, 2003; Hallam & Price, 1998). Later research has nonetheless shown that the same effect occurs when listening to other composers and therefore the improvement of spatial-temporal reasoning can be explained through the repetitive musical patterns of classical music that cause positive brain reactions.

Singing itself, an activity that mixes linguistic and musical information is produced through the same organs as natural speech, namely the vocal motor system. As an educational activity, singing can be simultaneously motivating, attention focusing, and simply enjoyable for learners of all ages (Good, Russo, & Sullivan, 2015). Moreover, singing increases the sense of togetherness and the use of songs in the classroom may result in positive effects on group dynamics. (Spychiger, Patry, Lauper, Zimmerman , & Weber, 1995; Wiltermuth & Heath, 2009) Songs emphasize stress and duration of phonetic elements and therefore singing applies as a good training opportunity for pronunciation, speech patterns, grammar, and vocabulary skills (Morley, 1991; Failoni, 1993).

An increasing number of studies have found a link not only between recognition of a word's phonological structure, articulatory rehearsal, and storage in short-term memory as well as consolidation in long-term memory, but also between songs, especially when implemented through singing, and memory (Rukholm, Helms-Park, Odgaard, & Smyth, 2018). Murphey (1990) introduced the "song-stuck-in-my-head" phenomenon as a melodic din that has an effect of involuntary musical and verbal rehearsal. A growing body of researchers has investigated the differences in learning passages through either spoken or sung conditions. It has been shown, that saying words out loud leads to faster learning and better retention (Ellis & Beaton, 1993; Papagno, Valentine, & Baddeley, 1991) but if the words are embedded in a song, there is even greater retention. In their study, Schön, et al., (2008) carried out an experiment where two different condition groups, speech condition group and sung condition group, were compared in a task where a text passage was learned. The outcome of the study was that consistent mapping of musical and linguistic information outperformed the condition group without musical information and the combination of musical and linguistic information enhanced the learning process, especially word segmentation. Ginsborg and Sloboda (2007) for their part showed that music and lyrics are recalled in an associative manner, meaning that the recall of music facilitates the recall of lyrics and vice versa. Rukholm et al. (2018) explored the relative influence of sung versus spoken lyrics on the acquisition and retention of vocabulary by adult learners and determined that songs and high elaborations (classroom vocabulary practice) are effective in facilitating the acquisition and retention.

In the psychological literature, the famous Working Memory model from Baddeley and Hitch (1974) presented a model of short-term memory, where a phonological loop stores verbally coded information and requires that information to undergo a process of articulatory rehearsal, either in vocal or in subvocal form. The rehearsed information has been proven to benefit the short- and long-term storage of information. This evidence is relevant when linked to the findings of songs promoting rehearsal of language and therefore aiding retention (Rukholm et al. 2018).

Music and songs in the classroom context are generally used to increase motivation and to develop different kinds of learning strategies. For instance, in English-speaking cultures, the learning of alphabets is often carried out through inserting the letters to the melody of the famous children‘s song “Twinkle, Twinkle, Little Star” to facilitate memorization. Accordingly, the advantage of songs to facilitate the recall of information has been studied widely both in the native language and in foreign language recall (Wallace, 1994; Rainey & Larsen, 2002; McElhinney & Annett, 1996). The experiments conducted by Wallace (1994) provided evidence for the argument that text memory can be supported by songs. In one of the experiments, participants were exposed to a text passage that was performed either in spoken or sung conditions. The results showed that participants in the sung condition group recalled a greater percentage of the words than participants in the spoken condition group. Moreover, the participants of the sung condition group used structural characteristics related to rhythm in order to support the recall.

Rhythm appears to be one of the strongest cues to promote memory support, as the study of Wallace and Rubin (1991) points out. In their experiment, the support of rhythm remained even though the melody was absent in the learning passages. Participants of the study profited from structural characteristics such as syllabic stress, line breaks, and phrasing of a spoken text when these characteristics were emphasized. From these outcomes could promptly be drawn perfunctory assumptions that the mnemonic support of song for the information recall is only caused because of rhythm and therefore the rhythmic speaking should hold the same mnemonic value as a song. Nevertheless, the structural characteristics of a song are not limited to the rhythm alone, as they also consist of melodic and tonal structure, which can also be factors that support memorization of the text. To provide evidence for this argument, the subsequent experiment of Wallace (1994) was carried out where the learning of text passage presented either as a song or as a spoken text with rhythmic intonation was compared. The results showed superior recall of the group that learned the text passage presented as a song.

Another factor providing evidence for the song advantage on text recall is the rate of presentation. When compared to speech, the text presented in songs occurs at a slower rate, which indicates more time is available for the encoding and rehearsal of the words (Good et al., 2015). The slow presentation rate gives an advantage for the learners also regarding the acquisition of pronunciation. Pronunciation is often stated to be one of the most difficult parts of language acquisition, because the phonological elements of the source and target languages may vary greatly. To be able to get understood while communicating, one must be able to pronounce the words correctly and therefore also be able to hear them correctly. Teaching phonetic skills through songs and singing is therefore considered to be one of the fastest ways to develop phonetic abilities since it is easy to find songs that emphasize a specific phoneme. Songs also exaggerate the stress and duration of the phonetic elements, which aids the memory and acquisition of the foreign language (Palmer & Kelly, 1992). In addition, Palmer & Kelly found out that the structure of the majority of songs being divided into divisions of four beats coincides with the stressed and non-stressed syllables of utterances facilitating memorization.

Further support for the singing benefit comes from the study of Ludke, Ferreira, & Overy (2014). They compared three different learning groups with either a singing condition, a speaking or rhythmic speaking condition. The participants were required to recall and reproduce short paired-associate foreign language phrases after a 15-min learning phase. The results showed that the singing condition group outscored the participants in the speaking and rhythmic speaking conditions. Additionally, in their study also the rate of presentation of the stimuli and the overall durations were carefully controlled across all three learning conditions.

Baills, Zhang, Cheng, Bu, & Prieto (2021) conducted a complementary study to explore, whether listening to songs and singing can improve second language pronunciation and vocabulary learning at beginning states of language acquisition. In the study, the participants received a 4-min training session to learn 14 words either in singing or listening condition. The results provided support for the song listening group outperforming the listening condition group in accentedness tasks, but in respect to the word recall, no advantage was found between the condition groups.

These outcomes addressed above correspond with the Involvement Load Hypothesis initiated by Laufer & Hulstijn (2001), which suggests the retention of unfamiliar words being dependent on the involvement load of a task, in other words, the extent of the need for search and evaluation. Tasks which induce a higher involvement load are more effective than those with lower involvement. This theory combines both the cognitive and affective components of the language acquisition process, which in the light of a substantial body of research supports the evidence that songs may facilitate elaborative processing and therefore promote better retention of words.

The introduced literature and research evidence suggests singing to be strong support for encoding and retrieval of information. Although a growing number of studies have been conducted to analyze the effect of music and singing on the various subdomains of foreign language learning, only very little attention has been focused, or the results have been contradictory on the effects of singing on vocabulary acquisition in foreign language learning.

6.2 Aim of the study

The conducted study of this paper aims to enrich the existing research body in the field of individual differences in language learning, as well as to expand the knowledge about the relationship between singing ability, musical aptitude, and vocabulary acquisition. Based on previous research findings, the effect of music on language learning has been commonly proved. Especially singers benefit from their musical experience in the domain of foreign speech imitation aptitude and pronunciation, as the results of Christiner and Reiterer (2015) report. The use of melody and songs as a memory aid advocates the advantage of recalling words when learning foreign languages, as the results of Ludke et al. 2014 verify.

In line with these research findings the aim of the conducted study is summarized below:

1) the study aims to investigate, whether singing a song facilitates vocabulary acquisition more than merely listening to the song or listening to the spoken stimuli of the song lyrics.

2) The study seeks to find out, whether individuals with high musical aptitude benefit from their musical experience in comparison to the individuals with lower musical aptitude.

Research questions and hypotheses

1. Does singing improve vocabulary acquisition more than merely listening to songs?

Hypothesis: It is expected that singing facilitates vocabulary acquisition more than listening to a song, thus that both singing and listening conditions overplay the spoken, music-free stimuli.

2. Does music aptitude correlate with vocabulary acquisition?

Hypothesis: It is expected that high music aptitude correlates with vocabulary acquisition.

The following empirical part will present the executed study which aims to enlighten the differences between listening to a song and singing a song and how these affect vocabulary retrieval.

EMPIRICAL PART

7 Research

As outlined in the theoretical part, pieces of evidence prove the connection of musical aptitude influencing foreign language learning. The earlier chapters sought to shed light on the background literature and current research findings that have been made so far. The empirical part presents the cross-sectional study that was performed to find out more information about the effect of singing performance on vocabulary acquisition in foreign language learning.

Before the study design and results are presented, the preceding pilot study is briefly demonstrated in the next chapters.

7.1 The Pilot Phase

Prior to the full-scale study, a pilot study was conducted to try out the study conditions and the experiment design. The pilot phase took over a period of six weeks in the summer of 2019. It was conducted in Vienna and a total of 12 young adults (7 males, 5 females) ages 22 to 32 (M=25,5) participated in the study. The preconditions for being included in the study were participants being native speakers of German plus no previous knowledge of Finnish.

In the short pilot study, the experiment song was based on the Finnish version of the internationally famous children’s song “Head, Shoulders, Knees, and Toes”, however, the melody was slightly altered to rule out the possibility that the song could be already familiar for the participants and the meaning of the words could be deducted from the previous knowledge of the song text.

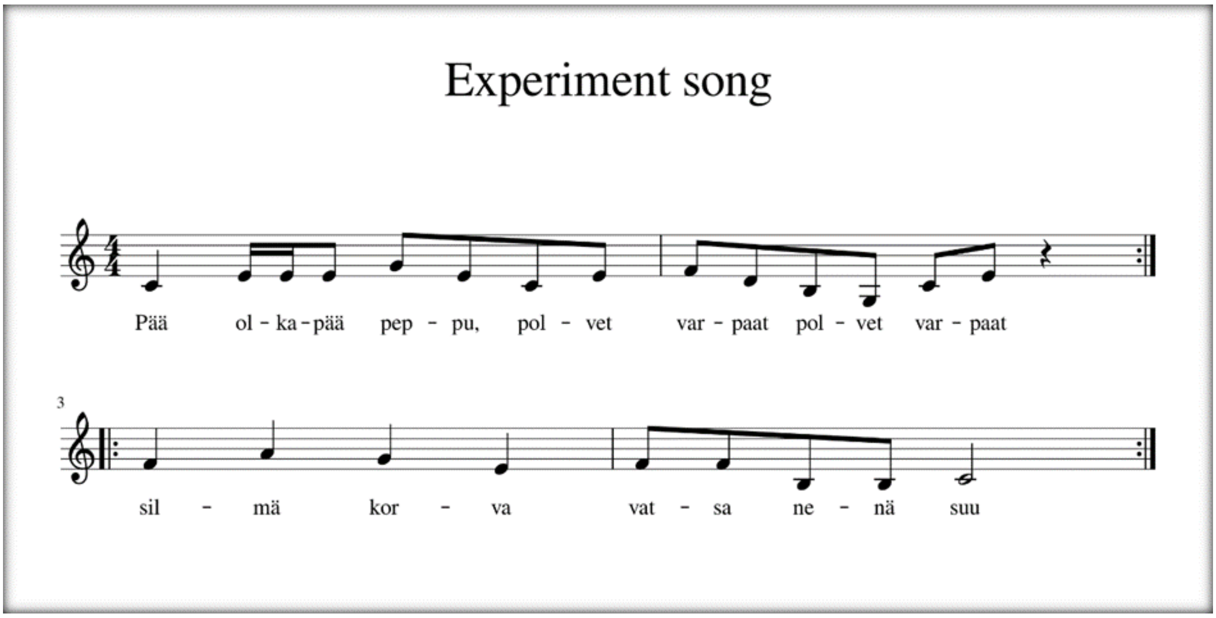


Figure 1 Experiment song Notation, Pilot Phase

In the pilot experiment, the participants were divided into two different condition groups and one control group, where the condition groups carried out language learning tasks corresponding to the condition they were separated into, either singing along or listening to the experiment song. The control group did not receive any musical input. After the learning phase, word recall tests were performed to question participants orally in two rounds, first through free recall recognition test and finally through cued recall recognition test.

The pilot phase aimed to find out the feasibility of the study design for the subsequent full-scale study as well as collect feedback from the participants of the pilot study. The pilot test revealed that both the experimental procedure and the estimated time frame for the experiment indeed turned out to be realistic and executable to be applied in the main study, however, some minor limitations came forth, and therefore some parts of the study design specified below were marginally altered to serve the study purposes.

First, in the pilot study, the singing scores were collected in person, which required participants to be willing to sing spontaneously in front of the interviewer, which caused the recording of the singing score to be more of an audition kind a situation. Even though none of the participants hesitated to accomplish the singing task, in the feedback interviews some of the participants reported that the task would have been easier, if they would have had the choice to send an audio recording of their singing in advance, believing that their singing performance suffered from the spontaneous situation causing them to be nervous and thus singing incorrectly.

Second, all the testing in the pilot phase was carried out in person, and the learning phase of the target words was implemented through paper cards, where each card presented one word with the corresponding picture. The cards were given to each participant at the beginning of the task where they had two minutes to learn the words. The use of concrete physical cards appeared unpractical and disorganized, which explains the demand to find a more functional and consistent alternatives for the execution of the learning phase. Later, the situation caused by the COVID-19 pandemic supported the experiment to be completely transformed into an online format.

Additionally, in the first learning phase, the time limit to learn the target words were altogether 2 minutes, which turned out to be too long a timespan for 10 words. Quite a remarkable amount of the pilot study participants managed easily to learn all the 10 words in two minutes which made the evaluation of the impact of different conditions difficult. Also, in word experiments, such as in the LLAMA B vocabulary acquisition test the time limit for word learning tasks was limited to 10 seconds per word (20 words in 2 minutes). Following the findings of the pilot study and other research papers, the time span of the first learning phase was reduced to 1 minute.

Lastly, the most uncertainty surfacing from the pilot phase was the type and realization of the singing condition group, where the assignment of the participants was to sing along with the experiment song, which was an altered version of semi-famous children's song. The variation of the melody and the order of the words in the song was performed to create a learning situation where the experiment song is unfamiliar to the participants both in melody and lyrics. Nevertheless, research shows that the benefits of the song are modulated by a song's predictability (Calvert & Tart, 1993; Kilgour, Jakobson, & Cuddy, 2000; Purnell-Webb & Speelman, 2008; Rainey & Larsen, 2002). That means that when the listeners are already familiar with the melody the song advantage becomes stronger. When the song is novel, differences between the condition groups i.e., sung vs spoken are inconsequential. It has been suggested that this occurs due to the dual-task of learning simultaneously both a novel melody and a novel text. In this case, the melody might operate more as a distraction than a stimulant for the learning process (Good et al.2015). Additionally, according to Calvert and Tart (1993), first, when the melody and rhythm of a song are familiar, the structural information including syllabic stress, line breaks, and phrasing, becomes more available and predictable to provide more recall cues for the passage.

These findings indicate that in this case, the novel melody might act even as a distractor rather than a support for the recall. In addition to that, feedback from the pilot phase participants support this argument, because some of the participants reported being overwhelmed when they were asked to sing along only after two rounds of listening to the experiment song. Participants noted also that they forgot to focus on the learning of the target words of the experiment since they had to simultaneously focus on the novel melody. To avoid this limitation in the full-scale study, the target words of the experiment were embedded into a familiar melody.

Yet another consideration regarding the singing condition group was that the test situation could have been more “natural”. According to the feedback received from the pilot phase subjects, many reported finding the test situation distressing because the singing task took place in front of the experimenter. It resembled for the subjects a situation similar to audition and the shared opinion was that it would have been more comfortable to sing for example in a situation where the song is taught by the experimenter or that the participants would learn the song in a group. Following these ideas, the original plan for the full-scale study was to carry out the experiment in a classroom kind of situation, ideally making the setting resemble the way choirs learn and practice a new song, where the focus is on the learning process and singing itself.

7.3 Full-scale study

Following the findings from the pilot phase, the full-scale study was conducted in the time range of eight weeks between December 2020 and February 2021. The participants were recruited partly through advertising in hobby choirs as well as in different groups of the University of Vienna. Additionally, participants were recruited through an advertisement post in social media canals explaining the topic and purpose of the study as well as the participant prerequisites. In total, 58 individuals volunteered to participate in the study. However, only 45 participants completed the whole study. Some of the dropouts only concluded parts of the study such as the questionnaire or part of the pre-tests, but never responded to further inquiries about completing the experiment. One of the dropouts refused to accomplish the singing score and was therefore excluded from the study.

7.4 Sample

The recruitment of subjects was controlled to keep the sample as homogenous as possible. Subjects were controlled for

1) age and mother tongue (German)
2) general education
3) linguistic background
4) musical background

5) singing background

Demographic background

In total, 45 native speakers of German (16 male, 29 female) were selected to participate in the study. At the time of data collection, the age of participants ranged from 19 to 39 years (M=26,09).

The candidates met the criteria to be selected in the study for being native German speakers, counting in individuals with the bilingual or multilingual backgrounds. Participants with knowledge of Finnish were excluded. Of all the 45 candidates, 38 reported German as their first language. 7 candidates reported growing bilingual, where the other first languages were Polish (3 candidates), Hungarian (1), Burgenland-Croatian (1), English (1), and Serbian (1). 30 out of 45 participants were citizens of Austria, 6 candidates were German citizens, 2 both German and Austrian, and 2 both Polish and German citizenship. The rest of the nationalities were Poland (1), Austria/United States (1), Austria/Hungary (1), Italy (1), and Serbia (1).

For a better understanding of the sample, figures 2 to 7 provide some illustrations.

Statistics

Age

N						
Valid	Missing	Mean	Median	Std. Deviation	Minimum	Maximum
45	0	26.09	25.00	4.315	19	39

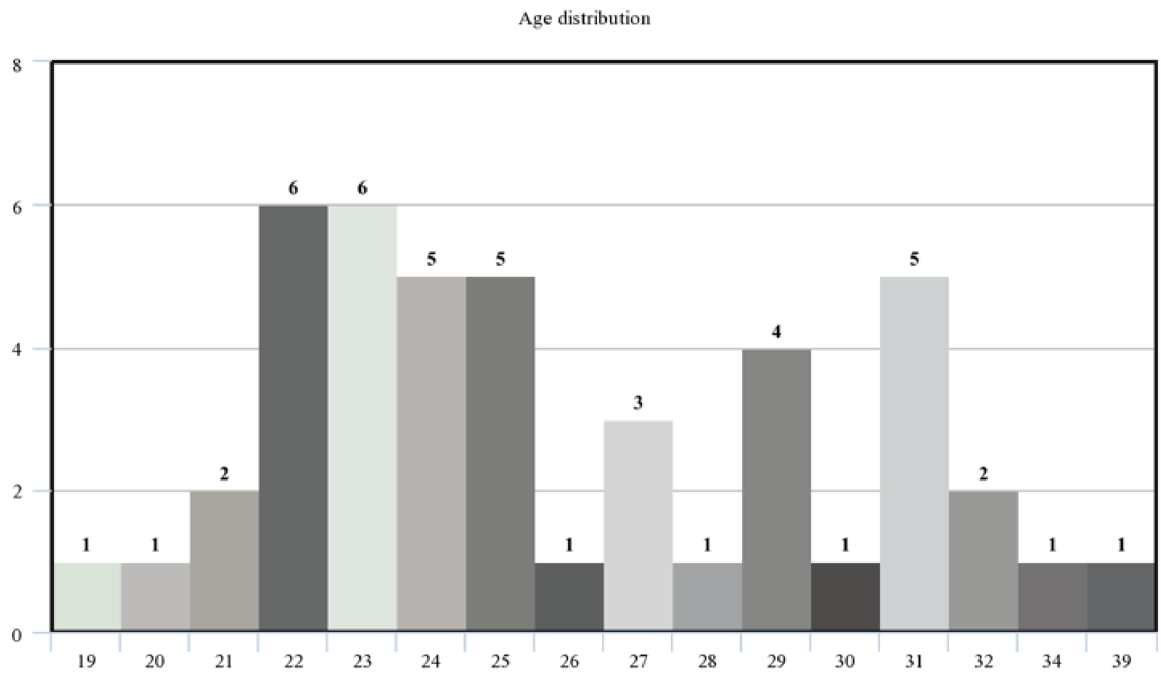


Figure 2. Age distribution of the sample

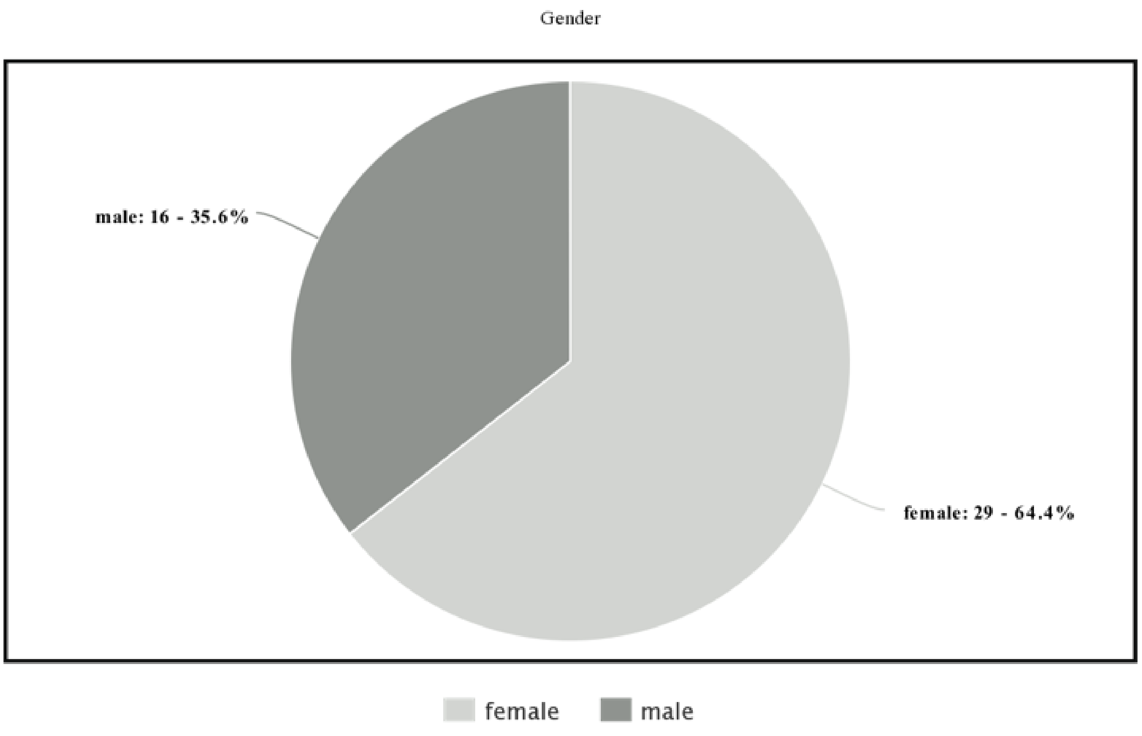


Figure 3: Gender. 16 males and 29 females participated in the study

Educational background

The most common educational background among the participants was a bachelor degree (40%), followed by A-levels (33,33%). 20 percent of the participants had absolved master‘s degrees and 4,44 percent had a doctorate. Only one participant (2,22%) had undergone vocational training. Participants were asked to fill in the field of their studies, which are summarized in Figure 4.

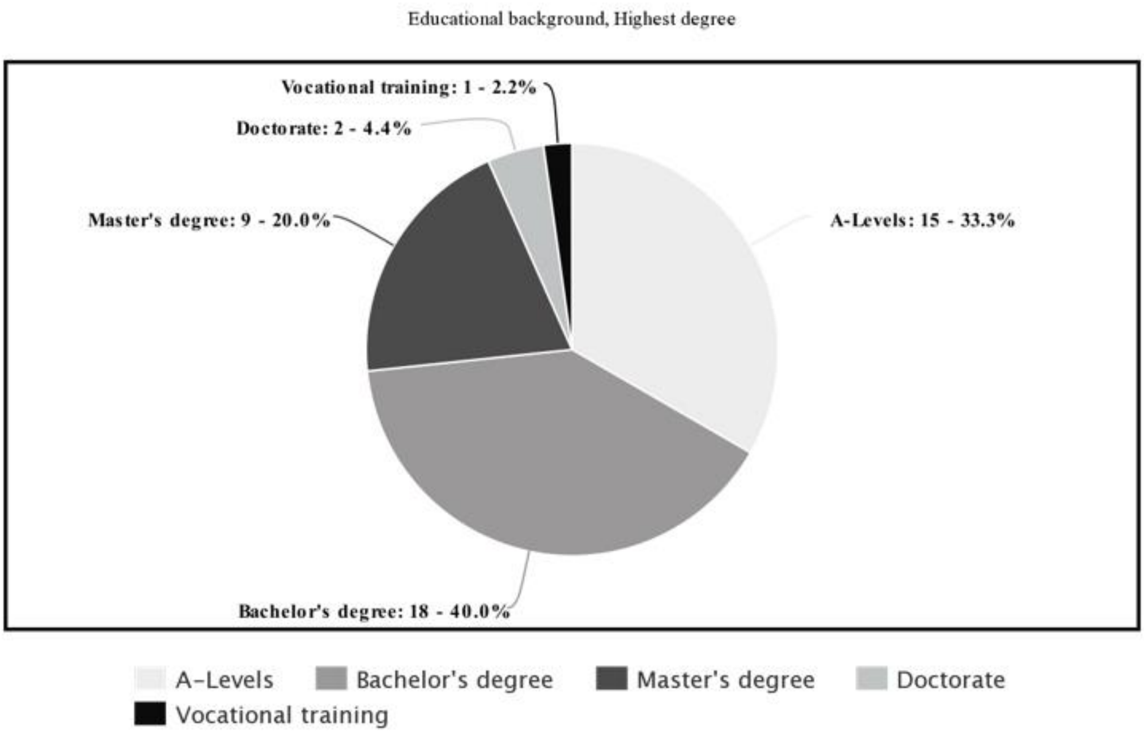


Figure 4 Educational background of participants: The highest degree

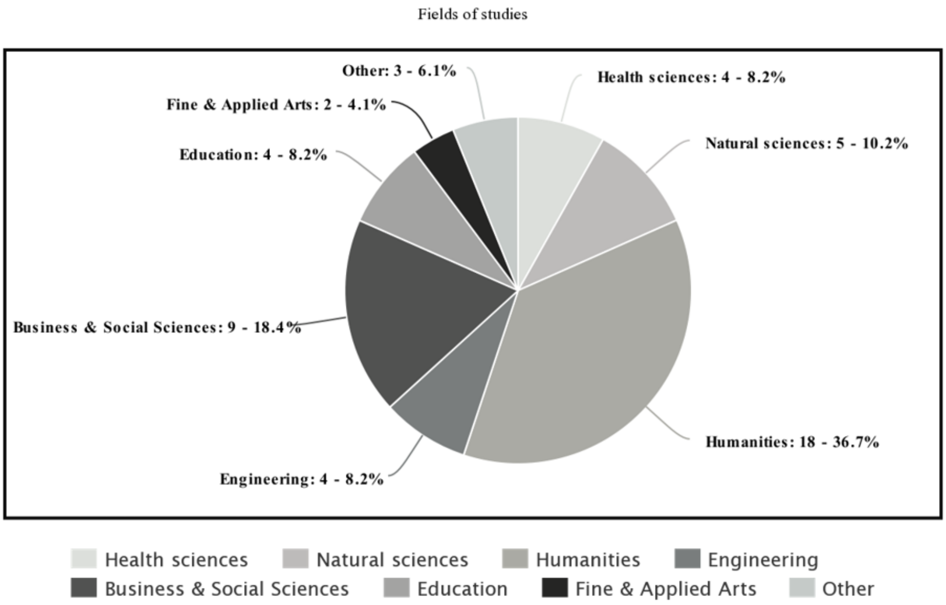


Figure 5 Fields of Studies

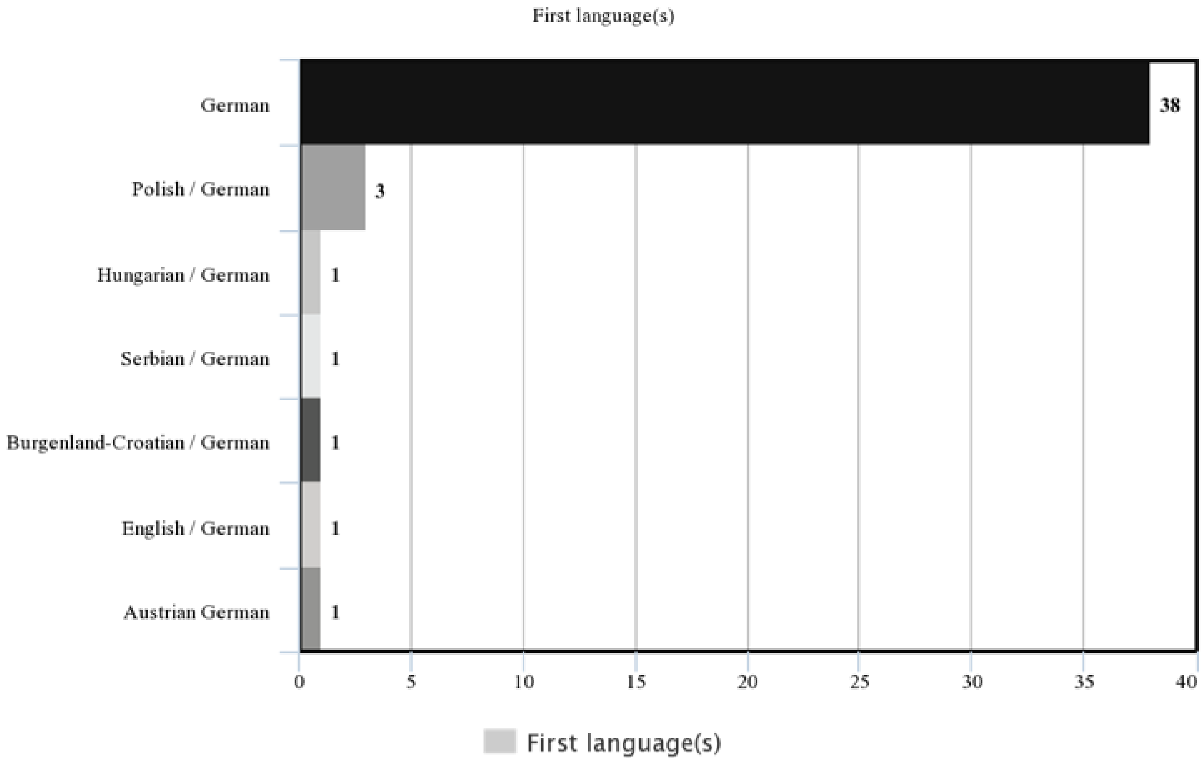


Figure 6 Linguistic Background: First Language(s)

Linguistic background

All 45 participants answered knowing at least one more language than their first language. 8 participants gave the information of knowing another additional language, whereas 15 participants answered knowing 2 additional languages. Similarly, 15 other participants agreed of knowing 3 additional languages. Knowing 5 foreign languages was the answer of 5 participants and 2 participants reported knowing 7 additional languages.

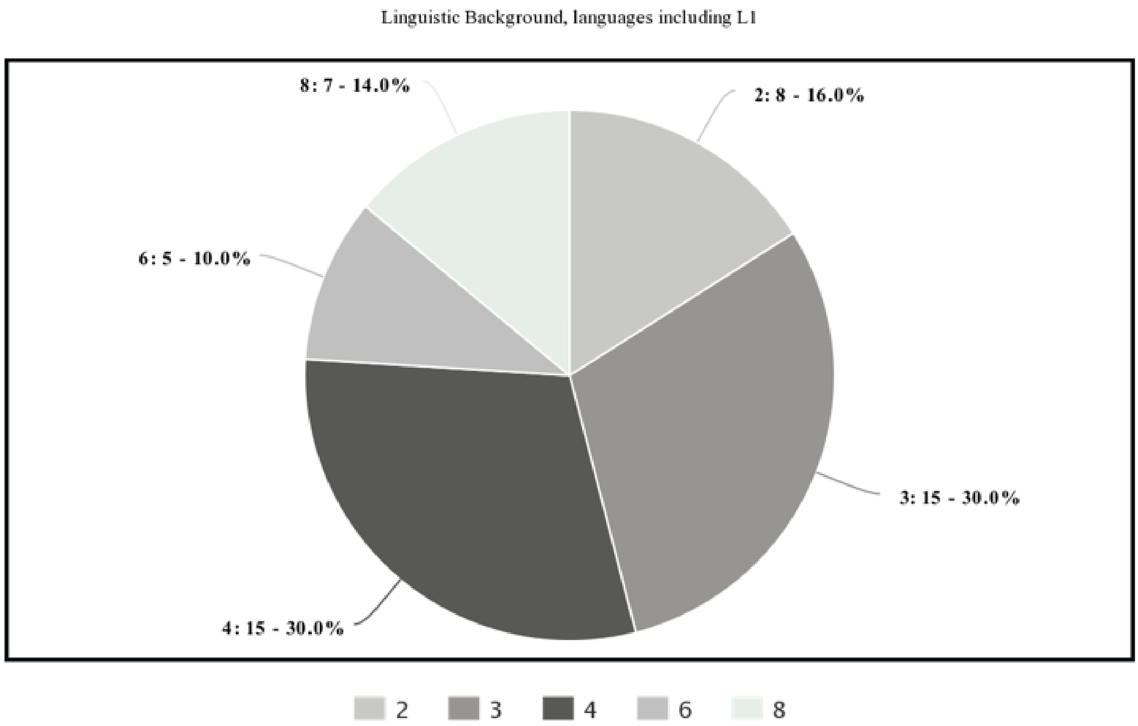


Figure 7 Languages including L1

When having a closer look at the participants‘ linguistic background, **English** (39 answers) was the most common second language, followed by **French** (19 answers), **Spanish** (15 answers), and **Italian** (7 answers). Other languages participants mentioned were **Hungarian, Russian, Swedish, Japanese, Irish, Hebrew, Kiswahili, Portuguese, Dutch,** and **Slovene.**

Musical background

One question in the basic questionnaire covered the incidence of *absolute pitch* among the participants. Absolute pitch refers to the ability to identify or categorize musical pitches without the aid of any external reference (Reymore & Hansen, 2020). This phenomenon that emphasizes auditory features such as discrimination and representation of pitch and timbre would presumably give a participant a possible asset in the learning process of new words based on the auditory nature of the task performed in the study. However, none of the participants announced possessing absolute pitch.

Based on the subjective answers of the participants in the questionnaire, 37 of 45 participants responded yes to the question of whether they play a musical instrument. The average age of onset of instrumental education between those participants was 8,8 and the most popular instruments among the participants were **piano** (21), **guitar** (15), and **recorder** (8). Although the majority (82,2%) of participants answer to playing an instrument, only 26 (57,7%) were still more or less actively practicing the played instrument.

Do you play a musical instrument?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	8	17.8	17.8	17.8
	Yes	37	82.2	82.2	100.0
	Total	45	100.0	100.0	

Another question performed in the questionnaire was "*Do you listen to music?"*. Every one of the participants gave yes as an answer to this question. Although this question was proposed more or less to warm up the subject for following sub-questions, the uniform answer is an explicit reference to the discussion of the universality of music – music in different forms is present in every culture and everyone enjoys music to some extent (Liu, Hu, & Schedl, 2019).

In the questionnaire, subjects were asked an open question to explain their music education. The majority of participants answered their music education based on different providers of music schools, mostly in the format of instrumental lessons. 12 participants had acquired to play instruments under the supervision of a private teacher. Altogether 15 participants mentioned that they either never received any music education or the education they absolved is limited to the music classes in the school context. Of the participants, 6 had absolved music education in choirs such as different children's choirs like *The Vienna Boys' Choir* (Wiener Sängerknaben) or *Vienna's Children's Choir* (Der Kinderchor der mdw ((Universität für Musik und darstellende Kunst Wien)) and 3 participants had acquired to play an instrument by themselves either through self-taught guides or YouTube-videos.

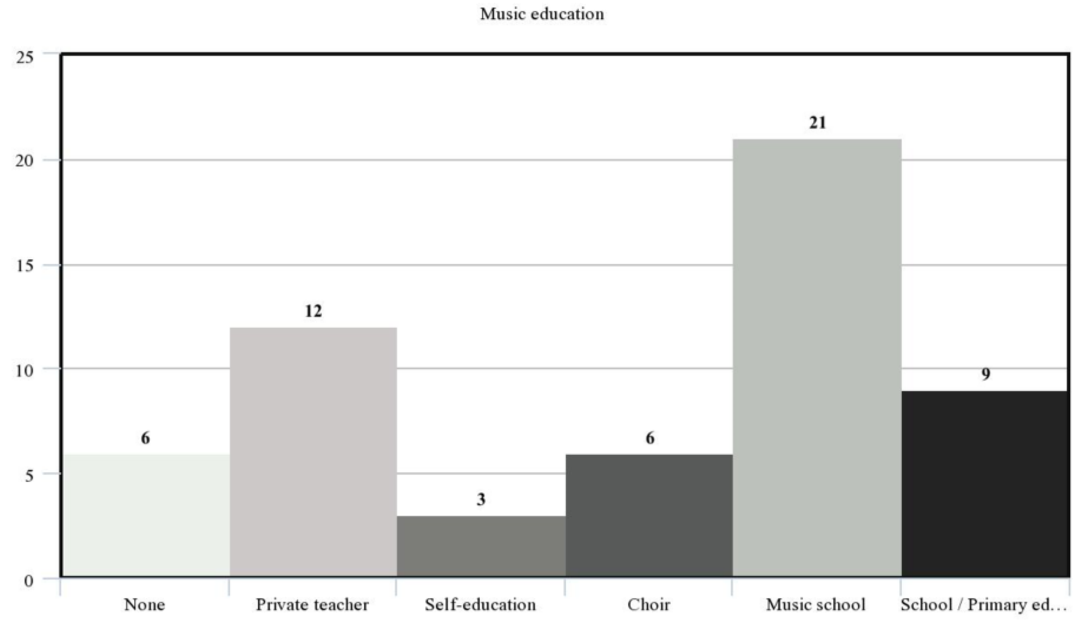


Figure 8 Music education of participants

Singing background of the participants

Of the 45 subjects who completed the questionnaire, 31 respondents (68,89%) replied yes to the question "*Do you sing?"*, whereas the percentage of non-singing subjects covered 31,1%.

Do you sing?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	14	31.1	31.1	31.1
	Yes	31	68.9	68.9	100.0
	Total	45	100.0	100.0	

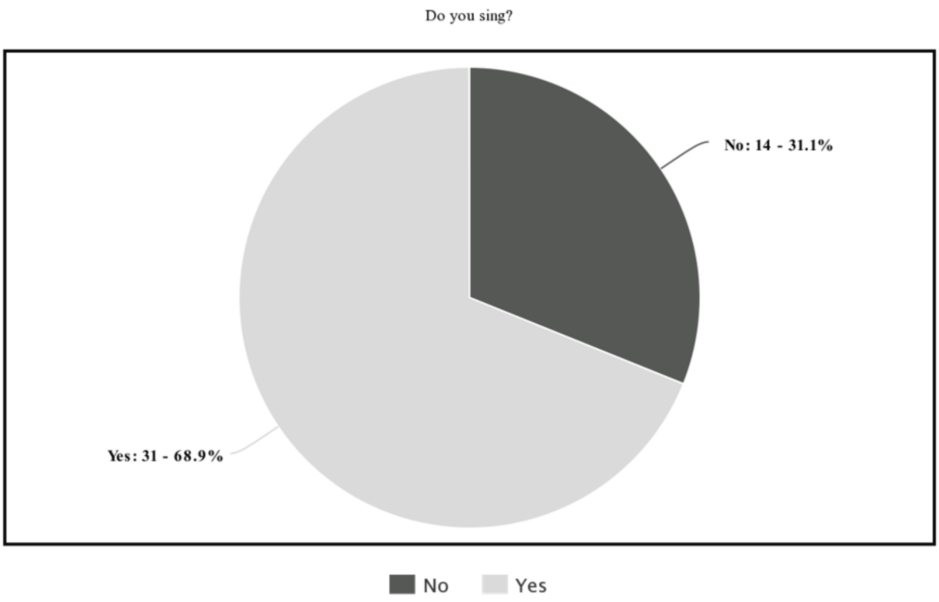


Figure 9 Singing background of participants; "Do you sing?"

The last question of the questionnaire covered the singing background of the participants. The participants were asked to explain in their own words the singing background. The singing experience of participants who responded yes to whether they practice singing can be divided into institutional contexts such as school, church, or university, to leisure activities like choirs, bands, and ensembles, singing lessons both in amateur and professional contexts and to the various settings without a specific context such as singing along to a song or radio.

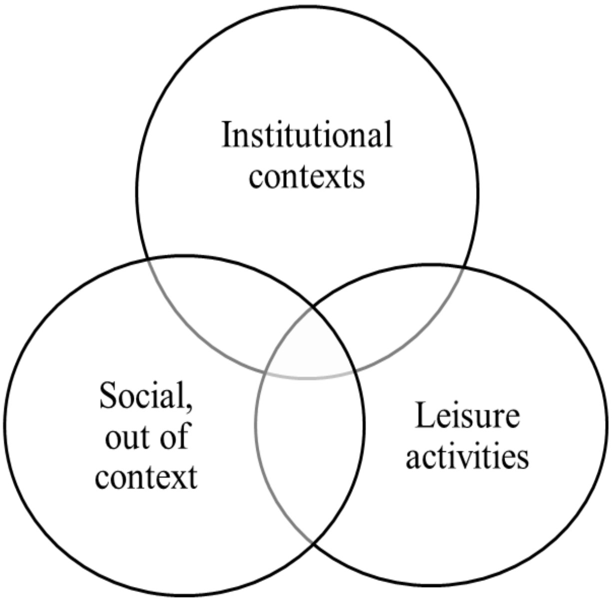


Figure 10 Singing background of participants

The open question about the singing background of the participants provided the most detailed answers. Below the answers are listed in a figure for a better understanding of the singing background of the respondents.

„I quite like to sing from time to time as a hobby“	„I like to „sing-around“ during my day or with friends“	„somebody always sang in our family (not professionally). It just always was present and singing together is lots of fun“
„First I started to sing to impress women, afterward I started to enjoy it by myself and got better with it over time“	„Singing is just a normal thing to me, something I do all day long. I sing when I'm playing the guitar, I sing along to songs I know, or when a melody comes to my mind. I have never had singing lessons or something, but I got precious input from people who know more about singing, like in the choir or church“	„just started for fun approximately 1 year ago, now I really like it“
„I have always enjoyed singing“	„My mother sang every day with us when we were little, many family members are in the choir“	„I have always loved singing and coming from a family what loves singing and making music themselves, I have been surrounded by music and singing since a very young age a lot. I have never sung professionally, but I enjoy singing to myself or I groups (of friends or family) very much“.
„It's the thing I love most and always was my main musical focus“	„My mom was a part of the semi-professional choir in Vienna for a long time and from her, I learned how to practice music“	„For fun and joy“
„Just for fun“	„At the moment I am singing what I want when I want, the way I want until Corona is finally over“	„In elementary school, I attended a singing class in my free time, but my singing does not sound good“

„Most of the time I sing for my children or to songs that I listen to on the radio“	„In my family, we often sing together on occasions like weddings, baptisms and so on“	„Just like singing at random moments, shower, etc“
„sang with my dad, who played guitar to it, sometimes we sang at home with him; otherwise Sunday Service and youth camps“	„4 years of singing education with a focus on classical singing in the conservatorium“	„no professional singing, only in the car every now and then“
„alone in the car“	„my first singing experiences were as a child in a school and with my family, later in music school as part of a band“	„I started singing lessons in high school together with choir practice, then I graduated in singing and after school, I joined a choir, where I stayed for 4 years“

7.5 Methods

The whole study including the pre-tests endured approximately one hour, the experiment part taking roughly 15 to 20 minutes per participant. Figure 11 provides an overview and the chronological order of the tests and questionnaires administered.

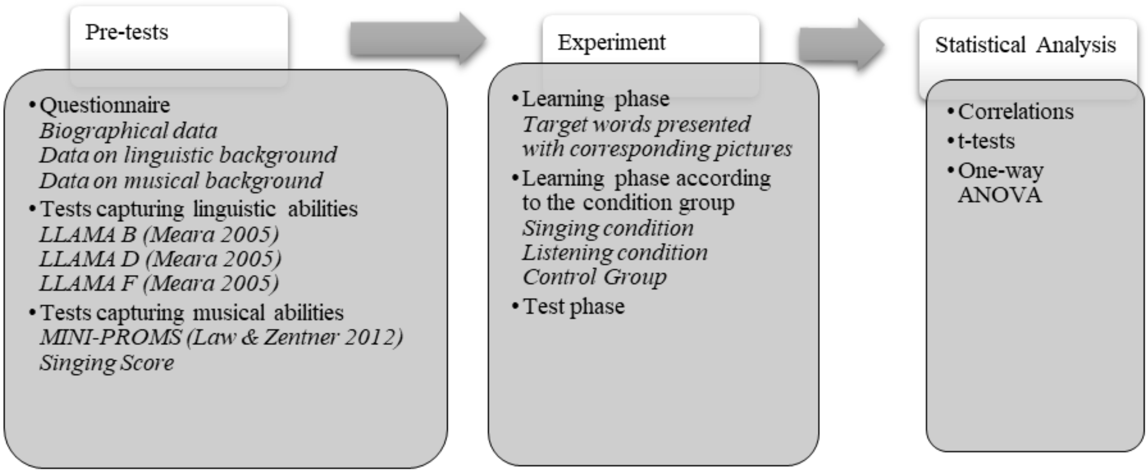


Figure 11 Experiment Procedure

Questionnaire

The participants were asked to fill out a basic questionnaire carried out through Google Forms to collect background information from the respondents. The survey consisted of demographic questions including age, gender, nationality, education, as well as the number and proficiency of languages respondents can speak. The second part of the survey covered information about participants' musical background, including questions about musical education, the possible experience of playing instruments, and singing background.

Singing score

To evaluate the singing performance of each participant, the participants were asked to record a short singing score, where the task was to sing "Happy Birthday" in freestyle the way the participants liked best, without any instrumental accompaniment. The worldwide known song "Happy Birthday" has been translated into innumerable languages and is performed on the order of a million times a day (Masataka, 2009). This piece which has been described as the most successful musical work due to the criteria of the most performed and most heard song, was composed by Mildred and Patti Hill in 1893 and revised in the 1930s. The song remained under copyright protection until the middle of the twentieth century.

The reasons for the choice of "Happy Birthday" as the object of the singing score are on the one hand due to its popularity; it is so well-known all over the world, that it is difficult to find an individual who would not have heard the melody at least once in life. Additionally, even if an individual would not possess a vast singing background, for many people, the singing of "Happy Birthday" is the only time they sing in public and at the same time, for some people, singing that piece is the only time they sing.

Despite "Happy Birthday" being familiar, it cannot be considered as a very easy song for a non-musician, since it consists of 25 notes that form a variety of different intervals, some of them being large jumps, where the biggest jump is exactly one octave. Considering the familiarity plus the intermediate level of the musical characteristics of the song, "Happy Birthday" suits excellently to be used in the evaluation of the participants' singing abilities.

Each singing score was evaluated anonymously by two experienced choir singers through the following criteria: intonation accuracy, sense of rhythm, and voice quality. With the concept of intonation, accuracy is meant the accuracy of pitch in playing or singing, as well as the act of singing or playing in tune (Kennedy, 1980, Swannell, 1992). Sense of rhythm refers to the ability to perceive and measure time and the category of voice quality incorporates timbre and warmth of the voice as well as of the use of the suitable breathing techniques. Each category was scored on the scale of 0-10, the maximal score, therefore, being 60.

Language aptitude tests

The LLAMA test battery merges the traditional and contemporary approaches used in the field of aptitude testing. The more traditional components i.e., tasks of associative word memory and grammatical inferencing are included in the test battery since LLAMA was originally loosely based on the MLAT (Bokander & Bylund, 2019). The newer component, the sequence recognition test has been assumed to assess implicit learning abilities and therefore included in the test battery. The linguistic stimuli utilized in the LLAMA tests are based on a northern Canadian indigenous language and accordingly allows the test battery to be language-neutral since it assumed that the test-taker has no extensive knowledge of the northwest British Columbian languages of Canada. Even though LLAMA tests are not validated, the test battery including its pilot project LAT (Meara, Milton, & Lorenzo-Dus, 2002) has been used widely in more than 40 studies over the last decade in many different topics.

In the current study of this paper, out of the four sub-tests three were administered to test out the vocabulary learning, sequence recognition, and grammatical inferencing skills of the participants.

LLAMA_B

LLAMA_B operates as a vocabulary learning module that tests the participants' ability to attach unfamiliar names to unfamiliar objects in a relatively short space of time. Learning vocabulary is one of the most important skills that affect how quickly and effectively an individual can learn a language. The words used in LLAMA_B are real words taken from a Central American language.

In the learning phase, the participant is presented with a set of 20 unfamiliar figures and has two minutes to learn the names. The program doesn't place any constraints on how the participant should learn the names. After the learning phase, the test continues to the testing phase, where the names of the figures are displayed one after another on the screen and the participant must identify the corresponding figure to the name. Each right answer is scored with 1 point the maximum score, therefore, being 20 (Rogers, Meara, Barnett-Legh, & Curry, 2017).

Figure 12 LLAMA_B, Vocabulary learning task

LLAMA_D

LLAMA_D is the only test that does not appear in MLAT, and it examines the skill to recognize repeated sounds of spoken language. The suggestion behind this test is according to Service & Kohonen (1995), that a learner who can recognize repeated stretches of sound is more likely to notice small variations in speech, which makes the isolation of individual words and variants of the words signaling morphology easier. Therefore, recognition of sounds in spoken language is considered as a measure of implicit learning (Speciale, Ellis, & Bywater, 2004). Moreover, according to Meara, good language learners seem to be better at recognizing words they have already heard. That leads to better word recall and facilitates the attachment of a meaning to the recalled word. In comparison to that, individuals unable to recognize already heard words will treat every heard word as a new word and this will slow the learning progress.

The sound sequences used in LLAMA_D are computer generated and the words occurring in the sequences are based on the names of flowers and natural objects in a British Columbian Indian language. In LLAMA_D the program will play a word for the participant to listen to and the participant's task is to decide if the played word is a new word or one that has been played before. Altogether there are fifty test items in all, and the progress of the test is shown by the bars underneath the response buttons. The maximum score in LLAMA_D is 20.

LLAMA_F

The LLAMA_F test analyzes the participant's grammar inferencing skills. According to Meara, many learners find it difficult to realize that languages use words in ways that are unusual in their first language. For learning a foreign language, it is important to notice varying patterns and be able to use them effectively.

This test begins with 5 minutes long learning phase, where the participant is supposed to study a data set containing a series of pictures illustrating different objects and shapes combined with a short sentence in an artificial language that describes each picture. After the learning phase, the program displays sixteen words on the blank buttons as well as new pictures, and the participant has to form sentences to describe each presented picture. Altogether, there are 10 test questions and the maximum score on the test is 20 points.

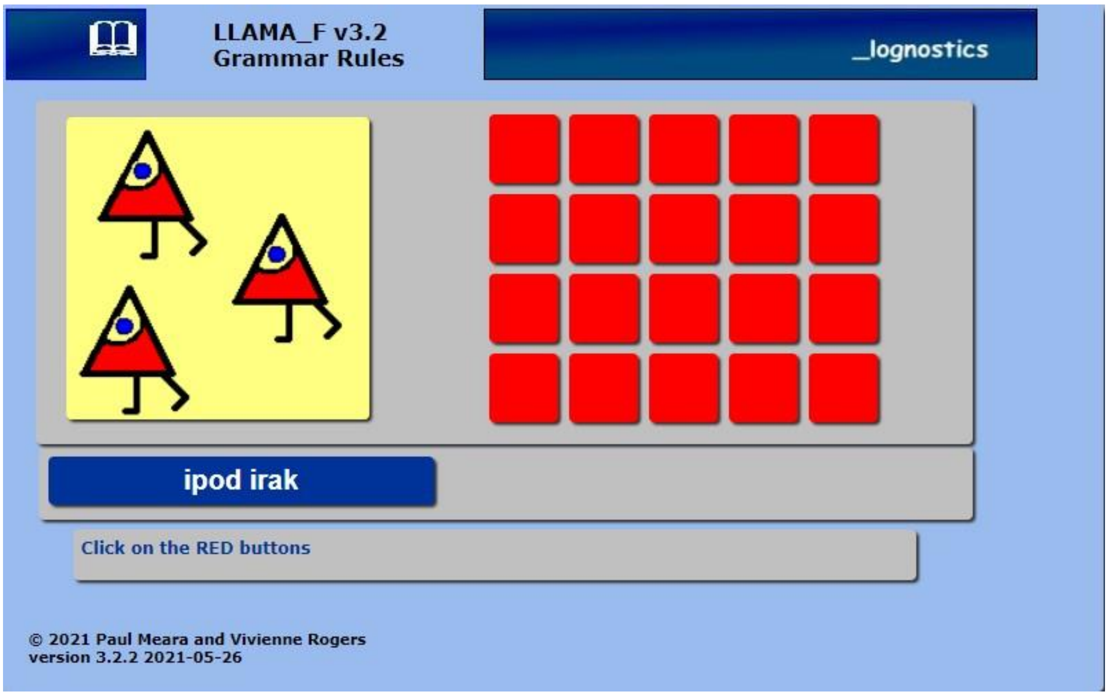


Figure 13 LLAMA_F, Grammatical Inference Task

Musical aptitude test

The musical aptitude of each participant was tested with the Mini-PROMS musicality test, which is a free online test based on the Brief-PROMS test (Profile of Music Perception Skills). Mini-PROMS is a shorter version of Brief-PROMS including a test battery that comprises the subtests of melody, tuning, tempo, and accent. Altogether it takes approximately 15 minutes to complete the Mini-PROMS.

Scoring of Mini-PROMS is based on the answer options "*definitely same"*, "*probably same", "I don't know", "probably different",* and *"definitely different*". Correct answers are rewarded with 2 points, correct but uncertain answers with 1 point. The maximum score in Mini-PROMS is 36.

In Mini-PROMS, the melody section assesses the test-takers melody perception skills. The section contains 10 items, where the test-taker hears first twice the same melody clip, which is followed by another comparison melody clip after a short interval. Test-taker must indicate whether the comparison melody differs from the initial melody or not. The maximal score in the melody section is 10 points.

The tuning section focuses to find out, how good the test-taker is detecting sounds that are out of tune. The participant is confronted with 8 test items produced by piano and as in the melody section, the procedure is similar: two sound clusters played one after another, followed by a comparison cluster sound where the task is to decide if the comparison cluster is the same as the initial one or not. The best score in the tuning section contains 8 points.

The accent section is constructed in the same way as the previous categories, investigating the perception of accent patterns in a rhythmic sequence. The accent patterns refer to the strong and weak elements in a musical rhythm. From the accent section, the test-taker can gain at best 10 points. The last category in Mini-PROMS covers tempo, where the test-takers skills to perceive differences in musical speed are examined. The maximum score in this section is 8 points.

At the end of the test, the program provides test-taker personalized performance feedback, where the results of each sub-tests are shown. The test also provides total test score, which is decoded as follows:

Basic < 18, Good 18-22, Excellent 23-27,5 and Outstanding > 28.

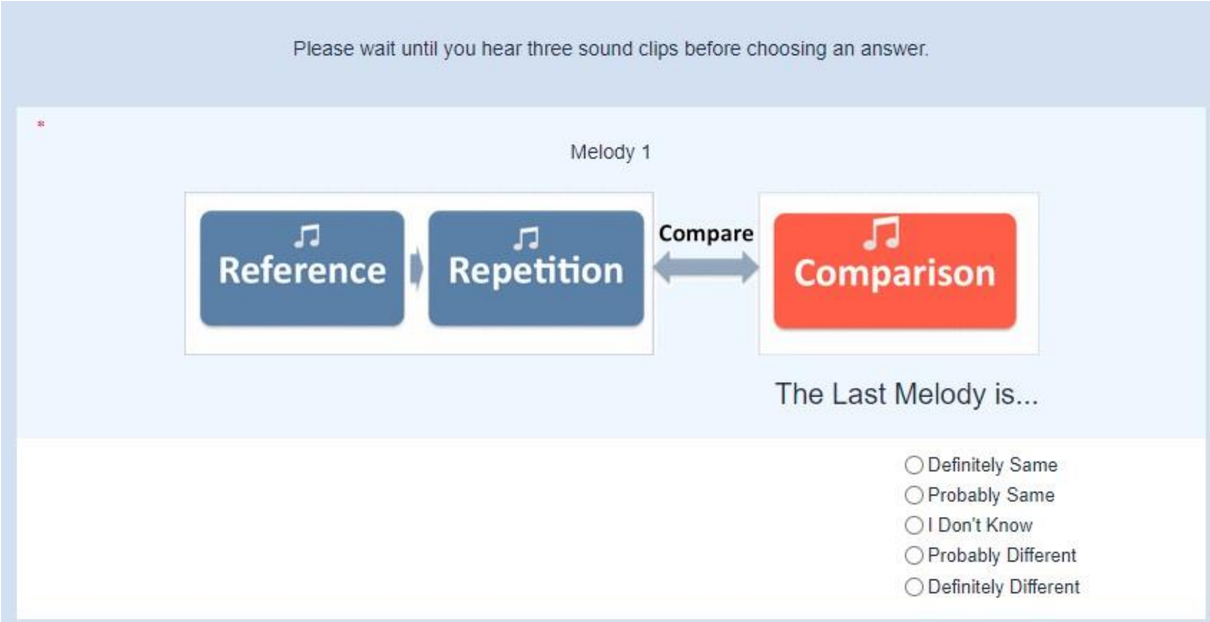


Figure 14 Mini-Proms

7.6 Procedure

The study consisted of the following parts: 1) basic questionnaire 2) pre-tests including the language aptitude tests and musical aptitude plus the singing score, and 3) the experiment part. Participants were asked to fill out a basic questionnaire online, as well as accomplish the pre-tests which likewise were accessible on the internet. The assignment for the singing score was to record the singing performance alone and send it beforehand to the interviewer.

In the experiment part, the participants were divided into three groups, the first group being the singing condition group, the second being the listening condition group, and the third group being the control group. Due to the safety measurements caused by the COVID-19 pandemic, the experiment was conducted online through Zoom-meetings. The approximate time for the execution of the experiment was ca 20 minutes. The experiment consists of three phases, where the first phase was the learning phase with visual stimuli. Participants' task in the first phase was to learn 10 Finnish words which were displayed on the screen with corresponding pictures referring to the meaning. The target words of the experiment are illustrated in Figure 10.

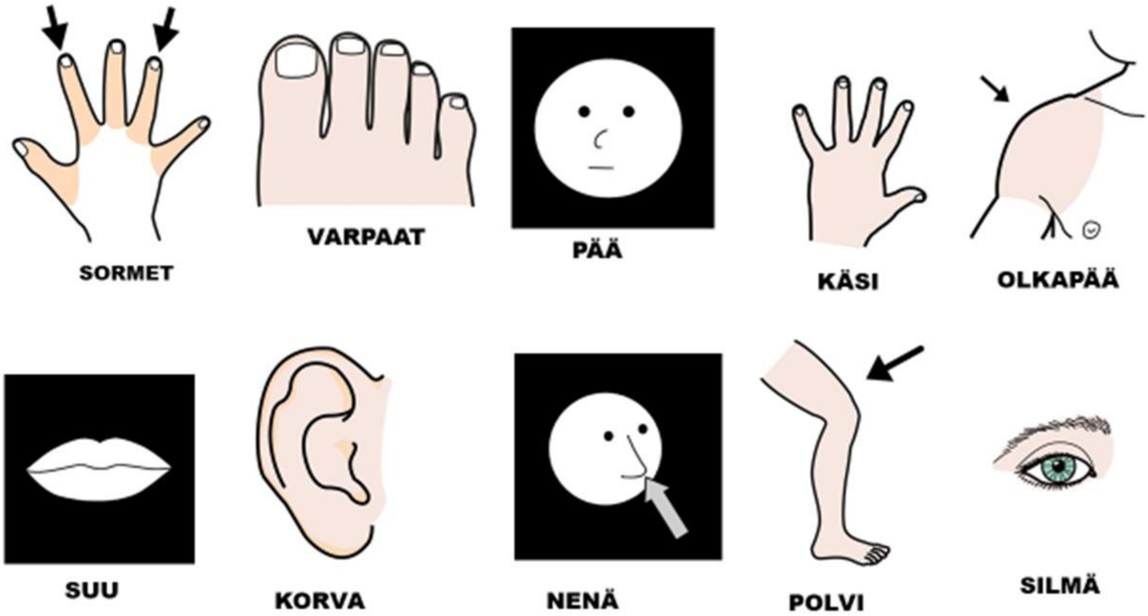


Figure 15 Target Words

The target words for the experiment were chosen for the study due to the assumption that body part words commonly form the basic vocabulary of a language, referring to words that in historical and comparative linguistics are typically associated with stability, universality, simplicity, and resistance to borrowing (Tadmor, Haspelmath, & Taylor, 2010). Finnish is also a Finno-Ugric language and differs therefore remarkably from participants' first language German, which belongs to the Indo-European language family. The 10 target words include 2 one-syllable words: *suu* ("mouth") and *pää* ("head"), 7 two-syllable words: *sormet* ("fingers"), *varpaat* ("toes"), *käsi* ("hand"), *korva* ("ear"), *nenä* ("nose"), *polvi* ("knee") and *silmä* ("eye"). The longest word *olkapää* ("shoulder") consists of three syllables and is a composita consisting of two words *olka* and *pää*.

Following the learning phase, both singing and listening condition groups heard the experiment song that repeated the target words. In the experiment song the target words were embedded in the melody of the famous children's song "Twinkle twinkle little stars". The experiment song was presented without musical accompaniment, sung by a native speaker of Finnish. In this phase, the pictures of the target words remained visible on the screen, but the written forms of the word were removed. Figure 11 represents the notation of the experiment song including the lyrics.

Figure 16 Experiment song, Full-scale study

The singing condition group heard the experiment song altogether 3 times, after that the experiment song was played 2 times more and the participants were asked to try to sing along. The participants were instructed to put their microphones off in order to focus on the learning task itself without being disturbed by the idea of being heard by the interviewer. The listening condition group participants heard the experiment song altogether 5 times, without being asked to sing. The control group heard the target words spoken by a native speaker 5 times after another without any melodical input. Altogether the learning phase duration was approximately 5 to 7 minutes.

After the experiment phase the participants' short-term recall of the learned vocabulary was tested immediately after the learning session. The participants were asked to repeat all the words they could recall, and if possible, to combine with the corresponding picture. The pictures of the target words were once again displayed numbered to the screen to facilitate the word recall. All answers were recorded and scored by three native speakers of Finnish following the phoneme score evaluation demonstrated in the next section.

Phoneme scoring

When compared with the whole word scoring, phoneme scoring has been suggested as an alternative due to the several advantages while evaluating phoneme and word perception. According to Billings, Penman, Ellis, Baltzell, & McMillan (2016), phoneme scoring provides an increased number of tokens in the same amount of test time, reduced variability in scores, information about specific phoneme perception errors, and reduced influence from top-down processing, such as effects of the listener's lexicon. Until these days there is no established practice concerning guidelines of phoneme scoring, which causes difficulty in comparing scores across clinics due to issues with scoring accuracy.

Following the phoneme scoring guidelines developed by Billings et al., the recorded participant interviews were evaluated through both word score and phoneme score. Three native speakers of Finnish listened and evaluated the interviews, where the rater's task was to assess the quality of each pronounced word. The scale for a rating in the word score was out of 10 possible, and the phoneme score was out of 49 possible phonemes. Recordings were provided electronically in MP3 format, in arbitrary order.

Finnish pronunciation

The Finnish language has a relatively regular pronunciation, which includes almost one-to-one correspondence between letters and sounds. Both vowels and consonants can be short when they are written with one letter, or long, when it is written with two letters. The length of a word is distinctive, which means that words may well differ only in the length of a sound.

In short words, the first syllable always takes the main stress, however stressing in Finnish is usually not very emphatic. Longer words might have secondary stress, usually on the 3rd or 4th syllable.

In Finnish, the vowel system consists of eight vowel phonemes that are given as /i/, /e/, /y/, /ö/, /ae/, /a/, /o/ and /u/. All vowel phonemes excluding /y/ and /ö/ occur in the target words of the experiment. The Finnish phoneme /ae/ that occurs in the target words *käsi* ("hand"), *silmä* ("eye"), and *nenä* ("nose"), is an open unrounded front vowel and is pronounced as English "a" in "bat". In the target words *pää* ("head") and *olkapää* ("shoulder"), the vowel /ae/ is long, where it has to be pronounced accordingly. Short phonemes in Finnish are written with one letter and long phonemes with two. Usually, the length of vowels and consonants is used to differentiate word meanings, but between the corresponding long and short vowels **ää-ä** there is no difference in quality. The long phonemes are pronounced as pure long vowels, not as if they were diphthongs. Additionally, the phonetic duration of a long vowel is approximately twice that of a short one (Karlsson, 2018).

What comes to the consonants occur in the target words, the consonants **p**, **t**, and **k** are pronounced without aspiration. The consonant **s** is commonly pronounced as the dark, thick sound that is close to [š], especially when it takes place in an environment of **u,** such as in the target word *suu* („mouth“). The consonant **r** is always pronounced trilled with the tip of the tongue, as in the target word *korva* („ear“).

7.7 Statistical Analysis

Statistical analyses were performed using SPSS Statistics 27.0 (IBM Corporation). First, correlations between all variables were conducted to find out possible positive correlations. Additionally, an independent-samples t-test was run between the condition groups to determine differences in the vocabulary recall scores. Finally, a one-way ANOVA was performed to compare the effect of the different condition groups.

RESULTS

8.0 Results

As previously mentioned, the statistical analyses were operated by SPSS Statistics 27.0 (IBM Corporation, New York, USA). For analysis, correlation tests were run to find out whether musical aptitude affects vocabulary learning. In addition to that, t-tests and a one-way ANOVA were run between the three different condition groups to find out, whether singing performance aids vocabulary learning tasks.

8.1 Foreign Language aptitude, Musical Aptitude, Singing score

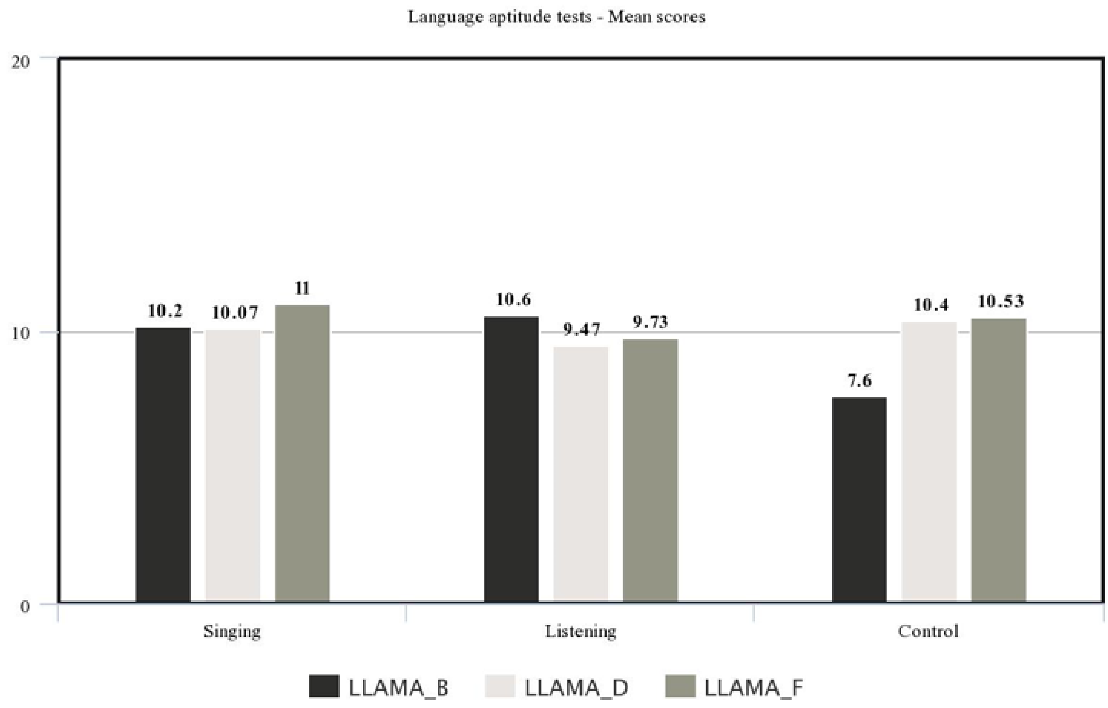
The following subsections provide an overview of the individual descriptive statistics of the main variables.

The mean of the vocabulary learning module (LLAMA_B) was 10,42; SD = 3,690. The statistics of the sound recognitions test (LLAMA_D) showed a mean of 9,47; SD = 4,934. In the grammatical inferencing module (LLAMA_F) the mean was 9,98; SD 4,181. Further frequencies of the language aptitude variable are illustrated in Table 1.
Among the language aptitude modules, there was no large variation in the results. On average, participants scored best in the grammatical inferencing task (LLAMA_F) the lowest scores were achieved in the vocabulary learning task (LLAMA_B).

Table 1 Descriptive statistics of Language aptitude tests

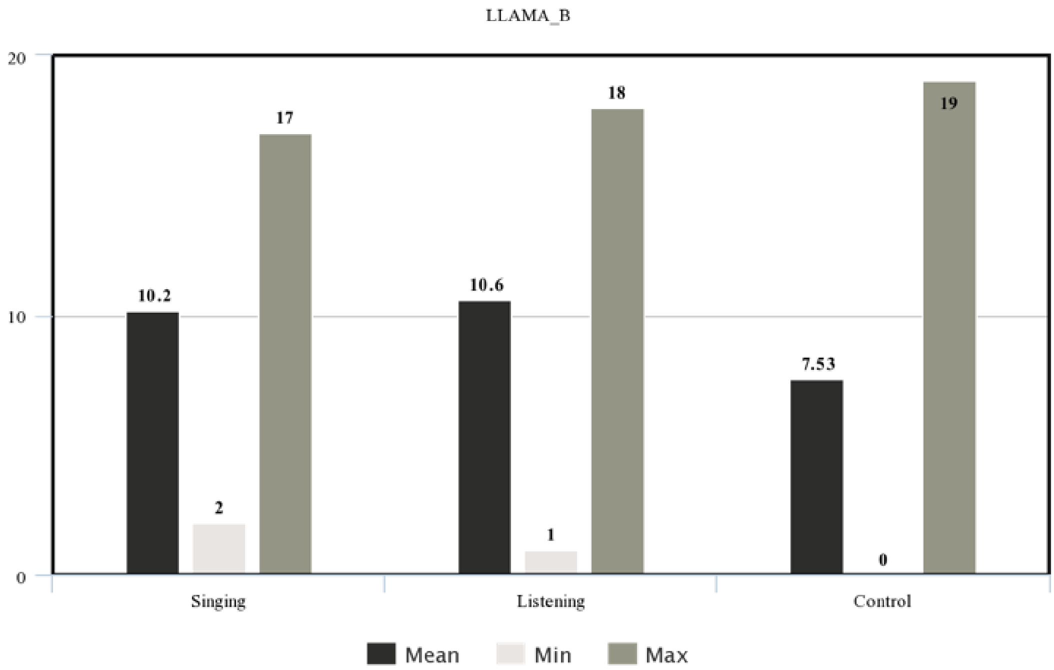
Statistics

	Mean	Median	Std. Deviation	Skewness	Minimum	Maximum
LLAM-AB	9.47	10.00	4.934	.129	0	19
LLAMA D	9.98	10.00	4.181	-.761	0	19
LLAMAF	10.42	11.00	3.690	-.383	3	18



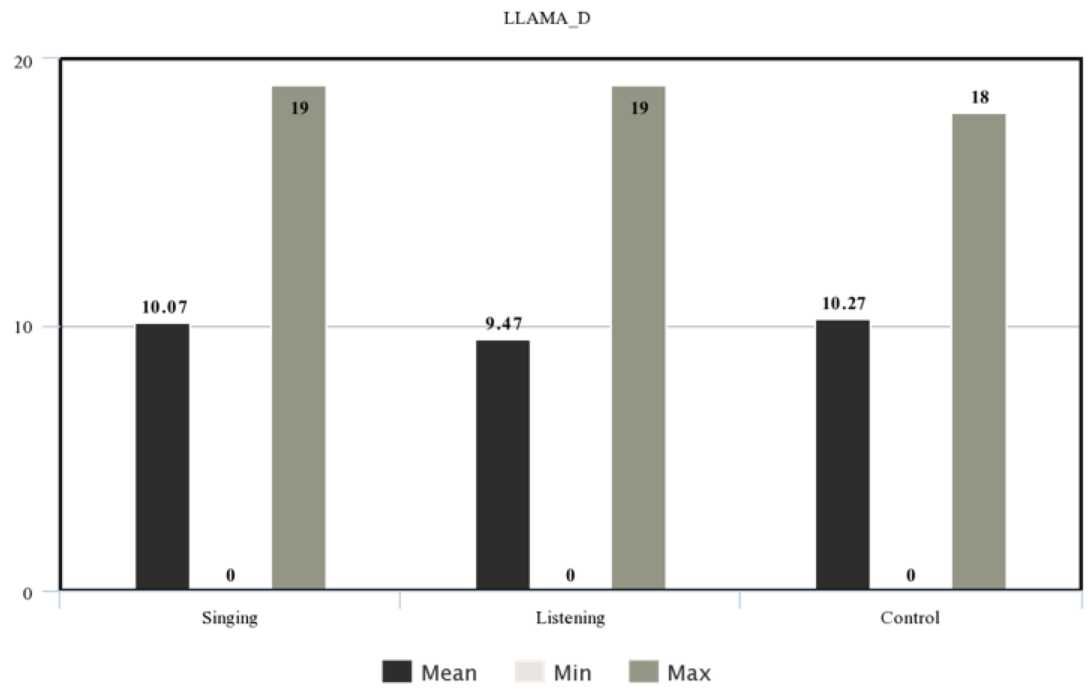
8.1.1 LLAMA_B

Results of the LLAMA_B represent the vocabulary learning skills of the participants. The maximum score to be achieved in the LLAMA_B module is 20. Among all participants, the average score of this sub-test was 9,47 points, the lowest score being 0 points and the maximum score 19. In the singing condition group, the average score was 10,20, 2 points being the lowest and 17 being the highest score. Among the participants of the listening condition group, the average score was 10,60, the lowest score being 1 point and the highest 18 points. In the control group, the lowest score of the vocabulary learning module was 0 points and the highest was 19 points, the average being 7,53.



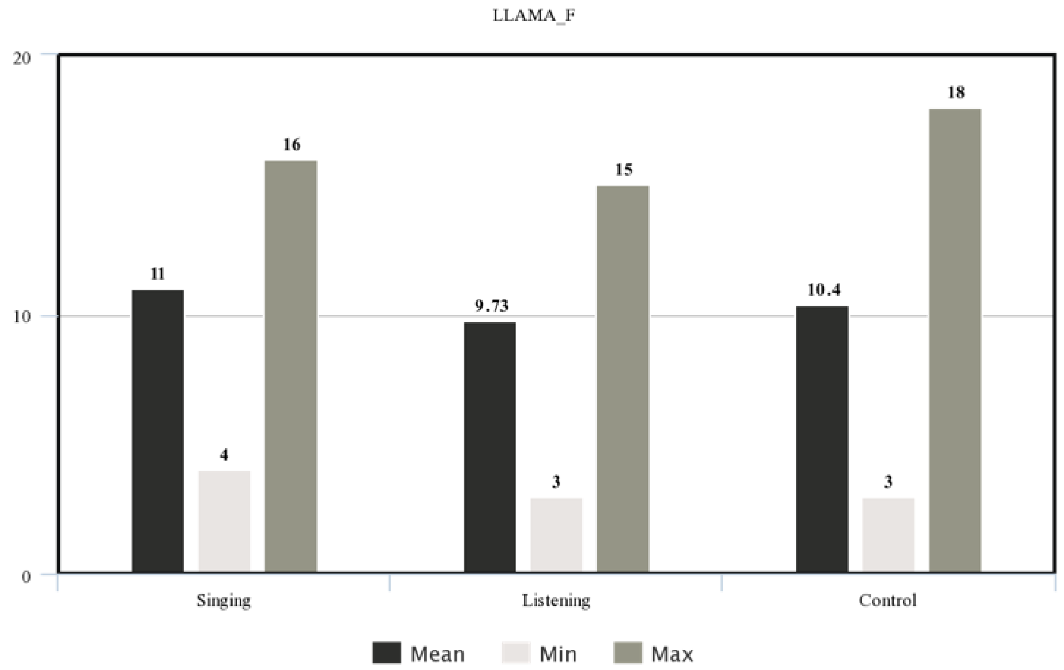
8.1.2 LLAMA_D

The LLAMA_D tests the participant's skill to recognize repeated sounds of spoken language. Similar to LLAMA_B, the maximum score to be achieved is 20 points. Among all the participants, the average score was 9,98, the lowest score being 0 and the highest 19 points. In the singing condition group, the mean average of the group was 10,07, the lowest score being 0 and the highest 19 points. The average score of the listening condition group was 9,47, the lowest score being 0 and the highest 19 points. In the control group, the lowest score was 0 points and the highest 18 points, the average score of the group being therefore 10,27.



8.1.3 LLAMA_F

The LLAMA_F module measures the grammar inferencing skills of the subjects. Likewise, to the other modules, the highest score to be achieved in LLAMA_F is 20 points. The average score among all the participants was 10,42, the lowest score achieved being 3 and the highest 18. In the singing condition group, the lowest score was 4 and the highest 16 points, the average score being 11. The average score in the listening condition group was 9,73, the lowest score being 3 and the highest 15. In the control group, the average score was 10,40, the lowest score being 3 and the highest 18 points.



8.2 Musical aptitude – MINI-PROMS

The musical aptitude of respondents was tested through MINI-PROMS musical aptitude test. The mean of the musical aptitude test among all participants was 23,45; SD=4,20. Among all test-takers, the lowest score achieved in Mini-PROMS was 13 points and the highest score was 32 points. Table 2 shows more descriptive details of the musicality variable. Of the initial sample, only 6 subjects (13,3%) scored outstandingly, meaning that their result from the musical aptitude was 28 points or higher, 36 points being the highest possible score in the test. The majority of all participants (22 subjects, 48,9%) received excellent test results, varying between 23 and 27,5 points. 12 subjects, (26,7%) scored with a good outcome, varying from 18 to 22,5 points, and lastly, 5 participants performed a basic score, which resembles the result of fewer than 18 points.

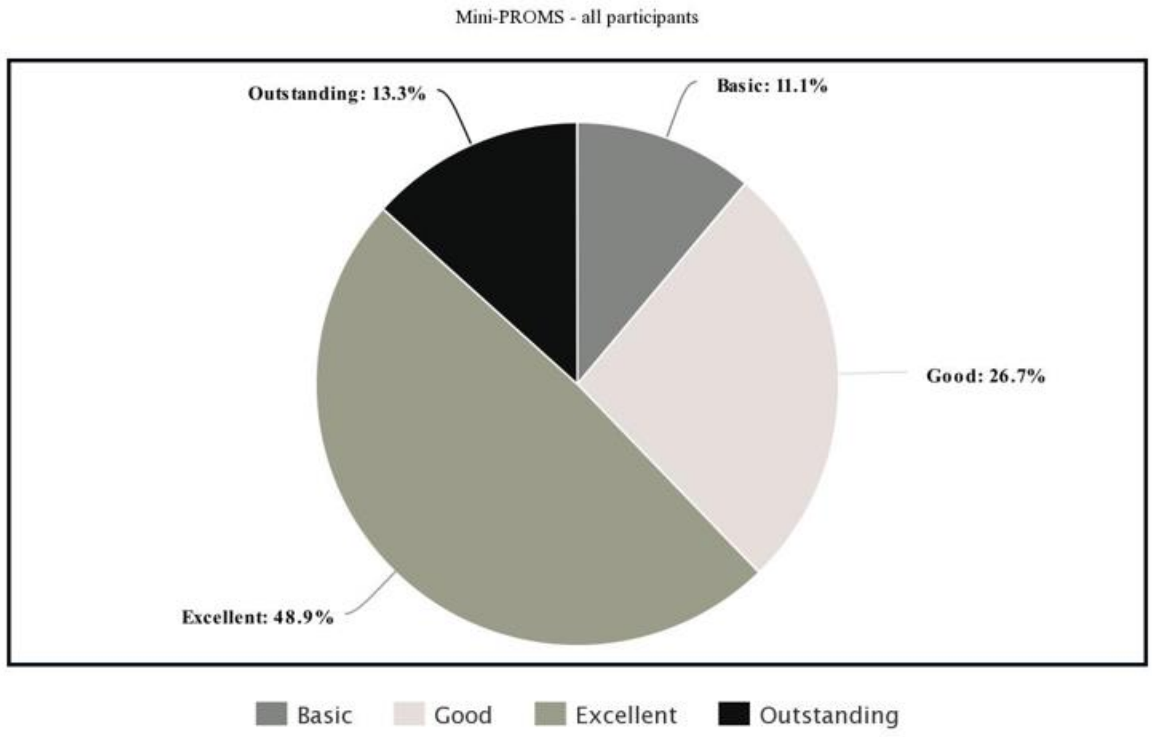
Table 2: Descriptive statistics

N	Valid	45
	Missing	0
Mean		23.4555556
Median		23.5
Mode		25.5, 21.5, 23, 26, 27
Std. Deviation		4.2087871
Skewness		-0.295757097
Minimum		13
Maximum		32

As illustrated in table 3, participants of the singing condition group performed best in the Mini-PROMS. The control group scored the lowest with a mean average of 21,80.

Table 3: Results Musicality

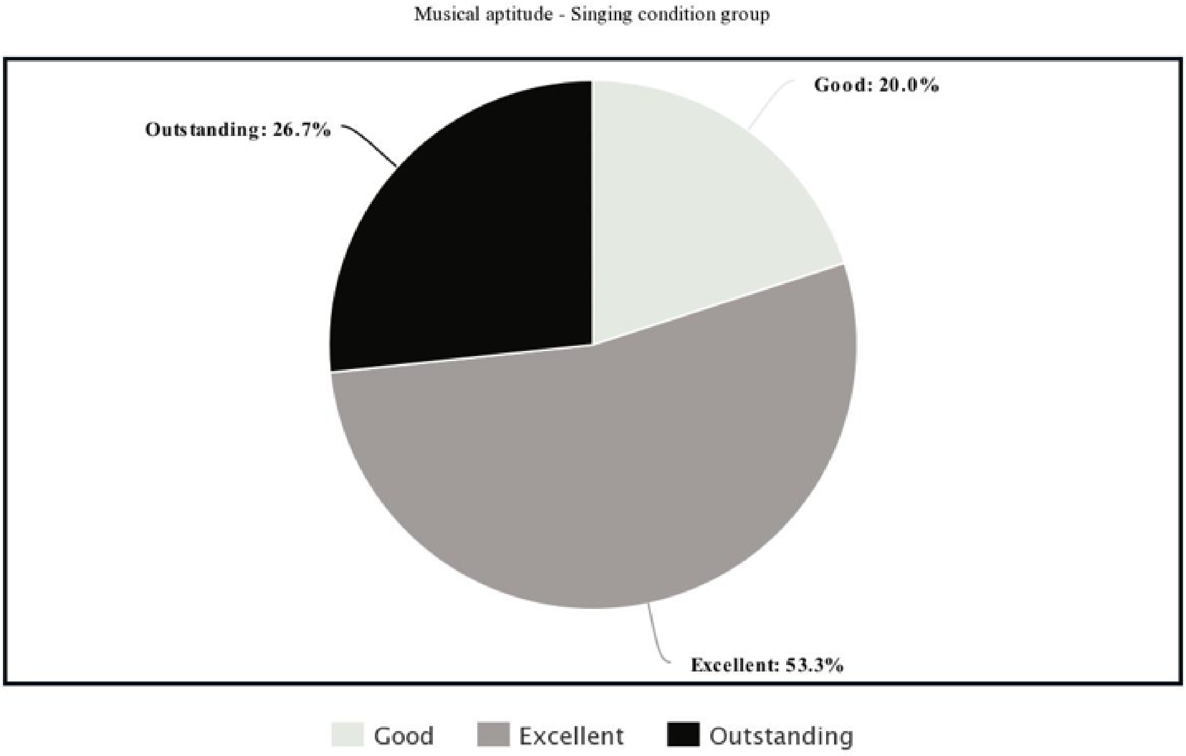
	Mean	Median	Min	Max
Singing condition group	25,56	26	19,5	30
Listening condition group	22,76	22,5	15,5	32
Control group	21,80	22	13	28



8.2.1 Musical aptitude - singing condition group

In the sample of 15 subjects who participated in the singing condition group, the mean performance score was 25,56; SD = 3.11, more statistical information is illustrated in Table 4. Over half (8 participants, 53,3%) received an excellent test result from the musicality aptitude test, whereas 26,7 percent, 4 participants performed outstandingly and 20 percent, 3 participants received a good music aptitude score. In this condition group, no one of the participants received less than 18 points. The lowest score in this group was 19,5 points and the highest 30 points.

Table 4: Descriptive statistics – Singing condition group		
N	Valid Missing	15 0
Mean		25.5666667
Median		26
Mode		24.5
Std. Deviation		3.13315602
Skewness		-0.338278669
Minimum		19.5
Maximum		30

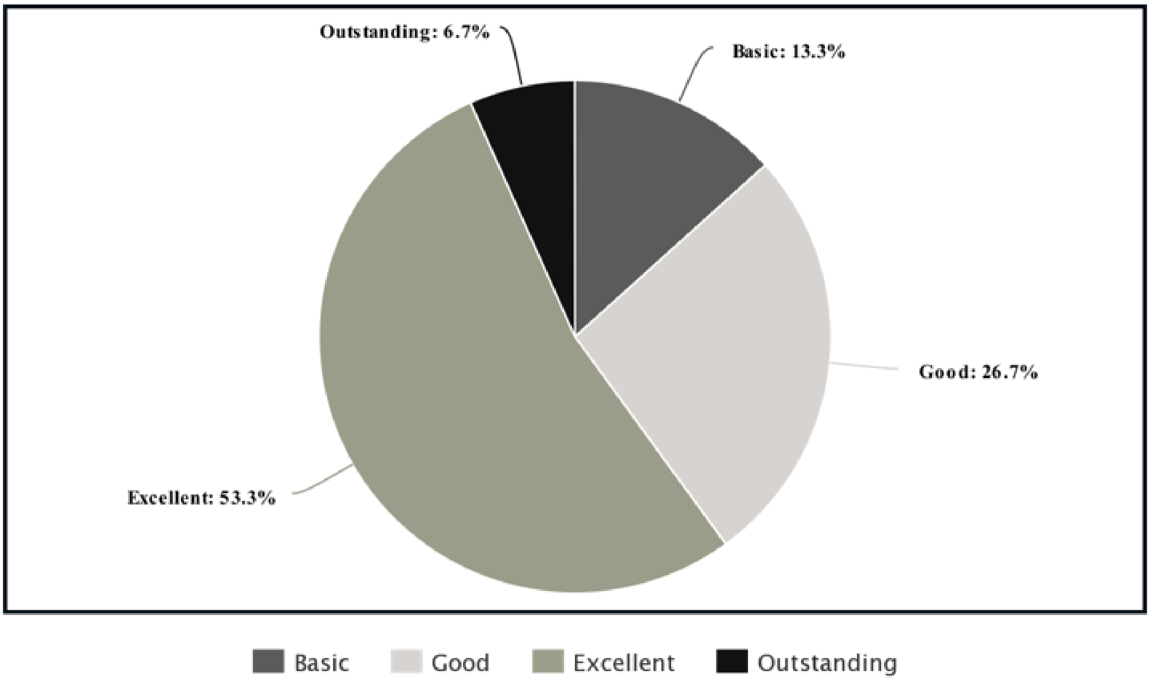


8.2.2 Listening Condition Group

The mean score of the listening condition group was 22,77; SD=4.11. A small majority of the subjects (8 participants, 53,3%), had an excellent musical aptitude score. Four participants (26,7 %) received a good musical aptitude score and only a couple of subjects had a basic score (2 participants, 13,3%). In this group, one participant (6,7%) had an outstanding score. The highest score among the listening condition group was 32 points and the lowest aptitude score was 15,5 points. Additional statistical information is displayed in Table 5.

Table 5: Descriptive statistics - Listening condition group		
N	Valid	15
	Missing	0
Mean		22.7666667
Median		22.5
Mode		22.5, 21.5, 25.5, 26
Std. Deviation		4.10951277
Skewness		-0.24649671
Minimum		15.5
Maximum		32

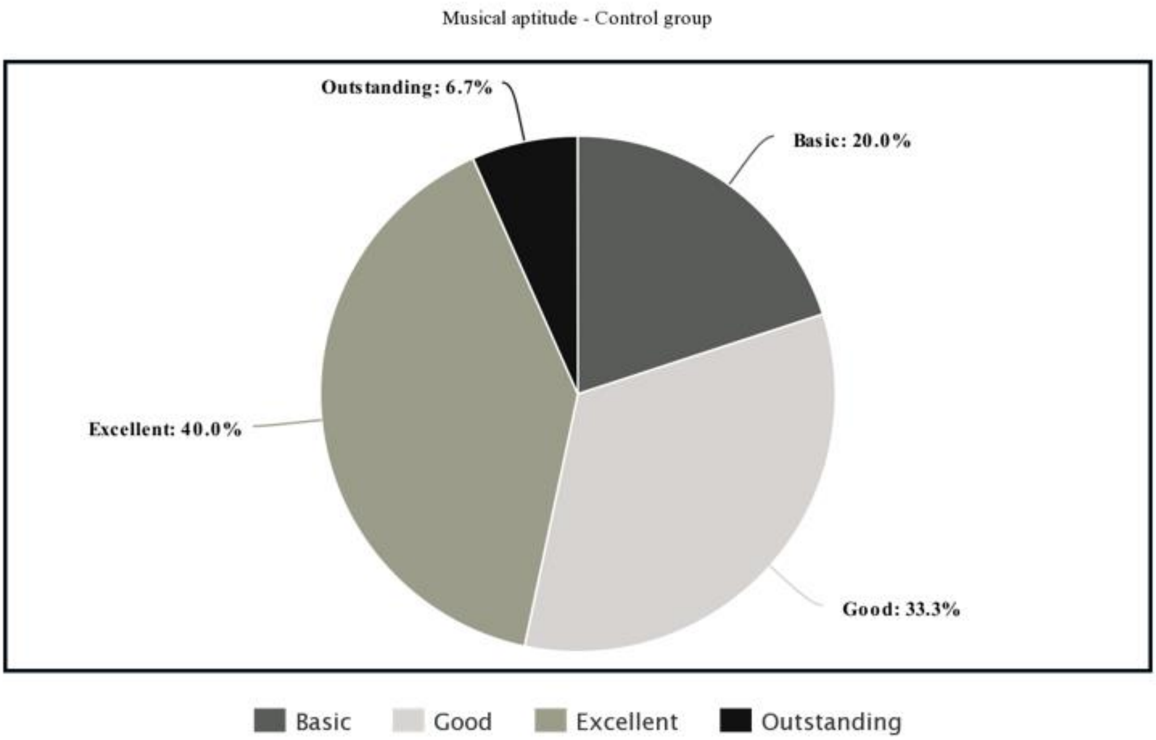
Musical aptitude - Listening condition group



8.2.3 Musical aptitude - control group

In the control group, the mean performance score was 21,8; SD= 4.341. The majority in this group received an excellent score with 6 participants (40%). 33,3 percent of the group, 5 participants achieved a good musical aptitude score and 3 test-takers (20%) had a basic musical aptitude based on the Mini-PROMS scale. Only one subject from the control group (6,7%) scored outstandingly. In this group, the lowest score received was 13 points, and the highest score was 28 points.

Table 6: Descriptive statistics - control group		
N	Valid Missing	15 0
Mean		21.8
Median		22
Mode		25.5, 23, 20.5, 27
Std. Deviation		4.34165867
Skewness		-0.311723553
Minimum		13
Maximum		28



8.3 Singing score results – all participants

The singing scores of the participants were evaluated by two experienced singers in three different categories: intonation accuracy, sense of rhythm, and voice quality. Each category was scored on a scale of 0 to 10 points, 0 being the lowest and 10 the highest possible score. Therefore, the lowest score that someone could receive was 0 and the highest 60. The mean score of all participants was 41,22, where the lowest score achieved was 20 and the highest 60. Additional statistical information is illustrated in table 7.

Table 7 - Descriptive statistics – Singing score, all participants		
N	Valid Missing	45 0
Mean		41.22
Median		42.00
Mode		48
Std. Deviation		9.317
Skewness		-0.408
Minimum		20
Maximum		60

Results Singing score	Mean	Median	SD	Min	Max
Singing condition group	48.33	48.00	5.802	41	60
Listening condition group	36.20	39.00	8.428	23	49
Control group	39.27	41.00	9.192	20	52

In general, the participants of the singing condition group performed the best in the singing score (48,33), followed by the control group (39,27). The lowest average score of all groups was in the listening condition group (36,20).

8.3.1 Singing score – singing condition group

The average of the singing score in the singing condition group was 48,33; SD = 5,80. The lowest score of the group was 41 and the highest 60, which is the maximum number of points one could achieve. More statistical information is illustrated in table 8.

Table 8- Descriptive statistics – Singing condition group		
N	Valid Missing	15 0
Mean		48.33
Median		48.00
Mode		48
Std. Deviation		5.802
Skewness		1.066
Minimum		41
Maximum		60

8.3.2 Singing score – listening condition group

The participants of the listening condition group had a mean of 36,20 in the singing score; SD = 8,42. The lowest singing score was 23 points and the highest 49 points. Descriptive statistics of this group are displayed in table 9.

Table 9 - Descriptive statistics – Listening condition group		
N	Valid	15
	Missing	0
Mean		36.20
Median		39.00
Mode		39
Std. Deviation		8.428
Skewness		-0.262
Minimum		23
Maximum		49

8.3.3 Singing score – control group

Finally, in the control group, the mean score of the singing task was 39,27, SD = 9,19. In this group, the lowest score received was 20 points, which was also the lowest score of all participants, and the highest score was 52 points. The descriptive statistic of this group is illustrated in more detail in table 10.

Table 10 - Descriptive statistics – Singing condition group		
N	Valid	15
	Missing	0
Mean		39.27
Median		41.00
Mode		41
Std. Deviation		9.192
Skewness		-0.618
Minimum		20
Maximum		52

8.4. Results – Experiment

In the experiment part itself, participants went through learning phases where they had to learn 10 words in Finnish about body parts. Depending on the condition group, the learning phases differed regarding the input format of the words. After the learning phase, participants were interviewed and asked to repeat the words they had currently learned. The interviews were recorded, and the answers were evaluated by three native Finnish speakers both in word-level and phoneme level. In word scoring, the lowest possible score one could achieve was 0 and the highest 10, whereas in phoneme scoring the minimum score to be achieved was 0 and the maximum 49. On the word scoring level, the mean average score among all participants was 5.64; SD = 1,77. The lowest score achieved was 2 words and the highest was 10 words. On the phoneme scoring level, the mean average score was 21,60; SD = 20,00. The lowest score received in the whole sample was 7 phonemes and the highest was 46 phonemes. Additional statistical information is displayed in table 11.

Table 11- Descriptive statistics - Experiment			
		Word scoring	Phoneme scoring
N	Valid	45	45
	Missing	0	0
Mean		5.64	21.60
Median		6.00	20.00
Mode		6	16
Std. Deviation		1.773	9.061
Skewness		-0.083	-0.529
Minimum		2	7
Maximum		10	46

As illustrated in table 12, on the word scoring level the singing condition group the best with the mean score of 6.00 words. Compared to that, the listening condition group had a mean score of 5.60 and the control group 5.13.

Table 12

Results Experiment – Word Scoring	M	MD	SD	Min	Max
Singing condition group	6.00	6.00	1.648	4	10
Listening condition group	5.60	6.00	1.724	2	9
Control group	5.13	5.00	1.959	2	9

Table 13

Results Experiment – Phoneme Scoring	M	MD	SD	Min	Max
Singing condition group	23.60	24.00	8.567	13	46
Listening condition group	20.33	19.00	8.121	7	39
Control group	19.73	15.00	10.215	7	40

The descriptive statistics of the results of the phoneme scoring regarding the different condition groups are displayed in table 13 According to that, the highest mean score was in the singing condition group, 23,60. In the listening condition group, the average score was 20,33 phonemes, and in the control group 19,73.

The next section focuses on the central research questions of this thesis. Hence, it displays an overview about if and to what extent singing performance aids vocabulary acquisition in foreign language learning. In addition to that, the role of musical aptitude is investigated.

8.4.1 Results – The Interrelation of Foreign language Aptitude, Musicality and Singing performance

Table 14 - Correlations

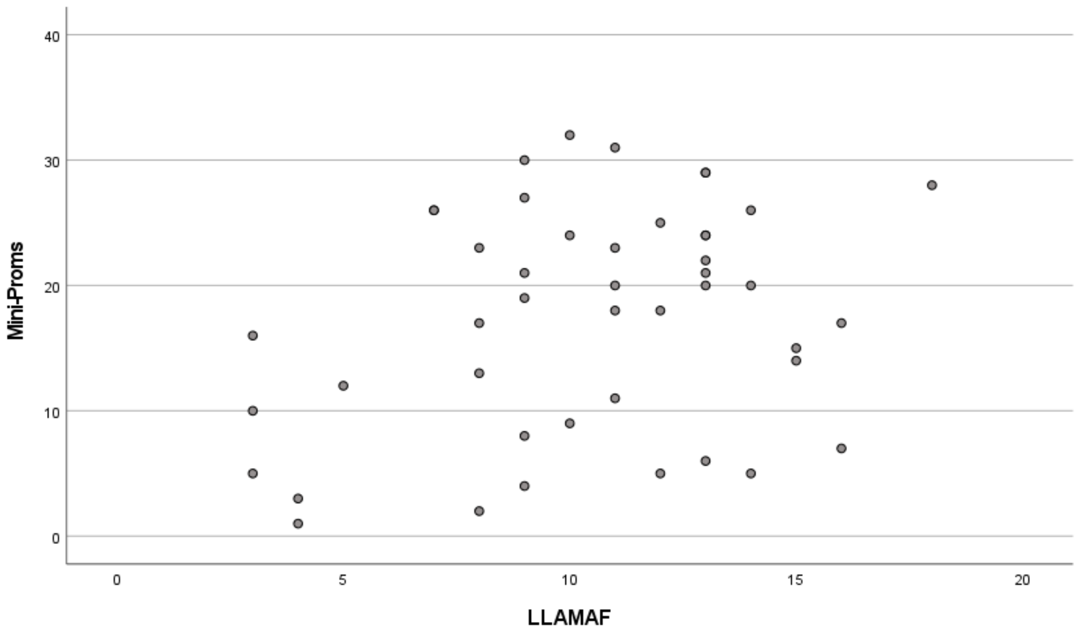
		LLAMAB	LLAMA D	LLAMA F	Mini-Proms	Singing score	Word scoring	Phoneme scoring
LLAMAB	Pearson Correlation	1	-.042	.296*	.080	.011	.180	.229
	Sig. (2-tailed)		.782	.048	.602	.945	.235	.130
	N	45	45	45	45	45	45	45
LLAMAD	Pearson Correlation	-.042	1	-.002	.202	.180	.060	.140
	Sig. (2-tailed)	.782		.988	.182	.237	.694	.358
	N	45	45	45	45	45	45	45
LLAMAF	Pearson Correlation	.296*	-.002	1	.300*	.269	.055	.023
	Sig. (2-tailed)	.048	.988		.046	.074	.721	.882
	N	45	45	45	45	45	45	45
Mini-Proms	Pearson Correlation	.080	.202	.300*	1	.356*	-.233	-.106
	Sig. (2-tailed)	.602	.182	.046		.016	.124	.490
	N	45	45	45	45	45	45	45
Singing score	Pearson Correlation	.011	.180	.269	.356*	1	.054	.140
	Sig. (2-tailed)	.945	.237	.074	.016		.723	.360
	N	45	45	45	45	45	45	45
Word scoring	Pearson Correlation	.180	.060	.055	-.233	.054	1	.923**
	Sig. (2-tailed)	.235	.694	.721	.124	.723		.000
	N	45	45	45	45	45	45	45
Phoneme scoring	Pearson Correlation	.229	.140	.023	-.106	.140	.923**	1
	Sig. (2-tailed)	.130	.358	.882	.490	.360	.000	
	N	45	45	45	45	45	45	45

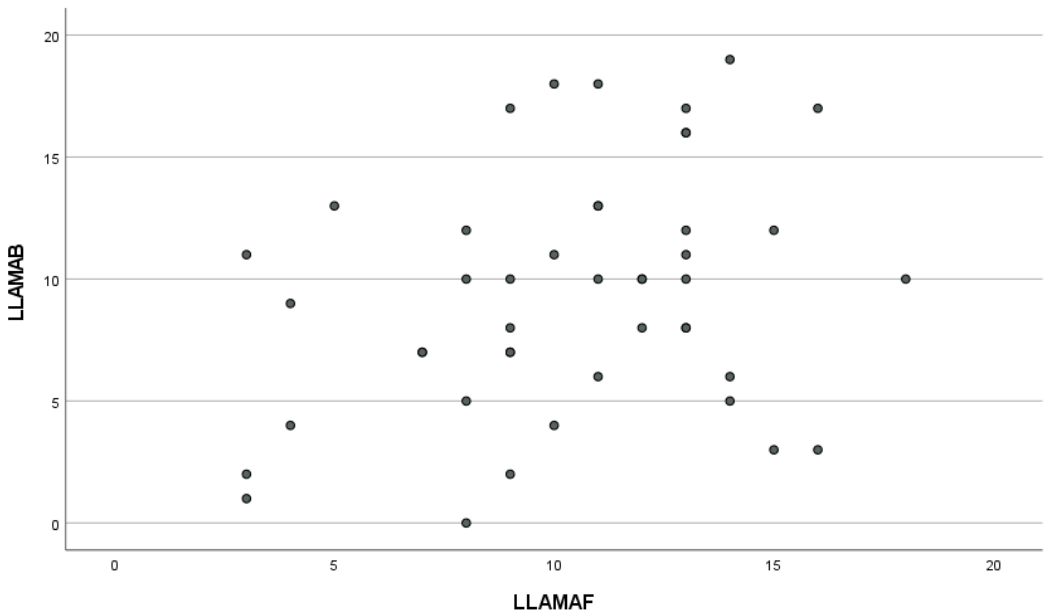
*. Correlation is significant at the 0.05 level (2-tailed).

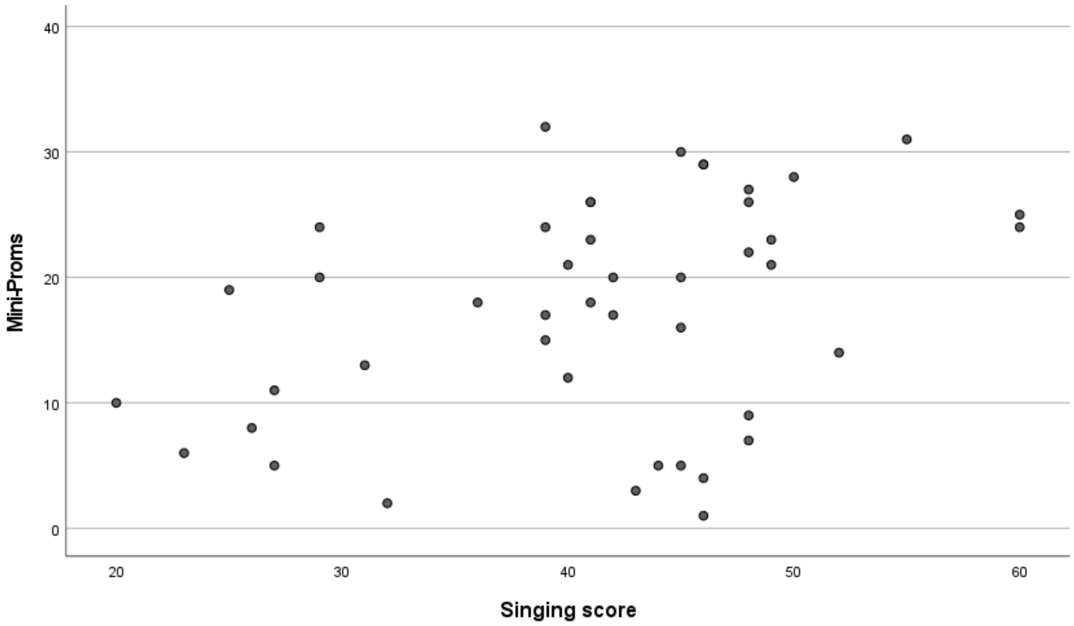
**. Correlation is significant at the 0.01 level (2-tailed).

The musical aptitude test (Mini-PROMS) and the singing score as well as the language aptitude tests were correlated with both vocabulary tasks (word scoring, phoneme scoring). No significant correlation was found between musicality and vocabulary acquisition and similarly, no significance was found between singing score and vocabulary acquisition.

Similarly, no relationship was found between language aptitude and vocabulary acquisition. Interestingly, there was a low positive correlation between musical aptitude and the grammar inference aptitude tests (LLAMA_F) (*p* = 0.300). Additionally, a low positive correlation was found between the vocabulary learning module (LLAMA_B) and the grammar inference aptitude test (LLAMA_F) (*p* =0.296). Table 14 summarizes the most important results of foreign language aptitude and musicality aptitude and scatterplots of the most meaningful correlations are displayed below.







8.4.2 t-tests

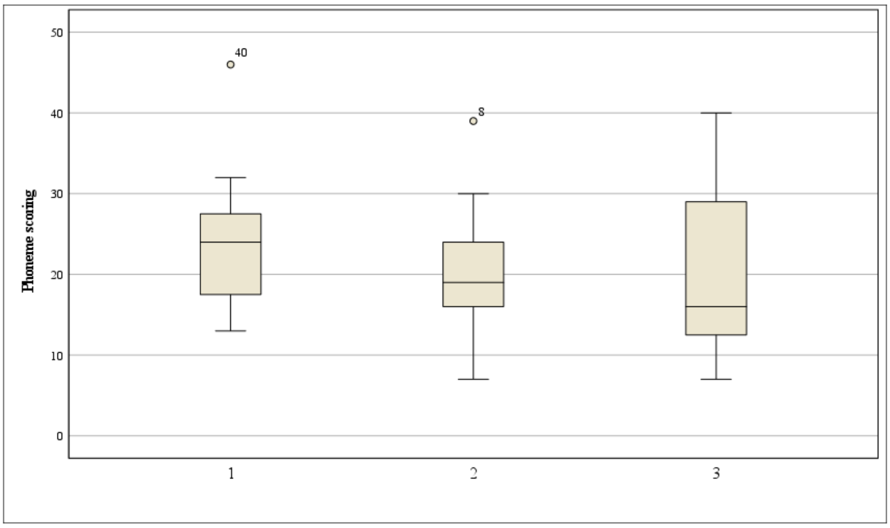
An independent-samples t-test was run between the condition groups (1 vs 2, 1 vs 3, 2 vs3) to determine if there were differences in the vocabulary scores. Data are mean ± standard deviation unless otherwise stated. The mean phoneme score of the singing condition group (23.60 ± 8.56) was higher than the mean phoneme score of the listening condition group (20.33 ± 8.12). The mean phoneme score of the control group (20.87 ± 10.59) was slightly higher than in the listening condition group.

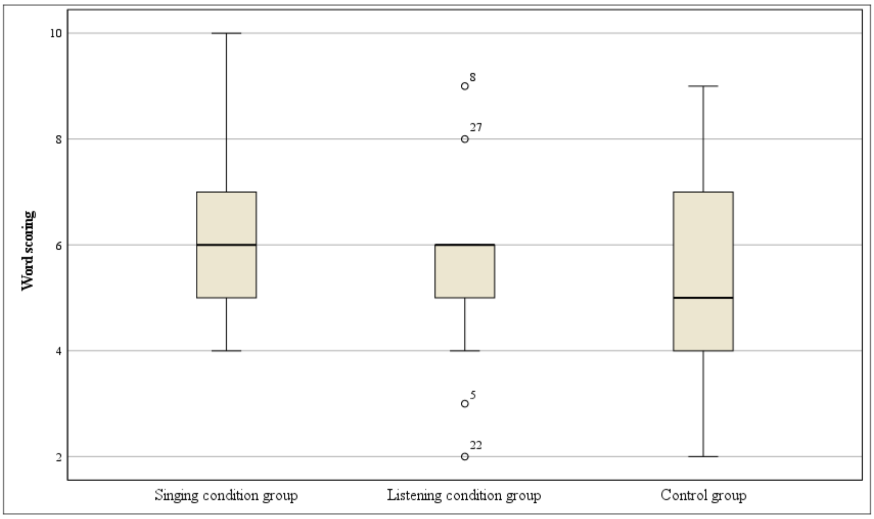
8.4.3 Normality and homogeneity of variances

The vocabulary task scores were normally distributed, as assessed by Shapiro Wilk's test ($p <$ .05) and visual inspection of the Normal Q-Q plots. There was homogeneity of variances for vocabulary scores between all condition groups in both vocabulary tasks, as assessed by Levene's test for equality of variances (p = > .05). Tables illustrating tests of normality and Normal Q-Q plots can be found in appendix.

8.4.4 Outliers

In the data of the vocabulary tasks, 2 outliers (39, 46) were detected in the data of phoneme scores and 4 outliers (2,3,8,9) in the data of word scores, as assessed by inspection of boxplots. After doublechecking the data, the possibility of the detected outliers being data entry errors was outlined.





To determine whether the outliers influenced results for a *t*-test on a vocabulary test, all the t-tests were conducted twice, with outliers (i.e., the entire sample) and without outliers. Even though the results were not entirely identical, the presence of outliers did not significantly affect the results of the analysis. Once all analyses were concluded, there was no significant difference between the vocabulary task scores for different condition groups. All the p-values of the analysis were above a significance level of .05, as illustrated in table 15 all comparisons with and without outliers are displayed in the appendix. Between condition groups 1 (singing condition) and 3 (control group) only the phoneme scores were examined because the outliers were only found from the phoneme score results of the singing condition group.

Table 15 - Comparison of results to define the effect of outliers

Comparison	Vocabulary task	With outliers	Without outliers
1 vs 2	Word scoring	.521	.874
	Phoneme scoring	.293	.221
2 vs 3	Word scoring	.689	.262
	Phoneme scoring	.878	.575
1 vs 3	Phoneme scoring	.444	.730

8.4.5 Statistical significances

The mean phoneme score of the singing condition group was 3.27 ± 3.05 (mean ± standard error) higher than the mean phoneme score of the listening condition group. Additionally, the phoneme score of the singing condition group was 2.73 ± 3.52 higher than the mean phoneme score of the control group.

There was no significant difference neither in the phoneme nor the word scores between the condition groups (p = > .05), as illustrated below. More detailed results summarized into tables can be found in appendix.

8.4.6 Independent Samples Test – Overview

Phoneme scoring

Condition group	**N**	**Mean**	**SD**
Singing condition	15	23.60	8.567
Listening condition	15	20.33	8.121
Control group	15	20.87	10.589

Sig. 2-tailed

Singing condition vs. Listening condition	Singing condition vs. Listening condition	Listening condition vs. Control group
.293	.444	.878

Word scoring

Condition group	**N**	**Mean**	**SD**
Singing condition	15	6.00	1.648
Listening condition	15	5.60	1.724
Control group	15	5.33	1.988

Sig. 2-tailed

Singing condition vs. Listening condition	Singing condition vs. Listening condition	Listening condition vs. Control group
.521	.326	.698

8.4.8 One-way ANOVA

Additionally, to compare the effect of the different condition groups, a one-way ANOVA was performed.

A one-way ANOVA revealed that there was not a statistically significant difference in the vocabulary acquisition between the three condition groups, word scoring task ($F(2,42) = 0.526$ ns) and phoneme scoring task ($F(2,42) = 0.581$ ns)

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Word scoring	Between groups	3.378	2	1.689	.526	.595
	Within groups	134.933	42	3.213		
	Total	138.311	44			
Phoneme scoring	Between groups	92.133	2	46.067	.550	.581
	Within groups	3520.667	42	83.825		
	Total	3612.800	44			

Table 16 - Summary of Research Findings

1. Does singing improve vocabulary acquisition more than merely listening to songs?

Result: Singing performance does not correlate with vocabulary acquisition.

The hypothesis of singing performance outscoring listening condition rejected.

2. Does music aptitude correlate with vocabulary acquisition?

Result: Music aptitude does not correlate with vocabulary acquisition.

The hypothesis of music aptitude correlating with vocabulary acquisition rejected.

DISCUSSION

9.0 Discussion

Unlike what was expected, the main findings of the present study did not confirm the hypothesis claimed at the beginning of this paper. The following chapters seek to discuss the main questions arising from the study findings in a more detailed manner as well as to continue the discussion of the benefits of musical activities implemented to foreign language learning.

9.1 Impact of musicality and singing performance on vocabulary acquisition

In the theoretical part of the present paper, the existing evidence of the link between musical aspects (i.e., musical aptitude, singing performance) and L2 acquisition has been already presented. Multiple studies point out significant correlations between musical skills and L2 proficiency, especially the domain of phonological and phonetic has been widely researched so far (Milovanov et al. 2008, Christiner & Reiterer 2013, 2015). However, other domains of foreign language acquisition, such as vocabulary acquisition, have not received such vast attention and the few research findings in respect to this thematic have been partly contradictory.

The results of the present study clearly point out that musical aptitude, as well as singing performance, had no effect on the performance in the vocabulary learning tasks in the frame of the current study design. These results are not uncommonly surprising, as former studies have also received similar conclusions, such as the study conducted by Baills et al. (2021). Their study examined the hypothesized positive effect of singing in training activities in foreign language educational programs and tested the effectiveness of song-based training for pronunciation and vocabulary learning. Whereas the results showed a clear link between singing and pronunciation, no advantage was found in terms of word learning.

However, other studies have indeed found supporting evidence of songs and singing on the text and vocabulary memorization (Rukholm 2011, Ludke et al. 2014, Good et al. 2015, Busse, Jungclaus, Roden, Russo, & Kreutz, 2018), whereas the present experiment did not find any significant difference in vocabulary learning between singing to the song and listening to a song. The lack of a beneficial effect of the singing condition might have been partly due to the way the experiment was forced to be conducted due to the restrictions caused by the COVID-19 pandemic, which is discussed more closely in the next chapter.

9.2 The effect of COVID-19 restrictions on the study

The COVID-19 pandemic affected the executed study in various ways. Due to the safety measures set up by the Austrian government, physical interaction was recommended to be limited to the minimum which ruled out the original plan to carry out a study where the participants would have been tested concurrently divided into the corresponding condition groups. The original concept was to replicate a situation where the training phase would be similar to the training phases in choir rehearsals, or alternatively situation resembling a learning situation in a music classroom. Due to the COVID-19 restrictions, the study design had to be changed remarkably, since each experiment had to be administered online via Zoom, and instead of testing groups, each participant was tested individually. In the margins of the restrictions of that time, it was also self-evident that an in-person study that would have included singing in a group with dozens of participants was impossible to be carried out.

The original idea to test participants in groups did arise from the findings in the literature, that choir singing has been associated with better mental and physical outcomes referring to the effects of choral singing: increase of positive feelings and reduction of negative emotional states (Caetano, Ferreira, & Mariotto, 2017). Whereas the increase of positive feelings promoted by choir singing is linked to factors such as improvement in quality of life, vitality, and cognitive stimulation, the reduction of negative emotional states is connected to a decrease in stress, anxiety, and depressive symptoms (Clift, Hancox, Morrison, & Hess, 2010). Furthermore, there is evidence of stronger positive effects of singing, when singing occurs in a group as opposed to alone. Schladt, et al. (2017) found out, that 20 min of either choir or solo singing is sufficient to increase happiness and decrease sadness and worry, but the positive effects of singing are more pronounced after choir singing compared with solo singing in the same participants.

Moreover, the main idea behind the testing groups came forth of the feedback provided by the participants of the pilot study, who reported having been suffering from music performance anxiety during the experiment. Music performance anxiety refers to the experience of marked and persistent anxious apprehension related to musical performance, that arises through underlying biological or psychological vulnerabilities (Kenny, 2011, Rowland & van Lankveld, 2019). Causes for music performance anxiety found in the study by Kenny (2011) are *pressure from self, inadequate preparation for performance, general lack of self-confidence, bad performance experience,* and *concern about audience reaction*, to name a few. Since the participants of the pilot study previously reported both the singing score part as well as the experiment part being stressful through the performance anxiety, this obstacle was planned to be solved in the full-scale study by creating a situation, where the participants would not have to sing alone, but in a group. The desired setting would have been one, where the participants focus on the task itself, learning the melody with the new text, instead of concentrating too much on the singing itself. Besides, the idea of holding the training phase in a group would have decreased the so-called *observer bias*, which refers to the situation, where a subject knows that they are being observed, which could cause them to act differently and therefore interfere with the experiment results.

Finally, when it was clear that the study had to be executed virtually, all possible measurements were done to establish a situation, where each subject would feel comfortable attending the singing condition group. For instance, it was made clear to the participants, that they don't need to keep their microphones or web cameras open during the learning phase. However, establishing a complementary study design where the subjects are tested in a group setting would possibly change the results of future studies.

9.3 Differences between the condition groups

When having a closer observation of the differences in the vocabulary tasks between the condition group, a modest tendency for the singing condition group to score better can be seen. Although there is no statistical significance, it is worthy of examining. Furthermore, the control group achieved the lowest scores, which would also speak for the benefit of musical input, which has been confirmed in the studies of Rukholm (2011) and Schön et al. (2007). For instance, in Rukholm's study, the treatment groups were exposed to a song either in a sung condition or read as a poem, without musical input. Participants exposed to the sung condition obtained higher scores than participants exposed to lyrics only. However, the contradictory results of the present paper added to the previous research findings would call for a more uniform experiment format in future studies.

As Baill et al. (2021) state, even though educators and researchers do agree on the value of singing in training activities in foreign language educational programs, there is a certain degree of novelty in the study designs teasing out the effects of singing. For future research orientations, strict control of factors that are known to influence results involving songs should be kept in mind:

1) Appropriateness of the materials in terms of the pedagogical value

In the past years, many different kinds of training materials have been implemented in the studies regarding the singing performance. In the study of Baill et al. (2021), audio-visual training materials were used where the participants learned the target words via training video where the instructor pronounced the words while performing a deictic gesture by means of pointing to the corresponding body part. The present study used similarly audio-visual material, but instead of training video and deictic gestures, the song was combined with target words presented in a PowerPoint presentation.

2) Degree of difficulty in terms of both language and melody

As discussed in the preceding sections, there is strong evidence stating that the benefits of a song increase, if the melody is already familiar to the listeners. If the melody is too difficult or novel, it might operate more as a distraction than a stimulant for the learning process (Good et al. 2015). When comparing the experiment designs of studies with results supporting the song benefit, it seems that the melodies have been very short such as in the study of Ludke et al. (2014), where the stimuli for the learning sessions consisted of short, 1-2 measures long rather simple melodic phrases. Alternatively, a melodic repetition can facilitate the retention of text, as used in the study of Rukholm (2011). In the study, the stimulus song was selected due to the fact it contains both verses and a refrain whose respective melodies are both simple and repeated throughout the song. In the experiment of this thesis, a famous nursery rhyme “Twinkle twinkle little star” was applied to ensure a moderate degree of difficulty.

In terms of the degree of difficulty regarding the linguistic input, it was important to select target words that would not be familiar to the participants in beforehand. Since subjects were controlled for not having knowledge of Finnish, which is also relatively distinct from subjects’ first language German, it was a considerably good choice for target language. Additionally, target words were body parts, which in terms of pedagogical value are easy to implement into the training material as shown by Baill et al. (2021). Moreover, the words were mostly two-syllable words.

3) Rhythmic characteristics

The discussion about the main reason behind the music/song advantage also speculates, whether it is the component of melody, rhythm, or number of repetitions that aids word retention. Therefore, the rhythmic characteristics of each condition group should be carefully controlled in future experiments. In the present experiment, the singing and listening condition groups did receive the exact same input in the form of song exposure, whereas the spoken input for the control group was not controlled in the terms of presentation rate or the rhythmic abilities.

4) Exposure time of the materials across the conditions

One obstacle to ensure equal training settings between the treatment groups is the exposure time of materials and the presentation rate, which is more closely discussed in the next chapter.

Table 17: Experiment results, Word scoring

Results Experiment – Word Scoring	M	MD	SD	Min	Max
Singing condition group	6.00	6.00	1.648	4	10
Listening condition group	5.60	6.00	1.724	2	9
Control group	5.13	5.00	1.959	2	9

Table 18: Experiment results, Phoneme scoring

Results Experiment – Phoneme Scoring	M	MD	SD	Min	Max
Singing condition group	23.60	24.00	8.567	13	46
Listening condition group	20.33	19.00	8.121	7	39
Control group	19.73	15.00	10.215	7	40

When comparing the results of tests measuring musicality and the singing ability, the singing condition group scores the best in both tests. As discussed previously, it is reasonable, that individuals with high musical aptitude also acquire better singing skills and probably have pursued musical activities. What could be counted as a limitation of this experiment is that through the recruitment of participants, individuals with a known history of singing experience were more likely to end up in the singing condition group.

Although the present study did not explicitly focus on the measurement of the pronunciation skills of participants, the closer inspection of phoneme scoring results combined with the singing score results shows that the singing condition group scored the highest. In comparison to the word scoring evaluation, where the words were accepted even if they weren't pronounced correctly, in the phoneme scoring evaluation the correct pronunciation did play a role. This would therefore support the growing evidence of singing facilitating imitation skills (Christiner & Reiterer, 2015) since the participants of the singing condition group also outperformed other participants in the singing score task.

Another interesting detail regarding the recruitment of participants and the results of musical aptitude scores is the dilemma of what musicality really comprises. As represented in the theoretical part, musical ability might remain undiscovered in individuals, who have not been practicing musical activities. According to Zentner & Strauss (2017), these individuals are referred to as *musical sleepers*. Subsequently, individuals whose musical proficiency languishes due to lack of practice, are referred to as *sleeping musicians*. During the recruitment phase of the current study, various participants claimed beforehand of not being musical or not being able to sing. Interestingly, once the musical aptitude test and singing score were performed, many of these self-defined "non-musicians" received relatively good results. This leads to the assumption, that it would have been worthwhile to ask participants to rate their musical aptitude in a form of self-evaluation before running the musical aptitude tests, for instance in the frame of the questionnaire.

Table 19: Results musical aptitude

Results Musicality	Mean	Median	Min	Max
Singing condition group	25,56	26	19,5	30
Listening condition group	22,76	22,5	15,5	32
Control group	21,80	22	13	28

Table 20: Results Singing score

Results Singing score	M	MD	Min	Max
Singing condition group	48.33	48.00	41	60
Listening condition group	36.20	39.00	23	49
Control group	39.27	41.00	20	52

9.4 Duration rate and number of repetitions of stimuli

One factor that has been linked to the debate on the effects of singing on the word recall is the duration of the learning material. As mentioned earlier, sung stimuli have significantly slower presentation rates than spoken stimuli. Among others, Kilgour et al. (2000) studied the effect of relative duration to the word recall. Even though the subjects in the sung stimuli condition scored better than the group of spoken stimuli, the advantage of recall for sung lyrics vanished when the presentation rate of sung & spoken material was equated. Furthermore, Wallace (1994) has argued that exposure to the melody of songs might increase the cognitive burden of the participant, but a sufficient number of repetitions seems to enhance the memorization process. Additionally, in the study of Calvert and Tart (1993) sung and spoken presentations of text were compared with the result of similar text recall scores between the condition groups when the materials were presented only once, but once the material was listened to repeatedly over weeks, the group who listened to the song received significantly higher scores than the group hearing the spoken text.

In the present study, the duration of presented stimuli was the same in the two groups receiving musical input (singing and listening condition groups), whereas the spoken stimuli presented to the control group were not controlled in respect of the duration. This would resemble the findings from Kilgour et al. (2000) of song advantage vanishing when the duration rate of the main target groups was controlled. Additionally, it could be speculated that also the control group would have scored better, if the duration of the presented stimuli for that group would have been adjusted to equate the singing stimuli.

Noteworthy is also to compare the present experiment to the study from Ludke et al. (2014), where the experiment design was similar to the one presented in this paper. In their study, the singing condition group showed superior overall performance compared with participants in the speaking and rhythmic speaking conditions. Indifference to the experiment design of the recent paper, the participants of Ludke's study underwent a 15-min "listen-and-repeat" learning period, separated into three 5-min learning sessions. The subjects had to practice 20 English-Hungarian paired-associate phrases one after another by first listening to the spoken English phrase, and then listening twice to the paired Hungarian phrase and repeating the Hungarian phrase aloud. In this experiment, the stimuli in the three learning conditions were controlled for overall duration and presentation rate. These findings could be interpreted so that maybe the effect of the singing performance is dependent on the training method, since Ludke et al. used a listen-and-repeat singing method using simple, previously unfamiliar melodies. Additionally, the duration of the learning phase was longer in their study, whereas in the present study the participants were exposed to the stimuli for approximately 5-7 minutes. More evidence for the argument of training duration being a crucial factor for the song benefit comes from the study of Coyle & Gomez Gracia (2014), where the effects of song-based activities on the L2 vocabulary acquisition were examined among preschool children. In their study, evidence for singing activities developing children's receptive knowledge of vocabulary was found. Here the subjects received three 30-minute lessons of a well-known children's song, which in general is a much longer presentation time the participants are exposed to than in the present study, and more alike with the study design of Ludke et al. (2014). This would imply for further experiments to implement training phases with longer presentation rates along with multiple repetitions. To illustrate the possible influence of duration rate, comparison of the main papers discussed in the chapter is displayed below:

Table 20: Comparison of song effect research studies

Study	**Duration rate**	**Results**
Present thesis	5-7 minutes	No positive effect
Baille et al. (2021)	5.20 minutes	No positive effect
Ludke et al. (2014)	15 minutes	Positive effect
Coyle & Gómez Gracia (2014)	30 minutes	Positive effect
Rukholm (2011)	30 minutes (repeated in one week)	Positive effect
Good et al. (2015)	4 learning sessions over a 2-week period	Positive effect

9.5 Explicit and implicit language learning

		LLAMAB	LLAMA D	LLAMA F	Mini-Proms
LLAMAB	Pearson Correlation	1	-.042	.296*	.080
	Sig. (2-tailed)		.782	.048	.602
	N	45	45	45	45

The correlation found between the subtests of LLAMA_B and LLAMA_F could be linked to the discussion about implicit and explicit language learning. According to Granena (2020), implicit learning refers to those cognitive abilities, that facilitate learning and processing of an L2 without conscious intention to learn or conscious awareness of the fact that one is learning. In other words, implicit inductive learning is defined as the ability to pick up patterns after experiencing them in the input without having to think about them consciously. Furthermore, implicit memory ability is described as the ability to retain and retrieve information effortlessly and incidentally (Fu & Li, 2021). In comparison to that, explicit learning is conscious and controllable, limited by attentional and working memory resources (Li, 2019). Tests that measure language aptitude, including LLAMA test battery, are often biased toward measuring cognitive abilities in the explicit domain, focusing on attention-driven cognitive processes such as explicit inductive learning ability, rote memory ability, and analytical ability (Granena ,2016). Correspondent to this, LLAMA_B being a vocabulary learning module and LLAMA_F testing the grammar inference skills, would categorize these subtests to the domain of explicit learning. In comparison to these, LLAMA_D is a recognition task, which usually is assumed to involve explicit memory retrieval. Nevertheless, it has been argued by Granena (2020) that recognition can be based on the assessment of *stimulus familiarity*, which is referring to the sensation that an item has been seen before without explicitly remembering any detailed contextual information. Moreover, studies from Wang & Yonelinas (2012) confirm that familiarity-based recognition memory judgments rely on the same process that supports conceptual implicit memory. Therefore, the correlation found between LLAMA_B and LLAMA_F could be explained by that these tests are based on explicit learning, whereas LLAMA_D measures implicit memory.

9.6 Musicality and Grammatical Aptitude

		LLAMAB	LLAMA D	LLAMA F	Mini-Proms	Singing score	Word scoring	Phoneme scoring
LLAMAF	Pearson Correlation	.296*	-.002	1	.300*	.269	.055	.023
	Sig. (2-tailed)	.048	.988		.046	.074	.721	.882
	N	45	45	45	45	45	45	45

Another correlation in the statistical analysis was found between the results of LLAMA_F and Miniproms, which supports the findings from earlier studies pointing out the possible relation between musical skills and grammatical aptitude, such as Gordon, Jacobs, Schuele, & Mcauley (2015), Kalcheva & Fonseca-Mora (2017), and Malzer (2018). For instance, in the study of Kalcheva and Fonseca-Mora (2017), two groups of adult learners were tested to find differences between professional singers and non-musicians. They found out that singing has a beneficial influence on grammar achievement. Additionally, the results of Malzer (2018) showed a strong, significant correlation between the musicality test scores and the grammatical test scores. Generally, studies on grammatical aptitude and musicality emphasize the involvement of similar brain regions in the processing of music and language. Processing music has been contrasted to resemble the processing of syntax information through the shared requirement of pattern detection. As mentioned in the theoretical part, music and language both contain syntactic systems that employ complex, hierarchically structured sequences (Kunert, Willems, Casasanto, & Patel, 2015).

Moreover, according to Flöel, de Vries, Scholz, Breitenstein, & Johansen-Berg (2009), the acquisition of grammar requires the extraction of rule-based information, which similarly is related to the processing of music.

The results of this paper added to the previous findings from Malzer (2018) underline the fact that musicality is not limited to the auditory skills and processing of sounds. Instead, in addition to the auditory parts, musical aptitude relates to both motor skills as well as to the processing of rules and patterns.

9.7 Influence of motivation and emotion to language learning

Another important factor to be discussed when going through the experiment results is the influence and presence of motivation in the language learning context. Motivation, one of the individual factors among language learners affecting language acquisition is a key aspect and prerequisite for successful language learning. As formulated by Reeve (2009), motivation gives behavior its energy and direction. Moreover, a strongly motivated language learner will learn the target language by any method, in comparison to an unmotivated learner, who might fail despite the best teaching method available.

In the recent research papers about motivation, Dörnyei (2003) introduces the concepts of intrinsic and extrinsic motivation, which are not necessarily observed as opposites of each other, but preferably seen as a continuum. Intrinsic motivation refers to a state where an individual achieves a positive affective phase of learning new things and has the motivation to discover the target language and the culture associated with it. In comparison to that, extrinsic motivation refers to a system of rewards and punishments by educators like parents or teachers.

A vast amount of research has been focusing on the various aspects of motivation in L2 acquisition. According to Gardner (1985), four elements must be present for a student to be considered motivated: a goal, a desire to achieve the goal, a positive attitude, and effort. These elements are referred to as *affective variables*, which compared to other, rather cognitive factors like intelligence or aptitude, are more focused on the emotional and mental part of language learning. Emotion, for its part, is closely intertwined with motivation, because emotion functions as an amplifier, providing the intensity, urgency, and energy to propel behavior (Tomkins, 1962). The relation between emotion, motivation, and language learning can easily be seen when emotion is observed more closely: behind human behavior, emotion is continually present varying in type and intensity. We are constantly experiencing some kind of emotion, which subsequently affects the actions we implement. Therefore, depending on the intensity of the emotion we experience, the greater is the impact on our activities.

Besides, emotions often carry impulses to act with them, according to the appropriate emotion, and these impulses to act might highly be appropriate or inappropriate for social functioning, including L2 interaction (Parkinson, 1995). For example, the feeling of being embarrassed usually produces the urge to withdraw and hide oneself, regardless of the reason of the embarrassment. Especially in foreign language learning, the learners often confront situations where they feel embarrassed due to the learning process, and even though it might be in the learner's best academic, cultural, and social interest to keep talking and learning, the strong emotion affects the actions of the individual.

Similar patterns of behavior can be observed when playing an instrument. The music performance anxiety (MPA) could be one of the reasons causing a strong emotion, which influences the learning process in a negative way. Literature agrees that state anxiety is higher before and during a public performance, i.e., in front of an audience or jury and it causes a wide range of affective, cognitive, somatic and behavioral symptoms (Guyon, Hildebrandt, Horsch, Nater, & Gomez, 2020). This could possibly be one reason affecting the results of the present paper, since the singing condition group members did confront the task of singing alone in a test situation.

Negative emotions may have an unwanted influence on individual's actions, but on the other hand, positive emotions facilitate and boost actions. The fact of music being universal and everybody enjoying music does provide many opportunities for beneficial implementation of music into learning settings. Many studies do confirm the motivational effect of music, such as the research of (Lowe, 1998), where students that learned music concepts and patterns daily along with their language study, scored significantly better in both foreign language study and music study in comparison to those students who did not study music. Similarly, in the study of Medina (2002), second-grade English language learners learned more words when the vocabulary was taught with the help of supplementary music and pictures. The group using only one or neither of the artistic enhancements scored significantly lower.

In the study of Murphey (1992) pop song lyrics were found to be effective in boosting language production. Pop songs contain high-frequency vocabulary referring to words that every learner needs to know plus pop songs have the tendency of handling universal themes that make them easy for people to identify with. As Duff (2002) states, popular culture in, in general, an avenue for constructing identities and it serves as a motivator for learners.

9.8 Pedagogical missions of music

The implementation of music into classroom activities has been one of the methods to boost language learning with the help of music. As previously illustrated, students enjoy music, they open up to it and they feel more relaxed in its presence, which all are factors related to motivation and emotion. The affective filter hypothesis created by Krashen (1982) explains that the most effective learning occurs in a setting of low anxiety and high self-confidence and motivation. Additionally, music lowers affective barriers and assists in making students more receptive to language learning (Engh., 2013).

In the past decades, a growing amount of alternative learning settings has taken place also in the field of language learning. Out-of-class foreign language learning can be conceptualized along a continuum from "self-instruction" to "naturalistic learning". For instance, online communities, language cafes and learning implemented in free time activities might provide different language learning opportunities, which are less hostile contexts than classrooms (Duff, 2002). From this perspective, implementing foreign language learning into a music class through singing could serve as a motivating, fun way to acquire language.

9.9 Effect of learning in a group

Returning to the experiment design of the present paper, the growing evidence of the positive effects of group singing would speak for the group singing activity being the factor aiding the language learning. Since group singing does yield a greater increase in the perception of greater well-being and the development of a strong sense of belonging, it simultaneously increases the level of motivation (Launay, Tarr, & Dunbar, 2016). Additionally, both singing synchronously in unison and singing in coordinated turns has been linked to increased feelings of belongingness to a group (Koudenbourg, Postmes, Gordjin, & van Mourik Broekman, 2015).

In the present study, the subjects were tested individually, which is a completely different setting than learning a song in a group. For future directions, it would be interesting to test whether there are differences in the results between different learning settings, such as individuals vs. group testing.

An overlapping topic to the motivation discussion is the influence of group performance. As already mentioned, whereas music and singing both have lots of positive effects on well-being and motivation, group singing does outscore solo singing in the benefits. However, the crucial factor behind the success of group singing might be the membership of the group instead of the singing activity. Stewart & Lonsdale (2016) compared choral singing to other relevant leisure activities, solo singing and playing a team sport to measure the self-reported well-being, entitativity, need fulfilment and motivation. According to the results, being part of a group might be a more important influence on the psychological well-being experienced by choral singers than the singing activity itself. Additionally, in the study of Launay et al. (2016) found no differing effects regarding well-being between group singing and other types of group activities. This would lead to the assumption, that the precise mechanisms behind the positive effects of group singing are based on the community cohesion instead of the musical aspect. Also, when comparing the recent papers in respect to the participant setting, the majority of those that received positive effects exposed the participants to the stimuli in a group setting. This, once again, encourages for further work to refine an understanding of what exactly the key aspect behind the song advantage really is.

10 Melodic intonation therapy

Another domain where music and melodies have been proven to help, is the treatment of aphasia patients. *Melodic Intonation Therapy* (MIT) refers to a rehabilitation program designed for aphasia patients, where high probability phrases and sentences are intoned to be used with patients with severe expressive deficits (Norton., 2009).

The origins of MIT are based on the observations of aphasia patients, where singing with lyrics is preserved whereas the same patients face difficulties when pronouncing the words otherwise. These observations raise questions about the representation of music and language functions in the brain, whether speech and singing are independent functions. According to the research findings, singing would be predominant in the right-hemisphere, and speech, on the contrary, is located in the left hemisphere. One common explanation for the preservation of singing is that song is a form of *nonpropositional speech,* which refers to conventionalized, context-dependent speech that consists of “ready-made” utterances which in a way can express emotion but are “intellectually dead” (Parsons, 2006). In comparison to that, normal speech is categorized as *propositional speech*, which refers to an intentional expression of thoughts. This means, singing songs does not involve syntactic parsing or the conscious formulation of new utterances to express a semantic message.

The effect of Melodic Intonation Therapy bases on the claims that melody activates right hemisphere language structures that were formerly dominated by the left hemisphere and that MIT increases the role of right hemisphere in interhemispheric language control. Also, the mechanism of MIT facilitates speech at the articulatory level, where the beneficial effect of singing works through the use of a slower tempo, a more precise rhythm and greater accentuation than in normal speech – the same factors that have been observed being beneficial in language learning tasks. The slower tempo also might lessen the neuromotor complexity of the articulatory process, as evidenced by causing positive outcomes on the treatment of stuttering (Hough., 2010). In addition to that, the syllable lengthening has been proved to facilitate phrase production in nonfluent aphasics while using MIT (Parsons, 2006).

11 What is behind the effect of music?

Returning to the very core of the thematic discussion regarding what really is the reason behind the benefits of combining music with foreign language learning, there is still no clear-cut answer to what exactly is the secret ingredient that aids language learning. Relying on the findings of the present study, it seems that singing, although beneficial with respect to pronunciation and motivational aspects of classroom activities, has nothing or very little to do with vocabulary acquisition.

According to studies conducted by Ludke et al. (2014) and Stahl, Kotz, Henseler, Turner, & Geyer, (2014) instead of the effect of singing affecting the word recall, the main component most crucial for learning and memory aid might be the melodic component, especially when the melody consists of a rather simple pattern of sounds. Another candidate for the beneficial effect of the listening to song and singing conditions could be the exposure time to the materials presented, since the number of repetitions does support the word recall.

Additionally, even though it is easily thought that singing a song would be more beneficial for L2 recall of vocabulary, one must keep in mind the individual differences of language learners: this generalization might hold for experienced singers or musicians, but participants who are not used to singing might suffer from music performance anxiety which decreases simultaneously the motivation to learn and distracts the learning process.

Moreover, the effect of motivation influencing the learning situation is worth of consideration. It is a challenging task for educational institutions to develop methods that implement musical activities in the sense where it increases the motivation of the language learners. Also, the effect of group activity seems to be another crucial factor playing a role in both the psychological wellbeing and motivation of language learners.

12 Limitations and Future directions

In general, the results of this study provide evidence for that even though music and language have been linked to each other and positive evidence of the effects of singing have been found to aid language learning, the domain of the impact is not vocabulary learning. Some minor correlations were found to support the link between musicality and grammatical inference, and similarly the findings between the subtests of language aptitude tests provide more evidence to the discussion about the implicit and explicit learning, as well as the factors of motivation and effect of group activity. However, some limitations should be noted.

First, a delayed post-test could have been carried out to find out whether the effects were persistent over time. Second, the whole sample suffers from being biased due to the fact that all participants were people who were willing to give a singing performance. Many potential participants were left out of the study, since they were not ready to record the singing score. Similarly, subjects in the singing condition group were partly individuals who were experienced singers and therefore, a more equal distribution of subjects as well as wider recruitment of participants would have been beneficial.

Third, the experiment setting might have had a remarkable effect on the results. The possible effect of testing groups instead of individuals as well as establishing a test situation in person instead of virtually could be tested in future studies. Also, it would be important to stick to a pattern of controlled experiment procedure with respect to the training materials. Especially the duration of the learning phase does encourage for further work. In addition to that, the spoken stimuli received by the control group could have been controlled regarding the duration and the rhythmicity.

Finally, a larger sample size might have allowed to detect the effects of individual measures more relevantly. It would also have been interesting to simultaneously test both the effect of song to the pronunciation and grammar aptitude, since the evidence for those already exists.

13 Conclusion

The present study confirms the well-known truth of music being meaningful and important for human beings. Even individuals, who do not identify themselves musical, enjoy listening to music, as ratified in the questionnaire of the participants – each respondent answered yes to the question, whether they listen to music. Also, the present study does confirm the beneficial psychological effects of singing, as the narrations of the participants of their singing background verify:

"*It's the thing I love most and always was my main musical focus"*

…

"*Singing is just a normal thing to me, something I do all day long."*

...

"I have always loved singing"

…

Meaningful melodic training combined with lyrics might be an asset for language learning settings, although it seems to be more advantageous in the domains of phonetic and phonological abilities and grammar inference, more than in vocabulary recall. When incorporating music and songs into classroom activities, language teachers and educators should consider learners' preferences and the ideal setting for the learning situation, as well as the level of difficulty of the song including the presentation rate.

Bibliography

Abril, C. R. (2007). I have a voice but I just can't sing: A narrative investigation of singing and social anxiety. *Music Education Research 9*, 1-15.

Ackermann, H., & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: A review and a hypothesis. *Brain and Language 89(2)*, 320-8.

Amunts, K., Schlaug , G., Jäncke, L., Steinmetz , H., Schleicher, A., Dabringhaus, A., & Zilles, K. (1997). Motor cortex and hand motor skills: structural compliance in the human brain. *Hum Brain Mapp*, 206-15.

Arleo, A. (2000). Music, Song and Foreign Language Teaching. *Cahiers de l´APLIUT 19(4)*, 5-19.

Ashida, S. (2000). The effect of reminiscence music therapy sessions on changes in depressive symptoms in elderly persons with dementia. *J Music Ther. 37(3)*, 170-82.

Ayotte, J., Peretz, I., Rousseau, I., Bard, C., & Bojanowski, M. (2000). Patterns of music agnosia associated with middle cerebral artery infarcts. *Brain: A Journal of Neurology, 123(9)*, 1926-1938.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation 8*, 47-89.

Baills, F., Zhang, Y., Cheng, Y., Bu, Y., & Prieto, P. (2021). Listening to Songs and Singing Benefitted Initial Stages of Second Language Pronunciation but Not Recall of Word Meaning. *Language Learning 71(2)*, 369-413.

Bengtsson, S. L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., & Ullén, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience*, 1148-1150.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *Front. Psychol. 2*.

Billings, C., Penman, T., Ellis, E., Baltzell, L., & McMillan, G. (2016). Phoneme and Word Scoring in Speech-in-Noise Audiometry. *Am J Audiol 25(1)*, 75-83.

Binns, H. (2016). Music: Why it affects us, how society uses it, and how this knowledge may benefit educators. In M. C. Fonseca-Mora, & M. Gant, *Melodies, Rhythm and Cognition in Foreign Language Learning* (pp. 7-23). Cambridge: Cambridge Scholars Publishing.

Bokander, L., & Bylund, E. (2019). Probing the Internal Validity of the LLAMA Language Aptitude Tests. *Language Learning 70(1)*, 11-47.

Brandstrom, S. (1999). Music Teachers' Everyday Conceptions of Musicality. *Bulletin of the Council for Research in Music Education, n141*, 21-25.

Busse, V., Jungclaus, J., Roden, I., Russo, F. A., & Kreutz, G. (2018). Combining Song - And Speech-Based Language Teaching: An Intervention With Recently Migrated Children. *Frontiers in Psychology 9*.

Caetano, K. A., Ferreira, I. M., & Mariotto, L. G. (2017). Choir singing as an activity to manage anxiety and temporomandibular disorders: Reports from a Brazilian sample. *Psychology of Music*, 96-108.

Callan, D., Callan, A., & Jones, J. A. (2014). Callan, D., Callan, A., & Jones, J. A. (2014). Speech motor brain regions are differentially recruited during perception of native and foreign-accented phonemes for first and second language listeners. . *Frontiers in neuroscience 8*.

Callan, D., Tsytsarev, V., Hanakawa, T., & Callan, A. (2006). Song and speech: Brain regions involved with perception and covert production. *NeuroImage 31(3)*, 1327-42.

Calvert, S. L., & Tart, M. (1993). Song versus verbal forms for very-long-term, long-term, and short-term verbatim recall. *Journal of Applied Developmental Psychology, 14(2)*, 245-260.

Carroll. (n.d.).

Carroll. (1958). A factor analysis of two foreign language aptitude batteries. *The Journal of General Psychology, 59(1)*, 3-19.

Carroll, J. B. (1962). The prediction of success in intensive of foreign language training. In R. Glaser, *Training and research in education* (pp. 87-136). Pittsburgh: University of Pittsburgh Press.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller, *Individual Differences and Universals in Language Learning Aptitude* (pp. 83-118). Rowley, MA: Newbury House.
Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. S. Parry, & C. W. Stansfield, *Language aptitude reconsidered* (pp. 11-27). Prentice Hall: Prentice Hall Regents.
Christiner, M., & Reiterer, S. (2013). Song and speech: examining the link between singing talent and speech imitation ability. *Frontiers in psychology, 4*, 874.
Christiner, M., & Reiterer, S. (2015). A Mozart is not a Pavarotti: singers outperform instrumentalists on foreign accent imitation. *Front. Hum. Neurosci*, 482.
Clift, S., Hancox, G., Morrison, I., & Hess, B. (2010). Choral singing and psychological wellbeing: Quantitative and qualitative findings from English choirs in a cross-national survey. *Journal of Applied Arts and Health 1(1)*.
Coyle, Y., & Gomez Gracia, R. (2014). Using songs to enhance L2 vocabulary acquisition in preschool children. *ELT Journal 68(3)*, 276-285.
Damstra-Wijmenga, S. M. (2009). The memory of the new-born baby. *Midwives Chronicle, 104*, 66-69.
Darwin, C. R. (1872). *The expression of the emotions in man and animals.* London: John Murray.
Davidson, J. W. (1993). Visual Perception of Performance Manner in the Movements of Solo Musicians. *Psychology of Music*, 103-113.
Davis, C., Kleinman, J. T., Newhart, M., Gingis, L., Pawlak, M., & Hillis, A. E. (2008). Speech and language functions that require a functioning Broca's area. *Brain Lang.*, 50-8.
Dörnyei. (2003). Attitudes, Orientations, and Motivations in Language Learning: Advances in Theory, Research, and Applications. *Language Learning, 53(S1)*, 3-32.
Dörnyei, Z., & Skehan, P. (2003). *The Handbook of Second Language Acquisition.*
Duff. (2002). Pop Culture and ESL Students: Intertetuality, Identity, and Participation in Classroom Discussions. *Journal of Adolescent & Adult Literacy, 45(6)*, 482-487.
Dunbar, R. (2004). *The Human Story.* London: Faber and Faber.
Elbert, T., Pantev, C., Wienbruch, C., Rockstsroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science*, 305-7.
Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 533-558.
Engh. (2013). Why use music in English language learning? a survey of the literature. *English Language Teaching (Toronto), 6(2)*, 113-127.
Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The Parahippocampal Place Area: Recognition, Navigation, or Encoding? *Neuron 23(1)*, 115-125.
Failoni, J. W. (1993). Music as Means To Enhance Cultural Awareness and Literacy in the Foreign Language Classroom. *Mid-Atlantic Journal of Foreign Language Pedagogy*, 97-108.
Falk, S. (2011). Melodic versus intonational coding of communicative functions: A comparison of tonal contours in infant-directed song and speech. *Psychomusicology: Music, Mind and Brain, 21(1-2)*, 54-68.
Fernald. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? *Child Development, 60(6)*, 1497-1510.
Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papousek, U. Jürgens, & M. Papousek, *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 262-282). Cambridge: Cambridge University Press.
Fernald, A. (1993). Approval and Disapproval: Infant Responsiveness to Vocal Affect in Familiar and Unfamiliar Languages. *Child Development 64(3)*, 657-674.
Flöel, A., de Vries, M., Scholz, J., Breitenstein, C., & Johansen-Berg, H. (2009). White matter integrity in the vicinity of Broca’s area predicts grammar learning success. *NeuroImage 47(4)*, 1974-1981.
Fonseca-Mora, C. (2000). Foreign language acquisition and melody singing. *ELT Journal 54*.
Fonseca-Mora, C. M. (2002). The Role of Musicality of Language in the Acquisition Process of English as a Second Language. Universidad de Michigan.

Fonseca-Mora, C., Toscano-Fuentes, C., & Wermke, K. (2011). Melodies that Help: The Relation between Language Aptitude and Musical Intelligence. *International Journal of English Studies, Vol.22*, 101-118.
Fraisse, P. (1982). Rhythm and Tempo. In D. Deutsch, *Psychology of Music* (pp. 149-180). Academic Press.
Friederici, A. D. (2017). *Language in Our Brain.* Cambridge: The MIT Press.
Fries, W., & Swilhart, A. A. (1990). Disturbance of rhythm sense following right hemisphere damage. *Neuropsychologia 28*, 1317-1323.
Fu, M., & Li, S. (2021). The associations between implicit and explicit language aptitude and the effects of timing of corrective feedback. *Studies in Second Language Acquisition 43*.
Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior, 26(4)*, 331-362.
Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century.* Basic Books.
Gardner, R. (1985). *Social psychology and second language learning: The role of attitudes and motivation.* London: Edward Arnold.
Gaser, C., & Schlaug, G. (2003). Brain Structures Differ between Musicians and Non-Musicians. *Journal of Neuroscience*, 9240-9245.
Ginsborg, J., & Sloboda, J. A. (2007). Singers' recall for the words and melody of a new, unaccompanied song. *Psychology of Music 35(3)*, 421-440.
Good, A., Russo, F., & Sullivan, J. (2015). The efficacy of singing in foreign-language learning. *Psychology of Music 43(5)*, 627-640.
Gordon, E. (1997). *A music learning theory for newborn and young children.* Chicago: GIA Publications.
Gordon, E. E. (1979). Developmental Music Aptitude as Measured by the Primary Measures of Music Audiation. *Psychology of Music*, 42-49.
Gordon, E. E. (1998). *Introduction to research and the psychology of music.* Chicago: GIA Publications.
Gordon, E. E. (2004). *Continuing studies in music aptitudes.* GIA Publications: Chicago.
Gordon, R. L., Jacobs, M. S., Schuele, C. M., & Mcauley, J. D. (2015). Perspectives on the rhythm-grammar link ans its implications for typical and atypical language development. *Annals of the New York Academy of Sciences, 1337(1)*, 16-25.
Gracco, V. L., Tremblay, P., & Pike, B. (2005). Imaging speech production using fMRI. *NeuroImage 26(1)*, 294-301.
Granena, G. (2016). Cognitive aptitudes for implicit and explicit learning and informationprocessing styles: An individual differences study. *Applied Psycholinguistics 37*, 577-600.
Granena, G. (2020). *Implicit Language Aptitude.*
Grape, C., Sandgren, M., Hansson, L., Ericson, M., & Theorell, T. (2003). Does singing promote well-being?: An empirical study of professional and amateur singers during a singing lesson. *Integr Physiol Behav Sci. 38(1)*, 65-74.
Green, L. (2008). Music, Informal Learning and the School: A New Classroom Pedagogy.
Grigorenko, E., Sternberg, R., & Ehrman, M. (2000). A Theory-Based Approach to the Measurement of Foreign Language Learning Ability: The Canal-F Theory and Test. *The Modern Language Journal*, 390-405.
Guyon, S. R., Hildebrandt, H., Horsch, A., Nater, U. M., & Gomez, P. (2020). Music performance anxiety from the challenge and threat perspective: Psychophysiological and performance outcomes. *BMC Psychology, 8(1)*, 87-87.
Hallam, S., & Price, J. (1998). Can the Use of Background Music Improve the Behaviour and Academic Performance of Children with Emotional and Behavioural Difficulties? *British-Journal-of-Special-Educations 25(2)*, 88-91.
Hebert, S., Racette, A., Gagnon, L., & Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain 126*, 1838-50.
Herschensohn, J., & Young-Scholten, M. (2012). *The Cambridge Handbook of Second Language Acquisition.* Cambridge: University Press.
Hough. (2010). Melodic Intonation Therapy and aphasia: Another variation on a theme. *Aphasiology, 24(6-8)*, 775-786.

Huron, D. (2001). Is music an evolutionary adaption? *Annals of the New York Academy of Sciences, 930,* 43-61.
Huron, D. (2006). Sweet Anticipation: Music and the Psychology of Expectation.
Hutchinson, J. M., Whiteley, H. E., Smith, C. D., & Connors, L. (2003). The developmental progression of comprehension-related skills in children learning EAL. *Journal of Research in Reading 26(1),* 19-32.
Hyyppä, M. T. (2011). Elinvoimaa yhteisöstä. *Tieteessä tapahtuu 29(8)*, 15-17.
Imfeld, A., Oechslin, M. S., Meyer, M., Loenneker, T., & Jancke, L. (2009). White matter plasticity in the corticospinal tract of musicians: a diffusion tensor imaging study. *Neuroimage*, 600-7.
Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, 101-44.
Ivanov, V. K., & Geake, J. G. (2003). The Mozart Effect and Primary School Children. *Psychology of Music 31(4)*, 405-413.
Jaffurs. (2004). The impact of informal music learning practices in the classroom, or how I learned how to teach from a garage band. *International Journal of Music Education*, 189-200.
James, D., Spencer, C., & Stepsis, B. W. (2002). Fetal learning: A prospective randomized controlled study. *Ultrasound in Obstetrics and Gynecology, 20,* 431-438.
Johnsrude, I. S., Owen, A. M., White, N. M., Zhao, V., & Bohbot, V. (2000). Impaired Preference Conditioning after Anterior Temporal Lobe Resection in Humans. *Journal of Neuroscience*, 2649-2656.
Jorgensen, E. R. (2008). The art of teaching music. 1-344.
Jusczyk, & Aslin, R. (1995). Infants′ Detection of the Sound Patterns of Words in Fluent Speech. *Cognitive Psychology 29 (1)*, 1-23.
Jusczyk, P. W. (1999). The beginnings of word segmentation in english-learning infants. *Cogn Psychol. 39(3-4),* 159-207.
Kalcheva, I., & Fonseca-Mora, C. (2017). Musical Experience in Singing and Grammar Achievement in English Language Acquisition.
Kämpfe, J., Sedlmeier, & Renkewitz, F. (2011). The impact of background music on adult listener: A meta-analysis. *Psychology of Music. 39,* 424.
Kang, & Williamson, V. J. (2014). Background music can aid second language learning. *Psychology of Music, 42 (5)*, 728-747.
Karlsson, G. (2018). *Finnish: A Comprehensive Grammar.* Abingdon: Routledge.
Keenan, R. J., Freymann, D. M., Stroud, R. M., & Walter, P. (2001). The signal recognition particle. *Annu Rev Biochem*, 755-75.
Kemler Nelson, D. G., Hirsh Pasek, K., & Jusczyk, P. W. (1989). How the Prosodic Cues in Mothers MIght Assist Language Learning. *Journal of Child Language 16,* 55-68.
Kennedy, M. (1980). *The Concise Oxford Dictionary of Music.* Oxford: Oxford University Press.
Kenny, D. (2011). Defining music performance anxiety.
Kilgour, A., Jakobson, L. S., & Cuddy, L. L. (2000). Music training and rate of presentation as mediator of text and song recall. *Memory & Cognition 28(5),* 700-10.
Kisilevsky, B., Hains, S., Lee, K., Xie, X., Huang, H., Ye, H., . . . Wang, Z. (2003). Effects of Experience on Fetal Voice Recognition. *Psychological Science 14,* 220-4.
Koelsch, S., Fritz, T., Cramon, D. Y., Müller, K., & Friederici, A. D. (2006). Investigating emotion with music: an fMRI study. *Hum Brain Mapp. 27(3)*, 239-50.
Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nat Neurosci. 7(3),* 302-7.
Koudenbourg, N., Postmes, T., Gordjin, E. H., & van Mourik Broekman, A. (2015). Uniform and complementary social interaction: distinct pathways to solidarity. *PLoS ONE, 10(6)*.
Krashen, S. (1982). *Principles and Practice in Second Language Learning.* Oxford: Pergamon.
Kunert, R., Willems, R. M., Casasanto, D., & Patel, A. D. (2015). Music and Language Syntax Interact in Broca's Area: An fMRI Study. *PloS ONE 10(11)*.
Laufer, B., & Hulstijn, J. (2001). Incidental Vocabulary Acquisition in a Second Language: The Construct of Task-Induced Involvement. *Applied Linguistics 22,* 1-26.

Launay, J., Tarr, B., & Dunbar, R. (2016). Synchrony as an Adaptive Mechanism for Large-Scale Human Social Bonding. *Ethology 122(10)*.

Law, L. N., & Zentner , M. (2012). M. Assessing musical abilities objectively: construction and validation of the profile of music perception skills. *PLoS One*.

Lecaneut , J. P., & Granier-Deferre , C. (1993). Speech stimuli in the fetal environment. In B. De Boysson-Bardies, S. de Schonen , P. Jusczyk, P. MacNeilage, & J. Morton, *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 237-248). Norwell: Kluwer Academic Publishing.

Legg, R. (2009). Using Music to Accelerate Language Learning: An Experimental Study. *Research in Education 82*, 1-12.

Levitin, D. (2012). What does it mean to be musical? *Neuron*, 633-7.

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *Neuroimage*, 2142-52.

Li, S. F. (2019). Six decades of language aptitude research. In Z. Wen, P. Skehan, A. Biedron, S. Li, & R. L. Sparks, *Language aptitude: Advancing theory, testing, research and* (pp. 78-96). Routledge Taylor & Francis Group.

Liégeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., & Chauvel , P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain*, 1853-67.

Liu, M., Hu, X., & Schedl, M. (2019). The relation of culture, socio-economics, and friendship to music preferences: A large-scale, cross-country study. *PLOS ONE 14(2)*.

Lowe. (1998). Teaching music and second languages: Methods of integration and implications for learning. *Canadian modern language review, 54(2)*, 218-238.

Luders, E., Toga, A. W., Lepore, N., & Gaser, C. (2009). The underlying anatomical correlates of long-term meditation: Larger hippocampal and frontal volumes of gray matter. *Neuroimage*, 672-678.

Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. *Memory & Cognition 42*, 41-52.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience 4(5)*, 540-5.

Malzer, D. (2018). Making Music and Learning Languages – Musicality and Grammar Aptitude. In S. Reiterer, *Exploring Language Aptitude: Views from Psychology, the Language Sciences, and Cognitive Neuroscience* (pp. 209-228). Springer.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' Cry Melody Is Shaped by Their Native Language. *Current Biology 19(23)*, 1994-1997.

Mark, M., & Reimer, B. (2003). A Philosophy of Music Education: Advancing the Vision. *Music Educators Journal*.

Masataka, N. (2009). The origins of language and the evolution of music: A comparative perspective. *Physics of life reviews 6*, 11-22.

McDermott, J., & Oxenham, A. (2008). Music perception, pitch, and the auditory system. *Neurobiology*, 452-63.

McElhinney, M., & Annett, J. M. (1996). Pattern of efficacy of a musical mnemonic on recaall of familiar words over several presentations. *Perceptual and Motor Skills 82(2)*, 395-400.

McPherson, G. (1993/1994). Evaluating Improvisational Ability of High School Instrumentalists. *Bulletin of the Council for Research in Music Education*, 11-20.

Meara, P. M. (2005). Llama Language Aptitude Tests. Swansea: Lognostics.

Meara, P., Milton, J., & Lorenzo-Dus, N. (2002). *Swansea language aptitude tests (LAT) v2.0* . Swansea: Lognostics.

Medina, S. L. (2002). "Using Music to Enhance Second Language Acquisition: From Theory to Practice". In J. &. Lalas, *Language, Literacy, and Academic Development for English language Learners.* Boston: Pearson Custom Pub.

Miller, G. F. (2000). Evolution of Human Music through Sexual Selection. In N. L. Wallin, B. Merker, & S. Brown, *The Origins of Music* (pp. 329-360). Boston: MIT Press.

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. A., & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: neural and behavioral evidence. *Brain Reseach 1194*, 81-89.

Morley, J. (1991). The Pronunciation Component in Teaching English to Speakers of Other Languages. *TESOL Quarterly 25(3)*, 481-520.
Moyer, A. (2014). Exceptional Outcomes in L2 Phonology: The Critical Factors of Learner Engagement and Self-Regulation. *Applied Linguistics 35*, 418-440.
Murphey. (1992). The Discourse of Pop Songs. *TESOL Quarterly, 26(4)*, 770-774.
Murphey, T. (1990). The song stuck in my head phenomenon: A melodic din in the lad? *System 18(1)*, 53-64.
Nardo, D., & Reiterer, S. (2009). Musicality and phonetic language aptitude. In G. Dogil, & S. Reiterer, *Language talent and brain activity. Trends in applied linguistics* (pp. 213-255). Berlin-NewYork: Mouton de Gryuter.
Norton. (2009). Melodic Intonation Therapy. *Annals of the New York Academy of Sciences, 1169(1)*, 431-436.
Numminen, A. (2005). Laulutaidottomasta kehittyväksi laulajaksi: Tutkimus aikuisen laulutaidon. Helsinki: Sibelius Academy.
Oxenham, A. J. (2012). Pitch Perception. *Journal of Neuroscience*, 13335-13338.
Palmer, C., & Kelly, M. H. (1992). Linguistic prosody and musical meter in song. *Journal of Memory and Language 31(4)*, 525-542.
Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language 30(3)*, 331-347.
Papousek, M., & Hwang, S. F. (1991). Tone and Intonation in Mandarin Babytalk to Presyllabic Infants: Comparison with Registers of Adult Conversation and Foreign Language Instruction. *Applied Psycholinguistics 12(4)*, 481-504.
Parkinson, B. (1995). *Emotion and motivation.* Longman.
Parsons, K. R. (2006). Preserved singing in aphasia: A case study of the efficacy of melodic intonation therapy. *Music perception; 24(1)*, 23-25.
Patel, A. (2010). *Music, language, and the brain.* Oxford: Oxford University Press.
Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annu Rev Psychol.*, 89-114.
Peterson, C. R., & Al-Haik, A. R. (1976). The development of the Defense Language Aptitude Battery (DLAB). *Educational and Psychological Measurement,* 369-380.
Pimsleur, P. (1966). *Pimsleur Language Aptitude Battery.* Rockville, MD: Second Language Testing Foundation.
Pinker, S. (2013). Language as and Adaption to the Cognitive Niche.
Purnell-Webb, P., & Speelman, C. P. (2008). Effects of Music on Memory for Text. *Perceptual and Motor SKills*, 927-957.
Rainey, D. W., & Larsen, J. D. (2002). The effects of familiar melodies on initial learning and long-term memory for unconnected text. *Music Perception 20(2),* 173-186.
Rauscher , F., Shaw, G., & Ky, K. (1995). Listening to Mozart enhances spatial-temporal reasoning: toward a neurophysiological basis. *Neuroscience Letter Feb 6;185(1)*, 44-47.
Reeve, J. (2009). *Understanding motivation and emotion.* Wiley.
Reiterer. (2018). *Exploring language aptitude: views from psychology, the language sciences, and cognitive neuroscience.* Springer.
Reiterer, S., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., . . . Ackermann, H. (2011). Individual differences in audio-vocal speech imitation aptitude in late bilinguals: functional neuro-imaging and brain morphology. *Front.Psychology 2*, 271.
Reymore, L., & Hansen, N. (2020). A Theory of Instrument-Specific Absolute Pitch. *Front Psychol.*
Riecker , A., Ackermann, H., Wildgruber, D., & Dogil, G. (2000). Opposite hemispheric lateralization effects during speaking and singing. *Neuroreport 11(9)*, 1997-2000.
Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics. 25.*, 46-73.
Robinson, P. (2012). Individual Differences, Aptitude Complexes, SLA Processes, and Aptitude Test Development.
Rogers, V., Meara, P., Barnett-Legh, T., & Curry, C. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association 1(1)*, 49-60.

Rowland, D. L., & van Lankveld, J. (2019). Anxiety and Performance in Sex, Sport, and Stage: Identifying Common Ground. *Frontiers in psychology, 10*, 1615.
Rukholm, V. N. (2011). Facilitating Lexical Acquisition in Beginner Learners of Italian through Popular Song.
Rukholm, V. N. (2015). Singing to Speak: An Examination of Adult L2 Learners and Vocabulary Learning through Song. *Italica 92(1),* 171-192.
Rukholm, V. N., Helms-Park, R., Odgaard, E. C., & Smyth, R. (2018). Facilitating Lexical Acquisition in Beginner Learners of Italian through Spoken or Sung Lyrics. *Canadian Modern Language Review/ La Revue canadienne des langues vivantes 74(1)*, 1-31.
Rysiewicz, J. (2008). Measuring Foreign Language Learning Aptitude. Polish Adaption of The Modern Aptitude Test by Carrol and Sapon. *Poznan Studies in Contemporary Linguistics*, 569-595.
Schellenberg, E. G. (2012). Cognitive performance after listening to music: A review of the Mozart effect. In R. A. MacDonald, G. Kreutz, & L. Mitchell, *Music, health, and wellbeing* (pp. 324-338). Oxford: Oxford University Press.
Schellenberg, E. G., Nakata, T., Hunter, P. G., & Tamoto, S. (2007). Exposure to music and cognitive performance: Tests of children and adults. *Psychology of Music 35(1)*, 5-19.
Schladt, T. M., Nordmann, G. C., Emilius, R., Kudielka, B. M., de Jong, T. R., & Neumann, I. D. (2017). Choir versus Solo Singing: Effects on Mood, and Salivary Oxytocin and Cortisol Concentrations. *Frontiers in human neuroscience 11*, 430.
Schlaug, G., Jäncke, L., Huang, Y. X., & Steinmetz, H. (1995). In Vivo Evidence of Structural Brain Asymmetry in Musicians. *Science 267(5198),* 699-701.
Schmithorst, V. J., & Wilke, M. (2002). Differences in white matter architecture between musicians and non-musicians: a diffusion tensor imaging study. *Neurosci Lett*, 57-60.
Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, 975-83.
Schön, D., Gordon, R., Campagne, A., & Magne, C. (2010). Similar cerebral networks in language, music and song perception. *Neuroimage 51(1),* 450-61.
Schön, D., Magne, C., & Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology 41*, 341-349.
Seashore, C. E., Lewis, D., & Saetveit, J. G. (1960). *Seashore measures of musical talents:*. Psychological Corporation.
Seliger, H. (1978). Implications of a Multiple Critical Periods Hypothesis for Second Language Learning. In W. C. Ritchie, *Second Language Acquisition Research. Issues and Implications* (pp. 11-19). New York: Academic Press.
Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics 16(2),* 155-172.
Shuter-Dyson, R. (1999). Musical ability. In D. Deutsch, *The psychology of music* (pp. 627-651). San Diego: CA: Academic Press.
Skehan, P. (1986). Where does Language Aptitude Come From?
Skehan, P. (1991). Individual Differences in Second Language Learning. . *Studies in Second Language Acquisition*, 275-298.
Skehan, P. (2002). Theorising and updating aptitude.
Sloboda. (1985). *The musical mind : the cognitive psychology of music.* Oxford: Oxford University Press.
Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics 25(2)*, 293-321.
Spychiger, M., Patry, J.-L., Lauper, E., Zimmrman , E., & Weber, E. (1995). Does More Music Teaching Lead to a Better Social Climate. In R. Olechowski, & G. Svik, *Experimental Research in Teaching and Learning* (pp. 322-336). Bern: Peter Lang.
Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2014). Rhythm in dis-guise: why singing may not hold the key to recovery from aphasia. *Brain 134,* 3083-3093.
Stern, E., & Neubauer, A. C. (2013). *Intelligenz – Große Unterschiede und ihre Folgen.* München: DVA.
Stewart, L., Walsh, V., Frith, U., & Rothwell, J. (2001). Transcranial magnetic stimulation produces speech arrest but not song arrest. *Annals of the New York Academy of Sciences, 930,* 433-435.

Stewart, N. A., & Lonsdale, A. J. (2016). It's better together: The psychological benefits of singing in a choir. *Psychology of Music 44(6)*, 1240-1254.
Swannell, J. (1992). *The Oxford Modern English Dictionary.* Oxford: Oxford University Press.
Swanner, D. L. (1985). Relationships between musical creativity and selected factors, including personality, motivation, musical aptitude, and cognitive intelligence as measured in third grade children. *Dissertation Abstracts International: Music Education*, 3464.
Tadmor, U., Haspelmath, M., & Taylor, B. (2010). Borrowability and the Notion of Basic Vocabulary. *Diachronica 27(2)*, 226-246.
Tarr, B., Launay, J., & Dunbar, R. I. (2014). Music and social bonding: "self-other" merging and neurohormonal mechanisms. *Front. Psychol.*
Thiessen, E., Hill, E., & Saffran, J. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy 7*.
Thompson, W., Schellenberg, E., & Letnic, A. (2012). Fast and loud background music disrupts reading comprehension. *Psychology of Music 40*, 700-708.
Tomkins, S. (1962). *Affect imagery consciousness: Vol. 1. The positive affects.* New York: Springer.
Tramo, M. J., Shah, G. D., & Braida, L. D. (2002). Functional Role of Auditory Cortex in Frequency Processing and. *J Neuropsychiol 87*, 122-139.
Trehub, S. (2003). The developmental origins of musicality. *Neuroscience 6*, 669-673.
Unwin, M. M., Kenny, D. T., & Davis, P. J. (2002). The effects of group singing on mood. *Psychology of Music*, *30(2)*, 175-185.
Uther, M., Knoll, M. A., & Burnham, D. K. (2007). Do You Speak E-NG-L-I-SH? A Comparison of Foreigner- and Indant-directed Speech. *Speech Communication 49*, 2-7.
Vaughan, E. D. (1977). Misconceptions about Psychology among Introductory Psychology Students. *Teaching of Psychology*, 138-141.
Vinter, & Perruchet, P. (2000). Implicit Learning in Children Is Not Related to Age: Evidence from Drawing Behavior. *Child Development, 71 (5)*, 1223-1240.
Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., & Ostergaard, L. (2006). It don't mean a thing... Keeping the rhythm during polyrhythmic tension, activates language areas (BA47). *Neuroimage*, 832-41.
Wallace, W. T. (1994). Memory for music: Effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20(6)*, 1471-1485.
Wallace, W. T., & Rubin, D. C. (1991). Characteristics and constraints in ballads and their effects on memory. *Discourse Processes 14(2)*, 181-202.
Wang, C., & Kageff, L. (1985). Effects of two modes of teaching on musical creativity of second graders. *MENC Southern Division Conference*.
Wang, W., & Yonelinas, A. P. (2012). Familiarity is related to conceptual implicit memory: An examination of individual differences. *Psychonomic Bulletin & Review, 19(6)*, 1154-1164.
Webster, P. (1983). An assessment of musical creative thinking in young children: Technical report. *Contributions to Symposium/83 Bowling Green Research*, 100-123.
Wermke, K., & Mende, W. (2009). Musical elements in human infants' cries: In the beginning is the melody. *Musicae Scientiae*, *Special Issue*, 151-175.
Wilson, E. O. (2012). *The Social Conquest of the Earth.* New York: Liveright Publishers.
Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychol Sci 20(1)*, 1-5.
Wing, H. D. (1970). Tests of musical ability and appreciation (2nd ed.). *The British Journal of Psychology*.
Winters, K.-L., & Griffin, S. (2014). Singing is a Celebration of Language: Using Music to Enhance Young Children's Vocabularies. *Language and Literacy 16(3)*, 78-91.
Yule, G. (2014). *The study of language.* Cambridge: Cambridge University Press.
Zentner , M., & Strauß, H. (2017). Assessing musical ability quickly and objectively: development and validation of the Short-PROMS and the Mini-PROMS: Assessing musical ability. *Annuals of the New York Academy Sciences*.

Appendix

Basic Questionnaire – Questions

Participation: I agree to participate in Liisi Törmäkangas' research study. The purpose and nature of the study has been explained to me. I am participating voluntarily. I give permission for my data to be used only for purposes of scientific analysis to which only the experimenter or the research team have access. Further analysis of data will strictly be anonymized (codes instead of names are used). I understand that I can withdraw from the study, without consequences, at any time, whether before it starts or while I am participating. I understand that anonymity will be ensured in the write-up by disguising my identity.

Name

Age

E-Mail address

Gender

Education: indicate the fields of your studies (either ongoing or completed)

Provide your highest degree

Nationality: (if double, give both)

Place of birth (town and country)

Place of current residence (town and country)

Mobility: in how many countries have you already lived (minimum stay: 2 months), name the countries

How many languages do you speak (including L1), indicate number of languages

Which languages? Please provide the proficiency level (1=min, 10=maximum, native level) and the age of onset of learning. Please also mark your mother tongue(s) as "L1" and your second languages as "L2". E.g. German, L1, at age 0, proficiency: 10)

Do you have absolute pitch? (the ability to identify or re-create a given musical note without the benefit of a reference tone)

Do you play a musical instrument?

If you play some instruments, which instruments?

At what age did you start playing a musical instrument?

What was the first instrument you played?

What music education did you absolve?

Please indicate how many hours you played the main musical instrument per day and week within the last year

Do you listen to music in your free time?

How much did you listen to music during the last year on average per day?

What music genres do you listen to?

Do you sing?

Please indicate how many hours you sing per week

Please explain in your own words your singing background

Experiment PowerPoint

Danke für das Teilnehmen an
meiner Master Studie!

Info

- Dieses Zoom-Meeting wird aufgezeichnet
- Nach Evaluierung der Ergebnisse wird diese Aufzeichnung gelöscht
- Name und Ergebnisse der Studie werden anonym bearbeitet
- Während des Zoom-Meetings, bitte das Audio eingeschaltet lassen
- Video ist optional
- Hast du Fragen?

Erklärung zum Inhalt der Studie

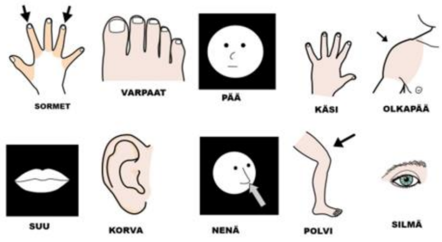
- Diese Studie dauert ca 20 Minuten
- Sie besteht aus 3 Teilen
  - ➢Lernphase mit Bilddarstellung
  - ➢Lernphase mit Tondarstellung
  - ➢Prüfungsphase
- Jeder Teil wird explizit vor Beginn erklärt
- Allfällige Fragen werden beantwortet

Teil 1/3

Teil 1 - Erklärung

- 10 Bilder
- 10 Finnische Wörter
- 1 Minute zum Einprägen

- Wenn du bereit bist, gehen wir zur nächsten Folie und die Zeit beginnt automatisch zu laufen
- Hast du Fragen, bist du bereit?



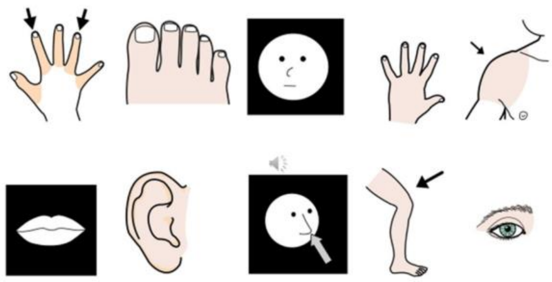
Ende von Teil 1

Teil 2/3

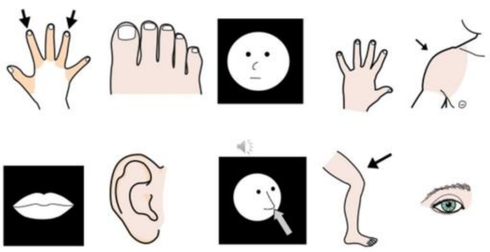
Teil 2 - Erklärung

- 10 Bilder (diesesmal finnische Wörter NICHT sichtbar)
- Lied mit finnischen Wörtern als Text zweimal hintereinander abgespielt
- Danach wird das Lied noch dreimal abgespielt, versuche dabei mitzusingen!

- Wenn du bereit bist, gehen wir zur nächsten Folie und das Lied beginnt
- Hast du Fragen, bist du bereit?

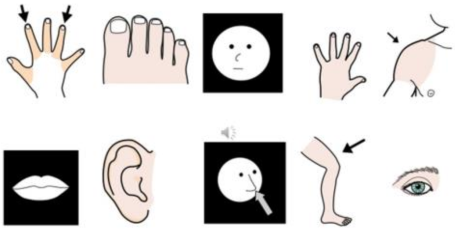


1/5

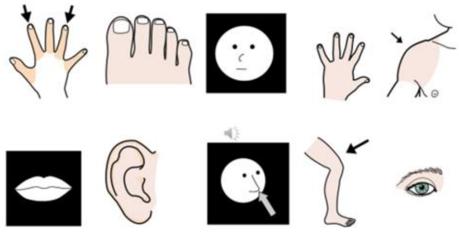


2/5

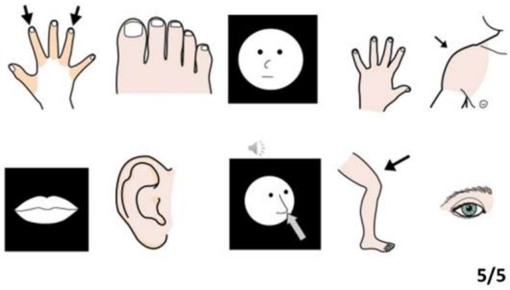
Jetzt hörst du das Lied noch dreimal, versuch dabei mitzusingen!



3/5



4/5

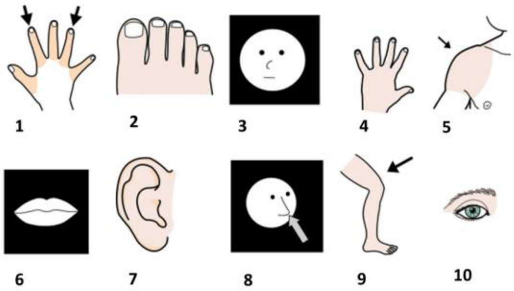


Ende von Teil 2

Teil 3/3

Teil 3 - Erklärung

- 10 Bilder
- Interview



Ende von Teil 3

Danke fürs Teilnehmen!
Liisi.tormakangas@gmail.com

Tests of Normality and Normal Q-Q plots

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Condition group	Statistic	df	Sig.	Statistic	Df	Sig.
Word scoring	1 Singing condition	.167	15	.200*	.915	15	.159
	2 Listening condition	.275	15	.003	.889	15	.064
	3 Control group	.199	15	.113	.937	15	.352

*. This is a lower bound of the true significance.

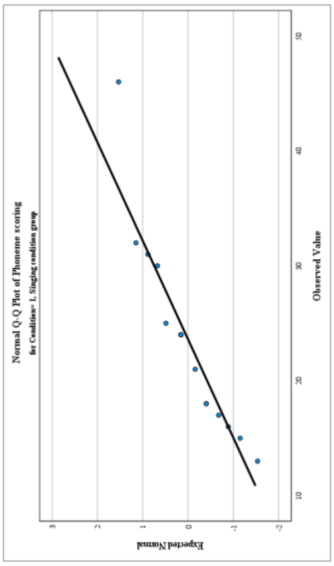
a. Lilliefors Significance Correction

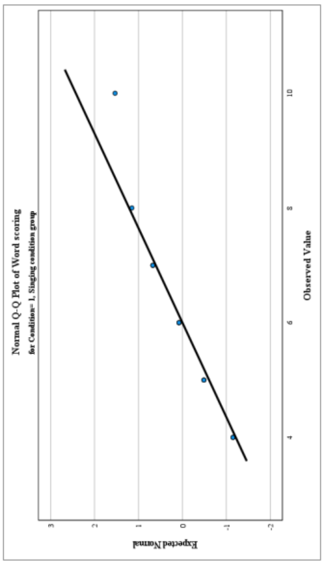
Tests of Normality

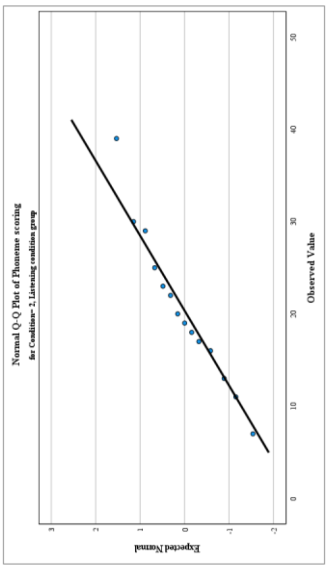
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Condition group	Statistic	df	Sig.	Statistic	df	Sig.
Phoneme scoring	1 Singing condition	.168	15	.200*	.900	15	.094
	2 Listening condition	.116	15	.200*	.969	15	.848
	3 Control group	.216	15	.057	.896	15	.082

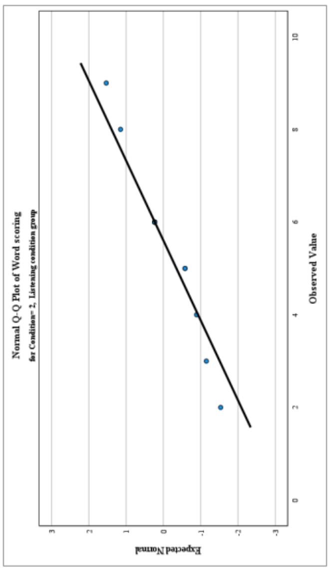
*. This is a lower bound of the true significance.

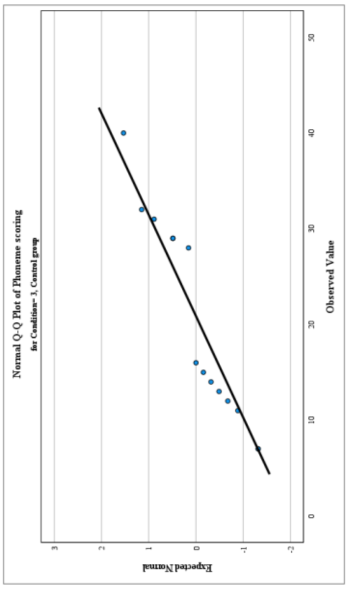
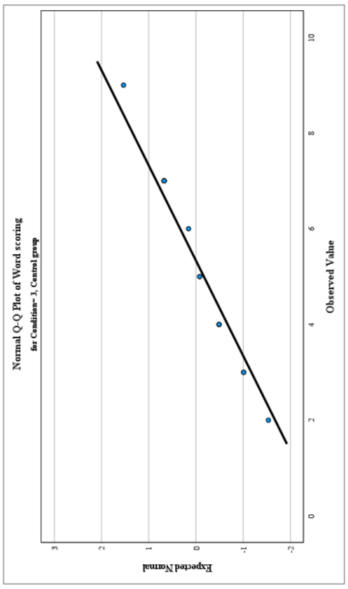
a. Lilliefors Significance Correction









Independent Samples Test – Phoneme scoring

Group Statistics

	Condition group	N	Mean	Std. Deviation	Std. Error Mean
Phoneme scoring	1 Singing condition	15	23.60	8.567	2.212
	2 Listening condition	15	20.33	8.121	2.097

t-test for Equality of Means

		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Phoneme scoring	Equal variances assumed	1.072	28	.293	3.267	3.048	-2.977	9.510
	Equal variances not assumed	1.072	27.920	.293	3.267	3.048	-2.978	9.511

Group Statistics

	Condition group	N	Mean	Std. Deviation	Std. Error Mean

	Condition group	N	Mean	Std. Deviation	Std. Error Mean
Phoneme scoring	1 Singing condition	15	23.60	8.567	2.212
	3 Control group	15	20.87	10.589	2.734

t-test for Equality of Means

		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Phoneme scoring	Equal variances assumed	.777	28	.444	2.733	3.517	-4.471	9.937
	Equal variances not assumed	.777	26.831	.444	2.733	3.517	-4.485	9.511

Group Statistics

	Condition group	N	Mean	Std. Deviation	Std. Error Mean
Phoneme scoring	2 Listening condition	15	20.33	8.121	2.097
	3 Control group	15	20.87	10.589	2.734

t-test for Equality of Means

		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Phoneme scoring	Equal variances assumed	-.155	28	.878	-.533	3.446	-7.591	6.525
	Equal variances not assumed	-.155	26.236	.878	-.533	3.446	-7.613	6.546

Independent Samples Test – Word scoring

Group Statistics

	Condition group	N	Mean	Std. Deviation	Std. Error Mean
Word scoring	1 Singing condition	15	6.00	1.648	.425
	2 Listening condition	15	5.60	1.724	.445

t-test for Equality of Means

		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Word scoring	Equal variances assumed	.650	28	.521	.400	.616	-.861	1.661
	Equal variances not assumed	.650	27.943	.521	.400	.616	-.861	1.661

Group Statistics

	Condition group	N	Mean	Std. Deviation	Std. Error Mean
Word scoring	1 Singing condition	15	6.00	1.648	.425
	3 Control group	15	5.33	1.988	.513

t-test for Equality of Means

		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Word scoring	Equal variances assumed	1.000	28	.326	.667	.667	-.699	2.032

		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
	Equal variances not assumed	1.000	27.066	.326	.667	.667	-.701	2.034

Group Statistics

	Condition group	N	Mean	Std. Deviation	Std. Error Mean
Word scoring	2 Listening condition	15	5.60	1.724	.445
	3 Control Group	15	5.33	1.988	.513

t-test for Equality of Means

		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Word scoring	Equal variances assumed	.393	28	.698	.267	.679	-1.125	1.658
	Equal variances not assumed	.393	27.449	.698	.267	.679	-1.126	1.660