Assessing and Understanding Individual Differences in Music Perception Abilities

Lily Ning Ching Law

Submitted for the degree of Doctor of Philosophy in Psychology

The University of York
Department of Psychology

October 2012

Abstract

A common approach for determining musical competence is to rely on information about individuals' extent of musical training, but relying on musicianship status fails to identify musically untrained individuals with musical skill, as well as those who, despite extensive musical training, may not be as skilled. To counteract this limitation, the working aim of this thesis was to develop a new test battery (The Profile of Music Perception Skills; PROMS) that measures perceptual musical skills across multiple domains: tonal (melody, pitch), qualitative (timbre, tuning), temporal (rhythm, rhythm-to-melody, accent, tempo), and dynamic (loudness). The development and validation of the PROMS are presented in studies 1 to 4. Overall, the PROMS has satisfactory psychometric properties for the composite score and fair to good coefficients for the individual subtests. Convergent validity was established with the relevant dimensions of Gordon's Advanced Measures of Music Audiation and Musical Aptitude Profile (melody, rhythm, tempo), the Musical Ear Test (rhythm), and sample instrumental sounds (timbre). Criterion validity is evidenced by a sizeable relationship between test performance and a composite of various indicators of musical proficiency as well as discriminant validity by a generic auditory discrimination task. The application of the PROMS in examining the structure of music perception mechanism is also presented. In particular, the relationship between music perception skills and non-musical abilities is explored in Study 4. The results suggest that the interrelationships among the various subtests could be accounted for by two higher order factors, sensory and structural music processing; the structural processing skill is related to short-term and working memory. Rhythm perception (rhythm and rhythm-to-melody subtests) also shows significant correlation with general mental ability. An Internet study with the PROMS was conducted to examine whether the findings of controlled studies can be replicated with a more diverse population and uncontrolled environment. Most of the findings of the controlled studies were replicated in the Internet study with several exceptions that are reported in Study 5. A brief version of the full PROMS is proposed as a time-efficient approximation of the full version of the battery.

Content

List of Figures

List of Tables

Acknowledgements

This thesis marks an important milestone in my career roadmap. It has essentially been an embarkation of a whole new life. First and foremost I would like to thank my supervisor, Dr Marcel Zentner for giving me the opportunity to work on this exciting project. I sincerely thank Dr Quentin Summerfield, Dr Tom Hartley and Dr Andy Hunt for their comments and advice as members of my thesis advisory panel. I thank Andy Hunt for continuing to give me support since my MSc, not only as an academic advisor, but also with useful advice about project management and presentation skills. I thank Dr Tuomas Eerola for making efforts to track down Karma's test and publications, as well as providing stimuli materials that were used in the thesis. I also thank Olivetti Belardinelli for willing to share Wing's stimuli that are almost inaccessible today.

I thank peer support from Alessio, Sally, Yue Yue, Elaine, Soojin, Rosa and KanXin who help me remember that I am not fighting alone in the PhD battle, and who have been willing to spend time helping to pilot-test my experiment. I thank Dan, Eliza, Vicky, Joy, Ian, Claire, Chris and Andy for dragging me out from my academic world to enjoy a simple life with cheers, laughter and music. I thank Cecilia, Taryn, Foongyam, Mr Sze and KK for their unconditional support from afar, and for keeping their friendship as warm as ever. I thank my family who are constantly praying for me, supporting and loving me unconditionally. I thank Michael, my soul mate, for being patient with me, taking care of the household whilst I am swamped with my research, and for his unlimited and unconditional support with everything. Lastly, I thank God for putting me in this place, at this time, in this year and with these people, they are the blessings in my life.

Financial Support

This PhD was supported by a scholarship from Holbeck Foundation Scholarship and the department of psychology.

Declaration

This thesis comprises the candidate's own original work and has not, either in the same or different form, been submitted to this or any other University for a degree. All experiments were designed by the candidate with advice from the supervisor. The candidate conducted all testing and analyses were conducted with advice from the supervisor. Some materials in the thesis have been written up as a publication manuscript with the supervisor.

Journal Publications & Conference Presentations

Work from this thesis has appeared in the following publications and conference proceedings:

Law, L. N. C., & Zentner, M. (2012). *A tool to assess individual differencesin perceptual musical ability*. Demonstration presented at The 9th International Symposium on Computer Music Modeling and Retrieval (CMMR): Music and Emotions, London, UK.

Law, L. N. C., & Zentner, M. (2011). *A tool to assess individual differences in perceptual musical ability*. Poster presented at the The Neurosciences and Music - Learning and Memory, Edinburgh, UK.

Law, L. N. C., & Zentner, M. (2010). *Individual differences in music perception ability*. Presented at the York Music and Science Colloquium, York, UK.

CHAPTER 1

Introduction

This chapter summarises the aims and objectives of the thesis by providing an overview of the research background and a brief overview of each chapter. First to be discussed is how music ability is developed and how we recognise it. Next is a discussion of why it is important to identify musical ability as well as its application in research and society. Then the currently available techniques for measuring musical ability will be evaluated and their limitations will be discussed. An overview of the thesis structure, along with the reasoning behind aspects of content and style, is also given.

1.1 The Development of Musical Ability

Musical cultures have existed for at least 35,000 years (Conard, Malina, & Münzel. 2009; d' Errico et al., 2003). Music is now recognised as an important and informative domain in which to study a variety of aspects of cognition, including expectation, emotion, perception, and memory (Scheirer, 1998). Several investigators are convinced that music can yield valuable information about how the brain works. They believe that the study of the brain and the study of music can be mutually revealing (Peretz & Zatorre, 2005; Tervaniemi, Ilvonen, Karma, Alho, & Näätänen, 1997), for example whether musicians or non-musicians differ in their abilities to pre-attentively group consecutive sound (Zuijen et al, 2004; Koelsch, Schröger, & Tervaniemi, 1999) or how the auditory stem relates to music timing perception deficit (Johnson et al., 2007).

In addition it has been found that music representation in the brain varies in individuals depending on their musical experience (Ohnishi et al., 2001; Hutchinson et al, 2003; Schneider et al., 2002). Although the left hemisphere of the brain is generally found to be dominant in dealing with verbal, analytical and executive functions such as language (Vigneau et al., 2006; Liégeois et al., 2004), and the right hemisphere has a more visual, spatial, emotional, holistic and intuitive mode of operation (i.e., music and art) (Joseph, 1988); individuals with musical experience tend to process music information in the left hemisphere just like language processing (i.e., musical analysis). This suggests music training contributes to a language-like system for coding and processing (Milovanov & Tervaniemi, 2011; Ohnishi et al., 2001; Marin, 1982, Wertheim & Botez, 1961; Gott, 1973). More discussion on the language-music link will be presented in Chapter 3.

However, the disassociation between language and music is also observed where brain-damaged composers who may lose their language ability and yet are able to continue to engage in musical activities at a professional level. The reverse effect is also noted in individuals who have bilatereral brain damage and suffer from severe and irreversible deficits in music perception and memory, but are still able to retain their language skill perfectly (Peretz, 2003). Therefore the link between language and music remains an interesting question in research to explore.

This special interconnection between the brain and music, like language, develops at a very early stage in life (Trehub, 2001). Healthy foetuses from 23-34 weeks gestation have reportedly had the capability to respond to sound stimulation, including mother's voice/song, speech and even sounds caused by body movement (Brezinka, Lechner, & Stephan, 1997; Zimmer et al., 1993). As the foetus develops and is eventually delivered from the mother, this ability to react to sound seems to strengthen. Using a heart-rate measurement technique[1], it was found that infants at 5 months were able to make temporal grouping discrimination (Chang & Trehub, 1977). In addition, infants between the ages of 8 and 11 months[2] were able to perform melody discrimination as investigated using a 'head-turn' procedure[3] (Trehub, Bull, & Thorpe, 1984; Trehub, 2001). A more recent study showed that babies from five months to two years old moved rhythmically to music (particularly to a beat rather than a melody) and babies smile more if they are able to successfully synchronise with the music (Zentner & Eerola, 2010). These studies seem to suggest the possible development of a special ability attributed to sensitivity to melody or rhythm, coinciding with individual, and collective, human development. This ability is referred to as "musical ability" in this

[1] Infants show greater cardiac deceleration when they show familiarity with the stimuli (Melson & McCall, 1970).
[2] Melody discrimination ability was also found in 2-month-old infants in a study by Papoušek & Papoušek (1981).
[3] Infants were trained to turn their head whenever they heard a "change" or when a new stimulus was presented (Eilers, Wilson, & Moore, 1977).

research - a predisposition to respond physically and emotionally to musical structure such as melody or rhythm. Whether the sensitivity towards music and sound at a young age is an innate ability is not known, however, as this ability increases throughout life as a result of development and training, the variable abilities between individuals are perceived as individual differences in musical ability (Sloboda, 2000; Sims & Nolker, 2002).

Amongst the variability, the group of individuals who perform better in specific tasks are often recognised as possessing a special "talent" or "intelligence". Gardner (1999) stated, "intelligence is a biopsychological potential to process information that can be activated in a cultural setting to solve problems or create products that are of value in culture" (p.34). Gardner's claim seems to illustrate well the judgement which society makes when we categorise a certain group of people possessing a special skill as *talented* or intelligent.

A British talent search program, Britain's Got Talent (BGT) has discovered several musically talented individuals such as Connie Talbot (7 year old, BGT 2007), and Susan Boyle (48 years old, BGT 2009). In fact, a self-taught armless pianist, Liu Wei won China's Got Talent 2010 for his spectacular piano performance using his toes. If the term "talent" and "intelligence" share the same "skill concept", these phenomena seem to fit with a scenario in which the musical talent within these winners was *activated* and discovered in an environment that the public recognized and appreciated, or to quote Gardner (1999) "products that are of value in culture" (p.34).

1.2 The Definition of Musical Ability

The evidence presented thus far has illustrated the development of music ability and the special status of the term "talent" referring to musical ability that stands out in individuals. In fact, this ability has been posited as "Musical Intelligence" which contrasts with other abilities such as linguistic, logical mathematical, spatial, bodily-kinesthetic, interpersonal, intrapersonal and naturalist (Gardner, 1985, 1999). Gardner (1985) stated:

> Evidently, there is no problem finding at least superficial links between aspects of music and properties of other intellectual systems…Yet, according to my own analysis, the core operations of music do not bear intimate connection to the core operations in other areas; and therefore music deserves to be considered as an autonomous intellectual realm. (p.126)

However, there is little clarity in the existing explorations of musical ability because of the subject's universality, and consequently there is difficulty in its elucidation. Recent qualitative research revealed that "musical ability" is perceived differently depending on the environment within which individuals are located, and their particular musical experiences or lack of them (Hallam, 2010). Nevertheless, the concept of ability was proposed as ranging from an understanding of exceptional ability as a result of enhancement of cognitive and physiological adaptation due to extended deliberate practice (Ericsson, Nandagopal, & Roring, 2005; Sloboda, Davidson, & Howe, 1999; Kemp, 1996; Hayes, 1981), environmental and intrapersonal catalysts, for example being given the opportunities and encouragement to learn (Sloboda & Howe, 1991), or the notion of innate giftedness (Gagné, 1999; Seashore, 1919a; Wing, 1948; Drake, 1954; Bentley, 1966; Gordon, 1965; Karma, 1973).

To illustrate, Sloboda and Howe (1991) interviewed 42 musically gifted children and 20 of the childrens' parents about their children's musical life prior enrolling to special music school. It was found that most of the children did not show any particular sign of musical potential, instead it was the parents who have taken an active role in supervising and encouraging the children. The authors concluded that musical ability is a result of social and motivational influences on learning and development. Sloboda and colleagues (1999) further pointed out that despite the possibility that inherent biological differences between people may make a contribution to differences in their eventual musical capabilities and that this can be examined more precisely by studying behavioural genetics (e.g., Plomin & Thompson, 1993), these links between biology and musical ability are likely to be complicated as there is no way to pinpoint the notion of a unitary "blueprint for music" (p.50) that is implied by the notion of innate ability.

Gagné (2003) opposed this idea where he believed a *gift* is an untrained natural ability, and thus he developed a metric model of levels of giftedness including mildly (1 in 10), moderately (1 in 100), highly (1 in 1,000), exceptionally (1, in 10,000) and extremely gifted (1 in 100,000). Gagné (1999) pointed out that the negative correlation between music achievement and amount of practice, shown in Sloboda and Howe's (1991) data, was an example of detecting innate ability that is dissociated from *training.* This is consistent with an understanding of musical ability in terms of "potential for learning music" *before* formal training and achievement (Shuter-Dyson, 1999, p. 627), or "the power to act but indicating nothing about the heritability or congenitalness of inferred potentiality" (Farnsworth, 1969, p.151).

1.3 The Significance of Identifying Musical Ability

Across disciplines, a number of scholars have shown interest in assessing individual differences in musical ability. One reason for this interest comes from a growing concern to understand the role of musical ability in non-musical faculties, ranging from motor skills, general intelligence, empathy to language processing, and reading impairments. For example, problems in rhythm perception have been found to relate to reading impairments, and there is reason to believe that training of rhythmic processing capacities could act as a remedy for dyslexia (Thompson & Goswami, 2008).

A second reason is that knowledge about the links between musical and non-musical traits could shed light on the perennial conundrum of music's evolutionary origins. For example, if musical abilities were found to relate to linguistic abilities, such as phoneme discrimination, this would support a music-as-language-corollary hypothesis of the origins of music. In turn, if the musicality measures were related to aspects of social functioning such as emotion recognition or empathy, this would support the social cohesion theory of the origins of music (Patel, 2008). However it must not be discounted that, as well as the link to motor skills and language stated above, the origins of music could in fact relate to the processes inherent in its emotional or empathetic processes.

Unfortunately, progress in understanding these relationships is currently hampered by the lack of an objective and standardized instrument to measure musical abilities. It is not that various aspects of music perception and production have not been extensively investigated—they have (e.g., Jones, Fay, & Popper, 2010). What has been missing is interest in the development of a psychometrically sound and construct-validated test, capable of diagnosing individual differences in musical ability.

1.4 Current Musical ability Assessments and Their Limitations

In the absence of an objective measurement tool for musical skills, researchers often use self-reported musicianship to estimate the presence of musical ability. Typically, a binary classification is used, comparing the performance of musicians versus nonmusicians on variables such as general IQ and mental abilities (e.g., Helmbold, Rammsayer, & Altenmüller, 2005); brain structure (e.g., Gaser & Schlaug, 2003); neural underpinning of musical sounds (Pantev, 2001), pre-attentive auditory processing (Koelsch et al., 1999), language processing (e.g., Lee & Hung, 2008; Wong, Skoe, Russo, Dees, & Kraus, 2007); vocal emotion recognition (e.g., Lima & Castro, 2011); memory (e.g., Williamson, Baddeley, & Hitch, 2010); motor skills (Meister et al., 2005); and even creativity (Gibson, Folley, & Park, 2009)—to cite just a few recent examples.

This practice is sensible, but has a number of limitations. First, being a "non-musician" does not, in and of itself, denote an absence of musical ability. The ability may be undiscovered, or circumstances may have prevented its development. Among the musically untrained, some people might reach a high level musical proficiency if given the time and opportunity to do so. These individuals are referred to as *musical sleepers*, because of their latent musical ability. Conversely, many years of musical training resulting in degrees and certificates are reasonable but not infallible indicators of above-average musical ability. Individuals whose musical proficiency languishes despite multiple years of training are referred to here as *sleeping musicians*. Due to the absence of a tool for identifying individuals that perform better (or worse) than would be expected from their extent of musical training, current research studies and findings may be biased by an unknown number of false negatives and positives.

Second, degrees and qualifications provide at best an estimate of generic musical accomplishment. Yet, once a link between general musical ability and another ability, trait or disorder is established, the next obvious question concerns the *type* of musical capacity that plays a key role in the relationship (e.g., tempo, pitch, rhythm, timbre, melody perception, or any combination of these). Such specific information is not only key to the scientific analysis of the relationship under examination, it could also have a role in devising treatment plans for a specific disorder using music materials. Third, using sophisticated instrumentation (i.e., neuroimaging scanner) to compare data, which at this point in time has unknown validity, seems inappropriate.

An alternative to inferring musical ability from musicianship status lies in devising objective tasks to assess musical capacity, such as a musical aptitude test. This idea is not new, indeed several authors from the last century have developed musical aptitude batteries. The details about these tests will be provided in Chapter 2 and their limitations are also briefly discussed here in order to present a clearer view of the current work's aims and objectives.

Despite the profusion of musical aptitude tests, they have not proved good enough for uses in contemporary research for the following reasons:

(1) These tests are generally considered obsolete and very difficult to access today as they were developed more than five decades ago and are not available even for the original publisher of the tests. Furthermore the sound formats of these tests generally require special machinery to operate (see Chapter 2, Table 2.4). Previous research has already characterised the tests in these terms over a decade ago (e.g., Murphy, 1999; Carson, 1998);

(2) Many of the tests were designed to measure children's generic musical aptitude; as such they are not suited for the assessment of adult interindividual differences in specific musical capacities or for examining questions related to the nature of music perception (e.g., Seashore et al., 1960; Wing, 1948; Bentley, 1966; Gordon, 1965; Karma, 1973);

(3) The stimuli material were poorly constructed and recorded, and inconsistencies in the number of stimuli and answer formats in the test design would make research findings difficult to interpret;

(4) Many of the test-batteries missed out crucial aspects of music perception (e.g., tempo, timbre, tuning);

(5) The procedures used for inferring test validity and reliability are tenuous by contemporary standards.

It is perhaps for this reason that investigators prefer to create their own tasks, depending on the nature of their research objectives, rather than relying on any of the reviewed test-batteries (e.g., Fabiani & Friberg, 2011; Geringer & Johnson, 2007). Highly specific, homespun tasks may serve the purpose of a given experiment very well, but are unsuited to assessing individual differences across a broad range of perceptual music skills. Also, they do not lend themselves easily to comparisons across studies, thereby preventing the incremental accumulation knowledge that is so important to the establishment and progress of any branch of science.

1.5 Research Aims and Strategy

Against this background, the current work is concerned with two areas. The first aim is to create a new music test-battery in order to fill the current gap in musical ability tests for normal or general adult populations. This test-battery is named *The Profile of Music Perception Skills* (PROMS) in this thesis. The PROMS aims to meet the following four criteria:

(1) The test should assess a broader range of specific perceptual musical skills than previous tests, which were usually confined to subtests measuring tonal memory and certain types of rhythmic skill.

(2) The perceptual skills targeted in the subtests should be measured with the greatest possible specificity.

(3) The test should be equally suitable for listeners differing in the extent and in the type of their musical background.

(4) The test should meet contemporary standards for test construction in terms of validity, reliability, and stimulus design.

The second aim of this research is to provide an understanding of the nature of music perception such as by examining the intercorrelations of various music perception domains. For example, whether abilities in subdomains of music perception are independent, or if they point to a general musical ability factor akin to Spearman's *g* (Spearman, 1904). Or whether musical ability as Gardner (1999) claimed, is totally independent from other non-musical abilities? These questions will be explored in the present thesis.

1.6 Thesis structure

This thesis is divided into four main sections to address the above issues as seen in Figure 1.1. (1) Stage 1: The development of the test-battery (2) Stage 2: The validation of the test-battery (3) Stage 3: The relationship between music ability and other non-musical abilities (4) Stage 4: The factor structure of the test. Appendices with audio files containing examples of the stimuli are also provided to enhance the clarity of the discussion.

Stage 1
DEVELOPMENT
Chapters 2, 3, 4

Stage 2
VALIDATION
Chapter 5

Stage 3
THE LINKS BETWEEN MUSICAL AND NON-MUSICAL ABILITIES
Chapter 6

Stage 4
TEST-STRUCTURE
Chapter 7

Figure 1.1. The four stages of the thesis structure without Introduction or Conclusion chapters

Stage 1: The Development of the Test-Battery

The development stage of the thesis provides an overview of the progress of the music test-battery's construction and revision, which is discussed in three chapters:

Chapter 2: This chapter provides an overview of the research background that is presented in this thesis. The concept of music applied in this work will be defined, then several music aptitude tests, auditory tests and music tests for special populations that were prominent during the last century and which have influenced the current work are reviewed.

Chapter 3: This chapter provides an overview discussion of the perceptual dimensions that are proposed and examined in Chapter 2, laying a background overview for the test construction in Chapter 4.

Chapter 4: This chapter provides the rationale for the selection of musical perceptual dimensions that are examined in this thesis and also provides the initial design and construction of the test-battery. Initial findings of the research work are reported.

Stage 2: The Validation of the Test-Battery

Chapter 5: This chapter introduces an improvement of the test design that was presented in Chapter 4, and describes the methodology that was used to validate the test-battery.

Stage 3: The Relationship Between Music Ability and Non-Musical Ability

Chapter 6: This chapter investigates the relationship between musical ability and non-musical abilities such as general mental ability, short-term memory and working memory, and auditory discrimination skill. This chapter also serves as a discriminant validation by examining whether the current test-battery measures music-specific skill rather than general cognitive or auditory skill.

Stage 4: The Factor Structure of the Test-Battery

Chapter 7: This chapter examines the factorial structures and the preliminary norm distribution of the test-battery with larger samples (Internet study), as well as to see whether the result found in controlled studies can be replicated with a more diverse population, and how uncontrolled testing environment and equipment may have affected the result

Discussions and Conclusions

Finally Chapter 8 summarises key results and conclusions from the previous four chapters. The usefulness and the limitations of the test-battery are discussed. Directions for future work that might further expand on the knowledge presented in the thesis and improve upon the techniques and system described are given.

CHAPTER 2

Literature Review

This chapter summarises the background of the research that is presented in this thesis. First, how sound and music is sensed by human ears will be described. Next, the heart of this research - "music" will be presented, beginning with the different definitions accepted in different cultures, followed by how music has evolved with technological advancements. Next, the definition of music that has been adopted in this work is stated, which indeed is a crucial part of the test-development. Finally a critical analysis of several music aptitude tests, auditory tests, and music tests for special populations that were prominent during the last century will be provided. Original objectives of the test-design rationales will be provided by quotations to accurately present the previous authors' intentions. By the end of the chapter, this information aims to give readers a clear overview of the research background that has influenced the current work.

2.1 Sensing Sounds and Music by Ears

Most people's experience with music is through music listening and this coincides with the focus of this thesis to create a music listening test. Therefore how sound is perceived by the human auditory system is first described here to illustrate how sounds are sensed by humans, and more importantly latter sections will discuss how humans organise these sounds into music. In particular, the way that humans perceive music differently especially for pitch and intensity-related components are dependent on how these sounds arrive at the ears and how the peripheral auditory system handles and processes the incoming information as meaningful music information (i.e. basilar membrane and Organ of Corti).

What is a sound? Sound is transmitted when vibrating particles (atoms or molecules) create waves at frequencies perceivable by the human ear and propagate through a medium (solid, liquid, or gas) from one location to another. People with normal hearing are able to hear sound waves by the peripheral auditory system, which starts from the ear canal at the outer ear and ends at a lightweight, thin and taut membrane which is the *ear drum (tympanic membrane)* of the middle ear. Although the middle ear is filled with air, a chain of three bones, called the *ossicles*, is used to carry acoustic vibrations from the eardrum to the *cochlea* (see Figure 2.1).

At the cochlea fluid travels through the basilar membrane in a wave motion, which triggers hair movement within the Organ of Corti that sits on top of the membrane. The movement of these hairs excites associated nerve fibres and it is these nerve fibres which carry auditory information to the brain. The amount of nerve fibres excited by the stimulus is determined by the extent of the hair movement (Figure 2.2). Therefore in order to cause auditory information to be sent to the brain, there must be a certain level of energy present. This means that below a certain threshold of intensity, variations in pressure at the opening of the ear canal will not cause sufficient stimulation of the auditory nerve fibers, this will result a listener to have no sense of sound through auditory system (Pickles, 1988).

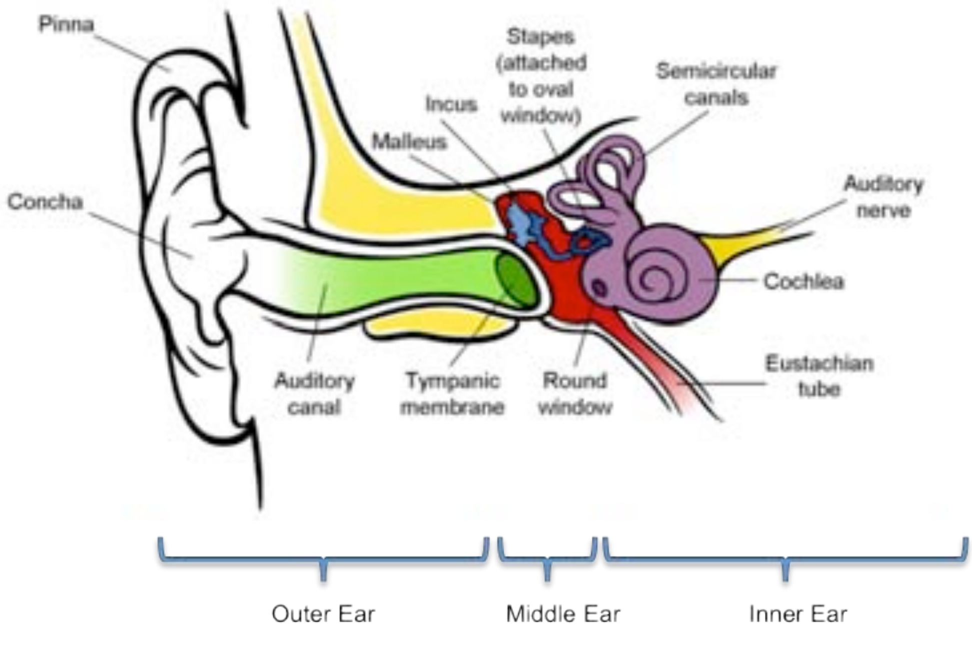


Figure 2.1. A schematic diagram of the peripheral auditory system. Image adapted from Brockmann (2012).

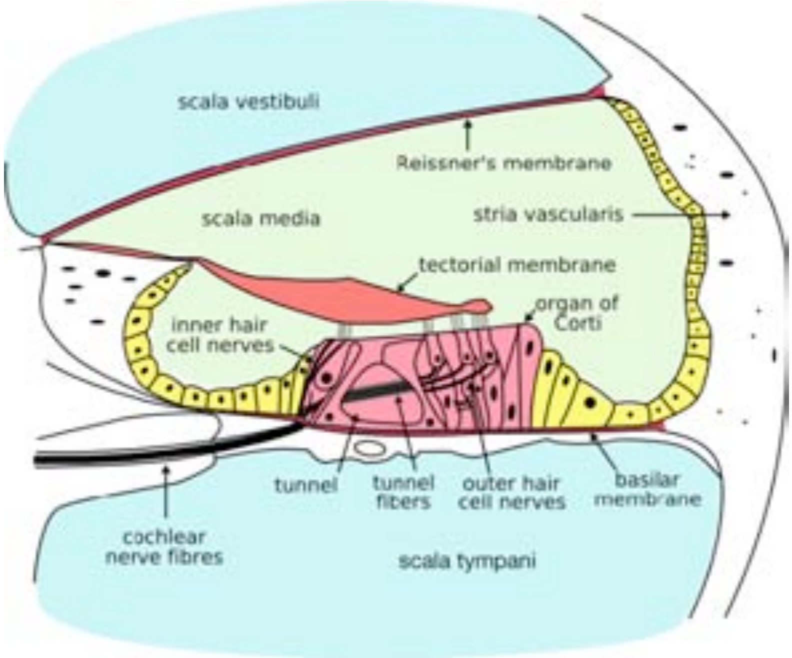


Figure 2.2. A schematic cross-section of the cochlea. Image from Ropshkow (2012).

2.2 Sound and Music

The main scope of this thesis is to create a test that is able to measure *music* perception ability, so naturally the concept of *music* must first be defined to provide a clearer focus of the work to be presented. There are three main areas that need to be addressed: (1) the concept of music, (2) whether auditory sound events such as noise and environmental sounds *are* music, (3) music perception versus music hearing.

2.2.1 The Concept of Music

The term "music" was originally derived from the ancient Greek *μουσική* (*mousike*), which refers to 'the art of the Muses'. However, the concepts and definition of music differ greatly in different cultures depending on the significance level of music to them (Nettl, 2005). For example the concept of music ranges from types of music such as classical or instrumental music; musical instruments or singing; a fine art; a product of communication and expressiveness like language; a combination of sounds with specific structures such as chords or rhythm; scores or music manuscripts; sounds that create emotion; social context such as concerts; or any sounds such as animal, environmental or noises (Nettl, 2005). Often, people adopt only one of these selections of concepts as being music as a whole. To this end, problems are shown in music psychology research where often Western classical music tends to be studied as most of the researchers are themselves skilled in this genre (Vink, 2001; Scheirer, 1998). This approach is problematic as music should be treated as an "aural tradition" rather than the "written tradition" often adopted as a notation system in Western classical music (Scheirer, 1998). Smoliar (1991) stated:

> The problem with a system like music notation is that it provides an *a priori* ontology of categories – along with labels for those categories – that does not necessarily pertain to categories that are actually formed as part of listening behavior. If we wish to consider listening to music as a cognitive behavior, we must begin by studying how categories are formed in the course of perception rather than trying to invent explanations to justify the recognition of categories we wish to assume are already present. (p. 50)

A similar argument is also supported by Serafine (1988):

> Traditionally, the elements of music are assumed to be tones and assemblages of tones called chords. Such a view critically determines how we conceive composition and perception. For example, tones may be considered the material with which a composer works, by arranging and conglomerating them, or tones may be considered the basic units processed by the listener. The present view, however, holds that tones and chords cannot in any meaningful and especially psychological way be considered the elements of music. Rather, tones and chords are viewed as the inevitable by-product of musical writing and analysis, and as such are useful, even necessary analytic tools with minimal cognitive reality. (p. 7)

The following section will clarify the relationship between sound and music in a non-classical music context. The music context that is used in this work will be reported.

2.2.2 Sound Versus Music

Music, a type of sound, can range from, but is not limited to, humans singing (e.g., Gregorian Chant; see Appendix 7), and playing musical instruments such as the piano or drums (e.g., *Spain* by Chick Corea; see Appendix 7). Nowadays, with the advantages of technology, musical instruments can be synthesized and sampled. Moreover, technology has not only created different types of "sounds" or timbres that can be used in music, it has also enabled people to use the raw, fundamental forms of sound with a single frequency, known as *pure tones*; as well as "a sound wave whose pressure varies in a random way over time" (Plack, 2005, p.26), known as *noise*; to create music. This non-traditional way of creating music using noises and environmental sound as music is described below.

The digital artists Ryoji Ikeda and Alva Noto use sinewaves as the key components of their music. Ikeda's album "*dataplex"* (2005, see Appendix 7) and Alva Noto's "*UniTxt"* (2008) are constructed of precisely sequenced snippets of noise and low rapid exchanges of sine waves at both extremely low and extremely high frequencies. Furthermore musicians such as Toshimaru Nakamura and Sachiko M use sinewaves at their most base level, often creating music that is not sensed by the human ear but *perceivab*le through the sense of touch, via the medium of a vibrating stimulus (i.e., very low frequency). Noise musicians such as Merzbow and Russell Haswell construct their work by building tense audio environments, focussing on the visceral presence of noise. Merzbow's album "*Pulse Demon:My Station Rock*" (1996) bears down upon the listener, taking advantage of the human inability to resolve multiple frequencies (see Appendix 7).

American composer, John Cage, eloquently illustrated that ordinary auditory events can be a musical composition with *4'33"* (1962) consisting of 4 minutes 33 seconds of silence from the performer, intended to allow the surrounding auditory events to create the composition. The Greek-French composer Xenakis (1992), who was renowned for applying Mathematical models in his music, termed his mathematical music creation method as "Stochastic Music". He defined "Stochastic Music" as "music constructed from the principle of indeterminism…The laws of the calculus of probabilities entered composition through musical necessity" (p.8). He even moulded music with architecture and heard music within elaborate versions of banal everyday occurrences:

> …natural events such as the collision of hail or rain with hard surfaces, or the song of cicadas in a summer field...are made out of thousands of isolated sounds; this multitude of sounds, seen as a totality, is a new sonic event. (p.9)

These composition methods can be reconciled with Bregman's (1990) Auditory Scene Analysis theory that humans are able to group sounds using a "schema-driven" process, meaning that listeners are able to transform auditory events into a "music pattern" based on their prior knowledge - either consciously or subconsciously. Gathering the concepts of these authors, music should be recognised as an *intended* piece of work that is composed using acoustic events, or a *flow of sounds* that is perceived and formed by the human mind either intentionally or subconsciously, which might evoke emotions. Music is born when it is composed *intentionally* by composers; but it can also be "formed" or "born" when listeners intentionally organise a series of auditory events into certain structures based on a schema-driven strategy, just as some people perceive birdsong, sea-waves or even vibrations as "music". This philosophy further expands the concept of music beyond the ability of "hearing". To give an example, Wigram (1995) speaks of a deaf-mute named Person Sutermeister of Berne about music:

> My main receiving station is my back. The sound penetrates here and flows through the whole trunk of my body, which feels like a hollow vessel struck rhythmically, resounding now louder, now softer, depending on the intensity of the music. But there is not the slightest sensation in my head and hands – the head is the least sensitive. (p. 17)

Dame Evelyn Glennie, a Scottish percussionist, also exemplifies someone who is able to enjoy and perform music without "hearing". Glennie continues to perform music internationally despite being found to be deaf at around the age of 8-12 (Brown, 1999). Glennie (1993) shared her experience of learning music without hearing:

> I spent a lot of time in my youth (with the help of my school Percussion teacher Ron Forbes) refining my ability to detect vibrations. I would stand with my hands against the classroom wall while Ron played notes on the timpani (timpani produce a lot of vibrations). Eventually I managed to distinguish the rough pitch of notes by associating where on my body I felt the sound with the sense of perfect pitch I had before losing my hearing. The low sounds I feel mainly in my legs and feet and high sounds might be particular places on my face, neck and chest. (p.1)

These examples suggest that deaf people respond to sound in a different way. With the loss of hearing, other parts of body become more sensitive to the vibration of sound waves. Using functional Magnetic Resonance Imaging (fMRI), Shibata (2001) compared the brain activity of 10 deaf people with 11 normal hearing people and found that the deaf people processed vibrations via touch in the auditory cortex which should only have been active during auditory stimulation. People with normal hearing did not show such brain activity.

This seems to suggest that sound and music do not need to be *heard* in order to be called "sound" and "music". Similarly, there are also sound artists who have worked on musical experience via visual perception only, commonly known as the visual music (e.g., Normal Mclaren, Steven Woloshen, Barry Spinello). Visual music is composed by translating sound or music into visual representation by specific devices or composers' interpretations (McDonnell, 2007). Whilst this thesis focuses on the perception of music in listeners with normal hearing, it is also important to acknowledge that music perception is not limited to the hearing ability examples presented above.

As the evidence above illustrates, the notion of music has surpassed its traditional concept as a product of harmonic[4] instruments such as the piano or inharmonic instruments such as the gamelan. Despite general auditory sound events being used for music for more than forty years (for example noise and environmental sounds), conventional music training courses do not offer this type of music widely in comparison to traditional Western European music particularly classical music, which has a longer history. Unfortunately, music composed using non-traditional musical instruments such as a computer is often overlooked or not recognised as "music" in scientific research. Consequently, pure tones and noise are often referred to as "auditory components" (Kidd et al., 2007) or "non-musical stimuli" (Zendel & Alain, 2009) rather than as part of a "music component" in scholarly research.

Against this background, the music context used in this research does not intend to be solely based on Western classical music (Vink, 2001), nor does this research attempt to examine specific cultures of music (e.g., Nettl, 2005; Brown & Jordania, 2011; Castellano, Bharucha, & Krumhansl, 1984). Rather, the aim of this work is to devise a test that prioritizes musical components that can be comprehended by individuals educated in different musical systems and styles, this will be further discussed in Chapters 3 and 4. In the following section, a substantial amount of previous studies where music aptitude tests have been developed will be reviewed, borne out of their notion of music and musical ability.

2.3 Previous Musical Aptitude Tests

A considerable amount of thought and attention has been paid to the definition and measurement of musical ability in the past two hundred years occurring in three phases: the phenomenological approach (1800 to 1910/1920); the psychometric approach (1920s to today); and musical meaning (1980s to today) (Gembris, 1997). A difficulty in determining what is musical ability is highlighted by Révész, (1953), "Experience shows that a person may possess a goodly quantity of such attributes and abilities with-out necessarily being in a position to grasp music in its autonomous forms and effects. It is quite another matter if these attributes are abilities are evaluated solely as

[4] 'Harmonic' describes instruments that produce harmonics approximately at integer multiples of their fundamental frequency.

symptoms of musicality" (p.131).

In addition, conflicting views on what musical ability might be inevitably has produced a variety of musical ability tests. Music syllabuses such as the Associated Board Royal School of Music (ABRSM), which was founded in 1889 (ABRSM, 2011), not only provide musical training, but also provide music exams as an assessment of apprentice training. However, this type of measurement is often not suitable for people who have not taken up a specific musical course, as Sloboda (1985a) highlighted:

> Whilst examinations presuppose intensive preparation of specific materials, tests of ability involve no foreknowledge of test content. Indeed, such tests are invalidated by extensive practice on the task they contain. This is because of the rationale which underlies their construction. (p.233)

Due to the lack of music tests that could be used to examine the musical aptitude of untrained listeners, several researchers (for example Carl Seashore, Edwin Gordon) were interested in creating tools that were not limited to measuring musical ability in trained individuals. The increased interest in developing a measurement for music ability has probably been generated as it is analogous to the study of Intelligence Quotient (IQ) or the idea of talent or aptitude being identifiable. Researchers attempted to use psychometric approaches to measure musical ability scientifically. Even though the exploration of musical ability began as early as 1770 by Christian Friedrich Michaelis (Gembris, 1997), 1883 by Carl Stumpf (2012[5]), this thesis only attempts to review musical ability research that was active during the last century (post 1900).

Carl Seashore is one of the prominent researchers who developed a musical ability test since 1919, and many tests by other researchers followed. These tests are different from one another as the authors had different beliefs about what constituted musical ability. These tests do, however, seem to reflect a common belief that musical aptitude is innate and can be discovered up to the age of 9. After the age of 9, musical aptitude stabilizes. Many of these early tests (Gordon, 1965; Seashore, 1919a; Drake, 1954; Wing, 1948; and Bentley, 1966) were therefore intended as group tests for children aged 8 or 9 years and above.

The following section begins with a detailed overview of several prominent music tests during the last century (see Table 2.4 for an overview), followed by a critical analysis of the presented studies. Prior to the discussion of these tests, the descriptions of different types of reliability methods and validity methods are also provided in Table

[5] This is the translated version of the 1883 book.

2.1 and Table 2.2. These tables act as a reference guideline for the characteristics and qualities of various reliability and validation methods that will be discussed throughout the thesis; in particular in the test-batteries reviewing section, as well as the results and discussion sections of the empirical studies (Chapter 4 to Chapter 8).

2.3.1 Reliability and Validity

Psychological testing is a process where a particular scale or test is administered to obtain a specific score. Subsequently, a descriptive meaning can be applied to the score on the basis of the normative findings (Meyer et al., 2001). The psychometrics property[6] of a psychological test is generally examined in two measures: Reliability and Validity.

Reliability refers to the “accuracy, dependability, consistency, or repeatability of test results” (Kaplan & Saccuzzo, 2009, p.10) and is used to describe different sources of measurement error in which each has a different meaning, for example *test-retest* and *internal consistency*.

Validity refers to the “meaning and usefulness of test results” (Kaplan & Saccuzzo, 2009, p.10). Test validity is conducted to assess to what degree a test measures what it purports to measure. Normally this is conducted using correlation analysis between the test and the criterion.

Table 2.1. Reliability Types and Their Descriptions

Reliability type	**Description**
Split-Half (Internal consistency)	In Split-Half reliability, a set of scores is randomly divided into halves. The results of one half of the test are then compared with the results of the other half. Some investigators prefer to calculate Split-Half reliability by comparing the scores for the first half of the items with a different score for the second half.
Odd-even	Odd-even is a type of Split-Half reliability, but instead of randomly assigning the test into halves, an odd-even system is obtained by comparing the odd-numbered items in the test against the even-numbered items. This is often used for items that get

[6] Psychometrics property refers to how well a psychological test measures the construct of interest

Reliability type	Description
	progressively more difficult.
Split-Half Spearman-Brown formula	Spearman-Brown is a corrected formula that is used to estimate what the correlation between the two halves would have been if each half had been the length of the whole test.
Kuder-Richardson (Internal consistency)	The Kuder-Richardson formula was developed by Kuder and Richardson (1937) to improve Split-Half reliability by simultaneously considering all possible ways of splitting the items. However, Kuder-Richardson reliability can only be used for dichotomous items (i.e., right or wrong answers). Kuder-Richardson 20 (KR_{20}) requires the calculation of the proportion of the people getting each item correct (*p*) and the proportion of people getting each item incorrect (*q*). Kuder-Richardson 21 (KR_{21}) employs the mean test score - an approximation of the sum of the *pq* product. KR_{21} is based on the assumption that all the items are of equal difficulty, or that the average difficulty level is 50%.
Cronbach's Alpha, α (Internal consistency)	Cronbach's Alpha reliability is regarded as the one of the most general methods for estimating internal consistency. Cronbach's can be used when the variance of items are in a dichotomous format and this method takes into account the inter-associations between all items in the scale. However, Cronbach's Alpha evaluates the extent to which the different items on a test measure the same trait, therefore, a low reliability is estimated if the test is designed to measure more than one trait. If the number of scales (or subscales) is small, alpha is likely to be low even if they are quite strongly associated (Streiner, 2003; Cortina, 1993; Loewenthal, 1996).
McDonald's Omega, ω (Internal consistency)	McDonald's Omega has only recently been used more frequently. McDonald's Omega is a measure of reliability based on the single factor model, or is a measure of the generalisability of the test items (McDonald, 1999).

Reliability type	**Description**
Test-Retest (Time sampling)	Test-retest reliability is when the same test is administered to the same sample on two different occasions. This type of analysis is more appropriate for measuring traits or characteristics that do not change over time. Generally the shorter the time gap between the tests, the higher the correlation and vice versa. Lower correlation (i.e., 0.43) might mean (1) the test has poor reliability, (2) the participant changed characteristics between the test times, (3) a combination of the two. The time interval between measures could range from a few hours to a few months but is recommended to be from 1 to 2 weeks (Pedhazur & Schmelkin, 1991). Any intervals that are longer than 2 weeks might assess substantial alterations within a person rather than random slight differences in test responding, and thus altering more accurate representations of the instruments' reliability (Anastasi & Urbina, 1997).

Note. The content of this table is taken from Kaplan & Saccuzzo (2009) unless stated otherwise.

The criteria of acceptability for reliability coefficients (internal consistency) depend on the test length and test format. Longer tests generally have higher reliabilities than shorter ones; test formats such as true-false formats compared to multiple choice are also likely to have lower reliability (Groth-Marnat, 2009). Kline (1993) recommended a minimum reliability coefficient of .80; The British Psychological Society's Committee on Test Standards recommended an acceptable coefficient of .70; Loewenthal (1996) and Cortina (1993) suggested that about .60 is acceptable if the scale is short (i.e., < 20 items) when there is good evidence for validity, and there are good theoretical and/or practical reasons for the scale. Several researchers recommended the minimum Cronbach's Alpa for a clinical tool is above .90, ideally .95 (Kaplan & Saccuzzo, 2009; Loewenthal, 1996), others suggest that Cronbach's Alpa above .90 indicates unnecessary redundancy or is "asking the same questions many different ways" (Streiner, 2003; McClelland, 1980, p.30).

Reliability coefficients, however, only evaluate the correspondence between a variable and itself. As a result, they cannot provide a reasonable standard for evaluating whether the test measures or examines what it claims to measure or examine (Meyer et al., 2001; Kaplan & Saccuzzo, 2009). For this reason, validation procedure should be in place in addition to reliability measurement to support the purpose of the psychological testing. The different types of validity procedures are reported in Table 2.2. Kaplan & Saccuzzo (2009) recommended that validity coefficients are not usually expected to be exceptionally high, and that they also depend on the reliability of the test and the criterion (see Table 2.3).

Table 2.2. Test-Validity Types

Type		**Description**
Construct–related validity. Construct–related validity is employed when there is no acceptable criterion that adequately defines the quality to be measured. Therefore construct validation requires assembling evidence about what a test means by showing the relationship between a test and other tests and measures.	Convergent validity	Refers to what extent the test is correlated with other measures that are designed to tap similar constructs.
	Discriminant/Divergent validity	Refers to the degree to which the test is not similar to criterion that is not supposedly unrelated.
Criterion-related validity. Criterion validity refers to how well a test corresponds with a particular criterion which it should be logically or conceptually related to.	Concurrent validity	Refers to the assessment of the simultaneous relationship between the test and the criterion.
	Predictive validity	Refers to the degree to which the test predicts scores based on the criterion measure.
Content-related validity. Content-related validity refers to the relevant representation of the conceptual domain the test is designed for in a particular assessment purpose.	-	-

Note. The content of this table is taken from Kaplan & Saccuzzo (2009). Kaplan & Saccuzzo (2009) have noted that validity is a unitary concept representing all of the evidence to support the intended interpretation of a measure. The categorisation presented in this table is for convenience illustrations purposes; this table does not imply that there are distinct forms of validity.

Table 2.3. How Reliability Affects Validity

Reliability of test	Reliability of criterion	Maximum validity (correlation)
1.0	1.0	1.00
.8	1.0	.89
.6	1.0	.77
.4	1.0	.63
.2	1.0	.45
.0	1.0	.00
1.0	.5	.71
.8	.5	.63
.6	.5	.55
.4	.5	.45
.2	.5	.32
.0	.5	.00
1.0	.0	.00
.8	.0	.00
.6	.0	.00
.4	.0	.00
.2	.0	.00
.0	.0	.00

Note. Table reproduced from Kaplan & Saccuzzo (2009, p.153). The first column displays the reliability of the test; the second column displays the reliability of the validity criterion; the third column displays the maximum theoretical correlations between tests, given the reliability of the measures.

Table 2.4. An Overview of Early Musical Abilities Research

Test	Sample	Reliability (Music Test)	Validity	Format
Seashore (1919b, 1967)	Ages 10 to 16	Internal consistency: .55 to .84 (Kuder-Richardson Formula 21) Test-Retest: Not reported	Convergent: Yes Criterion: Yes Predictive: Not reported	Single $33^{1/3\ rpm}$ Long Playing Recording
Wing (1948,1968)	Ages 8 to15	Internal consistency: .91 (Split-Half) Test-Retest: .76- .88	Convergent: Yes Criterion: Yes Predictive: Not reported	MP3 (Italian adaption by Olivetti Belardinelli, 1995).
Drake (1957)	Ages 7 to 23	Internal consistency: .91 (Split-Half Odd-Even) Test-Retest: Not reported	Convergent: Not reported Criterion: Yes Predictive: Not reported	12-inch $33^{1/3}$ rpm microgroove phonograph
Gordon (1965, 1995)	Ages 9 to 18	Internal consistency: .66-.95 (Split-Half)	Convergent: Yes Criterion: Yes Predictive: Yes	Compact Disc
Bentley (1966)	Ages 9 to 11	Internal consistency: .70-.90 (Kuder-Richarson) Test-Retest: .84	Convergent: Yes Criterion: Yes Predictive: Not reported	Ten-inch $33^{1/3}$ rpm disc record

Test	Sample	Reliability (Music Test)	Validity	Format
Karma (1973, 1984)	Ages 10 to 18	Internal consistency: .68 (Kuder-Richardson) Test-Retest: Not reported	Convergent: Not reported Criterion: Yes Predictive: Not reported	MP3
Gordon (1989, 1990)	Ages 17 to 19	Internal consistency: .83-.86 (Split-Half) Test-Retest: .79- .84	Convergent: Not reported Criterion: Yes Predictive: Yes	Compact Disc
Wallentin et al. (2010b)	Adult Population	Internal consistency: .69-.85 (Cronbach's Alpha) Test-Retest: Not reported	Convergent: Yes Criterion: Yes Predictive: Not reported	WAV and MP3

2.3.2 Seashore's Measures of Musical Talents [1919a, 1919b, 1938, 1960, 1967][7]

Carl Seashore, is one of the towering figures of musical ability research, well known for his pioneering work in the first half of the last century. Seashore (1938, 1967) stated that "everything that is rendered as music or heard as music may be expressed in terms of the concepts of the sound wave" (p.2), and reasoned that the physical aspects of sound waves (frequency, amplitude, duration and form) served as the basis for the psychological aspects of sound, namely pitch, loudness, time and timbre. With this philosophy in mind, Seashore developed a musical talent test that consisted of pitch, intensity, time, memory, consonance, and rhythm. He believed that perceptual skills in discriminating subtle differences within these dimensions should provide information for musical ability. Based on this view, a musicality test called *The Seashore Measures of Musical Talents* was then developed to measure these areas of musical skills. In contrast to the modern technology that is available today, Seashore utilized mechanical machinery to create his musical stimuli (see Figure 2.3.).

Figure 2.3. A figure showing Seashore working on the tonoscope for creating music stimuli. Image from Cary (1992).

[7] Brief descriptions of the different publications are provided here. More information can be found in the Reference chapter. 1919a (original) and 1960 (second revision) were the manuals for the music test-battery; 1919b and 1938 were the published books, 1967 was the unaltered republication of 1938.

The Seashore Measures of Musical Talents was designed for listeners from the fifth grade[8] level upwards as that is when a group measurement can be made satisfactorily, and it is at an early enough point to arrange musical education. It was also recommended the test be repeated in the eighth grade before high school (Seashore, 1919a). In addition, Seashore and colleagues (1960) also claimed that the test had been successfully used amongst adults. Seashore reasoned that the music talent tests are:

> …based on a thorough analysis of musical talent; they are standardized for content that does not need to be changed; they give quantitative results which may be verified to a high degree of certainty; they are simple and as nearly self-operating as possible; they are adapted for group measurements; they take into account practice, training, age, and intelligence; they have a two fold value in the concrete information furnished, and in the training and pleasure gained from the critical hearing of musical elements. (Seashore, 1919a, p.3)

The Seashore Measures of Musical Talents consisted of six measurements of musical ability: pitch, loudness, rhythm, time, tonal memory and consonance in the original version of the test (Seashore, 1919a). The consonance test was revised and later replaced by timbre test (Seashore, Lewis, & Saetveit, 1960). The revised version of the test was presented on a single $33^{1/3}$ rpm Long Playing recording, replacing the earlier 78 rpm phonograph records. The length of the music test was about 30 minutes, but an hour was needed for the whole procedure including instruction and demonstrations (p.4). The subtest structures of the test are reported in Table 2.5.

[8] "Grade" is a term used in the American Education system to represent different age levels in school; equivalent to England's use of "Year".

Table 2.5. Summary of Seashore's Test

Subtests	No	Sound Type	Specification	Instruction
Pitch	50	Pure Tone (created from a beat-frequency oscillator through a circuit producing pure tones that were lacking in harmonics and overtones) Frequency: 500Hz Duration: 0.6 seconds	Differing in frequency: 2Hz, 3Hz, 4Hz, 5Hz, 8Hz, 12Hz, and 17Hz	Listeners were asked whether the second pure tone[1] was higher or lower in pitch compared to the first pure tone
Loudness	50	Pure Tone Frequency: 440Hz Duration: 0.6 seconds	Differing in decibel: 0.5dB, 1.0dB, 1.5dB, 2.0dB, 2.5dB, and 4dB.	Listeners were asked whether the second pure tone* was stronger or weaker compared to the first pure tone*
Rhythm	30	Pure Tone Frequency: 500Hz Tempo: 92 crotches per minute.	Items 1-10: A series of five-note patterns in 2/4 time. Items 11-20: A series of six-note pattern in 3/4 time. Items 21-30: A series seven-note patterns in 4/4 time.	Listeners were asked whether the two patterns in each pair were the same or different.

[1] Please note that several authors have used the term “tone” to refer to pure tone or complex tone. In order to present the nature of the stimuli more clearly, the term “tone” is replaced with “pure tone*” or “complex tone*”; the asterisk denotes that the original term was used in the manual and instruction was the term “tone”.

Subtests	No	Sound Type	Specification	Instruction
Time	50	Pure Tone Frequency: 500Hz Duration: A tape timing device was used to control the duration of the tone automatically with a pre-set schedule for time intervals.	Differing in duration in seconds: .30s, .20s, .15s, .125s, .10s, .075s, .05s	Listeners were asked whether the second pure tone* was longer or shorter compared to the first tone
Melody	30	Hammond Organ Each ten items had three, four and five notes. The range of the stimuli was eighteen chromatic steps upward of middle C, with constant intensity and a controlled tempo.	There was one note different in the second melody sequence	Listeners were asked to identify the “different note” by its number
Timbre	50	Complex tone (multiple pure tones) Fundamental frequency at 180Hz and its first five overtones	The timbre of the tone was varied by reciprocal alteration in the intensities of the third and fourth harmonics. The smallest intensity change in the harmonics was 0.7dB and the largest intensity change was 10dB.	Listeners were asked whether each pair of stimuli had the same or a different complex tone* timbre quality (p.4).

Note. No= Number of Items. Stimuli descriptions are taken from Seashore and colleagues (1960, p.3-4)

Methodology. *The Seashore Measures of Musical Talents* was played to listeners via loudspeakers whereupon the listeners were asked to record their responses on the provided IBM sheet (a scoring sheet). Columns were labelled on the IBM sheet alphabetically; the listeners were guided by instructions to fill up the columns as the test went along. If the listeners were found to have a poor and doubtful score, they were permitted a second trial of the test with the original answer sheet removed. Unfortunately, the criterion for "poor or doubtful" score was not given in Seashore's manual. The scoring of the test was calculated by the number of correct responses either by hand or by the IBM test scoring machine.

Reliability. The reliability of the *Seashore Measures of Musical Talents* was estimated by internal consistency coefficients (Kuder-Richardson formula 21) as shown in Table 2.6 (Seashore et al., 1960, p.7). Test-retest was not reported.

Validity. Seashore and colleagues (1960) reported that the Seashore's test was found to correlate with external criteria, for example musical achievement, such as in studies by Bienstock, Lundin and Farnum (cited in Seashore et al., 1960). Unfortunately these cited studies were either unpublished material or the procedure and result of such validation were not described in Seashore's manual. Such lack of sound psychometric info would probably have prompted later authors to criticize Seashore's test about his approach of isolating elements of mental functioning as musical ability (the "atomistic" tradition) (e.g., Karma, 1980; Gordon, 1965). Seashore's test also reported to show low correlation with actual music performance, leaving the validity of Seashore's test unclear (Henson & Wyke, 1982; Wyatt, 1939).

In response to these criticisms, Seashore (1967) argued that the internal validity of *Seashore Measures of Musical Talents* was well established and that it was inappropriate to validate them with fallible external criteria. He further argued, "I have been bombarded all these years by the omni- busists[10] for this type of validation, but I have persistently refused action on the ground that it had little or no significance" (p.384). The tests, he says, "represent the theory of specific measurements insofar as they conform to the two universal scientific sanctions, on the basis of which they were designed; namely, that (1) the factor under consideration must be isolated in order that

[10] Those who believe music should be analysed as a 'whole' rather than isolating each of its elements.

we may know exactly what it is that we are measuring, and that (2) the conclusion must be limited to the factors under control" (p.383). For example, "When we have measured the sense of pitch, that is, pitch discrimination, in the laboratory with high reliability, and we know that pitch was isolated from all other factors, no scientist will question but that we have measured 'pitch' " (p.7).

Despite these criticisms, Seashore's test has nevertheless had an important influence on the subsequent music test-batteries which will be discussed further.

Table 2.6. Coefficients of Reliability for *The Seashore Measures of Musical Talents*

Test	Grades		
	Grades 4-5	Grades 6-8	Grades 9-16
Pitch	.82	.84	.84
Loudness	.85	.82	.74
Rhythm	.67	.69	.64
Time	.72	.63	.71
Timbre	.55	.63	.68
Tonal Memory	.81	.84	.83

2.3.3 The Wing Standardised Tests of Musical Intelligence [1948, 1968, 1962][11]

Development of musical aptitude tests continued following Seashore (1919a) such as the implementation of a completely new battery, called The *Wing Musical Aptitude Test* (Wing, 1948). Wing (1968) commented that the purpose of this battery was:

> …to pick out musical bright children at about the age of transfer to the secondary schools in order to give them the opportunity, if they wished to avail themselves of it, of coaching in an orchestra instrument; the test, therefore, attempts to measure both acuity of musical hearing and sensitivity to performance. (p.83)

The *Wing Standardised Tests of Musical Intelligence (*Wing, 1968) attempted to be more inclusive of musical components that were created from "real" musical instruments, focusing more on the music aesthetic or appreciation ability compared to Seashore's (1938) tests which focused more on sensory ability. Wing (1968) elaborated:

[11] 1948 (first) and 1968 (second) were the manuals for the music test-battery; 1962 was a published article in journal.

> Music appreciation, which is distinguished from musical ability both by musicians and by psychologists, is the power to recognize or evaluate artistic merit in music; it involves the deliberate aesthetic judgment of music as it actually exists in compositions rather than ability to solve problems connected with the elementary materials of which music is composed. (p.2).

Wing (1968) also coined the term, "musical capacity", referring to the combination of musical ability and musical appreciation, claiming these two skills are *"qualities of the whole mind"* (p.3). For this reason, Wing's test was categorised into two main characteristics: examining the musical ability (i.e., discrimination skill) and examining musical appreciation skill (i.e., judging the rendition quality of the stimuli). The test consisted of seven subtests: chord analysis, memory, rhythmic accent, harmony, intensity and phrasing, and were presented on nine 10-inch records. It took about 50 minutes to an hour to administer. The score was the number of items correctly answered, and participants were asked to guess when they were in doubt about their answers (p.50). The test structures of Wing's music test are presented in Table 2.7.

Table 2.7. Summary of Wing's Test

Subtests	No	Sound Type	Specification	Instruction
Chord Analysis	20	Piano	Each trial consisted of a chord that consisted of two to four notes.	Listeners were asked to determine the number of notes in a chord and marked with an X the number of notes on an answer sheet
Pitch Change	30	Piano	Each trial consisted of a pair of chords that consisted of three to six notes. The second chord were either identical as the first chord, or one note of the chord differed by semitone or a tone.	Listeners were asked to mark “S” if the two chords were the same, “U” if the altered note moved up, and “D” if the altered note moved down.
Memory	30	Piano	Each trial had two pairs of melody sequences that comprised of 3-10 notes. The second melody was either identical as the first melody, or one note was differed in a semitone or a tone.	Listeners were asked whether the second tune was the Same as or Different from the first, and marked the position of the altered note if their answer was “Different” (p.51).
Rhythmic Accent	14	Piano	Stimuli specifications were not described in the manual.	Listeners were presented with two tunes and asked whether the second tune was the Same or Different (the accentuated notes were in a different place). They were asked to choose the style of the playing that they thought suited the tune better if they noticed a difference (p.52).
Harmony	14	Piano	Music stimuli were selected from Bach chorales to periods up to	Listeners were presented with two tunes and asked

Subtests	No	Sound Type	Specification	Instruction
			comparatively modern styles, excluding Jazz music as “this would be unlikely to yield examples of really good harmony” (p.37). Two tunes were presented to the listeners, in which one of the tunes was reharmonised. The difficulty of the test depended more on the “nature than on the number of the faults in the reharmonised version” (p.37). The faults refers to (p.37): 1. Definitely harsh combinations, such as the fourth and the diminished fifth, replaced the softer thirds and sixths. 2. The chords were made thin by omitting the third or the fifth, or both, and using the octave instead. 3. The chords were made unbalanced by wrong doubling, e.g. of the third instead of the root 4. There was a monotonous use of the same chord 5. Harmonic interests, such as the seventh or the ninth, were absent 6. The chord progression were weak 7. The progression of the parts was unmelodic 8. Bass sequences were absent	whether the second piece was the Same or Different. Different refers to when the notes of the left hand were different during the second piece. They were asked, if they noticed a difference, to choose the style of the playing that they thought suited the tune better (p.52).

Subtests	No	Sound Type	Specification	Instruction
Intensity	14	Piano	Stimuli specifications were not described in the manual.	Listeners were presented with two tunes and asked whether the second tune was the Same or Different. The louder and quieter portions were in different places. They were asked to choose the style of the playing that they thought suited the tune better, if they noticed a difference (p.52).
Phrasing	14	Piano	Stimuli specifications were not described in the manual.	Listeners were presented with two tunes and asked whether the second tune was the Same or Different. Different here is explained as “different groups of notes may be played with short sharp strokes, or so that they follow on smoothly, etc.” (p.52). They were asked to choose the style of the playing that they thought suited the tune better if they noticed a difference (p.52).

Note. No= Number of Items. Stimuli descriptions are taken from Wing (1968, p.51-52, 90-92)

Reliability. Reliability coefficients for the seven subtests were estimated with 100 music teachers using Split-Half analysis and were found to have coefficients of .90. Unfortunately the criteria of the music teachers and the test-procedure were not provided in the manual. The test-retest procedure was also tenuous as it exceeded more than the recommended time interval of 2 weeks (see Table 2.1). The result of the test-retest is reported here nevertheless. The correlation of test-retest was estimated two years apart with volunteering 19 year old listeners, r = .88; however a slightly lower coefficient was found in non-volunteers and children, r = .76 (Wing, 1968, p.87).

Validity. Wing (1968, p.88) reported the test was given to 223 junior musicians under training in the Royal Marine School of Music[12] who were then graded by their instructors into three groups: average, above average and below average. Wing reported "There was a positive and significant correlation between these gradings and the test result…" (p.88), unfortunately the exact coefficient value was not reported in detail.

2.3.4 The Drake Musical Aptitude [1957]

About a decade following Wing's first revision of the musicality test (Wing, 1948), Drake (1957) devised a more compact test that consisted of only *Musical Memory* and *Rhythm*. Drake claimed that these two skills are predominantly innate, independent from each other, and do not require learning and acquiring; more importantly that, "both components were determined through intensive and systematic analysis of the skills shown by successful performers in various fields of music" (Drake, 1957, p.4). The purpose of this test was to "provide measures of musical aptitude; to predict achievement in musical training" (p.2). He further emphasized that these tests are "limited to a measure of general musical aptitude…The tests do not measure a person's interest in music; neither do they measure highly specific aptitudes for particular kinds of musical performance" (p.11).

Drake (1957) claimed that the memory factor is essential in predicting musical talents. He even went so far as to say that "it can be stated with assurance that no great musician has ever had less than a phenomenal musical memory" (p.4). Rhythm perception, on the other hand, is a fundamental ability in estimating musical talents as it

12 G. de C. Newton, Selection of Junior Musicians for Royal Marine School of Music. An Evaluation of H.D. Wing's Test (Senior Psychologist's Department, British Admiralty, 1959), reported in Wing (1968).

generates the ability to maintain accurate rhythm under a number of highly distracting situations (i.e., dynamics, performance articulations). The *Rhythm test* was designed to measure “a bona fide response of maintaining a pre-determined tempo” (p.13).

The test-battery was recorded on a 12-inch, $33^{1/3}$rpm microgroove phonograph record. The entire battery takes 80mins (or two 40-minute sessions) to administer. It can be administered with participants ranging in age from elementary school children to adults. The score of the Drake test was based on the calculation of the test score that converted to percentile ranking through use of the norms based on 2390 participants for the musical memory subtest and 412 participants for the rhythm subtest (p.8, p.22). Drake’s test consists of two subtests, *Musical Memory* and *Rhythm* and each of these subtests consists of two forms, respectively A and B. The *Musical Memory* forms are equivalent; Form B of the rhythm test is more difficult than Form A. The condition whether to choose Form A or B is presented in Table 2.8.

Table 2.8 Testing Procedures for Drake Test

	Musical Memory	**Rhythm**
Trained Subjects (five or more years of musical training)	FORM A (or B)	FORM B only
Untrained Subjects (less than five years of musical training)	FORM A and B	FORM A and B

Note. Table reproduced from Drake (1957).

Musical Memory. The *Musical Memory* subtest consisted of 12 reference two-bar melodies with piano sound. Each listener was presented with a reference melody followed by a few comparison melodies with changes of key (K), time (T), note (N) or remaining the same (S). Listeners were asked to compare the comparison melodies with the reference melody and write down their answer (K, T, N or S) on an answer sheet. The number of comparison melodies increases in a stepwise pattern from two to seven melodies. For example, for each of the first two standard melodies there are two comparison melodies. For each of the subsequent pairs of standard melodies there are three comparison melodies. This pattern continues so that for the 11th and 12th standard melodies, there are seven comparison melodies (p.6).

Rhythm. The *Rhythm* subtest consists of 50 items in each form*,* as illustrated below*:*

Form A: Listeners were presented with a metronome clicking sound and a voice counting "one", "two", "three", "four". The voice and the metronome clicks stopped at the count of "four" but listeners were asked to continue counting (i.e., "five", "six", "seven" and so on) until a voice said "stop". Listeners were asked to write down the number of the count at the point when the voice said "stop". (p.7) – the number of the count when the voice said "stop" was always between 9 and 16.

Form B: Form B was similar to Form A except there was a distraction during the "silent part" which listeners were asked to ignore (p.5).

Reliability. The reliability of Drake's test was estimated using the "Odd-even" Split-Half method with several group ranges of 42 to 826 listeners. The reliability coefficients for musically trained listeners were found to be higher, preponderantly in the .90's as opposed to untrained listeners with the coefficient preponderantly in the .70's and .80's (Drake, 1958, p.17). Drake stated, "This evidence suggests that scores should be treated with greater caution when counselling with or selecting among untrained subjects" (p.11). He further commented, "These data indicate that test reliabilities tend to run higher for homogeneous musical groups than for wide heterogeneous group. One possible explanation for this unusual finding may be that answers obtained from homogeneous musical groups are less influenced by guessing" (p.17-18).

Validity. The validity coefficients of Drake's test were obtained from teachers' ratings prior to the test results of their students. These were obtained with 23 groups of teachers and students (9 to 103 participants in each group) using a seven-point rating scale based on three considerations: (1) the extent of the accuracy of the rater or judge to evaluate the individuals taking the test, (2) the ability of the rater to define true ability separate from length of musical training, general intelligence, personality, and effort, (3) the accuracy of the rater in using the rating scale provided. However, the questions of the rating scales were not provided in Drake's manual. The correlation between teachers' rating and Drake's test ranged from .31 to .91 (see Table 2.9), at which "a majority attaining a value greater than .58" (Drake, 1957, p.17).

Table 2.9. The Correlations Between Drake's Test and Teacher Ratings of Students' Performances

Group Number	N	Musical Memory A+B	Rhythm A+B	Rhythm A	Rhythm B
38	46	.67	-	-	-
39	26	.66	-	-	-
12 (School A)	103	.50	-	-	-
12 (School B)	47	.75	-	-	-
12 (School C)	59	.39	-	-	-
12 (School D)	11	.82	-	-	-
7 (Teacher A)	19	.32	-	-	-
7 (Teacher B)	24	.51	-	-	-
7 (Teacher C)	26	.77	-	-	-
28	66	.45	-	-	-
36 (ages 12-20)	38	.59	-	-	-
36 (ages 16-20)	41	.34	-	-	-
36 (ages 15-20)	39	.36	-	-	-
11	50	.91	-	-	.83
31	12	-	.78	.73	.78
17	26	-	.85	.82	.76
14	27	-	.59	.59	.67
24	9	-	-	.78	.83
5	43	-	.62	.58	.57
31	13	-	.41	-	-
31	58	-	.57	.45	.41
13	31	-	.31	.36	.47
25	25	-	.54	.31	.60

Note. Table reproduced from Drake (1957, p.16).

2.3.5 The Measures of Musical Abilities (MMA) by Arnold Bentley (1966)

A decade later, Arnold Bentley developed *The Measures of Musical Abilities (MMA)* at the University of Reading, England, in 1966. Bentley developed a new test because "the measurement of musical ability has not yet progressed beyond a rather rudimentary and unsatisfactory stage…we may be able to recognize it, or think we can, but we cannot as yet define it" (Bentley, 1966, p.18). He further argued that

> whether musical ability is a single whole or whether it is analysable is still very much a matter of speculation. But because no satisfactory means have been found of measuring the whole, this is no valid reason for not trying to discover what we can about some of the parts of the whole. (p.19)

Bentley's test was primarily designed for use with children of elementary school age. He claimed that this test battery was not designed to measure all attributes of musical aptitude, but rather to look into "the investigation of such abilities as are basic and essential to progress in active music-making as vocalist and instrumentalist- i.e., to the performance of music…we must devise tests that will reveal not so much what we think children should be able to do, but what in fact they can do" (Bentley, 1966, p.20). Bentley saw the need for a test that could be taken by even very young children. Thus, Bentley limited the numbers of tones in the tonal memory test to four on the grounds that it is within the span of digits of one hand as opposed to Wing's test that ranged up to 10 tones (Bentley, 1966).

The test-battery was presented on both sides of a ten-inch $33^{1/3}$ rpm disc record which lasted about twenty minutes (excluding asking the listeners to fill in their personal information). The score of Bentley's test was the number of items correctly answered. Bentley also asked his participants to choose the answer that they thought may be correct when they were not sure (p.54). The test structure of Bentley's test is presented in Table 2.10:

Table 2.10. Summary of Bentley's Test

Subtests	No	Sound Type	Stimuli Specification	Instruction
Pitch Discrimination	20	Pure Tone (generated by oscillator) Base Frequency: 440Hz Duration: 0.6 seconds	Differing in frequency: 3Hz to 26Hz.	Listeners were asked whether a second pure tone* was higher (U – for 'up') or lower (D – for 'down') or the Same (S) in pitch compared to the first pure tone* and to write their answer on an answer sheet.
Tonal Memory	10	Pipe Organ Tempo: 120 crotchet notes per minute	Each melody consisted of five notes. The positions of the altered notes were randomly distributed between the first and the fifth.	Listeners were asked to listen and count the number of the notes in the stimuli during the first playing, and were then asked to write down the position of the altered notes (whole-tone or semitone) that changed on second playing.
Rhythmic Memory	10	Pipe Organ Tempo: 72 pulses per minute	Each item consisted of a rhythmic pattern that contained one of four pulses or beats. The positions of the changed pulses were randomly distributed between the first and the fourth.	Listeners were asked to write "S" if the pattern was unchanged during repetition or the position of the beat that changed, if they noticed a difference.
Chord Analysis	20	Pipe Organ Duration: 3 seconds	Each item consisted of a group of two to four notes.	Listeners were asked to listen and identify the number of notes in each chord and write them down on the answer sheet.

Note. No= Number of Items. Stimuli descriptions were taken from Bentley (1966, p.57-62, 76-78)

Reliability. The test-retest procedure of Bentley's test also appeared to be tenuous by contemporary standards as the time interval between Time 1 and Time 2 was more than 2 weeks (see Table 2.1 for the recommended criteria). Nevertheless, the test-retest reliability was found to have the coefficient of .84 when the test was given to 90 boys & girls between the ages of 9-11 with a period of four months after the first test (Bentley, 1966, p.89). In a study conducted by Young (1979, p.77), the internal consistency reliability (N = 504) was reported using the split-half method, the Pitch test was found to have the coefficient of .65; Tonal Memory, .83; Chord Analysis, .74; Rhythmic Memory, .61 and the Composite Test Battery was .83.

Validity. A music exam that consisted of sight singing, singing melody from memory and melodic and rhythm dictation was used to validate Bentley's test. A correlation of .94 was found between the exam test and Bentley's test in seventy 11-12 years old (Bentley, 1966, p.86-87). Such high correlation (.94) is rather suspicious, showing a huge degree of content-overlap in Bentley's test and music exam items. Despite this suspicious correlation, Bentley's test was validated by Young (1979) using another music test. In a study that was conducted by Young (1979, p.79) with 504 junior high school students, it was found that the composite of Bentley's test had a moderate correlation with Gordon's Musical Aptitude Profile (MAP), $r = .58$ (see next section).

2.3.6 Gordon's Musical Aptitude Profile (MAP, 1965, 1967, 1995)[13]

Around the same period, Edwin Gordon who is known for his contribution in developing various levels of music tests, devised his first music test in 1965 - the Gordon's Musical Aptitude Profile (MAP).

The MAP (Gordon, 1995) focuses on basic factors in musical aptitude, for example aural perception, kinaesthetic musical feeling, and musical expression rather than looking into details of musical achievement. The test was designed for elementary and high school students "to act as an objective aid in the evaluation of students' music aptitudes so that the teacher can better provide for all students' individual musical

[13] 1967 was the subsequent journal publication about the test that provided discussion about its psychometric properties; 1995 is the 3rd revision manual that comes with a digitized version of the test.

needs" (Gordon, 1995, p.9) where he defined musical aptitude as the "potential to audiate" (p.56). Gordon further commented:

> Audiation, the foundation of music aptitude, is the ability to hear and **to comprehend** music for which the sound is not physically present (as in recall), is no longer physically present (as in listening), or may never had been physically present (as in creativity and improvisation) (p.8).

MAP consists of three divisions: Tonal Imagery, Rhythm Imagery, and Musical Sensitivity, with seven subtests (total 250 items). Melody and Harmony are the subtests of Tonal Imagery; Tempo and Meter are the subtests of Rhythm Imagery. Musical Sensitivity has three subtests: Phrasing, Balance and Style. An appraisal of musical expression and musical creativity are also included in this battery. The first edition of the test was recorded on high fidelity magnetic tape, which can be used on an ordinary tape recorder and the latest version (3rd version) of the test is stored on a CD format.

The scores for MAP subtests were calculated by transforming the raw score to a "standard score" (with a mean of 50 and a standard deviation of 10) that was provided in the manual (Gordon, 1995, p.48). The general rule of Gordon's test was that listeners were not to guess if unsure of the answer. The design of the test is as described in Table 2.11. The listeners were provided with an answer sheet to record their responses for all subtests. When unsure about the answer, they were asked to question mark (?) the columns.

Table 2.11. Summary of Gordon's Musical Aptitude Profile (MAP)

Subtests	No	Sound Type	Specification	Instruction
Tonal Imagery: Melody	40	Violin	The first phrase was known as the "musical question" followed by a "musical answer". The musical answer could be a melodic variation of the musical question or it could be a different melody. The melodies were composed in various tonalities (major, minor, Dorian, atonal, etc.) and meters (duple, triple, paired meter, unpaired meter).	Listeners were asked to mark "L" when the musical answer was a melodic variation of the musical question; and mark "D" when the musical answer was a different melody to the musical question.
Tonal Imagery: Harmony	40	Violin (Upper) and Cello (Lower)	The test design was similar to the melody test except that the upper melody stayed the same for the musical question and musical answer whilst the lower melody could be a variation of the musical question or totally different.	Listeners were asked to decide whether the musical answer was "like" or "different" to the musical question.
Rhythm Imagery: Tempo	40	Violin	The ending of the musical answer could be slower, faster or exactly the same as the ending of the musical question.	Same as the Harmony test
Rhythm Imagery: Meter	40	Violin	Usual meters (duple, triple, etc.) and unusual meters (paired, unpaired) were used.	Same as the Harmony test

Subtests	No	Sound Type	Specification	Instruction
Musical Sensitivity: Phrasing	30	Violin and Cello	Stimuli specifications were not described in the manual.	Listeners were asked which rendition was performed with better musical expression (1 or 2). The “correct” answer for the musical sensitivity subtests was chosen by the agreement of at least nine out of the ten carefully selected musicians.
Musical Sensitivity: Phrasing	30	Violin and Cello	Stimuli specifications were not described in the manual.	Listeners were asked which rendition was performed with better musical expression (1 or 2). The “correct” answer for the musical sensitivity subtests was chosen by the agreement of at least nine out of the ten carefully selected musicians.
Musical Sensitivity: Balance	30	Violin	Stimuli specifications were not described in the manual.	Listeners were asked which ending of the two best concluded the phrases with respect to both tone and rhythm.
Musical Sensitivity: Style	30	Violin	Stimuli specifications were not described in the manual.	Listeners were asked which tempo of the two phrases best suited the phrase.

Note. No= Number of Items. Stimuli descriptions were taken from Gordon (1995, p.1-2, 44-47)

Reliability. The reliability of the Gordon (1967)'s subtests was generally around .70's and .80's, and approximately .94 for the complete test (N=241). However the type of reliability analysis used was not reported. Like Wing (1968) and Bentley (1966), Gordon (1967) did not conduct test-retest procedure within the recommended 2 weeks time interval (see Table 2.1 for the recommended criteria). Rather, listeners were invited for a retest after three years of intensive instrumental training, the correlation was shown to be moderately strong, r = .77 (Gordon, 1967, p.54).

Validity. The validity coefficients of Gordon's MAP were obtained after the third year of the study (Gordon, 1967, p.54). It was measured by the rating of (a) a tape record of students' performances they had prepared in advance *with* teacher help; (b) a tape recorded of students' performances they had prepared in advance *without* teacher help; (c) a tape record of students' performances of sight-reading material. Teachers were also asked to rate students' musical progress compared to other students in the group (d). A musical achievement test was also designed to assess the ability of identifying music notation that associated with melodic, rhythmic and harmonic passages that was presented on a tape recording and the knowledge of musical terms and sign (e). In addition, two judges independently rated the students' recorded performances (f). The validity coefficients of Gordon's MAP (p.54) are presented in Table 2.12. Unfortunately, the criteria of choosing the judges, as well as the number of participants for the validation procedure was not provided in Gordon's manual or publications.

Table 2.12. Validity Correlation Coefficients of MAP

Tasks	Correlation Coefficients
Selection Prepared with Teacher Help[(a)]	.58
Selection Prepared Without Teacher Help[(b)]	.70
Sight Reading[(c)]	.70
All Performances Combined[(a+b+c)]	.68
Teachers' Rating[(d)]	.35
Achievement Test[(e)]	.71
All Criteria Combined [(a+b+c+d+e)]	.75
Judges' evaluations of students' tape-recorded performances[(f)]	.85-.87

2.3.7 Gordon's Advanced Measures of Music Audiation (AMMA, 1989, 1990, 2001a, 2008)[14]

Other than MAP, Gordon also developed the Advanced Measures of Music Audiation (AMMA) test that was specifically designed for high school students and college/university music and non-music majors. The purpose of AMMA was "to enable administrators to establish objective and realistic expectations for the music achievement of high school and college/university music and non music major" and "to efficiently and diagnostically adapt the music teaching in classrooms, ensembles, and private instruction to fit the individual musical differences among students" (Gordon, 2001a, p.13).

AMMA consisted of 30 questions played with an electric piano. The test was stored on a CD and took 20 minutes to administer. Similarly with MAP, two short musical phrases were presented to the listeners, a musical question followed by the musical answer. Listeners were asked to decide whether the pairs were the same or different. If different, they were also asked to decide whether the difference was attributed to Tonal or Rhythm Change (Gordon, 2001a). The scores (raw and percentile) for AMMA were generated automatically by the accompanying CD program.

Reliability. Split-Half reliability coefficients for 225 students were found to be .83 for tonal, .86 for Rhythm and .88 for Total score (Gordon, 1990, p.7). Test-retest with Grade 11/12 students (N=70) for Tonal was r = .79; r = .80 for Rhythm and r = .84 for the Composite score (Gordon, 2008, p.1).

Validity. The longitudinal predictive validity of the AMMA was administered using the students' AMMA scores and three judges' ratings of students' recorded performance of an etude piece. Again, the judges' backgrounds were not described. The correlation coefficients between the scores and ratings are presented below (Gordon, 1990, p.10).

[14] 1989 was the first book publication about the test; 1990 was the book publication about the predictive validation of the test; 2001(a) was a book publication about the differences between Gordon's series of test-batteries; 2008 was the computer version of the manual that came on a CD.

Table 2.13. Correlations between Judges' ratings and AMMA scores

AMMA	Judges		
	1	2	3
Tonal	.74	.76	.70
Rhythm	.71	.74	.69
Total	.80	.81	.76

Note. N=114

2.3.8 Musical Aptitude As The Ability to Structure Acoustic Material by Kai Karma (1973, 1975, 1980, 1984)[15]

In contrast to his precursors, Karma (1973) did not compartmentalize musical ability into separate domains (e.g., pitch, rhythm). Rather drawing on Gestalt psychology, Karma defined musical aptitude as “the ability to structure acoustic material” (Karma, 1984, p.28). Karma (1984) claimed that this approach to musicality measurement is “culture free” because musical styles, tonality, rhythm, harmony sensations are built on the basis of a psychological ability to structure.

He further suggested that “musical ability can be defined as the ability to conceive auditive patterns, i.e. sets of relations between tones” (p.28), which is similar to “spatial ability”, an auditive structuring ability. Karma also suggested “The repeated theme forms a primitive hierarchic structure: the whole is formed by grouping of groups” (p.28), a pair of systems which forms a process which he believed to be at the heart of general and basic musical properties. The basic structural idea of Karma’s (1984) test is summarised below:

Structure ability. There were 31 pairs of trials with repetitions of a simple 2-6 tone pattern to form a sequence (sequence length was 0.68 to 0.74 seconds with 0.16 seconds pause in between, Karma, 1975, p.8 - 9); the repetitions follow one another without pause and there is no indication of start and ending points. The listener was asked to find the repeated pattern mentally. Karma (1984) commented “the correctness of this mental grouping can be controlled in many ways: perhaps the best format is to repeat the pattern once more after a pause in the same or in a changed form. The alternatives

[15] 1973 was the first journal publication about the test (background theory and pilot study); 1975 was the subsequent journal publication (test-construction and results); 1980 was the subsequent journal publication (summary and conclusion); 1984 was the subsequent journal publication about the overview of the test.

are thus ‘same’ and ‘different’ (see Figure 2.4).

The difficulty of the items was adjusted by changing the sequence of the notes. The score of Karma test was based on the number of items correctly answered. Karma’s test was designed for listeners from seven years old to adult and more effectively to be used as a group test. In one of Karma’s studies (1980), it was found the difference between seven-year-olds and adults was about 10-15%. Karma reported that the correlation between his test and music training were modest, r = .01 and .23 (Karma, 1984, p.29).

Reliability. The reliability of the Karma’s test (Karma, 1975, p.13) was estimated with 300 listeners using Cronbach’s Alpha, r = .55 (latest revision). Test-retest for the latest revision was not reported, however previous revisions were found to be moderately correlated, r = .57 to .68.

Validity. Karma obtained the validity of his test by asking the instrument teachers to estimate their pupils’ musical aptitude without taking into account the students’ sensory capacities or the present skill attained by training (concurrent validity). The correlation between the two variables was r (54) = .53 (Karma, 1975, p.18).

Figure 2.4. An example of Karma's test. The top figure shows the base stimulus and test stimulus are the “same” whilst the bottom figure shows that they are “different” because of the different position of the accent. Image reproduced from Karma (1984).

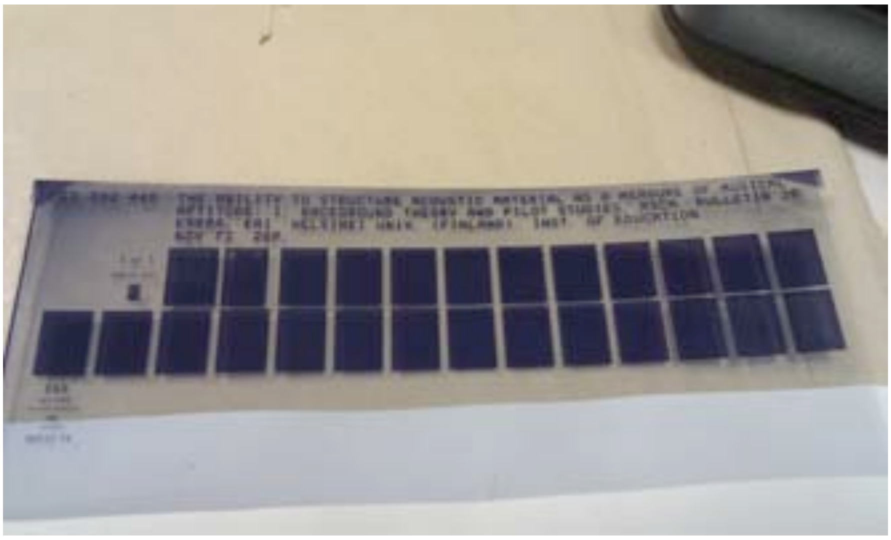


Figure 2.5. An example of Karma's publication that required special equipment to view (Karma, 1973)

2.3.9 Musical Ear Test (MET)

The Musical Ear Test (MET) is a contemporary music test that measures skills in melody and rhythm perception exclusively (Wallentin et al., 2010a). As such it is similar to Gordon’s AMMA, with the exception that its rhythm test uses simple variations in note durations using percussive sound whereas in Gordon’s test rhythmic variations are embedded in a melodic context. MET consists of 104 trials: 52 trials in Melody and another 52 trials in Rhythm that takes 18 minutes to complete in entirety. Both tests have equal numbers of Same and Different trials. The scores for MET are calculated by the percentage of the number of items correctly answered.

Melody. In the test, each melody contains three to eight tones, with tempo at 100 bpm. 13 of the Different trials contain pitch violation, and half consist of pitch and contour violation. There are 25 trials in non-diatonic tones, 20 trials in a major key and seven in a minor key (p.189).

Rhythm. Rhythm sequences consisted of four to eleven beats, using the sound of a woodblock. All rhythm sequences were at a tempo of 100 bpm. The Different trial consisted of one rhythm change, and the inclusion of triplets in 21 trials varied the complexity. 31 trials contained even subdivisions of the beat. Only 37 trials of the rhythm sequence started on a downbeat (p.189).

Reliability. The reliability coefficient was estimated using Cronbach's Alpha (N=60), which was found to be .85 for the full test; .82 for the melody subtest, .69 for the rhythm subtest and for 60 listeners (Wallentin et al., 2010b, p.705). Test-retest reliability was not reported.

Validity. A strong correlation coefficient of .89 was found between MET and the Imitation test which was used as a tool for evaluating the progress of students' performance at the 'rhythmic' departments of the academies of music in Denmark. It is also worth noting that the Imitation test has not been formally evaluated (Wallentin et al., 2010a, p.189).

2.4 Previous Auditory Ability Tests

Like individual differences in musical abilities, individual differences in auditory abilities are measured by a broad range of auditory discrimination and recognition tasks, and the aims of this research is to identify the number and nature of distinct auditory abilities, generally analysed using factor analysis. The auditory ability test-batteries were commonly used in problem with speech recognition with the hypothesis that problems with speech recognition are a consequence of limited temporal or spectral processing ability (Kidd et al., 2007).

2.4.1 Individual Differences in Auditory Abilities

Influenced by Karlin's (1942) research on auditory abilities, the Test of Basic Auditory Capabilities (TBAC) was developed to investigate listeners' skills with spectral-temporal patterns (Watson, Johnson, Lehman, Kelly, & Jensen, 1982a; Watson, Jensen, Foyle, Leek, & Goldgar, 1982b). This test battery was used in several studies (e.g., Christopherson & Humes, 1992; Watson & Miller, 1993; Drennan & Watson, 2001; Surprenant & Watson, 2001; Jakobson et al., 2003).

The TBAC consists of eight subtests. Six of the subtests were taken from an earlier 22-subtest version of the battery. These six subtests investigate the auditory ability area of tonal stimuli in three areas: Single-Tone discrimination (Test of frequency, intensity, and duration discrimination); Temporal Pattern Discrimination (Tests of Rhythm, Temporal Order, Tonal-Pattern Discrimination using multitone sequences (Johnson, Jensen, & Watson, 1980). Two extra tests, developed by Dubno and Levitt (1981), were added to the TBAC battery to investigate speech sounds. These speech tests looked at both syllable order and nonsense syllable identification. This battery was later replicated using three additional speech tests (identification of sentences, words and consonant vowels [cv]) (Surprenant & Watson, 2001). This study (Surprenant & Watson, 2001) found three factors; speech recognition; non-speech discrimination and temporal-order discrimination, supporting Karlin (1942) that performances between speech and nonspeech are independent.

TBAC was further expanded by Kidd, Watson and Gygi (2007), and became known as the Test of Basic Auditory Capabilities, Expanded (TBAC-E). This battery consists of 19 subtests, including the original TBAC with additional tests including: [1] Ripple-Noise Discrimination (Yost *et al.*, 1978); [2] Detection of Amplitude Modulation in Gaussian Noise, (Viemeister, 1979) and [3] Detection and Discrimination of Temporal Gaps (Glasberg & Moore, 1989; Snell *et al.*, 2002; Tyler *et al.*, 1982), as well as a recognition test for familiar non-speech sounds (e.g., dogs barking, doors slamming, cars starting (Kidd et al., 2007)

Factor Analysis. Kidd et al. (2007) conducted a principal component analysis using arcsine-transformed overall percentage correct scores for each of the 338 subjects and all 19 subtests. Four factors emerged from the analysis: *Amplitude modulation* factor, *Familiar sound* factor, *Loudness-Duration* factor and *Pitch-Time* factor. For the purpose of this research, only *Loudness-Duration* factor and *Pitch-Time* factor will be discussed.

Factor 1: Loudness- Duration:

Five subtests contributed to the Loudness-Duration factor: Intensity, Duration, Syllable ID, Pulse Train and SAM 200. This factor or commonality was proposed as the basis in the sensitivity to the overall energy change.

Factor 2: Pitch and Time Factor:

Six subtests contributed to the Pitch and Time Factor: The Pitch and Ripple-Noise Discrimination Tests, Temporal Order for Tones, Embedded Test Tone, Gap Discrimination, and Syllable Sequence. This factor emerged two types of pattern processing: spectral and spectral-temporal.

2.5 Musical Ability Tests for Special Populations

Musical ability test-batteries that developed by previous research (Section 2.3) were only suitable for healthy people with normal hearing. In order to cater the needs for special populations, some researchers have developed music batteries that customised to these group of people for a more accurate measurement. In particular, two test-batteries will be described in the following section: (1) The University of Washington Clinical Assessment of Music Perception Test (CAMP), a test battery that developed for people who have hearing deficit; and (2) Montreal Battery of Evaluation of Amusia (MBEA), a test-battery that designed for listeners who suffer from musical impairment.

2.5.1 The University of Washington Clinical Assessment of Music Perception Test (CAMP)

The University of Washington Clinical Assessment of Music Perception Test (CAMP) is a computerized Music Perception Test that assesses Pitch Direction Discrimination, Melody Recognition and Timbre Recognition ability of cochlear implant patients (Kang et al., 2009). The Pitch and Melody subtests stimuli were both complex tones that used digitally synthesized synthetic piano sounds. Listener scores on these subtests were based on the percentage of the correct answers.

Pitch. In the Pitch subtest, each listener was presented with tones in the range of 1 to 12 semitones. This subtest used a two-alternative forced, 1-up 1-down adaptive testing method (Levitt, 1971). Listeners were asked to indicate whether the first or second note had the higher pitch by clicking on the corresponding button.

Melody. The Melody subtest consisted of 12 commonly known melodies, including, “Happy Birthday”, “Jingle Bells”, “London Bridge” and other well-known tunes. All melodies were eighth notes played in an isochronous manner within octaves from middle C. Each melody was played three times (with melodies presented in a random order), and listeners were asked to identify the melodies by selecting the respective title from the answer choice.

Timbre. The Timbre subtest consisted of eight instruments of four major instrument classes (strings, brass, woodwinds, and percussion). All instruments played an identical sequence of five notes, C4-A4-F4-G4-C5. Similarly to the Melody Test, each melody was played three times, and again presented in a random order, and Listeners asked to identify the timbre by selecting the labelled icon of the instruments from the given answer choice.

Reliability. Forty-two cochlear implant users and 10 normal-hearing listeners took part in this study. The intraclass coefficients for pitch, melody and timbre subtests were .85, .92 and .69. Test-retest reliability was strong for the melody and the pitch subtests, and moderate for the timbre subtests (see Figure 2.6).

Validation. Melody and Timbre subtests were found to have moderate correlation with the Consonant Nucleus Consonant test (CNC), r = .47 to .50; Pitch subtest was also found to have moderate to strong correlation with CNC, r = -.66. It was also found that each music subtest moderately correlated with Speech Recognition Thresholds (SRTs) in Steady State Noise and Two-Talker Babble tests, r = -.42 to .58 (p.415 - 416).

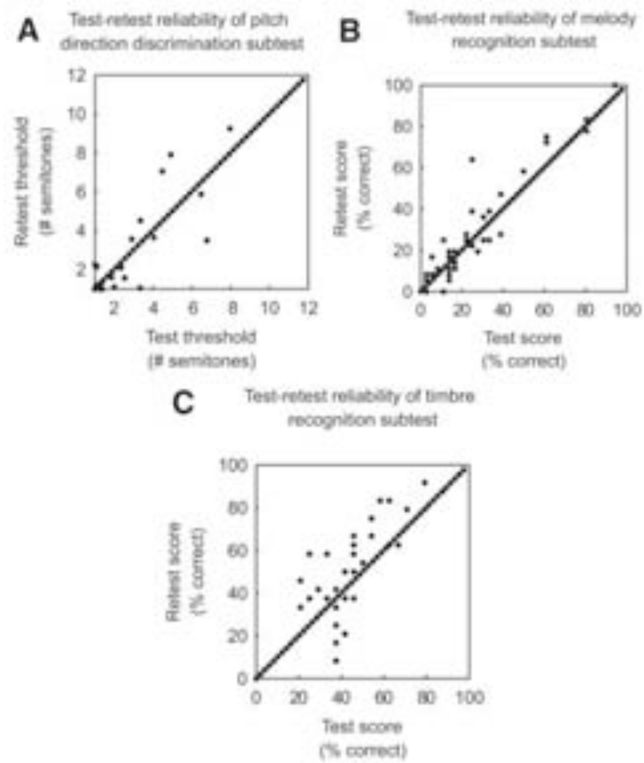


Figure 2.6. Test-Retest Correlation of CAMP. Pitch (A) and Melody (B) subtests shows strong correlations and Timbre subtest (C) show moderate to strong correlation. Image from Kang et al. (2010).

2.5.2 The Montreal Battery of Evaluation of Amusia (MBEA)

Whilst most people are able to enjoy music, a small percentage of listeners are unable to do so. One *amusic* described his/her experience of music listening as: "…music sounds like a rattling of pots and pans" (unattributed, as cited in Sacks, 2007).

This group of people suffer a musical impairment called *Amusia* or 'Tone-Deafness'. The term, 'Tone-Deaf' has been a 'misconception' amongst many people. For example, it was found there were 17% of Canadian undergraduates who believed themselves to suffer from this disorder (Cuddy, Balkwill, Peretz, & Holden, 2005) whilst there are actually only 4% of the population who actually suffer from this unusual musical impairment (Peretz & Hyde, 2003)

Tone-Deafness, more specifically known as Congenital Amusia, is a musical disorder in which a person is not able to discriminate between two musical notes, fails to recognize familiar melodies, and cannot sing in tune. Furthermore, this musical defect is not caused by lack of musical training, hearing loss or lack of environmental stimulation (Ayotte, Peretz & Hyde, 2002). The possibility that certain people are born with amusia has been envisaged for more than a century (Allen, 1878; Geschwind, 1984). Not until recent years a special test-battery- the Montreal Battery of Evaluation of Amusia (MBEA) was developed to screen criteria for amusia (Peretz, Champod & Hyde, 2003).

The MBEA (Peretz et al., 2003) consists of six tests of functioning musical components: contour, interval, scale, rhythm, meter and memory tests. All six tests are 30 novel melodies that follow the Western Tonal System. The computer-generated version of this test is played on an electric piano sound, generated by a sample playback digital synthesizer (Roland Sound Canvas SC50) and a MIDI sequencing program (Sequencer Plus Gold) triggers the sound. The whole test lasts approximately one hour and a half.

In one of the Amusia studies using the MBEA, it was found that the level of successful performance was quite high for all tests, with 90% correct in each test (i.e., 27 of 30 correct responses) (Peretz et al., 2003). Despite the high level of performance across the test, more than 80% of the listeners did not obtain perfect scores for each individual test.

Reliability. Test-retest reliability was assessed with a subgroup of 28 fire workers (training) 4 months after the initial session. The correlation between the two session was moderately high, $r = .75$, $p < .01$.

Validation. Gordon's subtests (melodic and meter imagery from the MAP test-battery) were used as a validation for the MBEA. It was found that both tests are positively correlated, r = .53. These two tests appear to have more in common with the Melody test, r = .41, compared to the Meter test, r = .23. Gordon's meter test also correlated with the MBEA's rhythm test, r = .43 (Peretz et al., 2003, p.68)

2.6 Summary of Previous Music and Auditory Tests

Musical ability research post 1900 began with Carl Seashore's work in 1919 (with a revised version in 1960), which had a great influence upon the development of follow-up research. Seashore's test focused on the sensory aspect of the auditory ability, indeed, his work is close to detecting the Just Noticeable Difference (JND) ability in humans when the physical aspect of auditory events changes (e.g., frequency, harmonics and amplitudes; Stern & Johnson, 2010). Despite Seashore's "atomistic" approach receiving much criticism over the years for not being "musical" (i.e., Wyatt, 1939; Karma, 1973), I feel this accusation is not justified because it often reflects a particular lack of understanding of Seashore's test. Criticisms of Seashore were often due to researchers having stereotypical and clichéd' attitudes towards what constitutes music. Often their concept of music was restricted to the application of vocal and traditional musical instruments only. However, as discussed in section 2.2, the concept of music has evolved over time and it has become regular practice to use non-traditional musical instruments (e.g., computers) as part of contemporary music making; non-conventional ways of playing musical instruments (e.g., playing the strings of the piano instead of playing on the piano keys); the use of basic auditory events as creative elements (e.g., pure tones and noises) and composing music in non-rigid structures (e.g., restricted to certain harmony and phrasing rules) are also accepted as part of contemporary music. Despite Seashore's test being regarded as the "ancient" music ability test, it can also be thought of as the most applicable test of all time because of its non-expiring musical component - the pure tone which is the basic component of all sounds.

Development of musical aptitude tests flourished in the subsequent decades. Wing (1948) developed the first music ability test which used standard musical instruments such as piano, and became known as the *Standardised Tests of Musical Intelligence* or *The Wing Musical Aptitude Test*. This took a different approach focusing more on musical aesthetic or appreciation ability rather than sensory acuity. Instead of making simple judgments of "same" or "different" when listening to a pair of stimuli, like Seashore's tests, Wing also asked his listeners to determine whether the altered stimulus in a "Different pair" (pitch, intensity, grouping change) sounded better than the reference stimulus.

Another decade passed in the 1950s before the *Drake Musical Aptitude Tests* (1954) was developed. Drake (1957) stated that musical aptitude "is *not* a unitary trait" (p.19) and concluded that musical aptitude comprises three factors - Musical Memory and Rhythm, are the more important factors, and, to a lesser extent, musical aptitude relies on Pitch Discrimination. It is worth noting that Drake's (1957) test of rhythm did not involve various rhythm structures which appear in other conventional rhythm tests. Rather, Drake's rhythm test was more like a tempo or timing test where listeners were asked to keep time as the stimuli played.

Another decade passed, and, in the 1960s, Bentley and Gordon developed their musical ability tests, respectively in 1966 and 1965. Bentley had the intention of developing his own test because there was no previous test devised specifically for younger children (i.e., Seashore and Wing's tests were designed for children from eight years old and above). Bentley (1966) argued, "Tests primarily for older children may or may not be appropriate for sorting out children several years younger" (p.43). Bentley's tests can be seen as a combination of Seashore and Wing's tests, examined pitch, tonal memory, chord analysis and rhythm perceptions. However, the stimuli-material design was simpler as Bentley's test was designed primarily for young children. For example, Bentley limited the number of tones in the Memory test to 4 tones simply because that is within the span of digits of one hand as opposed to Wing's test which ranged up to 10 tones (Bentley, 1966).

Edwin Gordon is probably one of the most significant contributors in the musical ability research domain as he developed at least nine different musicality tests over the years. All these tests focused on different ages and groups of listeners. Amongst the nine tests, Musical Aptitude Profile (MAP) and Advanced Measures of Music Audiation (AMMA) were designed for adolescents and adults. MAP is a three hour test (or three 50-minutes sessions) focusing on Tonal Imagery, Rhythm Imagery and Musical Sensitivity, mainly judging whether pairs of sound stimuli are the same or different. AMMA applied a similar method to MAP. However, instead of asking whether the pairs of stimuli are the same or different, listeners are asked to judge whether the differences are due to tonal or rhythm transformations. The difficulty level for AMMA is higher than MAP as it is intended for high school or university students.

Because the MAP and AMMA are available in sound formats that can be easily played by today's audio technology (see Table 2.4), they are more frequently used by modern researchers than other music tests (e.g. Bugos et al., 2007; Hayward & Eastlund Gromko, 2009; Peretz et al., 2003)

Kai Karma, a Finnish researcher argued that previous tests, such as Wing (1948) and Gordon (1965), were heavily based on culture or training, and that they told us little about untrained potential (Karma, 1984). He then developed a test based on the idea of ability in structural discrimination rather than categorising musical elements as specific domains (Melody, Rhythm) as previous tests did. He believed this method to be "objective and free from the effects of culture and training" (p.28). Despite this distinct belief in the idea of musical aptitude, Karma's (1973) test also employed the 'Same and Different' paradigm design as most of the previous studies (i.e., Seashore, 1919; Wing, 1948; Gordon, 1965). Karma's test, constructed in Finnish, remains untranslated and appears to have been used primarily in Finland.

The development of musical aptitude tests almost came to a complete halt in the subsequent three decades (see Table 1 for details). It was not until 2010 that a test conforming to contemporary standards of test construction was published. This test, called the Musical Ear Test (MET, Wallentin et al., 2001a), is quite similar to Gordon's AMMA, with the exception that its rhythm test uses simple variations in note durations using percussive sounds, whereas in Gordon's test rhythmic variations are embedded in a melodic context (Wallentin et al., 2010a). Although this test marks a welcome return to musical skill tests, several limitations might limit its use. By singling out one specific kind of rhythm skill and tonal memory, only a fraction of all the skills used in the perception of music are measured. Test-retest coefficients for the MET have not been reported.

In addition to musical ability research, an auditory ability measurement test, namely the Test of Basic Auditory Capabilities (TBAC, Kang et al., 2009), was developed by Watson and colleagues in 1982, and expanded by Kidd et al. in 2007. This battery is a follow-up study of Karlin's work (1942). TBAC examines auditory skills which include musical properties (e.g., pitch, timbre, rhythm) as well as generic auditory components such as speech and environment sounds. This auditory research primarily focused on individual differences and factorial studies in auditory abilities. Several "musical-related" factors emerged these studies: *Pitch-Quality*, *Loudness*, *Auditory Integral for Perceptual Mass, Loudness-Time*, *Loudness-Duration* and *Pitch-Time*. The aims of this type of auditory research targeted factorial studies of auditory features, and therefore they did not discuss the concept of musicality.

The Montreal Battery of Evaluation of Amusia (MBEA, Peretz et al., 2003) was intended to screen for Amusia or musical deficits, which is the opposite aim of all the other musicality tests which seek to uncover musical talents, rather than demonstrate their absence. Since 'amusia' is a disorder whereby listeners are not able to discriminate between two musical notes, this test focuses more on melody and tonal memory perception. MBEA is available as a web-test and has received good feedback.

Another established test that has been developed for a special population such as patients with cochlear implants was The University of Washington Clinical Assessment of Music Perception Test (CAMP). This test was designed to be easy in order to accommodate the impaired hearing ability of the patients.

There is evidence of several research efforts to create a tool to measure musical ability. The general aims of these tests were to measure children's potential to learn music (musical aptitude) before providing them with any suitable music education or training. This measurement was generally administered for children 8-9 years old, as that is believed to be the age in which musical ability becomes stable, although, there is little scientific evidence of such claims. In order to cope with the large number of children being tested, most of these tests were administered as a group test (Seashore, 1919; Wing 1960; Drake, 1953; Gordon, 1965; Bentley, 1966; Karma, 1973). The musical skill areas that have been investigated are Tonal Memory, Pitch, Rhythm, Time/Tempo, Meter, Timbre, Loudness, Harmony, Phrasing and Style. Some authors exercised an "atomistic' approach (Seashore, 1919), while others used real instruments either in monophonic or polyphonic form (i.e., Wing, 1939; Gordon, 1965).

2.6.1 The Limitations of Previous Tests Stated by the Original Test-Authors

The notion of using an intelligence test (such as Intelligence Quotient - IQ) as a method of predicting performance ability (e.g., Ree & Earles, 1992) has had a parallel impact on musical ability tests. For example Wing (1968) claimed that his test was designed to "pick out musically bright children" (p.83) or could be used "as an ability detector" (Wing, 1962, p.45); Drake (1957) stated the test was designed "to predict achievement in musical training" (p.2); Seashore (1919) stated the tests "are based on a thorough analysis of musical talent" (p.3).

As pointed out by Bentley (1966), the concept of musical ability is rather vague and yet to be understood. If any of the previous musical ability tools was to be used as the sole evaluation of an individual's musical abilities, the administrators of these music tests might have underestimated what music is and how different components of music structures stimulate the brain and transform these signals into an ability or skill. Several music aptitude test authors have also acknowledged the limitations of their tests and generally agreed that their tests cannot be used to measure musical ability as a whole. For example, Seashore (1919) stated, "These five measures do not constitute a complete survey of musical talent…but they do measure specific and fundamental traits of musical mind" (p.6). Seashore (1967) further commented, "Thus, if we measure the sense of pitch and we find that record made is in the 99 centile, the conclusion is not that the child is musical, but that he has an extraordinary sense of pitch, that he is superior in one of the scores of talents essential to musical success. He may be utterly incompetent in other talents" (p.304).

Drake (1954) stated his tests "…do not measure all factors of either the inherent or acquired types (e.g., specific skills depending on motor speed or coordination); neither are they measures of creative or interpretative abilities" (p.13).

Bentley (1966) stated, "We have not been concerned with the whole of musical development and education, nor with musical ability as a whole, but only with such aspects of it that we were able to measure in young children" (p.142).

Gordon (1995) stated, "Test scores, when considered without regard for human judgment and extra-musical factors, are of limited usefulness in the assessment of music aptitude and in the prediction of success in music endeavours" (p.8).

Karma (1980) also commented, "According to the experience of the present writer, the most obvious and probable danger in this phase is that the way the structures are composed and the way the test is given make the results more or less measures of intelligence instead of musical aptitude" (p.15).

Wallentin et al. (2010a) also stated "One possible limitation of this test is that it loads heavily on working memory" (p.94).

2.6.2 Limitations of Previous Tests as a Standardised Musical Ability Measurement

Despite the profusion of musical ability tests, they have not proved good enough for use in contemporary research for several reasons. First, most of these tests were conceived for use in music education to measure music aptitude before music training exposure, rather than for assessing interindividual differences in adults or for examining questions related to the nature of music perception (e.g., Seashore et al., 1960; Wing, 1948; Bentley, 1966; Gordon, 1965; Karma, 1973; see Table 2.4 for an overview).

For example, Karma (1984) - who drew on Gestalt psychology - did not compartmentalise musical ability into separate domains (e.g., pitch, rhythm). Karma's test perhaps serves as a good test to measure musical aptitude, but such test design does not allow us to identify individual perception ability, which was the aim of this thesis. Similarly, Bentley's (1966) tests were designed *primarily* for younger children where the test was intended to be very easy, and thus is unsuitable for the target adult audience of the current thesis. Other test limitations are described below:

2.6.2.1 Sound quality and sound consistency of stimuli material

The use of human performers in the recording of the auditory test materials has led to stimuli with undesirable inconsistencies in timing, timbre and intensity between standard and comparison trials, or even slips in the performances (e.g., Wing, 1948, Gordon, 1965, 1989). Audio-quality of the stimulus material is no longer up to the standard of contemporary listening habits (e.g., Olivetti Belardinelli, 1995). To the contemporary ear, many of the audio-samples used in previous tests sound impure or distorted, either due to limitations in recording techniques used during those days or to the quality of the audio material having degraded over time. This is problematic with scientific measurement requiring absolute control for better result interpretation.

2.6.2.2 Stimuli material design

Some of the previous batteries measured a combination of skills rather than the specific skill purportedly targeted by a given subtest. For example, in an attempt to make stimuli more "musical" than those devised by Seashore, Wing's "rhythmic accent test" (Wing, 1948) and Gordon's tempo test (Gordon, 1965) are presented in a melodic form, although the particular perceptual properties they assess relate to timing rather than melodic skills. Although Seashore had been criticized for using "musical atoms" rather than actual music in his test, some of his successors overlooked that the more musical features are added to a given stimulus, the harder it becomes to unambiguously attribute performance to one skill rather than to a combination skills (e.g., to timing and melodic skills rather than to timing skills only in the current example).

More importantly, most of these tests were constructed in accordance with Western classical music features, or in circumstances where the authors have shown biases towards certain types of musical styles. For example, Wing (1968, p.35) has shown a strong harmony preference towards Bach chorales (which was used in his Harmony subtest) and that he categorised as "very full and rich" and "some of the best examples of harmony in any style". Wing (1968) even purposefully excluded Jazz music in his stimuli because "this would be unlikely to yield examples of really good harmony...". This idea would surely spark passionate debates today as Jazz music is considered to have strong harmony elements (Levine, 1995; Geem, 2009; Nettles & Graft, 2002). For this reason, the reliability and validity standards that were obtained by these "cultural-based" tests were only limited to people who were exposed to Western music environment, and cannot be generalised for normal population that have diverse cultures of music (Gembris, 1997; Nettl, 2005).

2.6.2.3 Test design

There were also problems in the overall design of the batteries, either due to an unequal number of stimuli or their duration within a subtest (see Table 2.14), variations in answer format across subtests (e.g., Seashore et al, 1960, Bentley, 1966; Wing, 1968), or insufficient control of response bias and guessing patterns, which are today commonly addressed by coefficients such as *d'* (Macmillan & Creelman, 2005).

2.6.2.4 Reliability and validity

Several of these tests were devised before advanced methods of scale building became a routine requirement for psychological research. Thus, the procedures used for inferring test validity and reliability are tenuous by contemporary standards (see Section 2.3.1). Thus, reliability estimates were based on obsolete indicators of internal consistency, sometimes test-retest reliability was not examined (see Table 2.4), and with the possible exception of Gordon's batteries, the validation procedures were not described in sufficient detail to allow robust inferences about the tests' actual validity (Carson, 1998). For all these reasons, it is not surprising that most of the previous music aptitude batteries are no longer used today and indeed are very difficult to access (Carson, 1998).

There are more recent music-related test batteries that are based on rigorous principles of test construction and validation. However, these test-batteries were devised to capture deficits rather than individual differences in musical perception skills within the normal range. For example, the Montreal Battery Evaluation of Amusia (MBEA) was developed to assess amusia (Peretz, Champod & Hyde, 2003). Another battery, the Clinical Assessment of Music Perception (CAMP), was developed to evaluate the music perception of adults with cochlear implants (Kang et al., 2009). A recent test, the Musical Ear Test (MET), exclusively measures skills in melody and rhythm perception (Wallentin et al., 2010a). It is perhaps for this reason that investigators prefer to create their own tasks (e.g., Fabiani & Friberg, 2011; Geringer & Johnson, 2007), but by doing so their work cannot easily be compared across studies, and thus prevents the incremental accumulation knowledge that is vital to the progression of any branch of science.

The following chapter will provide an overview of the perception dimensions that were proposed and examined in this chapter, setting a background overview for the test-construction in this thesis.

Table 2.14. Summary of Subtests of Previous Music Battery

Subtest	Seashore et al. (1960)	Wing (1968)	Bentley (1966)	Drake (1957)	Karma (1975)	Gordon (1989, 1995)	Wallentin et al. (2010a)
Melody	30	30	10	12	-	15 (AMMA)	52
Rhythm	30	-	10	-	-	15 (AMMA)	52
Pitch	50	30	20	-	-	-	-
Loudness	50	14	-	-	-	-	-
Accent	-	14	-	-	-	-	-
Tempo	-	-	-	-	-	40 (MAP)	-
Timbre	50	-	-	-	-	-	-
Phrasing	-	14	-	-	-	30 (MAP)	-
Duration/Time	50	-	-	50	-	-	-
Chord Analysis	-	20	20	-	-	-	-
Harmony	-	14	-	-	-	40 (MAP)	-
Meter	-	-	-	-	-	40 (MAP)	-
Style	-	-	-	-	-	30 (MAP)	-
Balance	-	-	-	-	-	30 (MAP)	-
Structure	-	-	-	-	31	-	-

Notes. The number represents the “number of trials” in each subtest

CHAPTER 3

Basic Music Research Background for the Profile of Music Perception Skills (PROMS)

Chapter 1 and Chapter 2 have laid out the research background of the current work, as well as presenting the limitations of previous tests to be used as a standardised musical perception ability measurement tool.

Based on the literature review of previous tests, this chapter provides an overview of the perception dimensions that were proposed and examined in Chapter 2, setting a background overview for the test construction in the next chapter. First, a general understanding of music perception and cognition will be discussed, followed by a discussion about the perception dimensions from previous research. The perception dimensions that were used in the musical ability test, the definitions of the perception dimensions, how we perceive them, and how they are used in music, are reported. A brief discussion of the role of music perception in language will also be described.

3.1 Music Perception and Cognition

A sound, once thought as a stimulus that is physical, physiological and psychological (Stumpf, 1883, 2012), has a parallel influence on music that is configured from sounds, to be studied as theories of perception and cognition. Music perception and cognition is an area of cognitive psychology that serves to determine the mental mechanisms underlying our appreciation of music, either through music listening or other music activities (Justus & Bharucha, 2002). Understanding music perception and cognition allows us to comprehend how our mind attains awareness when perceiving incoming sounds, and how we interpret these sounds as music.

Several music perception domains have been proposed from previous musical ability research, namely pitch, melody, rhythm, meter, tempo, accent, timbre and consonance (see Chapter 2 for more details). Some research defines this category slightly differently based on the computational model of music feature extraction techniques: namely dynamics, timbre, harmony, register, rhythm, articulation and structure (Eerola, 2012). For the ease of linking the discussion with previous musical ability tests, the following discussion uses the more common music perception domain as proposed in Chapter 2. Seashore (1938, 1967) in particular grouped these perceptual domains as four fundamental aspects in music: the tonal, the temporal, the qualitative and the dynamic (p.4):

> The tonal aspects are primarily the outgrowth of pitch and timbre; the dynamic are usually reduced mainly to intensity; the temporal rest basically upon time but are greatly modified by intensity; the qualitative rest primarily upon timbre, but this is greatly modified by pitch, intensity, and time in sonance. (p.76).

3.1.1 Pitch Perception

The role of pitch in music is often considered as one of the most important dimensions in music perception (McDermott & Oxenham, 2008; Rasch & Plomp, 1999). The reason being that sounds with clear pitches are used in various combinations as the building blocks of music: for example, a combination of multiple pitches at *various* time periods is known as melody; likewise harmony is a combination of different pitches that occur *simultaneously* within a time period. In addition, a pitch system (i.e., diatonic-chromatic and the 12-tone systems), an interval (the pitch difference between two notes), and an octave (a frequency interval corresponding to a doubling in frequency) are also examples of music dimensions that based on organisation of pitches. For this reason, the background of basic pitch perception (pure tone and complex tone) is first described here, the application of other organisations of pitch perception will be discussed in the subsequent sections.

Pitch perception was one of the main perception dimensions frequently used in examining musical ability (e.g., Seashore et al., 1967; Wing, 1968; Bentley, 1966). The American National Standards Institute [ANSI](2004) defines pitch as "that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high" (p.1994). The ANSI's emphasis of the "auditory sensation…from low to high" (p.1994) suggests that pitch is a *subjective* (psychological) attribute of *sensation* experienced by a listener, therefore "pitch" does not refer to any physical property of a sound (Mathias, 2010, p. 9).

However, ANSI's definition of pitch purely as a one-dimensional quality (low to high) contradicts the notion in which other researchers regard pitch as being a multi-dimensional quality. For example Seashore (1938, 1967) suggests pitch is "qualitative" *and* that it denotes "highness and lowness" (p.53). Similarly, Shepard (1964) used the terms *pitch chroma*[16] as referring to the pitch "quality" or "colour", and *pitch height* to distinguish the octave differences (see Figure 3.1). This idea coincides with the scientific pitch notation system of using the combination of letter names (pitch chroma) and numbers[17] (pitch height) when representing pitch information (Young, 1939). For example, C_4 and C_3 share the same chroma (both are "C" note), but are different in height (C_4 is 12 semitones[an octave] higher than C_3).

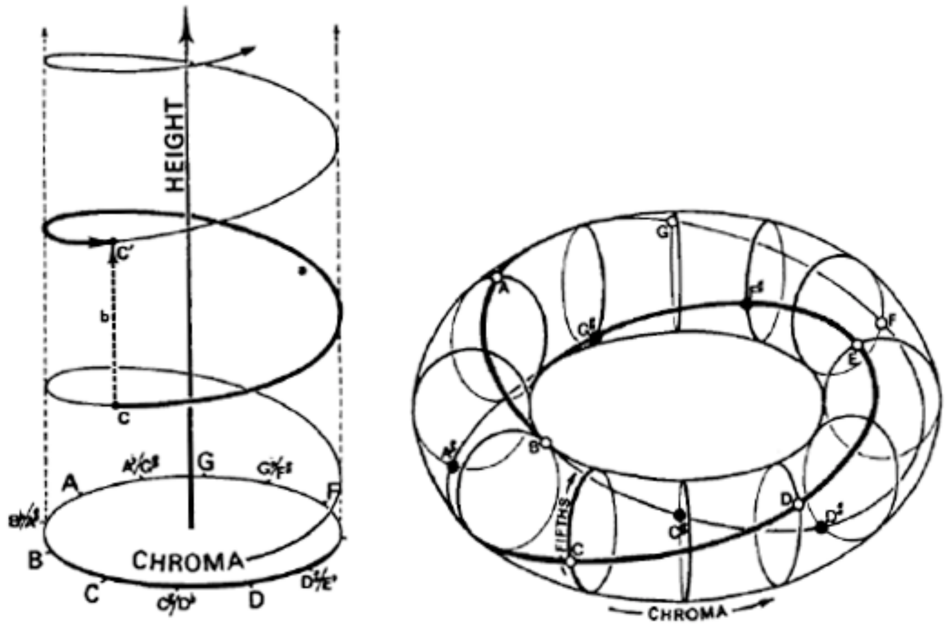


Figure 3.1. Two examples of pitch perception illustrating height and chroma. The left helical illustration demonstrates the pitch height and chroma. This represents the increment/decrement in pitch chroma spanning an octave. The verticality represents pitch height, or, pitch perception in octaves; the interval formed in 12 semitones (an octave 0). The right illustration depicts a four dimensional torus integrating a double helix to accommodate the concept of the circle of fifths. Image from Shepard (1982).

16 Chroma - originally from *khroma* (Greek) meaning "colour".
17 Reference C_0 – The number 0 represent the lowest possible audible frequency, at about 16Hz (Young, 1939), the subsequent number represent 12 semitones higher.

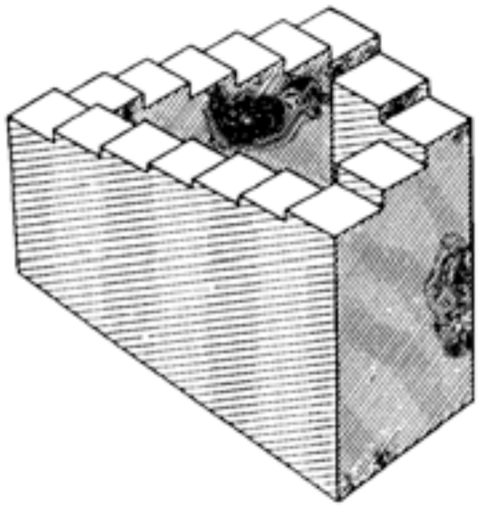


Figure 3.2. A visual presentation of perceiving pitch in a circular manner where there is no ending nor beginning. Image from Shepard (1964).

How do we perceive a pitch? As introduced in Chapter 2, humans perceive the vibration of a physical soundwave as a sound. When the soundwave vibrates at certain repetition rate, humans perceive it as a pitch. Earliest theory proposed that pitch perception of a harmonically complex tone[18] as being the result of the change of a fundamental frequency (Ohm, 1843; von Helmholtz, 1877). However this concept has been revisited by further research where it was found pitch is still "perceivable" even *without* the fundamental frequency. This phenomenon is referred to as *periodicity pitch or effect of the missing fundamental* (Licklider, 1954; Schouten, 1938; Schouten, Ritsma, & Lopes Cardozo, 1962), where it was proposed that listeners are able to perceive pitch information of a complex tone from other subsequent harmonics that produce the same common repetition rates as the fundamental frequency.

Listeners with normal healthy hearing generally can hear between 20Hz to 20kHz. However the upper limit may drop to 16khz by the age of 20, and this will continue to drop to 8000Hz by the age of 60, but the frequency in the lower limit is less affected compared to the upper limit (Kitterick, 2008). A musical note (e.g., C, D, E) is also often referred to as its fundamental frequency in its physical attribute (e.g., A_4=440Hz), and the musical note range of traditional musical instruments generally are much narrower, with a fundamental frequency of around 27Hz to 4186Hz (Seashore, 1967).

[18] Harmonically complex tones refer to signals that are composed of a set of pure tones whose frequencies are integer multiples of the lower frequency (*f, 2f, 3f, 4f;* 100Hz, 200Hz, 300Hz, 400Hz).

This range coincides with the range over that most listeners can perceive pitch by humans' auditory system. However, the possibility of using a wider frequency range in music is possible with the advent of modern technological hardware as a music creation tool, which in theory enables creation of an infinite frequency range. Such pitch restraints in musical instruments become obsolete where some musicians, such as Toshimaru Nakamura and Sachiko M, have created music inaudible to the human ear but "*perceivable*" by the vibration of the sound waves (refer to Chapter 2 for more details).

Musically trained listeners are generally found to have better pitch discrimination ability than musically untrained listeners: not only were musically trained listeners able to detect tiny pitch change more accurately, they were also able to detect pitch changes faster than non-musicians (Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2004; Kishon-Rabin, Amir, Vexler, & Zaltz, 2001). However, mismatch negativity (MMN) showed that there was no significant difference between musically-trained and non-musically trained listeners when they were asked to read a book and ignored the sound stimuli that played to them, suggesting musical expertise may exert its effects merely at attentive levels of processing (conscious accumulation of information from the environment), and not necessarily already at the preattentive level (Tervaniemi et al., 2004).

In general, pitch processing ability in music (e.g., single tone, melody, interval or harmony) was found to relate to language skill (Patel, 2008); in both adults (Schön, Magne, & Besson, 2004; Slevc & Miyake, 2006) and children (Magne, Schön, & Besson, 2006; Anvari et al., 2002). The musical pitch processing related language skill ranges from syntax (Patel et al., 1998; Patel, 2008), semantic (Besson & Faita, 1995; Koelsch et al., 2004), phonemes (Ross, Choi, & Purves, 2007; Gromko, 2005), intonation (Jiang et al., 2010; Ayotte et al., 2002; Foxton et al., 2004b; Patel et al., 1998; Lochy et al., 2004), second language ability (Slevc & Miyake, 2006; Milovanov & Tervaniemi, 2011) and prosody of speech (Magne et al., 2006; Thompson et al., 2004). To give an example, it was found music training improved the phonological awareness that involves the ability of discriminating, detecting and manipulating of linguistic sounds (Gromko, 2005) that is highly predictive of early reading ability (Bus & van IJzendoorn, 1999), this improvement also showed in children with dyslexia (Overy, 2003). In addition, because prosody in speech "has both a linguistic and an emotional function…and can be broadly defined at the abstract, phonological level, as the patterns of stress and intonation" (Magne et al., 2006, p.200), trained musicians (both adults and children) were found to perform better than non-musicians at identifying emotions (e.g., sadness, fear) that conveyed by spoken sentence (Thompson et al., 2004).

In summary, pitch processing skill is not only important to understand the function of various elements in music such as melody, interval or harmony (which will be discussed in details in the following section), it also allows us to understand its essential role in language processing.

3.1.2 Melody Perception

The word "melody" originates from the Greek word "*Melōidía*", meaning singing or chanting. Plack (2005) stated, "If you can show that a sound can produce melodies, then you can be sure that it has a pitch" (p.133), highlighting the role of pitch in melody perception. Melody is considered to be the "most salient voice in a piece of music (Purwins et al. 2008, p.173), and it is proposed as being one of the fundamental skills in music ability and hence it was often used in past musical ability research (Seashore, 1919a; Wing, 1948, Bentley, 1966; Gordon, 1965; Wallentin et al., 2010a). Melody can be seen or heard, to a degree, as pitch structures moving in time with relation to proximity, similarity and continuity between perceived events (Purwins et al. 2008). These structures can be a series of simple repeated patterns, continuity of pitch direction, up or down, or a group of harmonic structures (Plack, 2005). Human minds are able to group these structures based on their psychophysical characteristics, and form them into a melody or musical tune (Kim, Chai, Ricardo, & Barry, 2000; Garner, 1974; Garner & Clement, 1963; Restle, 1973; Vitz &Todd, 1969).

Melody sequences are thought to be encoded in two forms*: a contour code* and *interval code* (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004; Dowling, 1978; Deutsch, 1969). A melody contour is often used to describe how melody patterns "flow" in a musical passage, and can be described as a graphical representation of the up and down patterns of pitch changes (Foxton, Brown, Chambers, & Griffiths, 2004; Fujioka et al., 2004; Herndon, 1974). The concept of melody contour in music, is parallel to the concept of "intonation" in speech (Patel, 2008). On the other hand, the interval representation in a melody sequence (the exact ratio of pitch between successive tones), unlike contour, is specific to music only as a basic form of musical scales and harmony (Fujioka et al., 2004).

Melody perception is stronger depending on the recognition of the contour pattern in melody sequences, compared to recognising the intervals or the actual pitch changes between notes (Trehub, 2001; Bartlett & Dowling, 1980; Dowling, 1978). Specifically, it is easier to detect contour differences when the pitch differences occur at points of pitch-direction changes, in contrast to the points where pitch direction is maintained (Dyson & Watkins, 1984). This implies that the actual pitches of a melody sequence are perceived within a contour framework, and that contour perception is therefore necessary for successful performance on the actual pitch task (Foxton et al., 2004; Dyson & Watkins, 1984).

Listeners' tendencies of relying on contour cues in melody sequences are even apparent in melody transposition studies (Edworthy, 1985a; Dowling, 1978; Dowling & Fujitani, 1971). In these studies, listeners were asked to discriminate between a original melody (standard stimulus) and a comparison melody that was either (1) the exact transpositions where the intervals between the notes were retained, (2) the same contour but a different pattern of intervals, (3) a different contour from the original melody. The results showed that listeners often became confused by melodies that had the "same contour but a *different pattern* of intervals", showing their response was dominated by contour similarity (Dowling, 1978). However, the discrimination ability improved if there was a interrupted task or longer delay inbetween the discrimination tasks (i.e., which can be filled by other melodies or some distracting tasks), implying that listeners' preference on contour cue on melody sequences declined in favour of exact pitch intervals, as if the mental representation of pitch relation was being consolidated in terms of a sequence of interval categories (Dowling, 1991; Dowling & Bartlett, 1981; Dewitt & Crowder, 1986). This also suggests that a contour information is more efficiently retrievable in short-term memory, whilst pitch *interval* information is more effective in long-term memory tasks (Dowling & Bartlett, 1981). In general, the difficulty of melody recognition in transposed keys also increases if it is a distant key (Krumhansl & Kessler, 1982).

The expectation of the continuity of melody are two-fold according to the Implication-Realization (I-R) model (Narmour, 1992): *bottom-up* and *top-down* auditory pathways. Bottom-up melody expectancy is an innate ability as it acts as the biological grounding of grouping principles or organisation of the auditory scene (more details in Section 3.1.9); top-down melody expectancy is a learned ability due to long term exposure to music (Purwins et al., 2008).

Melody perception ability, can vary depending on listeners' background where musicians were found to be better than non-musicians. Bartlett & Dowling (1980) stated:

> Anyone with a good ear for music can sing, whistle, or hum familiar tunes correctly, that is, with the appropriate intervals among the notes. Yet, it is surprising for nonmusicians to extract precise interval information from unfamiliar melodies on a single hearing. (p.501)

With the advantages of musical training experience, musicians can retrieve the musical information stored in their memory during musical training (Foxton et al., 2004; Williams, 1980), process and ultimately extract the heard information into a meaningful context (Cuddy, Cohen, & Mewhart, 1981; Fujioka et al., 2004), as well as being able to expect the nature and timing of foreseeable melody events (Pfordresher, 2005; Margulis & Levine, 2004; Schellenberg et al., 2002; Carlsen, 1981; Jones, 1978). Although musicians tend to be better at melody recognition, they often fail to recognise unconventional melody structures when the “music theory rules” learned in previous training are violated (Cuddy et al., 1981). On the other hand, despite listeners with little musical training generally finding explicit music tasks more difficult than trained musicians (i.e., interval recognition), they are still able to respond to the relational information that is provided by a melody sequence (Cuddy et al., 1979; Dewar et al., 1977).

In summary, melody perception seems to involve rather complex musical processing due to the nature of musical pitch coordination that create different musical elements such as intervals, contours, scales, keys etc. More importantly, perception can vary between individuals as these elements change, for example, if these elements changed in parallel with a musical rule that musicians are familiar with, they would find this more advantageous than non-musicians.

3.1.3 Rhythm and Meter Perception

Rhythm perception is often seen as one of the fundamental music skills and thus was investigated in most previous musical ability research (e.g., Seashore, 1919a; Bentley, 1966; Gordon, 1989; Wallentin et al., 2010a). It is considered as “an interesting domain in which to explore participants’ discovery and processing of abstract generalizations based on concrete, temporally organized acoustic stimuli” (Fitch & Rosenfeld, 2007, p.44).

Despite rhythm perception receiving a considerable amount of attention in the empirical and theoretical work in the psychology of music, there has been no systematic and general agreement of the definition of rhythm itself. The problem lies in the fact that rhythm itself is not independent, and its structure is often the culmination of duration, time, grouping and meter. Nevertheless, some researchers have attempted to define rhythm in the context of grouping and meter (or pulse finding) (Seashore, 1967; Lerdahl & Jackendoff, 1983; Fitch & Rosenfeld, 2007; Huss et al., 2010; Purwins et al, 2008). For instance, Huss et al. (2010) commented: “Rhythm in music reflects at least two core aspects of temporal organisation, periodicity or metrical structure, and the patterning of musical events into similarly-structured groupings, or phrase structure” (p.675).

The first element, *grouping,* is concerned with phenomena that extend over specific durations, refers to the segmentation of music at various levels from a group of a few notes up to a large scale of music elements (Clarke, 1999). Bregman’s *Auditory Scene Analysis* (Bregman, 1990) proposed the idea that sound events can be perceived as a group based on their similarities, which is inspired by the Gestalt principle. For example, if a series of notes with the same time value occur repetitively, it will be heard as a “group”. The ability to remember “many small groups” is as strong as the ability to remember “an individual sound event without grouping”. This means that when one is listening to a series of notes, if he or she is able to group it rhythmically, they can remember as many bars[19] as one listening to an individual sound without rhythmic structures (Seashore, 1967).

Meter the second element, has received the most consideration since 1980 - a dominant influence of Western music metrical structure and of popular music (Clarke, 1999). Because meter perception is provoked by other rhythmic elements such as beat and accent, some aspects of this area will be discussed in this section. However, more specific details of these dimensions will be discussed in later sections.

Lerdahl and Jackendoff (1983) explained meter perception as a process of detecting and filtering *phenomenal and structural* accents as to discover the underlying periodicities: *phenomenal accents* refers to the points of local intensification caused by physical change of sound property such as intensity, note density, timbre or duration; and *structural accents* refers to the points of arrival or departure caused by tonality or cadence.

McAuley (2010), Eck (2002) and Povel (1981), on the other hand, explained meter

[19] bar is also know as “measure” in American music education system

perception based on the notion of beat. “Beat” refers to the strong position of a metrical framework (Grahn & Brett, 2007), sometimes it is also know as the terms “pulse” (Snyder & Krumhansl, 2001; Handel, 1989) or “tactus” (McAuley, 2010; Lerdahl & Jackendoff, 1983). Therefore, if listeners tap their feet along on the strong position of the beat in a metrical framework, they are tapping “on the beat”, a skill called *beat induction* (Eck, 2002). On the other hand, if listeners tap their feet along at the weak position of the metrical framework, they are tapping on the “syncopated” beat, also know as the “upbeat”[20]. Therefore, meter perception is an ability to sense the strong and weak beat “that arises from interactions among hierarchical levels in a pattern having nested periodic components” (Eck, 2002, p.1). Povel (1981) proposed that there are two steps of meter perception: (1) to detect the regular occurring accents of the beats; and (2) to identify the individual events as specific subdivisions of these beats into a smaller number (usually only two or three) of equal parts. Perceiving beat in music can be explained by (1) rule-based algorithms (i.e. beat is perceived based on the patterns of the first few audio events) and (2) oscillator model (i.e., the raw audio signal is segregated into different sub-bands, where the energy envelope of each output is then associated with a bank of comb resonators covering the tempo space) (Purwins et al. 2008, p.172; Eck, 2002).

This perception of beat can be perceived even in isochronous sequences with identical sonic events or perceptually “flat” sound events (Bolton, 1894), and that the beat rate does not necessarily correspond to the denominator of the time signature (Drake, Gros & Penel, 1999). Studies found both trained and non-trained musicians were able to accurately tap a low-ratio meter series such as 2:1 (crotchet-quaver-quaver) and 3:1 (dotted crotchet- quaver-quaver-quaver) (Povel, 1981). However the capability of reproducing more complex beats such as fifths, sixths, sevenths and eighths seems to be a trained factor that can only be perceived by musicians (Sternberg, Knoll, & Zukofsky, 1982).

The ability to perceive “beat” in rhythmic structure seems to be an untrained ability as recent research shows that babies from five months to two years old were found to move rhythmically to music (particularly to beat rather than melody), and they smile more if they are able to synchronise with the music more successfully (Zentner & Eerola, 2010). This perception of beat in music is also robust where studies has found

[20] sometimes it is also known as “off beat”

the listeners are able to synchronize to it even with music passages that have a moderate tempo fluctuation for expressive performance (Large & Palmer, 2002; Drake & Palmer, 1993). Despite the metric system having an essential role in Western music, it is subject to learning and effects of enculturation (Hannon & Trehub, 2005; Iverson & Patel, 2008), and that it is not always a feature or has a subordinate role in other styles of music, such as African, Indian, Javanese and Chinese music, and Gregorian chant (Carterrett & Kendall, 1999; Merriam, 1981; Becker, 1979; Kubik, 1962, 1979; Patel, 2008; Crocker, 2000).

Similarly in language research, the term “rhythm” also refers to systematic patterning of sounds in terms of timing, accent and grouping (Patel, 2008). In fact, to learn a new language requires more than just mastering its phoneme, vocabulary, and grammar; each language also has a rhythm that is part of its sonic structure that consists of patterns of timing, and accentuation that characterise the flow of syllables in sentence. And this is what makes each language unique (Patel, 2008).

In general, excelling in rhythm perception in music was found to facilitate language skill. For instance, the ability to perceive word boundaries in languages (e.g., the “foot” in English, the “syllable” in French or Spanish, the “mora” in Japanese) and phonological learning in language is a similar skill to perceiving metrical and rhythm grouping structure in music; and that perceiving the durations in musical rhythms is parallel as perceiving “categories” in linguistic (Clarke, 1987; Schulze, 1989) supporting the role of rhythm perception in language processing (Alcock et al., 2000; Huss et al., 2010; Cutler, 1996; Echols, 1966; Goswami, in press; Corriveau & Goswami, 2009; Auer, Couper-Kühlen, & Müller, 1999).

3.1.4 Tempo/Timing Perception

Tempo, meaning “time” in Italian, is mostly used in a musical context to indicate the pace or speed of a piece of music (McAuley, 2010). It is normally measured in Beats Per Minute (BPM), and is represented in this format (i.e., ♩ = 120). Tempo perception, in musical terms, is the perception of whether the pace of the musical structure undergoes any changes, that is to say noticing if it is getting slower or faster. Or, in other words, changes of tempo in music essentially change the time interval between two sound events within a given time. It is illustrated in the figure below.

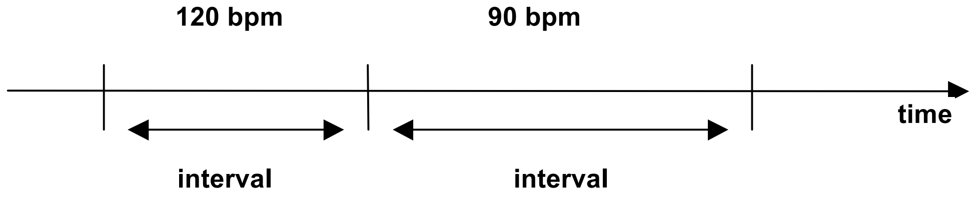


Figure 3.3. Overview of tempo structure. When the tempo is faster (120bpm), the distance/time between the two events is short, but when the tempo is slower (90bpm), the distance/time between two events is longer.

Tempo/timing perception has rarely been discussed in previous musical ability research, in fact only Gordon (1965) investigated this perception dimension. Psychoacoustical literature suggests that when listeners are asked to compare two time intervals (empty or filled by a sound), they are able to perceive the difference between the time intervals when it is at least 6% to 10% different from the standard duration (Interonset Interval (IOI) = 500ms) between the central processing of the onset and offset of the stimulus (Abel, 1972; Allan, 1979; Creelman, 1962; Getty, 1975, 1976; Small & Campbell, 1962; Woodrow, 1951). Similar evidence has been found when listeners were asked to detect the change in duration of *one* or *two* of the intervals contained in regular and rhythmic sequences (Drake, 1993; Drake, Botte, & Gérard, 1989; Gérard, Drake, & Botte, 1993; Halpern & Darwin, 1982; Hirsh, Monahan, Grant, & Singh, 1990; Monohan & Hirsh, 1990; van Noorden, 1975). Interestingly, the level of sensitivity is much higher, with variations of just 2% being detectable, when the listeners are asked to compare the *rate* or *tempo* of two isochronous sequences, which suggests that when listeners are listening to isochronous sequences they are able to extract multiple cues - one of which is the interval between two events; and the other is the interval between the two sequences (Michon, 1964).

Similarly, Drake and Botte (1993) also proposed two strategies for such tempo perception: (1) '*multiple interval-by-interval comparison (MIBIC)*', which is the duration comparison of each interval in the first sequence with the duration of each interval in the second sequence; (2) the '*multi-look*' (ML) - creation of a memory trace of the *average* duration and also the *degree of dispersion* of the intervals in the first sequence heard by the listener (Figure 3.4). When a series of notes is presented to the listeners, they first analyse the durations of each interval in the first sequence, then they analyse the durations of each interval in the second sequence *(MIBIC strategy)*. After that they will compare the duration difference between the two sequences, based on their memory of the first sequence being compared with the second *(ML strategy)*. That means the level of the sensitivity would be greater with the memory trace when there are more intervals in the first sequence (Drake & Botte, 1993). Drake and Botte's (1993) model was supported by later studies (Rousseau & Rousseau, 1996; McAuley & Kidd, 1998; McAuley & Jones, 2003).

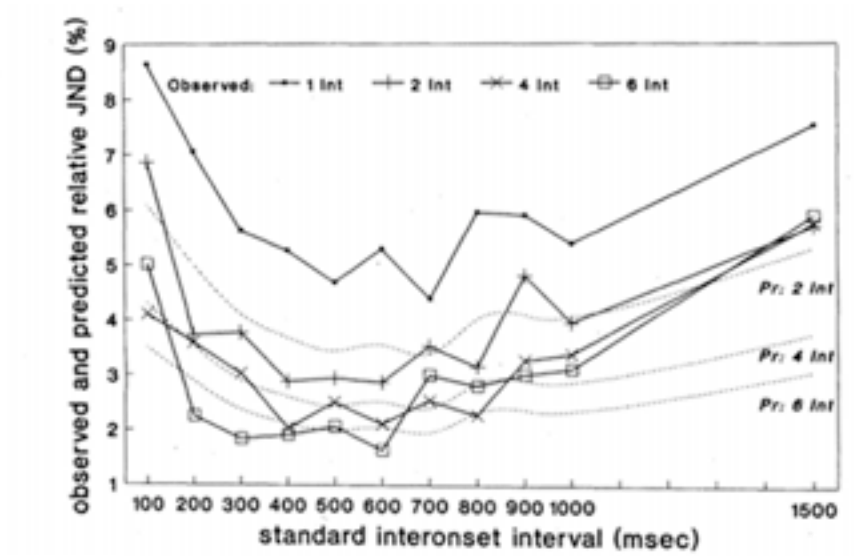


Figure 3.4. Mean relative JNDs for standard tempi ranging from 100-1500ms IOIs for sequences containing 2, 3, 5, 7 tones. Dotted curves are prediction for the multi-look mechanism. Image from Drake & Botte (1993).

Listeners are found to have a preference for a certain range of tempi, typically centered at around 100-120bpm (500-600ms)[21] (Moelants, 2002; McAuley et al., 2006), and this is also the range where people are the most accurate at making duration judgements (Eisler, 1976). Listeners often tap the tempo that are of simple divisions and multiplications of the main tempo (McKinney & Moelants, 2006; Drake, Jones & Baruch, 2000; Handel & Oshinsky, 1981; Parncutt, 1994). However, the tempo rate does not always correspond to the denominator of the time signature (Drake, Gros & Penel, 1999). Tempo perception can be affected by the articulation of a music passage, in particular, listeners tend to perceive staccato stimuli as increasing in tempo (Geringer, Madsen, MacLeod, & Droe, 2006), and slower tempo for music that has dynamic ac-

[21] This range appeared to be similar with the average duration between stressed syllables in languages (Dauer, 1983)

cents (McKinney & Moelants, 2006). In addition, listeners are better in discriminating slower tempi than faster tempi, and that they perceive the tempi of music stimuli to be slower than it actually is (Madsen, 1979). This explains that the listeners displayed greater discrimination ability when tempo decreases rather than when it increases, which resulted in the tendency to speed up during an actual performance. Musicians seem not to be aware of their tendency to increase the tempi while playing, or in other words, they do not perceive the tempi changes when they are increasing the tempo, which results in them often unintentionally speeding up during a performance.

Studies show that musicians have lower Just Noticeable-Differences (JNDs) (more sensitive) in tempo perception than non-musicians, which suggests that musical training may have a specific effect on the tempo perception (e.g., Drake & Botte, 1993, Figure 3.5; Madsen, 1979). Despite it being found drum machines or computer sequencers have shown to produce timing error (mean tempo deviation of 3.5%, with a standard deviation of 4.5 %), this deviation is lower than JND of most people (see Figure 3.5), therefore it often goes unnoticed, even in professional drummers (Perron, 1994). Based on the kinematic model of expressive timing (Honing, 2003, 2006b), musicians never perform the temporal structure in a 'perfectly regular' fashion, but rather perform with both intentional and unintentional variability for musical timing expression explicitly linking the law of physical motion in the real world to expressive timing in music performance (Drake & Palmer, 1993; Sloboda, 2000; Large & Palmer, 2002).

Nevertheless, ability to perceive tempo deviation plays an important role to keep in time whilst performing, which is considered a key musical performance skill (Gordon & Martin, 1993/1994; Drake, 1957). For instance, it was found musicians who exhibit great tempo perceptual sensitivity also exhibit stable time-tapping ability (Repp & Doggett, 2007).

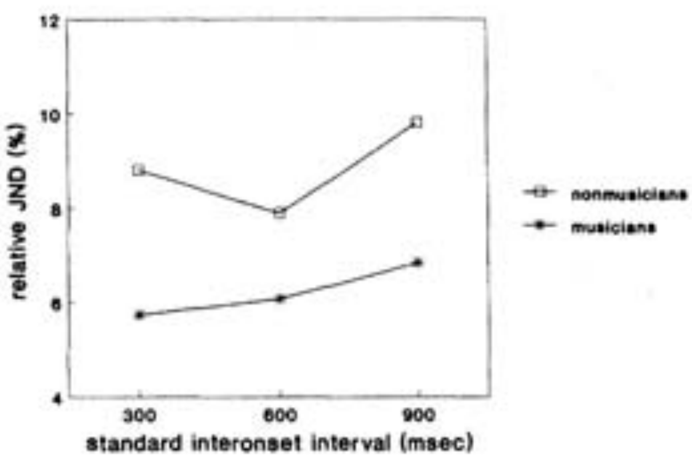


Figure 3.5. Difference in JNDs between musicians and non-musicians. Musicians showed to have lower (better) JND as compared to non-musicians. Image reproduced from Drake & Botte (1993).

3.1.5 Accent Perception

Accent perception was often neglected in musical ability research and in fact, only Wing (1968) investigated this perception ability. Although the definition of accent is varied across music-theoretic literature (Friberg & Battel, 2002), accent is defined "as a relatively salient event – that is, an event that attracts the attention of a listener" and that can be used as "an integral part of acoustic communication" (Parncutt, 2003, p.164).

There are two main categories of accent identified: *immanent accent* and *performance accents* (Parncutt, 2003; Lerdahl & Jackendoff, 1983). The distinction between these two relies on whether the accent expression is presented on a musical score or not. For instance, immanent accent refers to the *evidence from the score* itself and generally can be divided into four types: grouping, meter, melody, harmony and reductional accents (Parncutt, 2003, see Figure 3.6 for an example). If a performer plays the score nominally, the positions of these structures will be perceived as accented (Friberg & Battel, 2002). In contrast, performed accents are *added accents* from a performer, although it is also primarily used as a reinforcement of immanent accent. For example, in piano music, performed accent can be perceived when there are changes in timing, dynamics, articulation, and pedalling (Bisesi, Parncutt, & Friberg, 2011).

Accent perception is an important perceptual skill to identify different musical styles and expressions. As highlighted by Parncutt (2003), "…it is essential for clear understanding of the sound signal that the listener not only correctly decode the individual acoustic events (as given syllables or notes) but also get a feel for their importance relative to each other, to facilitate the inference of underlying structure and meaning" (p.164). For example, in most classical music (e.g., Baroque music) performers normally play the accent on the downbeat, in contrast, jazz musicians often stress the notes on the upbeat to give a "swing" feel - a distinctly different approach to performance expression between the two musical styles by using accent. To further elaborate Parncutt's comment on the note of "get a feel of their importance relative to each other"- the usage of accent to emphasise the musical features has been reported to project stronger emotional feeling to the listeners. For instance, Lindström (2003) reported whilst tense notes are normally related to the emotion of anger, when emphasis is put on a relatively tense note, it enhanced the intensity of the associated anger. Thus, this ability has an important role in musical performance particularly at a professional level (e.g., Geringer & Johnson, 2007; Sloboda, 2000).

The expression of emotion in music using accent, has striking parallels with the expression of emotion in speech (Gabrielsson & Juslin, 1996; Juslin & Sloboda, 2010). For instance, an *immanent* accent that performed from a music score, is parallel to reading from a written text in language. Of the five varieties of *immanent accent* in music that proposed by Parncutt (2003), *grouping* and *reductional* accents have clear relationships with speech.

Grouping accents in speech refers to the beginnings and endings of serial groups on hierarchical levels, for example phrases, sentences, paragraphs or even a whole story. *Reductional* accent refers to events that are important for syntactic and semantic reason, for example, an unusual word that appears in a text for the first time will normally be semantically important. Although *melodic* accent does not seem to have clear connection with speech in the Western countries, it is important in tonal language such as Chinese Mandarin and Cantonese in articulating the pitch of syllabus in tonal language (Parcutt, 2003).

Performed accent, the second model that was proposed by Parncutt (2003), on the other hand, draws attention to important syllables, words or phrases by manipulating the timing, loudness, pitch, or timbre parameters in a speech, highlighting the role of accent perception in speech communication and expression (Parncutt, 2003).

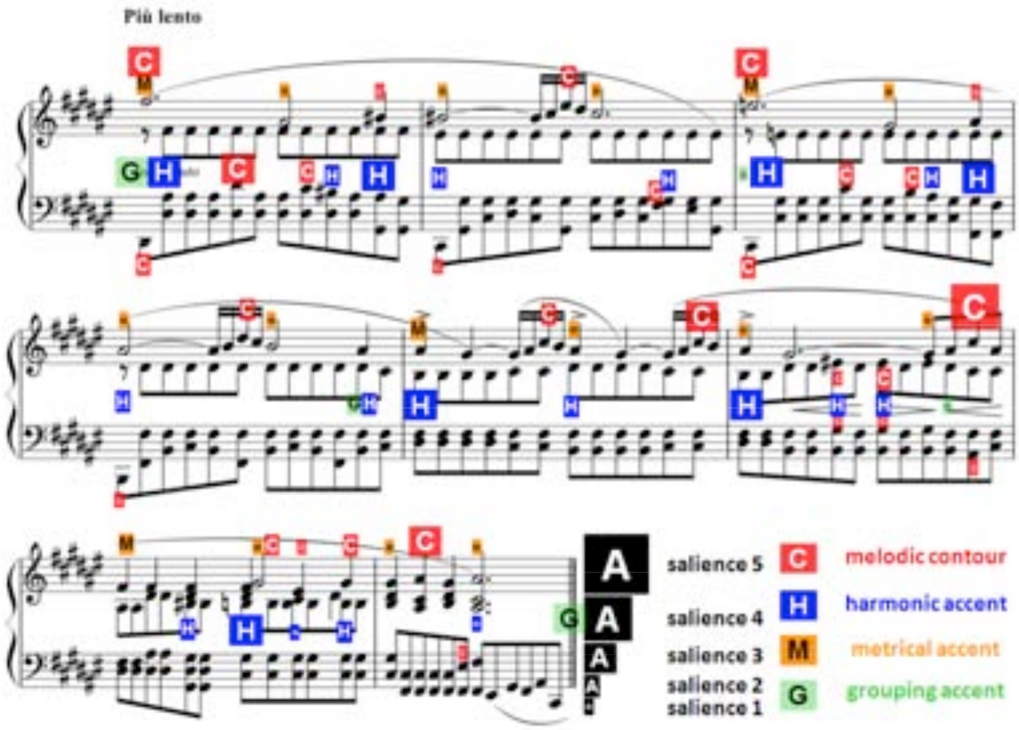


Figure 3.6. An example of immanent accent from Chopin's Prelude *Op.28 No.13.* The salience rating between 1 to 5 to each accent is indicated by the size of the square. Image from Bisesi, Parncutt, & Friberg (2011)

3.1.6 Timbre perception

Timbre perception is rather a neglected field in past musical ability research. It was examined at the primary stages of musical ability research by Seashore (1919a) and latterly by Kang et al. (2009) but little evidence was found inbetween these periods.

One of the major cues that enables listeners to distinguish between two sounds is the uniqueness in "colour" or "quality" of a sound, normally known as timbre. Earlier research on timbre perception by Grey (1977) suggested that timbre is multidimensional, unlike other auditory properties such as loudness; which is one dimensional and whose property changes can be represented by a single unit such as a decibel (dB). Grey (1977) asked listeners to rate the dissimilarity between pairs of sounds of recorded musical instruments and found three dimensions from these ratings; namely (1) the *spectral energy distribution*; (2) the presence of *synchronicity*; (3) the presence of *low-amplitude, high-frequency energy* (see Figure 3.7).

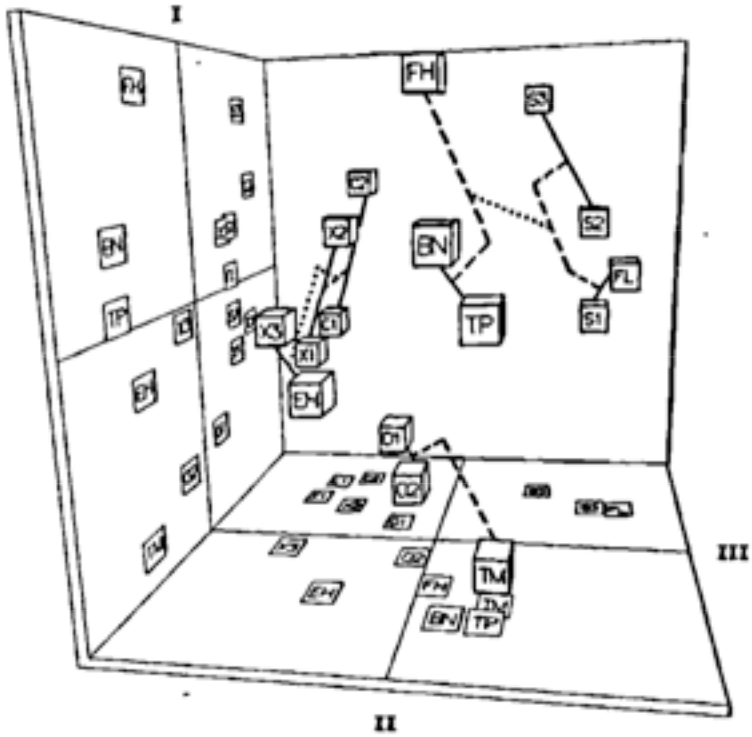


Figure 3.7. Timbre representation in three dimensions: Dimension I is the vertical axis, representing the spectral energy distribution of the sound. Dimension II is the horizontal axis, representing the spectral synchronicity (onset and offset of harmonics) of the sound; Dimension III is the depth axis, representing the preceding high frequency and low amplitude energy (mostly inharmonic). Abbreviations for musical instruments: O1, O2= oboes; C1, C2= clarinets; X1, X2, X3= saxophones; EH= English Horn; FH= French Horn; S1, S2, S3= strings; TP= trumpet; TM= trombone; FL= flute; BN= bassoon. Image adapted from Grey (1977)

The definition of timbre is often vague with regards to its multidimensional characteristic, Bregman (1990) even went so far as to comment, “We do not know how to define timbre, but it is not loudness and it is not pitch” (p.93). Despite this vagueness, it is commonly agreed that the timbre of a complex tone (i.e., musical instruments) is dependent upon its spectral (harmonics) and temporal envelope features (see Hall (1937) for a review). For instance, when a piano tone is played in reverse, it gives a totally different timbre, showing how timbre can be altered with different temporality even within the bounds of the same spectral property (Risset & Wessel, 1999).

New musical instruments are constantly being introduced that strive to show the vital role of timbre in music. “Imagine an electronic instrument that is small and portable, and can faithfully reproduce the crisp notes of a Steinway piano, the sweet sound of a Stradivarius violin or the brilliant tone of a trumpet” (SNS, 1994, para.1). Different synthesis techniques were explored when scientists attempted to create a musical instrument-like sound. For instance, Frequency Modulation (FM) synthesis was used in the creation of the Synclavier (early form of digital synthesizer), the Yamaha

DX-7 in 1983. In addition to the built-in sounds which are intended to imitate real musical instruments, users were also able to create their own sounds or timbres by programming the synthesizer (Chowning & Bristow, 1987).

A musical instrument mimicking technique- *sampling* was later introduced. This technique differed from previous synthesis and modelling techniques by storing recorded musical instrument sounds in digital memory, ready for playback. Therefore, this technique is arguably the most precise technique for reproduction of real instrument sounds (see Figure 3.8). However sampling setups are very expensive as high capacity computer memory is required as well as high-end microphones. Furthermore, studio and musician hire is costly, especially since high quality musicians are required to produce high quality samples (i.e., VSL, 2002). Sampling techniques also cannot cope well with anything but playing back exactly what was recorded in the sampling process, and thus is not very effective for live work where much real time control is needed by the performer (see Figure 3.8).

Another synthesis technique- *physical modelling,* first implemented by Hiller & Ruiz (1971), involves acoustic analysis of an instrument's behaviour and implements various mathematical models and equations to reproduce the instrument's sound (Wells, 2006; Cosi, De Poli., & Lauzzana, 1994; Castagne & Cadoz, 2003).

Digital synthesis and digital sampling solve the problem of there being a finite number of sounds in the natural world by introducing endless timbral possibilities. The evolution of timbre creation and analysis give further evidence of the important role that timbre plays in music.

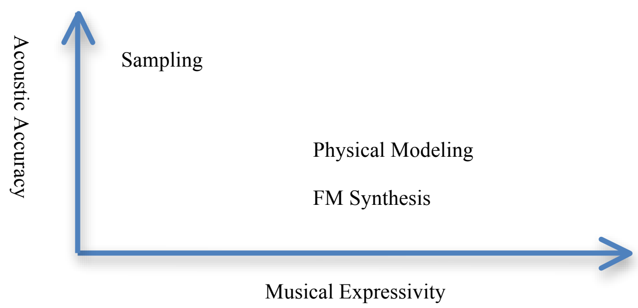


Figure 3.8. The characteristics of different synthesis and sampling techniques. Acoustic accuracy refers to how closely the timbre of the synthesis technique resembles actual musical instruments. Musical expressivity refers to the control of expression shape components such as phrasing or articulation change in voice or string instruments. Physical Modelling and FM synthesis have the advantage of controlling the expressivity components whilst Sampling produces better acoustic resemblance. Image reproduced from Risset & Wessel (1999).

Timbre perception enriches one's musical experience, and is indeed often used as a cue to identify expressive intent or emotion in music (Eerola, Alluri, & Ferrer, 2008). Recent research also suggested that the perception of emotion in music is independently affected by the timbre of musical instruments, even after controlling for other acoustic, cognitive and performance factors (Hailstone et al., 2009). This perception of emotion that is affected by timbre does not seem to be influenced by musical expertise (Hailstone et al., 2009), however, it was found the auditory brainstems of musicians respond more to their principal musical instruments (Strait et al., in press). Timbre perception skills are not only useful for discriminating between sounds or musical instruments; recent evidence suggests that timbre perception has also had a role in identifying dynamic feature perception in musical instruments (Fabiani & Friberg, 2011).

The role of timbre perception in language is highlighted by Patel (2008), "…there is no question that the primary dimension for organised sound contrasts in language is timbre" (p.50). Human voice is in fact the supreme instrument of timbral contrast. Maddieson (1984) conducted a survey of languages and revealed that human voice is capable of producing timbres corresponding to around 800 distinct phonemes, and this represents only phonemes know from extant languages. Therefore, understanding timbre in language would allow us to examine the relationship between musical and linguistic timbres of a culture. Patel and Iversen (2003) found that tabla drummers were able to recognise a variety of timbral contrasts between drums sounds, to the extent that they can find diverse ways of mapping these onto timbral contrasts in speech accurately. This study provides an example of how language exploits its special phonetic inventory to capture a musical timbral contrast.

3.1.7 Consonance and Dissonance (Tuning) Perception

The consonance subtest originally from Seashore's battery was taken out during the second revision and replaced by the timbre subtest in 1960. The reasons given by Seashore et al. (1967) were

> While this measure has been found very significant, in the diagnosis of talent it has certain defects which have been remedied in the forthcoming revision. These difficulties were that (1) the instructions, giving the directions for observation, were too involved, especially for children; (2) the judgment "Better" or "Worse" suggest agreeableness; and (3) there was a tendency to judge in terms of likes or dislikes. (p.132-133)

Consonance and dissonance or tuning perception is triggered by the particular qualities of musical intervals (Justus & Bharucha, 2002). Musical interval is related to the frequency ratio between two tones. For example, in an equal temperament system, the frequency ratio of 1.260 between two tones has the interval of major third (see Table 3.1). A major triad in a root position is the combination of three tones that has the intervals of major third (frequency ratio of 1.260) and minor third (frequency ratio of 1.189). The term "mistuned" or "out of tune" generally refers to when the frequency ratio of two or more tones is violated in the tuning system, giving a perception of "incorrect tuning" or dissonance (Burns, 1999). Tuning perception can thus be defined as "the ability of a listener to detect a mistuned harmonic in an otherwise periodic tone is representative of the capacity to segregate auditory entities on the basis of steady-state signal cues" (Hartmann, McAdams, & Smith, 1990, p.1712).

There are generally two theories explaining the cause of mistuned perception. Moore et al. (1985) proposed that when a harmonic partial is mistuned from a complex tone, (1) it can be heard as an separate entity, and (2) it gives the sensation of beating and roughness. The former generally occurs in lower harmonics and the latter in higher harmonics. In a follow-up study by Moore et al. (1986), listeners were asked to judge whether they heard 'a single complex tone with one pitch' (p.479) or a 'complex tone plus pure tone which did not "belong" to the complex' (p.479) in order to investigate the effect of fusion of mistuned harmonics in a complex tone. It was found that despite a mistuned harmonic, sufficient to be heard as a separate tone (mistuning about 1.3%-2%) being present, it had a significant contribution to the pitch perception of a complex tone as a whole. Hartmann, McAdams and Smith (1990) later identified that this is due to neural synchrony as listeners demonstrated losing the ability to segregate the mistuned harmonic at high frequencies (2.2kHz and 3.5kHz) when the synchronous neural firing vanishes. Later research by Darwin and Ciocca (1992) also found that if the onset time of the mistuned components is shifted 300ms prior to the complex tone, it is segregated from the complex tone and the contribution to the pitch perception as a whole disappears. Studies show that although sounds are generally heard as a single perceptual stream, when the components of the sound elements have identical onset and offset times and the amplitude modulates in the same way, long and steady-state sounds show a tendency towards sound segregation (Bregman, 1990; Moore, 1982).

Musicians are found to be sensitive to bitonal sensory roughness (dissonance) that is unnoticed by non-musicians (Plomp & Levelt, 1965; Wolpert, 1990, 2000). Recent neuroscience research suggests that musicians' brains (N2) respond strongly when listening to a consonant chord compared to a dissonant chord (whilst non-musicians used the same strategy for both sounds) suggesting that musicians are able to process the consonance chord according to its harmonic criteria (Minati et al., 2009). Burns (1999) commented that when using "… harmonic intervals composed of complex tones with rich spectra, such as those produced by most musical instruments, subjects in a laboratory situation can use the beats and roughness associated with sensory dissonance to distinguish intervals that are mistuned from small-integer frequency ratios, that is, from just intonation" (Burns, 1999, p.245). These findings support the *Auditory Scene Analysis* theory by Bregman (1990) who suggested that listeners are able to group auditory events in a meaningful way based on their prior knowledge.

To excel in tuning perception facilitates individuals to achieve and maintaining a "sound" that is appropriate in the desired context (Toulson et al., 2008). For example, to produce a more "blending" sound in music, a consistent tuning system is normally preferred within a group of musical instruments. In contrast, to achieve a more "stand out" or "dissonance" quality, composers or music performers would intentionally choose to combine a group of sounds that are harmonically not related in order to produce that desired effect.

Table 3.1 . Interval Comparison in the Equal Temperament System

Interval Name	Solfeggio	Letter notation	Equal Temperament	
			Frequency ratio	Cents
Unison	Do	C	1.000	0
Minor Second	Ri/Ra	C# /Db	1.059	100
Major Second	Re	D	1.222	200
Minor Third	Di/Me	D#/Eb	1.189	200
Major Third	Mi	E	1.260	400
Perfect Fourth	Fa	F	1.335	500
Augmented Fourth/Tritone	Fi/Se	F#/Gb	1.414	600
Perfect Fifth	So	G	1.498	700
Minor Sixth	Si/Le	G#/Ab	1.587	800
Major Sixth	La	A	1.682	900
Minor Seventh	Li/Te	A#/Bb	1.782	1000
Major Seventh	Ti	B	1.888	1100
Octave	Do	C	2.000	1200

Note. This table is adapted from Burns (1999, p.216).

3.1.8 Loudness Perception

Loudness (dynamic) perception, the sensitivity of perceiving the intensity change of a sound or piece of music, seems to be a less notable area in music research. It is often discussed in hearing research to measure how loudness varies due to age or hearing disability (e.g., Moore & Glasberg, 2004), rather than how it is used in a musical ability context. In fact, Seashore (1967) was the only researcher who employed a loudness test in measuring musical ability and highlighted the role of loudness perception ability in a musical context. He stated (1967), “Intensity discrimination measures the ability to hear differences in loudness and is therefore a measure of a person’s capacity for using loudness differences in every dynamic aspect of music…” (p.85). Seashore also proposed that the importance of loudness perception ability extended to timbre perception (p.76).

This “unpopularity” of loudness perception in music is probably due to loudness being a subjective concept compared to other attributes such as pitch, melody and timbre which can easily be determined or visualised graphically on a musical score. For example, pitch and melody are actually visualised as “musical notes” on a score, and timbre can be seen as the instrumental arrangement of a piece of music. Although loudness, or dynamic, is illustrated with dynamic expressions on a music score, for example, ***p*** ("piano" in Italian) refers to the meaning of 'soft' and ***f*** ("forte" in Italian) refers to 'loud', there is still no objective measurement of how "loud" is loud or how "soft" is soft.

Plack (2005) defined loudness as "the perceptual quantity most related to sound intensity" (p.115). The *objective measurement* of sound quantity and the *subjective experience* of sound intensity are not the same things, likewise the terms dB, dBSPL, and loudness have different meanings.

A decibel (dB) *alone* does not indicate any power, but rather is the logarithm unit of measure of the ratio between two power values, generally expressed as the power levels change from a fixed power as a reference

"Sound Pressure Level (SPL)", also called sound level, is a logarithm measure of the actual sound pressure of a sound *relative to a fixed reference value*, and is often measured in decibels (dB). For instance, the standard reference level of sound pressure level is 0 dB SPL, which is the threshold of hearing. A whisper-quiet library generally has a sound level of 30 dB SPL; and city traffic (inside a car) produces about 85 dB SPL (Carol, 2007). Sound pressure level is normally measured with a Sound Level Meter (see Figure 3.9). Loudness, on the other hand, is a *subjective* magnitude of sound (i.e., when the sound level is above about 40 dB SPL, a 10dB increase in level is required to produce a sensation of doubled loudness) (Plack, 2005).

Seashore (1967) suggested that loudness discrimination ability is frequently used in two motor capacities: *Matching Intensities* and *Differentiating Intensities*. *Matching Intensities* refers to the ability to control the intensity of a sound in order to match or reproduce the given intensity of a tone. *Differentiating intensities* refers to an ability similar to that of Matching Intensities but using voice and instruments. This capacity plays an important role in being able to produce artistic deviation in loudness or musical phrasing to shape and delineate musical ideas (i.e., in relation to their importance to the musical message), and to create expressive and emotive qualities that add drama to a musical moment (Olsen, Stevens, & Tardieu, 2010; Coutinho & Cangelosi, 2011).

Studies show that children are able to make judgments of loudness very easily at the age of four (Riley & Mckee, 1963; Williams, Sievers & Hattwick, 1932) compared to other tasks such as rhythm or pitch. This is not surprising as the concept of "loud" and "soft" are commonly used in daily life in a non-music context, such as in speech or environmental sound. Loudness perception in musical context also seems to be unrelated to musical experience, and also is unaffected by other music elements such as pitch, rhythm and harmony (Geringer, 1993). Similarly in a study by Karlin (1942), it was also found that different pitches or frequencies do not affect loudness discrimination ability. What strikes perception researchers is that non-music majors were found to perform better in loudness judgment compared to music majors (Geringer, 1995). It was also found that non-music majors tend to focus on dynamic change in a piece of music compared to musicians who focus more on other music elements (pitch, rhythm) compared to dynamic/loudness (Flowers, 1984). This seems to fit Bregman's (1990) Auditory Scene Analysis (ASA) theory which states that listeners choose what they want to hear based on their prior experience. Loudness perception is also a dimensional perception skill (i.e., is the sound louder or softer?). It has been reported that listeners were better at detecting intensity change when the stimuli were presented in the negative direction (decrescendo) (Moore, 1981; Geringer, 1991). This finding may have explained the tendency for listeners to be less sensitive when a piece of music gets louder compared to when music gets softer in a performance.

Another usage of loudness perception is for sound localisation. For instance, listeners localize sound based on two main cues, interaural time difference (ITD) and interaural level difference (ILD), also known as interaural intensity difference (IID). This skill is not only used for listeners to localise the direction of sound, but is also often used in sound engineering to "position sound" or "musical instruments" in "the right space" to create a virtual placement of live music performances, for example, instrument tracks with higher volume levels (louder) are often perceived as "closer" to the audience or listeners. Likewise, conductors and music producers play a similar role to sound engineers by having the skill of being able to use loudness perception ability to control different musical elements/ instruments. This allows them to coax listeners to focus in on the specific parts of the music. For example, the soloist in a music album or live performance normally plays louder or is 'brought forward in the mix' to draw the attention of the listeners. Therefore, understanding loudness perception ability might be one of the crucial steps in understanding more about musical perception ability.

In conclusion, the application of loudness perception in daily life, including in a musical context, is probably greater than other musically specific attributes such as melody and rhythm. Only Seashore (1967) mentioned the application of loudness dynamic in a musical performance context and this has been supported by recent research that has found loudness perception to be important in detecting the dynamics of musical instruments (Fabiani & Friberg, 2011), further strengthening the importance of loudness in a musical context.

Figure 3.9. An example of using the sound level meter to measure the sound pressure level in the headphones.

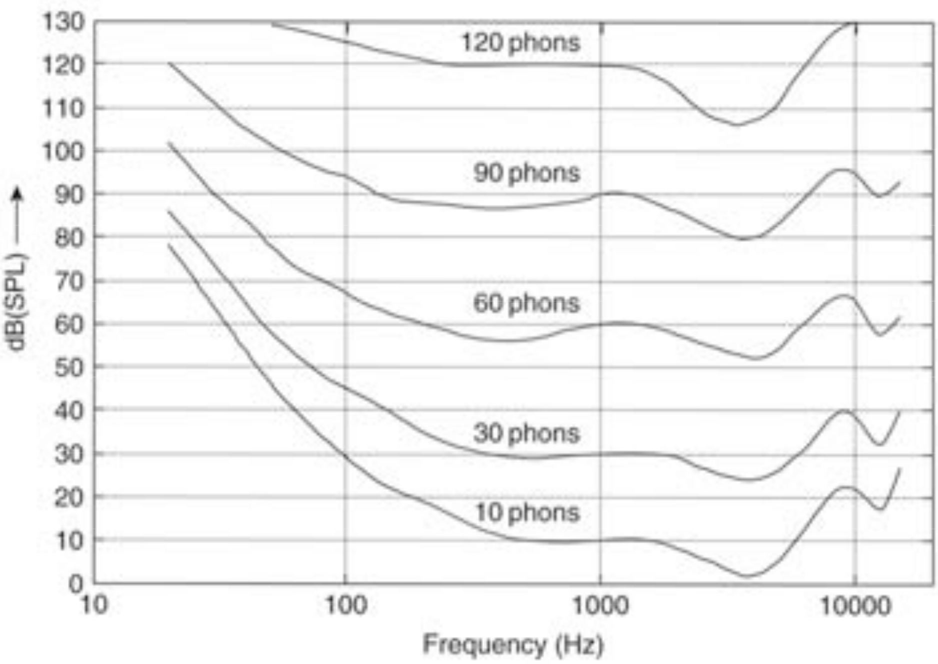


Figure 3.10. Equal loudness contour for the human ears. The loudness of human listening varies over different frequencies. More intensity is required to be able to be able to perceive at a lower range (<100hz) compared to a higher range (>1000Hz). Image from Howard & Angus (2006).

3.1.9 Structural Perception

In contrast to his precursors, Karma (1973) did not compartmentalise musical ability into separate domains (e.g., pitch, rhythm). Rather, drawing on Gestalt psychology, Karma defined musical aptitude as “the ability to structure acoustic material” (Karma, 1984, p.28) where he argued that all musical styles, tonality, rhythm, harmony sensations are built on the basis of a psychological ability to structure. Because most music perception research does not generally categorise music based on the structural processing notion proposed by Karma (1973), the following discussion is based on models from Bregman (1990) and Purwins et al (2008) which have used structural processing notions to discuss how humans process auditory and music information. This is an alternative theory on music processing in contrast to music perception research that relies on psychophysical changes of music components as presented in sections 3.1.1 to 3.1.7).

The Auditory Scene Analysis (ASA) theory by Bregman (1990) proposed that listeners tend to group auditory events based on their unique structures, which is closely linked to Karma’s concept of structure perception and will be discussed below.

The human auditory system, particularly the cochlea, has the function of breaking down sound waves, originating from different sources, into different frequencies, amplitudes and durations. Different sound components are grouped together and then assigned to the appropriate sound source. Bregman (1990) has termed this whole process Auditory Scene Analysis (ASA) - a perceptual process that decomposes complex acoustic signals. The auditory grouping and segregation could be achieved by two distinguishable processes - *Primitive Process* and *Schema-Driven Process*.

Primitive process prioritises innate experience; it requires neither past learning nor voluntary attention. "The primitive processes of scene analysis seem to employ a strategy of first breaking down the incoming array of energy into a large number of separate analyses" (Bregman, 1990, p.641). Effects of primitive segregation are often symmetrical. By way of contrast, a schema-driven process is based on regularity and familiarity. For instance, a drummer tends to focus on percussive grooves whilst listening to a song; backing vocalists pay more attention to the backing vocals in a recording; we spot ourselves in a picture of hundreds of people. Bregman stated that the result of a schema-driven process is always asymmetry. However, this type of auditory scene analysis could be switched from a primitive to a schema-driven process. "Effects could be different if listeners are trying to focus their attention on the tones instead of one of the streams, the effects of frequency separation are different. The frequency separation of the high from the low tones need only exceed some small amount (a few semitones in the case of two alternating tones) before the target sequence can be followed by attention" (Bregman 1990, p. 643). The ASA's (Bregman, 1990) and Purwins et al. (2008) models of structure grouping are presented below:

Proximity. Elements that are placed close together tend to be perceived as a group. For example, pure tones with similar frequencies (pitch); complex tones with similar fundamental frequencies tend to be grouped into single streams (close distances between auditory features, e.g., frequency, amplitude, onset time).

Similarity. Components with similar attributes tend to be grouped together (similar to proximity, but refers to the properties of a sound and is normally multidimensional, e.g. the instruments in the Strings family [e.g., violin, viola, cello] have similar timbres).

Good Continuation. Continuation occurs whenever the elements of the pattern establish an implied direction. (Components move smoothly in frequency [Scale] or amplitude over time [Crescendo]).

Closure. We tend to make our experience complete a space that is not completely enclosed (good continuation with interruption inbetween, for example we could still perceive the continuation of a scale even with a short musical *rest* inbetween two notes)

Common Fate. Objects tend to group together if they vary together over time (frequency components with similar changes in time, e.g. ensemble playing or choirs singing).

Figure & Ground. We perceive one object as the primary source, and regard all others as background/secondary source (e.g., sound tracks with dialogues [figure] and background music [ground]).

3.2 Summary

"Music notation and theory suggest that pitch and time play a more important role than loudness and timbre in the building of musical structures" (Parncutt, 2003, p.165). The above discussion has revealed that the perception dimensions examined by previous musical ability research were also mainly based on pitch and time properties, either on their own such as pitch or rhythm perception, or as a combination of pitch and time such as melody perception. However, tempo perception - a time-based perception was rarely investigated. The importance of perceiving loudness change was also often overlooked, where timbre perception, tuning/consonance perception, and accent perception were rarely investigated despite their importance in musical ability being highlighted by several studies. Karma (1973) on the other hand, suggested that the heart of perception ability lies in structure ability, hence he did not compartmentalise test-stimuli based on perception domains as other authors have done. In summary, the above discussion has provided evidence of the essential roles of these musical perceptual dimensions in music, as well as presented its close relationship with language and speech processing.

The following chapters will provide the rationale in the selection of musical perception dimensions that are used in this thesis based on the perception review in this chapter, as well as documenting the initial design and construction of the test-battery.

CHAPTER 4

Test Structure and Stimuli Creation, and Some Preliminary Findings

Chapter 1 provided an overview of the research background of this thesis as well as the aims of the research. Chapter 2 discussed the concept of sound and music as well as reviewing a substantial amount of musical ability research that was prominent within the last century. Chapter 3 presented an overview discussion of the music perception dimensions that were examined in the previous musical ability.

This chapter serves to present the development of a music test battery. There are two studies described in this chapter. Study 1 was a pilot study using a same-different paradigm, along with 2AFC and 3AFC as the answer choice. The rationale of the subtest selection as well as the test construction is also presented. The main purpose of Study 1 was to present the test-construction as well as to provide a preliminary overview of the test-battery psychometrics properties. Study 2 discusses the revision of the test-battery with a new test-paradigm, the "double reference playback with confidence rating scheme" that takes into account better stimuli encoding and confidence rating in the answer choice. A new subtest was also introduced in Study 2 (accent subtest). Findings and implications of these two studies are reported.

4.1 Study 1 (Preliminary Study)

Chapter 1 and 2 introduced the background of musical ability research and provided evidence of how little understanding there is of musical ability and the lack of a suitable musical ability assessment tool for contemporary research. However, there is no agreement on how musical ability might be best measured with objective tasks. This is in part a result of the complexities involved in defining ability as discussed in Chapter 1, and to an even greater extent, the complexities of music.

As discussed in Chapter 2, the assumptions that were used in music psychology tended to refer to Western classical music (Vink, 2001). Others (i.e., non-researchers) may think that The Beatles' songs or current forms of popular music are epitomes of "music", however these are only selective exemplars of an almost endless spectrum of musical varieties, as ethnomusicologists will readily point out (Nettl, 2005). For example, the use of functional harmony varies strongly across musical styles and systems (Carterett & Kendall, 1999). Central to the works cited, it plays a negligible role in Indian classical music, or Central African drumming music. In addition, since the days of Debussy and Varèse over a century ago, much of modern Western art music is either atonal, or relies only occasionally on functional harmony. Musical systems and styles also vary considerably in the emphasis they place on the tight rhythmic structuring referred to as meter (Brown & Jordania, 2011; Carterrett & Kendall, 1999; Stevens & Tolbert, 2009; or see Chapter 3, section 3.1.3).

It is also difficult to measure musical ability that involves aesthetic input such as "creativity". This is because creativity is often limited to a special genre or field, with specific directions of interest, in regards to training that individuals receive, as well as other environmental aspects such as the influence of an inspiring personality, the specificity of physical environment, general intellectual background and the fundamental disposition of the individuals (Révész, 1953, p.143).

Some researchers proposed a series of universals in music processing, that include perceptual principles of grouping and segmentation (see Chapter 3, Section 3.1.9), hierarchical organization and relational processing (tonal and temporal), musical expectancies, implicit knowledge of musical structures, temporal expectancies, synchrony and entrainment, and multimodal processes and integration (movement perception, i.e., dance) (Stevens & Tolbert, 2009).

Thus, a fundamental question in developing tasks for assessing musical ability is whether the tasks are supposed to test the comprehension of a specific, culturally evolved musical system, or the ease of processing elementary patterns of rhythm and sound found across various musical systems and traditions. The aim of this thesis was to devise a test that prioritizes the latter. More specifically, this research only examines one of the identified musical abilities (Boyle, 1992) - *music perception (or listening) ability,* which serves as the fundamental ability of many music activities such as performance (Prinz, 1990; Geringer & Johnson, 2007; Sloboda, 2000).

In order to fill the current gap in musical ability tests for normal or general adult populations, a new music test-battery should meet the following four criteria:

(1) The test should examine a wider ranger of musical perceptual skills than previous tests, which tended to focus on tonal memory and rhythmic skills only
(2) The perceptual skills measured in the subtests should be examined with the greatest possible specificity.
(3) The test should be applicable to all listeners regardless of their music backgrounds
(4) The test's validity, reliability and stimulus design should be equivalent to those of contemporary standards.

These goals made it necessary to confine the musical material to relatively elementary patterns of sound, pitch and rhythm. At first, fully-formed music may seem to be the natural stimulus material to test musical ability, but there are three difficulties in using complex musical stimuli. First, complex musical stimuli consist of several perceptual features at once, which potentially undermine the *specificity* of a subtest. Second, they cannot be neutral with regards to musical *system* or *style*, thereby conferring an advantage to listeners that are familiar with the type of music being used. Third, tasks involving extensive or complex musical material may be easier to process for *musically trained* individuals compared to untrained individuals, thereby giving an undue advantage to the former. Thus, a good performance may simply reflect the extent of training or familiarity with music of an individual subject rather than inherent musical ability in the sense of a potential for effective learning and processing of music.

Although the relevance of confining musical material to relatively elementary patterns of sound, pitch and rhythm as the processing of music, may not be instantaneously obvious, the psychological literature is replete with examples showing that performance on tasks tapping elemental, lower-level skills is related to higher-level abilities. For example, single letter knowledge and phoneme discrimination are sensitive predictors of broader measures of linguistic proficiency, such as reading ability (e.g., Hulford, 1990, Kirby, Parilla, & Pfeiffer, 2003; Muter, Hulme, Snowling, & Stevenson, 2004; or see Chapter 3). Similarly, the Raven Progressive Matrices test is one of the most sensitive measures of general mental ability, including level of educational achievement, numerical ability and language proficiency (e.g., Frey & Detterman, 2004; Jensen, 1998; Pajares & Kranzler, 1995). Yet, the trials consist of abstract visual patterns that would not necessarily strike one as obvious items for the measurement of general intelligence.

In the selection of musical dimensions, those that are relatively universal were prioritized over others that are highly salient in certain types of music but play only a negligible role in others (Brown & Jordania, 2011; Nettl, 2005; Stevens & Tolbert, 2009). Thus, tasks which tapped perceptual sensitivity to functional harmony or to the tight rhythmic organization referred to as meter were not examined, rather tasks tapping perceptual sensitivity to tempo, timbre, rhythm, pitch, and melody were prioritized.

4.2.1 Method

4.2.1.1 Listeners

24 listeners in total with self-reported normal hearing took part in Study 1. Twelve of these listeners volunteered, whilst another 12 listeners were students recruited from the University of York, U.K. in exchange for course credit or cash rewards of 5 pounds. Other demographic details were not collected.

4.2.1.2 Stimuli

There were nine subtests in total. Two of these were considered to be essential (melody and rhythm), five had rarely been investigated (timbre, pitch, loudness, tempo, rhythm-in-melody), and two were novelties in this study [tuning (chords) and tuning (orchestra)].

For each of the nine subtests, 18 trials were devised with an equal number of "Same" and "Different" trials. *Same* trials are those where the standard stimulus and the comparison stimulus are identical. *Different* trials are those where the comparison stimulus differs from the standard stimulus. Trials were purposefully designed to represent different levels of difficulty (easy, moderate and complex). Figure 4.1 illustrates the overview of the subtest design. All audio files were produced in lossless, uncompressed wav format with 44.100kHz, 16 bits. All sound samples used in this test were edited and normalized to achieve uniformity in loudness (except the loudness subtest, see below). This loudness level is termed as "standard stimuli loudness" in this thesis, and will be used later in order to present the stimuli materials more clearly.

Study 1 employed the standard test type: discrimination design with Two (2AFC) and Three Alternative Forced Choice (3AFC) as each of these designs is widely used in perception research due to its simple form (e.g., Seashore, 1919; Irwin & Francis, 1955;

Wing, 1968; Gordon, 2001; Peretz et al., 2008). The gaps between standard and comparison stimulus were 2000ms.

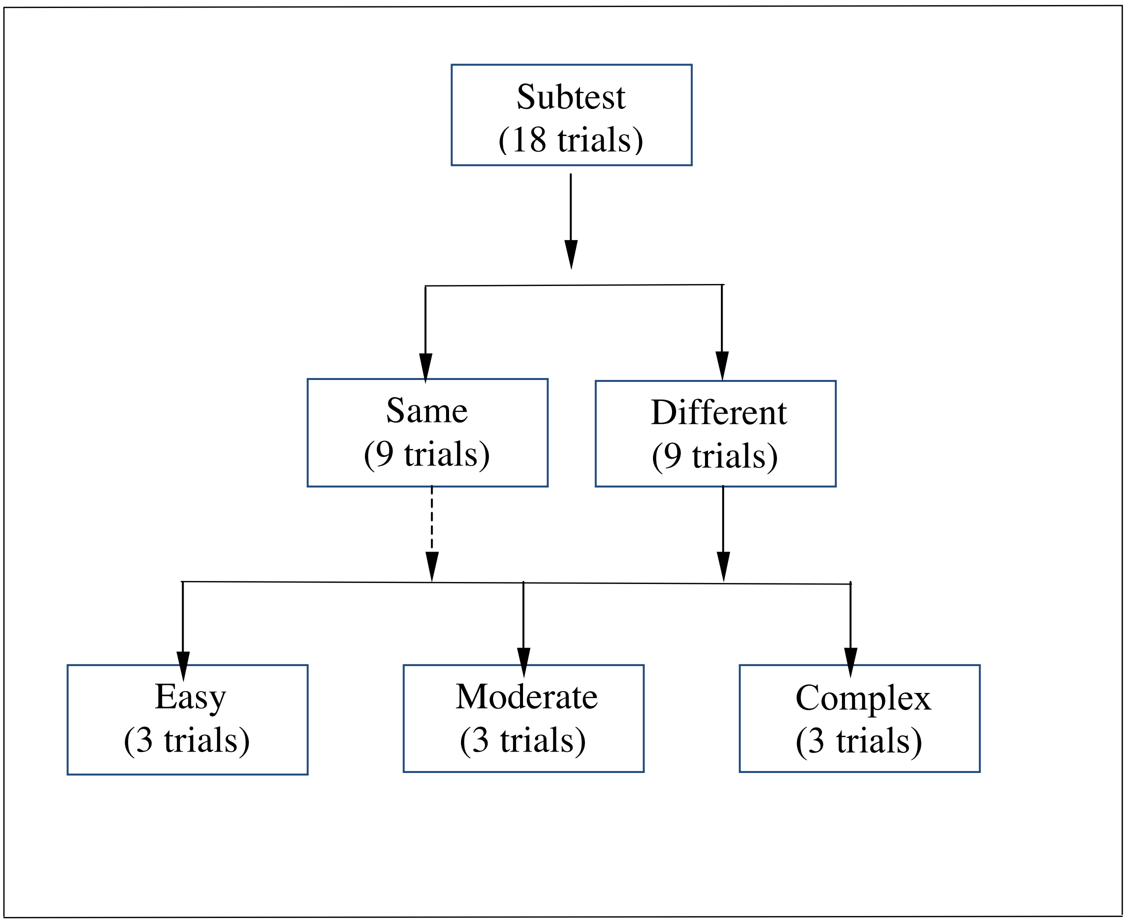


Figure 4.1. Overview of the subtest design of the music test-battery. Each subtest consists of 18 trials: half of them were *same* trial; and the other half were *different* trials. These trials are then categorised into three different difficulty levels: easy, moderate and complex, with three trials each. Also note that the pitch, loudness, tuning, tempo, and timbre subtests' *same* trials were not categorised into difficulty levels as they do not require as much memory attention as the melody or rhythm subtests.

Table 4.1. Overview of Stimuli Design (Study 1)

Task	Difficulty Level (Different Trials)		
	Easy	**Moderate**	**Complex**
Melody	Simple rhythm with inharmonic notes on down beat	Regular rhythm and melody structures with a consonant passing note.	Higher note densities with a short consonant passing note.
Rhythm	Simple rhythm structure with mostly crotchets and quavers	Moderate rhythmic structure with mostly crotchets, quavers and semiquavers	Complex rhythmic structure, mostly quavers and semiquavers
Rhythm-in-Melody	Simple rhythm structure with mostly crotchets and quavers	Moderate rhythmic structure with mostly crotchets, quavers and semiquavers	Complex rhythmic structure, mostly quavers and semiquavers
Timbre	Comparison of instruments that are harmonically significantly different	Comparison of instruments that are harmonically relatively similar	Only one note of the chords was replaced
Pitch	Difference of 20-30 cents	Difference of 15-25 cents	Difference of 5-10 cents
Tuning (Chords)	Mistuned 30-50 cents	Mistuned 20-30 cents	Mistuned 10-20 cents
Tuning (Orchestra)	Mistuned 30-50 cents	Mistuned 20-30 cents	Mistuned 10-20 cents
Tempo	Difference of >7bpm	Difference of 3-5bpm	Difference of 1bpm
Loudness	Difference of >5dB	Difference of 2-3dB	Difference of 0.5-1dB

4.2.1.2.1 Melody

This subtest consisted of melody sequences of 9 to 21 tones that were generated using the "Harpsichord" voice in Steinberg's *Hypersonic* (Steinberg, 2009). The difficulty of the trials was manipulated by increasing note density and change in the melody tonality since previous research found that listeners are more sensitive to tonal melody than atonal melody according to the structural processing model and melodic expectancy model (see Dowling, Kwak, & Andrews, 1995; Schulze, Dowling, & Tillmann, 2012; Honing, 2005; Bregman, 1990; Pearce & Wiggins, 2006; Temperley, 2008; or see Chapter 3, Sections 3.1.2 and 3.1.9). For instance, the *easy* trials had lower note densities, in both *same* or *different* trials. Whereas in the *different* trials, the *easy* item contained at least one pitch violation or an inharmonic note that would break the similarity stream in the sequence (Bregman, 1990). The *moderate* trials contained higher note densities compared to the *easy* trials and the alteration of the note on the test-stimulus was less apparent as harmonically related notes were used. The *complex* trials were designed to be difficult where the note densities were higher (mostly semiquavers) and the alteration of the test-stimulus was also a harmonically related note, making it hard to notice (see Figure 4.2). Listeners were presented with a pair of melody sequences and asked whether the second melody was the same or different compared to the first melody.

Melody: Easy

Melody: Complex

Figure 4.2. Two examples of Melody trials. The asterisk * represents the alteration in the comparison stimulus. The top figure shows the *easy* trial where the alteration is implemented with an inharmonic note. The figure underneath shows the *complex* trial with a higher note density and the alteration placed on the passing note position with a harmonically related note.

4.2.1.2.2 Rhythm

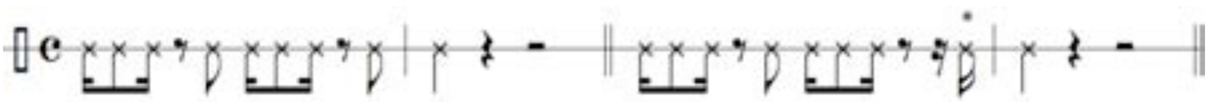
As discussed in Chapter 3, rhythm itself is not independent, and its structure is often the culmination of duration, time, grouping and meter. The rhythm context that is employed in this subtest is that it keeps an absolute duration of the sound events, but with an equivalent duration of a different rhythmic structure (Purwins et al., 2008). For example, a crotchet (one sound event) is equivalent to two quavers but their total duration is the same. The rhythm subtest consisted of rhythmic sequences of 7 to 17 notes that were generated using the “Rim Shot” voice in Steinberg’s Hypersonic (Steinberg, 2009). The note densities and the changes in the rhythmic groupings varied the difficulties of the trials (Large & Palmer, 2002). For instance, the *easy* trials were generated by simple rhythmic structures using mostly crotchets and quavers; where the rhythmic change on the different trials was achieved by adding or subtracting one or more notes on the downbeat[22]. The *moderate* trials were slightly more complicated, being derived from

22 Downbeat refers to the first beat of a bar or a group of notes.

crotchet, quaver and semiquaver rhythmic patterns and with the alterations occurring mostly on the *passing note*. The *complex* trials, on the other hand, comprised various rhythm patterns and the alteration was normally short. Examples of the *easy* and *complex* trials can be seen in Figure 4.3. Listeners were presented with a pair of sound clips and asked whether the second rhythm of the second sound clip was the same or different compared to the first sound clip.

Standard-stimulus Comparison-stimulus

Rhythm: Easy



Standard-stimulus Comparison-stimulus

Rhythm: Complex

Figure 4.3. Two examples of Rhythm trials. The asterisk * represents the alteration change in the comparison stimuli. The top figure shows the *easy* trial consists of a simple rhythm with the alteration occurring on the downbeat. The figure underneath shows a *complex* trial where the rhythm is more complicated and the alteration has a shorter duration (i.e., semiquaver).

4.2.1.2.3 Rhythm-in-Melody

The rhythm-in-melody subtest was influenced by Gordon's AMMA to inspect rhythm perception in a melody form. As a melody subtest, this subtest consisted of melody sequences of 8 to 21 tones generated using the "Harpsichord" voice (again in Steinberg's Hypersonic). The difficulty of the trials was varied by the complexity of the rhythms. For example, the *easy* trials consisted of simple rhythmic patterns, mostly quavers and crotchets. The alteration on the *different* trial was made on the downbeat of the melody sequence. The *moderate* trials were similar to the *easy* trials but contained more semiquavers and quavers. The rhythmic patterns of the *complex* trials were more syncopated and the alteration occurred normally on the weak beat. Listeners were presented with two melody sequences and asked whether the rhythm of the second melody was the same or different compared to the *rhythm* of the first melody.

Rhythm-in-Melody: Easy

Rhythm-in-Melody: Complex

Figure 4.4. Two examples of rhythm-in-melody trials. The asterisk * represents the alteration in the comparison stimuli. The top figure shows the *easy* trial consists of a simple rhythm with the alteration occurring on the main beats. The figure underneath shows the *complex* trial where the rhythm is more complicated and the alteration has a shorter duration (i.e., semiquaver).

4.2.1.2.4 Timbre

Seashore et al. (1960) were the only researchers to have used the timbre test in their musical ability test-battery. However, the validity of their timbre test came into question when international professional orchestral musicians scored significantly poorer compared to the normal population data presented in the 1960 revision of the manual (Henson & Wyke, 1982). The authors further argued that Seashore's timbre test was not suitable for measuring musical ability as the high level of discrimination required by the timbre test was of little consequence for the professional musicians. Therefore, in contrast to Seashore et al. (1960) who used pure tones in his timbre test, this thesis aimed at emulating the sounds of original instruments as closely as possible because the timbre of musical instruments was found to have multidimensional spectral properties (Grey, 1977; see Chapter 3 for a review).

To this end, the timbre subtest used the sound samples from the Vienna Symphonic Library (Vienna Symphonic Library GmbH, 2002). For each stimulus the timbre subtest consists of chords of four notes (C_4, E_4, G_4, C_5) to produce a rich timbre with a possibility for making very subtle changes. The duration of each chord was 1.5 seconds. The difficulty was varied by means of subtle changes to the instrumentation in each chord. In the *easy* trials, the comparison was between two chords played by different families of instruments such as horn versus strings. In the *moderately* difficult trials, two chords were played by instruments from similar families such as Woodwind versus Horns (both blowing instruments). In the *complex* trials, only one note was replaced (e.g., a Horn sound is replaced by a Piano sound; see Figure 4.5). Listeners were presented with a pair of sound clips and asked whether the second sound was the same or different compared to the first sound.

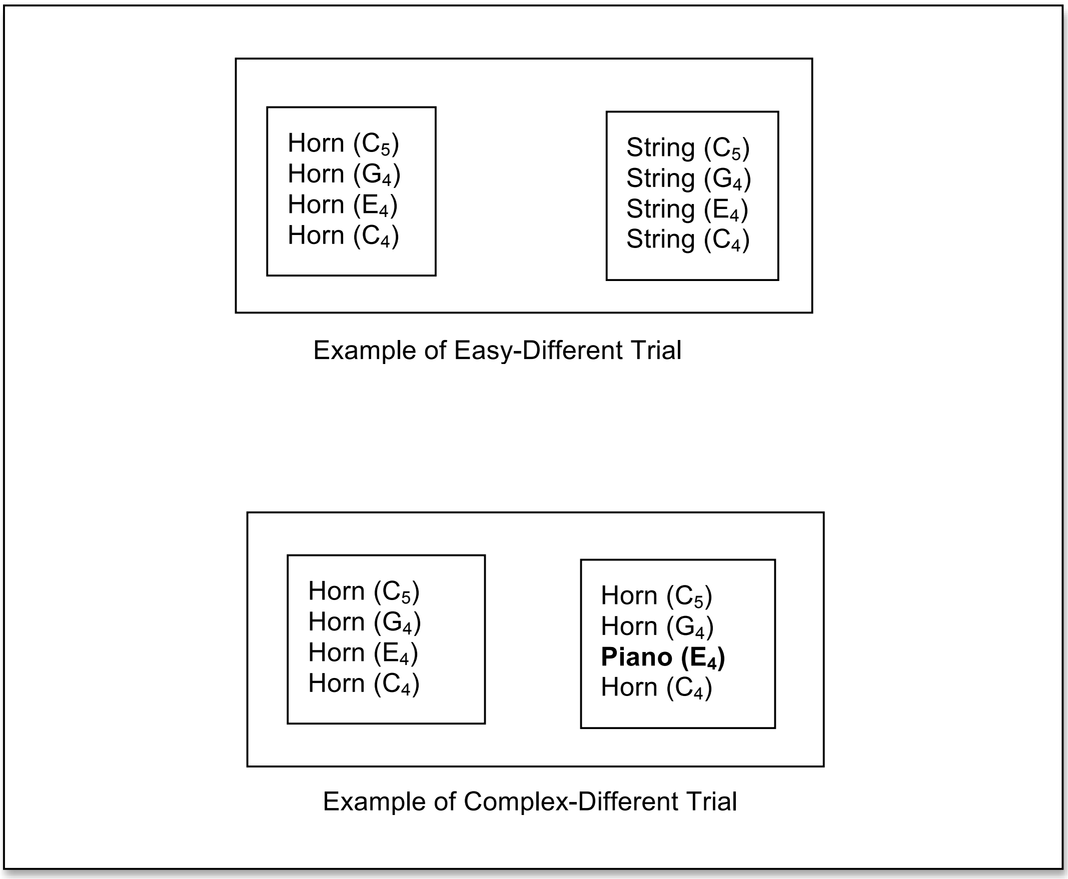


Figure 4.5. Example of a timbre subtest. The easy-different trial consists of two groups of instruments from different families. The *complex-different* trial is from the same family with one note altered.

4.2.1.2.5 Pitch

As discussed in Chapter 3, the concept of pitch can be ambiguous depending on its application and physical attribute. Therefore, sinusoids, or pure tones, were selected for this subtest as the use of complex tones can result in pitch changes being perceived due to harmonics (complex tone) rather than the fundamental frequency (Licklider, 1954). All the stimulus tones were 2000ms sinusoids with a 100ms on/off linear ramp to de-emphasize the salience of the on and off-sets. The difficulty level of the pitch subtest was manipulated by varying the degree of the frequency difference the between base and test stimulus (range of 5 to 30 cents, or 1 to 12 Hz, pivoting at the frequency of 440Hz). 440Hz (A4) was used because it is the standard pitch used as a reference-tuning note in an orchestra (Martin, 2008; ISO, 1975), and also appeared to be the typical pitch range of music that most healthy listeners of all ages are sensitive to (Seashore, 1967; Schellenberg & Moreno, 2009). Listeners were presented with a pair of tones and asked whether the second tone was “lower”, “higher” or the “same” compared to the first tone.

4.2.1.2.6 **Tempo**

As discussed in the previous chapter, tempo perception is often neglected in music perception research despite the fact that the ability to perceive tempo deviation plays an important role in keeping time whilst performing, which is considered a key musical performance skill (Gordon & Martin, 1993/1994; Drake, 1957). A short music clip was composed as the stimuli for this research to give the flexibility to incorporate tempo changes without giving cues to the listener often evident when audio files are “time-stretched”[23]. Production techniques (mixing and mastering) were used on the clip in an attempt to achieve professional quality production. The composition was created in Logic Pro 9 (Apple, 2011), using an available built-in sound source/VSTi[24].

Sound Source:

2 Steps BoxerBeat	Bass Pad	Electro Elevate Bass
80sDanceBassSynth	DnBAuro	HouseDance

The composition is a simple house/dance groove with strong rhythm beats. The drum

23 Tempo changes with time-stretching techniques generally result in slight pitch changes due to the physical properties of the sound file being manipulated.
24 VSTi- Virtual Studio Technology Instrument, a program for emulation of synthesizer sounds.

track uses *HouseDance* (straight drumbeat) and *2 Steps BoxerBeat* (funky drum grooves). The bass sound/pattern uses *80s DanceBassSynth*, *BassPad and Electro Elevate Bass,* which creates a dance pattern groove in the bass frequencies. Finally the composition added *DnB Auro* (pad) as the harmony of the composition. Only two bars of the music were used because previous studies have shown that listeners have an error correction strategy in their internal clock that will correct the tempo retrospectively (e.g., Repp, 2000). If a longer sound clip was used, it would be difficult for the listeners to remember and compare the second stimulus with the previous clip as their "correction system" would change their internal clock/tempo, leaving them without a comparison.

In contrast to Gordon's previous tempo test that employed various tempo changes in each stimulus, this study used a constant stimulus tempo (one overall tempo in each stimulus) to avoid confusing the listeners. All tempo stimuli were within the range 110bpm to 130bpm as 120bpm is the preferred tempo range for most listeners (e.g., Moelants, 2002). Listeners were presented with a pair of music clips and were asked whether the second clip was "slower", "faster" or the "same" as the first clip.

4.2.1.2.7 Loudness

Loudness perception was another perception dimension also previously neglected, despite Seashore (1967) as well as Fabiani and Friberg (2011) reiterating its essential application in music (see Chapter 3). In order to keep the loudness test consistent with other subtest, all tones were 2000ms sinusoids at 440 Hz with 100ms on/off linear ramp to de-emphasize the salience of the on and off-sets. (similar with pitch subtest). The intensity range of the loudness stimuli ranged from 3dB higher than the standard stimuli loudness level, to 6dB lower than the standard loudness level. Therefore the lowest intensity of the loudness stimuli and the highest intensity gave a decibel difference of 9dB. The difficulty level of the trials was varied by the intensity difference of the standard and comparison trials: 6 to 7 dB (easy), 3 to 5 dB (moderate), and 1 to 2 dB (complex). A standard of 440Hz maintained the consistency of this test battery. As mentioned in the pitch subtest, A= 440Hz is known as the standard pitch, therefore it would be easier to maintain consistency in order to compare the listener sensitivity at the same range. A=440Hz is also in the middle register where it was found to be the sensitive range for most listeners (Seashore, 1967; Schellenberg & Moreno, 2009). Listeners were asked whether the second tone is "softer", "louder" or the "same" as the first tone.

4.2.1.2.8 Tuning (Chord)

The tuning (chord) subtest was a novel test introduced into this thesis. The tuning perception in this context refers to concurrent sound perception of a chord. As in the timbre subtest, each stimulus consisted of C4, E4, G4 and C5 to form a C chord and had duration of 1500ms. This combination of diatonic harmony is relatively universal (Cook, 2009) thereby attenuating the possibility of misunderstandings about "correct tuning" due to listeners' backgrounds The Piano sound samples from the Vienna Symphonic Library were used. The difficulty level of the test-trials was varied by subtle manipulations to the E note (differing in frequency 10-50 cents) (see Figure 4.6). Listeners were presented with a pair of piano sounds and asked whether the first or second sounds were more "mistuned" or "discordant".

C
G
E ← (10-50cents)
C

Figure 4.6. Example of the Tuning subtest. The difficulty of the tuning subtest is based on the "tuning" level of note E4, from 10-50 cents.

4.2.1.2.9 Tuning (Orchestra)

Dissonance perception in a piano sound has certain limitations and is limited only to the same instrument. Therefore, it was considered necessary to add a second dissonance perception task, in which these limitations are controlled. A small chamber piece (a small group of instruments playing) was used to perform the "Death and Maiden" by Franz Schubert. The term "orchestra" instead of "chamber" is used in this context because it is a more common term understood by general population.

The tuning (orchestra) subtest comprises of four sections of instruments: Violin Section, Violin 2 Section, Viola Section and Cello Section from the Vienna Symphonic Library. The difficulty level of the test-trials was varied by subtle manipulations to the Viola tracks (differing in frequency 10-50 cents). Listeners were presented with a pair of music excerpt and asked whether the first or second music excerpt was more "mistuned" or "discordant".

4.2.1.3 Procedure

LTware™, which was implemented in Matlab (R2009b) was used to run the experiment. The code for *LTware* is available in Appendix 1. Listeners were tested in groups of one to three individuals in a quiet room on individual computers (iMac PowerPC G5, 1.8 GHz). All computers were equipped with high quality headphones, Audio Technica ATH-M40FS. The sound pressure meter was used to measure headphones sound levels to be 60dBA. Three practice trials were presented to the listeners preceding the actual trials to make sure that the listeners had a good understanding of how to operate *LTware*. The experiment took about one hour.

Figure 4.7 An example of the Melody subtest in *LTware*. Listeners press “Start” and hear two consecutive melodies. The sequential trials play automatically following the listeners answer selection. When the test is completed, listeners press “Submit” then click “next” for the subsequent subtest.

4.2.2 Results

4.2.2.1 Scoring and descriptive statistics

Performance on each trial was evaluated as being correct (1 point) or incorrect (0 point) and the number of correct responses calculated the scores of Study 1. Table 4.2 shows the descriptive scores for mean and standard deviation. Because there are 18 trials in each subtest the maximum that any listener can obtain is 18/18, and the minimum 0/18. The chance performance 3AFC subtests (for pitch, tempo and loudness) were 33.3% (6.00); 2AFC subtests (melody, rhythm, rhythm-in-melody, timbre, tuning [chord] and tuning [orchestra]) were 50.0% (9.00).

Table 4.2. Descriptive Summaries for the Music Battery ranked in order of Mean score (out of 18 max) for Subtest; and out of 162 max for the Composite Score (Study 1)

Test	Mean	Standard Deviation
Timbre	14.75	2.23
Tuning (Piano)	14.67	2.33
Rhythm	14.54	2.38
Rhythm-in-Melody	13.25	1.82
Tuning (Orchestra)	12.88	2.17
Loudness	11.67	3.62
Tempo	11.54	4.74
Melody	11.42	2.43
Pitch	8.83	3.31
COMPOSITE	113.54	17.09

Note. N=24. Number of trials per subtest = 18

4.2.2.2 Internal consistency reliability

Reliability coefficients are given in Table 4.3. The rhythm-in-melody subtest shows poor internal consistency reliability as it is well below the acceptable .60 (recommended criteria is available in section 2.1, Chapter 2). The explanation for this might be due to listeners not being able to show consistency when rhythm is changed within a melody form. Modest reliability coefficients were also shown for the tuning (orchestra) subtest and melody subtest. The poor psychometrics of tuning (orchestra) subtest score might be due to interference from unnecessary variables such as the melody and rhythm structure of the stimuli, rather than focusing purely on the tuning property. Other subtests show reasonable moderate to high inter-item correlation, which is considered acceptable given the relatively modest number of trials in the subtests (section 2.1, Chapter 2).

Table 4.3. Internal Consistency Reliability (Study 1)

Test	Cronbach's Alpha
Tempo	.89
Loudness	.81
Pitch	.72
Timbre	.70
Rhythm	.60
Tuning (Chord)	.56
Melody	.44
Tuning (Orchestra)	.33
Rhythm-in-Melody	.11
COMPOSITE	.92

Note. N=24

4.2.3 Discussion

Study 1 was a pilot study to examine the psychometric properties of the novel music stimuli. Although the internal consistency reliability for the composite score was excellent, the range of the reliability coefficients across subtest scores was very wide, from very poor internal consistency reliability (rhythm-in-melody subtest, r = .11) to a good reliability coefficient (tempo subtest, r = .89). Because the test-battery employed a mixture of Two (2AFC) and Three Alternative Forced Choice (3AFC), the estimates relating to means, standard deviations, and psychometric properties are hard to compare across subtests due to the different designs of answer choices. Confidence rating and guessing problems were also insufficiently controlled in this study.

The next study aimed to improve the stimuli materials of Study 1 as well as to explore a new test paradigm that could provide confidence rating and consistent answer choice across subtests.

4.2 Study 2 (Double Reference Playback with Confidence Rating Scheme)

In order to pilot test the psychometrics of the new test-paradigm, the following strategy was used. First, the experiment was divided into two sessions where each session only took 30 minutes. The reasons for this approach were (1) to improve listeners' concentration with a shorter test as listeners reported the one hour test as being too long, (2) to pilot test the psychometric property of the new test-paradigm before implementing it with a complete test. Therefore two groups of listeners were recruited for Study 2.

39 listeners (Group 1) were first recruited to pilot-test the new-test paradigm with the tempo, timbre, melody and accent subtests. This new test-paradigm was shown to be reliable, therefore another 39 listeners (Group 2) were recruited for the second half of the subtests, namely rhythm-to-melody, rhythm, tuning and loudness subtests.

4.2.4 Method

4.2.4.1 Listeners

A total of 78 listeners participated in Study 2. They were students and staff from the University of York, U.K., who participated in exchange for either course credit or a cash reward of 5 pounds. Because of the length of the test, 39 listeners were allocated to one part of the test (Group 1: melody, accent, timbre, tempo subtests) and the other 39 to the second part (Group 2: rhythm, rhythm-to-melody, pitch, tuning, loudness subtests). Listeners in Group 1 were six males and 33 females (mean age = 20 years, SD = 2; range 18-27). Listeners in Group 2 were seven males and 32 females (mean age = 21 years, SD = 3, range 19-31). Of all the listeners, 45 had received relatively extensive musical training averaging 9.5 years (Group 1, n = 23; Group 2, n = 22), whereas 33 had received no instrumental lessons or any other form of music education (Group 1, n = 16; Group 2, n = 17).

4.2.4.2 Stimuli

As Study 1, for each of the nine subtests, 18 trials were devised with an equal number of "same" and "different" trials. In the Same trials, the standard stimulus and the comparison stimulus were identical. In the Different trials, the comparison stimulus differed from the standard stimulus. Trials were purposefully designed to represent different levels of difficulty. All audio files were exported to MPEG Audio Layer III (MP3) with 44.100 kHz, 16 bits, 128 kbps using Steinberg Nuendo 4 in order to achieve optimal sound quality while keeping file sizes low for smooth data loading, as the sounds were delivered via a web platform (Limesurvey version 1.87, http://www.limesurvey.org/[25]). All sound samples used in this test were edited and normalized to achieve uniformity in loudness (except the loudness and accent subtests; see below). In the following sections, each of the tasks and stimuli are described in detail.

The stimulus material was also revised on the basis of (a) imbalances in subtest difficulty, and (b) a psychometric analysis of poorly performed trials identified in Study 1 (Table 4.4). Specifically, subtests that were comparatively too difficult or easy were revised by replacing some of the most difficult (or easy) trials with trials of a more moderate level of difficulty. Furthermore, trials with unsatisfactory item-to-total correlations in each subtest were revised or replaced with new trials. More details of the revision are provided in the main text.

Table 4.4. Summary of Stimuli Revision in Study 2

Subtest	Revision
Melody	Multiple rhythm patterns of the melody sequences were revised to constant rhythm only (quavers).
Rhythm	Several *easy* trials showed to be too easy (100% correct response), hence the difficulty level for *easy* trials was increased.
Rhythm-in-Melody	This subtest was removed due to poor psychometric property (α = .11) and replaced with the Rhythm-to-Melody subtest.
Timbre	The *moderate* trials and *complex* trials were revised to more difficult levels as the correct responses for these trials were above 50%.
Pitch	The *easy* trials were adjusted to 40 cents and above (original was

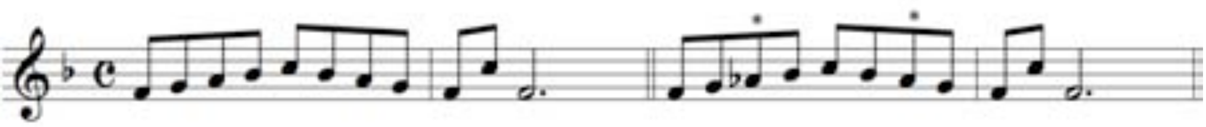
[25] Limesurvey is an open source online survey platform

Subtest	Revision
	between 20-30 cents) as the correct responses were around 50%.
Tuning (Orchestra)	This subtest was removed due to poor psychometric property (α = .33).
Tuning (Chord)	The *moderate* trials and *complex* trials were revised to more difficult trials as the correct responses were above 80%.
Tempo	Two stimuli instrumentation (monolayer and dual layers) were added in addition to the multilayers stimuli instrumentation. The trials with lower item-to-total correlations were removed to accommodate the additional trials.
Loudness	The *moderate* trials were adjusted to between 2 to 3 dB (original was 1.5 to 3dB) as the correct response for 1.5 dB was very low (24% correct response).
Accent	This new subtest was added to the test-battery

4.2.4.2.1 Melody

In contrast to Study 1, all melodies were recomposed in constant rhythms (quavers only) instead of multiple rhythm patterns in order to avoid rhythm cues in melody perception. Some trials were transposed so that the pitches of the stimuli were within the range around G3 to C5, which is the middle range of an 88-keyboard/piano. Examples of *easy* and *complex* melody structures are given in Figure 4.8. Stimuli were composed and delivered with the “harpsichord” timbre from Logic Pro 9 (Apple, 2011) for its neutrality.

Example of Easy-Different Trial



Standard-stimulus Comparison-stimulus

Example of Complex-Different Trial

Standard-stimulus Comparison-stimulus

Figure 4.8. Example of melody trials. An *easy* trial consists of a tonal melody (upper part) as opposed to a *complex* trial, which is atonal (lower part). *Represents the alteration in the comparison-stimuli.

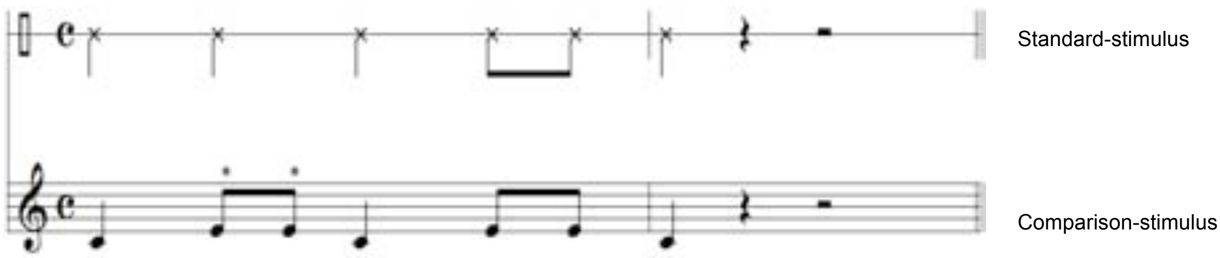
4.2.4.2.2 Rhythm

The *easy* trials of the rhythm test in Study 1 appeared to be too easy (100% correct response), therefore the difficulty level of the rhythm test was slightly increased. For example, the note density of the *easy* trial was slightly increased from 7 notes to 9 notes. The rhythm subtest was delivered with the “rim shot” voice from Logic Pro 9 (Apple, 2011) for its pure percussive, clear and crisp timbre.

4.2.4.2.3 Rhythm-to-Melody

The rhythm-in-melody subtest from Study 1 was not used in this study due to its poor psychometric property, so this subtest was replaced with the new rhythm-*to*-melody subtest. This is a novel subtest that targeted listeners’ ability to recognise a rhythmic pattern when it is no longer provided in its original form (i.e., in non-pitched percussive form), but embedded in a melody. This subtest aimed to examine the listeners’ ability to remember or attend to rhythmic structure when there is an external variable such as pitch grouping competing with the original rhythmic grouping system. Thus, listeners must solely attend to the rhythmic structure of a melody without being influenced by its pitch contour. Specifically, listeners were asked whether a rhythmic pattern, presented initially on a percussive instrument, was the same or different in a subsequently presented melodic context (and vice versa). The intensities of all notes were held constant. All melodies in this subtest were tonal to avoid diverting listeners' attention from rhythm to atypical melodic features (see Figure 4.9).

Example of Easy-Different Trial



Example of Complex-Different Trial

Figure 4.9. Example of rhythm-to-melody trials. An *easy* trial consists of a simple rhythm (mostly quavers and crotchets), as compared with a *complex* trial, which consists of a more complicated rhythm (quavers and semiquavers). All melodies (comparison-stimuli) are tonal. *Represents the alteration in the comparison-stimuli.

4.2.4.2.4 Accent

This subtest aimed to investigate a further temporal ability other than the standard rhythm ability. Previous research often used a "meter" test to examine additional temporal ability. However, as explained in the Chapter 1 and Chapter 3, this idea was abandoned in this thesis because meter perception is rather a culturally based perception, which conflicts with the aim of this thesis to make the test accessible to listeners from any culture. Therefore, accent perception was chosen instead of a meter test.

Accent perception, as reviewed in Chapter 3, is a perceptual skill to identify musical styles and expressions; an important aspect necessary to achieve a higher level of musical ability. Only Wing (1968) attempted to look into this area of musical ability, but unfortunately Wing's rhythmic accent subtest was confounded as described in Chapter 2. This research attempted to improve on Wing's rhythmic accent test by eliminating the confounding melodic features, as well as by keeping constant the intensities for accented and unaccented notes. The absolute note durations (rhythms) were identical between standard and comparison stimuli. The accented notes were normalized at the "standard stimuli loudness" as the other subtests, and the intensity of the unaccented notes differing at 3dB lower. In the *easy* trials, intensity changes were applied to most sound events so as to increase the probability of detecting the alteration. In the *moderate* and *difficult* test trials, there were fewer intensity changes, which required more subtle perceptual skills to be identified (see Figure 4.10). Like in the standard rhythm subtest, stimuli were composed with the "rim shot" voice from Logic Pro 9 (Apple, 2011).

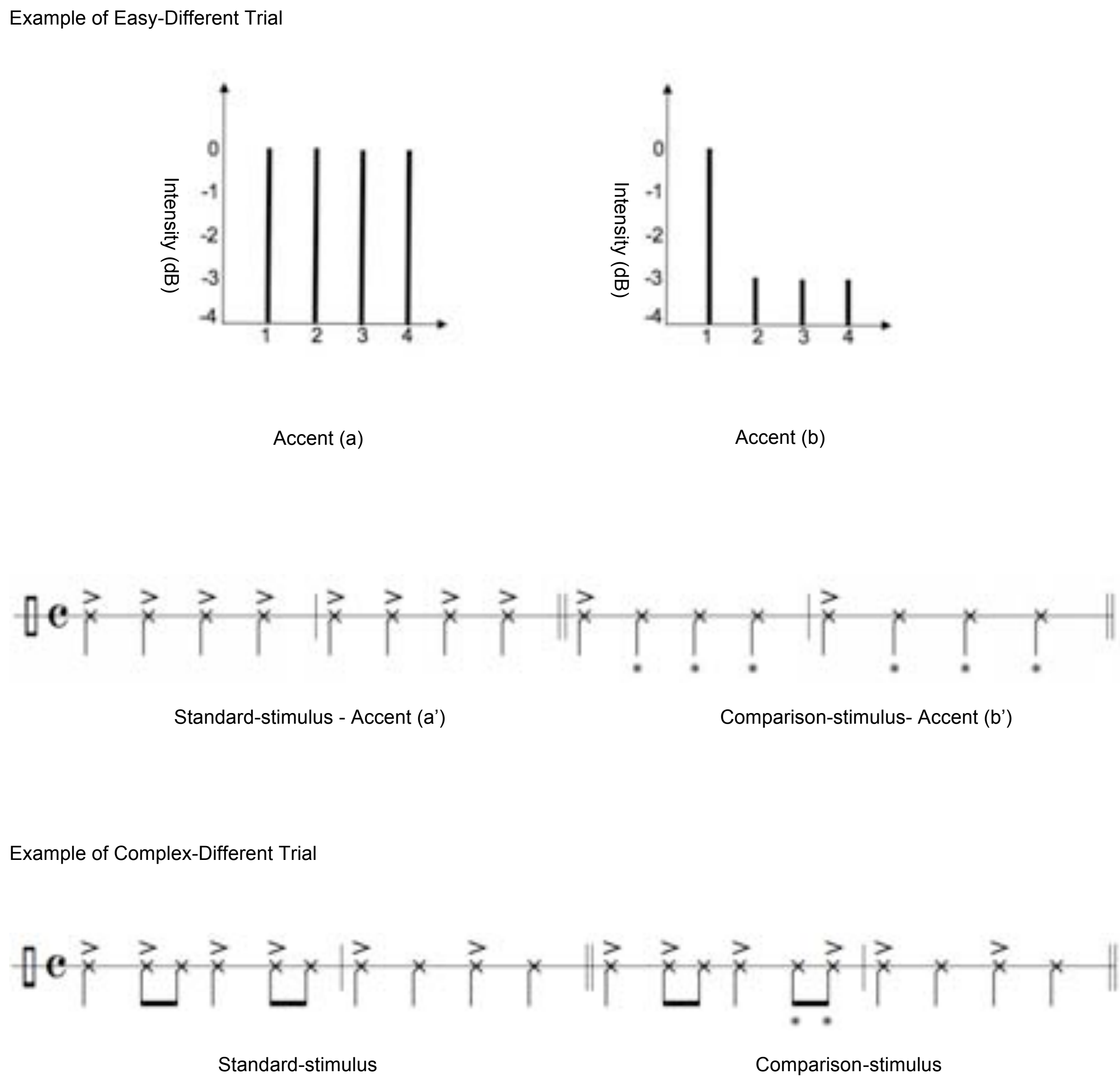


Figure 4.10. Example of accent trials. The top figure shows the intensity differences of the accented and unaccented notes. As the top figure shows, the intensities of the accent notes (a) are represented by the sign > in the figure below [i.e., Accent (a’)]. Accent (b) shows the unaccented notes (second, third, and fourth beats) are -3 dB lower than the accented note, which can also be seen in the comparison-stimulus Accent (b’). The example of a complex trial shows the alteration affecting only one or two events. * Represents the alteration in the comparison-stimuli.

4.2.4.2.5 Timbre

The *easy* trials of the timbre test were kept as before - where listeners were asked to compare two chords played by instruments from different families such as horn versus strings. The *moderate* trials were revised. In Study 1, listeners showed high discrimination skills for the *moderate* trials (85% of the listeners answered correctly) that compared similar families of instruments such as Woodwind versus Horns (both of these are blowing instruments). Therefore in this study, the replacement for the *moderate* trials occurred in only one of the four voices (e.g., woodwind C_4, woodwind E_4, woodwind G_4, and woodwind C_5 against woodwind C_4, *violin* E_4, woodwind G_4, and woodwind C_5). The *complex* trials were also revised. In Study 1, listeners showed high discrimination skills for the *complex* trials (50% of the listeners answered correctly) - for example the trial that consists of Horn Section (Horn C_3, Horn E_3, Horn G_3, Horn C_4) versus Horn Section+Piano (Horn C_3, *Piano* E_3, Horn G_3, Horn C_4). Thus it was replaced by the Same-family instrumentation (e.g., a viola sound is replaced by a violin sound; see Figure 4.11).

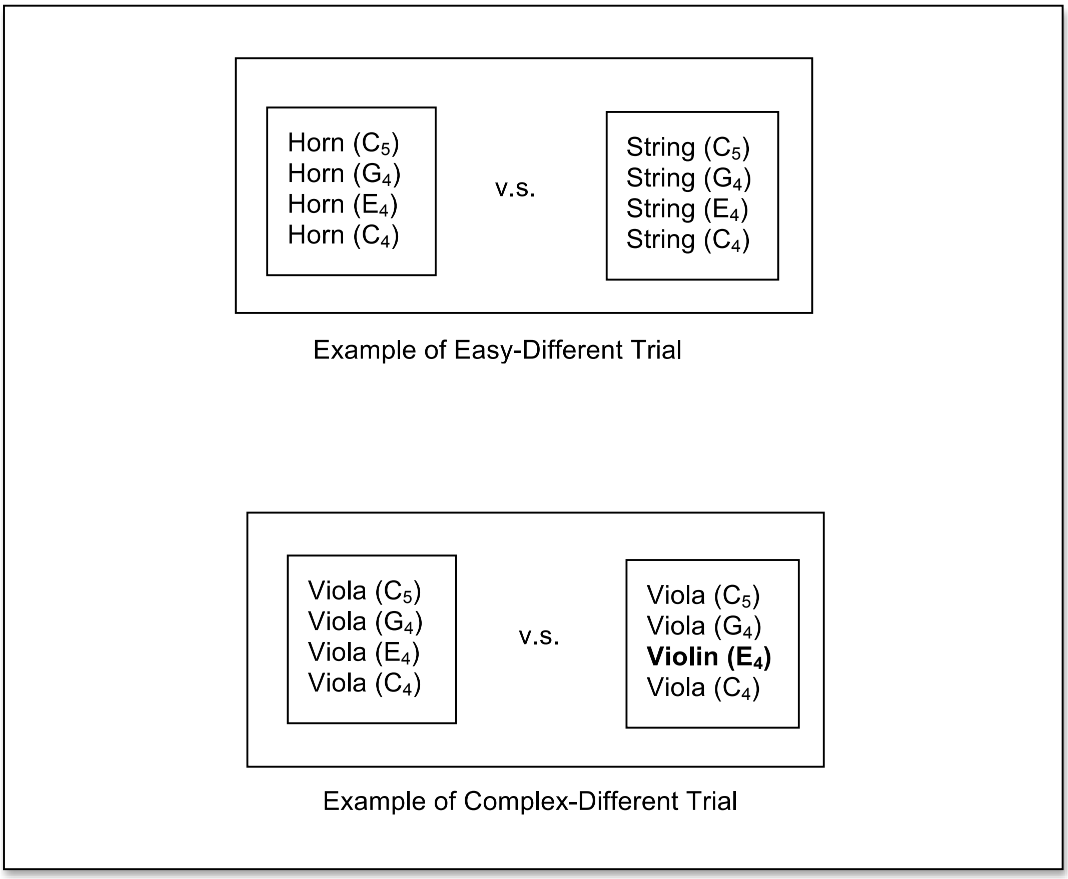


Figure 4.11 Illustration of timbre subtest. The *easy* trial consists of two groups of instruments from altogether different families. In the *complex* trial the instrument (which is changed on only one note) is taken from the same family (strings).

4.2.4.2.6 Pitch

The *easy* range of the pitch subtest was increased to 40 cents and above as listeners from Study 1 showed low discrimination in the items with 20-30 cents pitch difference (where only 33% of the listeners answered correctly).

4.2.4.2.7 Tempo

Several listeners from Study 1 reported that it was difficult to maintain concentration after repeatedly listening to the same tempo stimuli across trials. In order to attenuate the risk that preference for a given instrument or rhythm might affect the performance in this test (e.g., Dahl & Granqvist, 2003) as well as to address the reported issues from the listeners, stimuli with differing rhythmic structures and timbres were composed. The timbres were drums, bass, harmony, and melody (multilayers); conga and shaker (dual layers), and rim shot voice (monolayer). The trials with lower item-to-total correlations were removed to accommodate the additional trials.

4.2.4.2.8 Loudness

The *moderate* trials were adjusted to differences in intensity of 2 to 3 dB (original was 1.5 to 3 dB), because the correct responses for the original *moderate* trials were very low, with only 24% of the listeners answering correctly.

4.2.4.2.9 Tuning (Chord)

Study 1's tuning (orchestra) subtest showed poor psychometric properties hence only the tuning (chord) subtest was kept in this study. As explained in Chapter 4, the poor psychometrics of the tuning (orchestra) subtest might be due to tuning perception being affected by unnecessary variables such as melody and the rhythm structure of the stimuli, hence the subtest was kept simple with a basic chord design. Just as with the tuning (chord) in Study 1, the difficulty level of the test trials was varied by subtle manipulations of the E note (differing in frequency 10-50 cents; see Figure 4.12. In particular, the frequency differences in the moderate items were changed to between 15-20 cents, and the complex trials to 5-10 cents (original was 10-20 cents) as listeners from Study 1 showed high discrimination skills in the moderate trials consisting of differing frequencies of 20-30 cents (83% of the listeners answered correctly in Study 1).

C5
G4
E4🡐(10-50 cents)
C4

Figure 4.12. Illustration of tuning trials. The difficulty of tuning trials is manipulated by the extent to which the note E4 is shifted out of its proper frequency (from 10 to 50 cents).

4.2.4.3 Procedure

All tests (including the demographic questionnaires, see Appendix 4) were computerized and delivered via *LimeSurvey* software (version 1.87, see Figure 4.13). Listeners were tested in groups of two or three individuals in a quiet room, with a fence that visually isolated listeners from one another. Prior to the music-listening part, listeners were asked to fill in the basic demographic questionnaire and the music background questionnaires. Subsequently, they were asked to put on headphones for the music test (Senheiser HD-25-1 II, Figure 4.14). To facilitate encoding of the standard stimulus, the standard stimulus was presented twice, followed by the comparison stimulus. To make the distinction between the standard stimuli and the comparison stimulus clearer, a 1.5 seconds interval between standard stimulus and its repetition were used, followed by a 2.5-s interval preceding the onset of the comparison stimulus. Three practice trials were presented to the listeners preceding the actual trials. Roving design (the first stimulus in each trial was randomized independently from that in the other trial) was employed to discourage listeners from relying on a fixed internal reference in their memory. Therefore, listeners are forced to rely on perceived differences in each trial rather than across all trials (see Micheyl, Kaernbach, & Demany, 2008). The order of the subtest presentation was counterbalanced by reversing the sequence for half of the listeners.

The listeners' task was to decide whether the comparison stimulus was the same or different compared with the standard stimulus. In contrast to Study 1 which employed a mixture of discrimination design with Two (2AFC) and Three Alternative Forced Choice (3AFC), this study attempted to use a more consistent answer choice across all the subtests so that the result was comparable across subtests. Another reason for this choice was because previous research suggested that explicit music tasks, such as asking listeners to judge whether the pitch of a tone is higher or lower, were advantageous to listeners trained to perceive such a change (Bigand, & Poulin-

Charronnat, 2006). This type of method gives little information about the nature of musical ability as a large part of human cognition occurs at an implicit level with a possible continuum between unconscious and declarative knowledge (Underwood, 1996). Therefore the answer choice used in this study has changed to “Same or Different” without asking listeners to give specific judgments such as “higher or lower” and “softer or louder” as used in Study 1. This ensures the task only involves fundamental musical intuitions understandable equally by musically trained listeners as well as untrained listeners.

Therefore, in this study, multiple answer options were provided, involving levels of confidence, namely, “definitely same,” “probably same,” “probably different,” “definitely different,” and “I don’t know”. This design was implemented to capture listeners’ responses more accurately. The problem with the previous Same-Different paradigm in Study 1 was that it was difficult to identify whether listeners had chosen their answers based on their perceptions or if it was a random guess. For example, when a listener chose the “Same” answer in a “different” paradigm, it was difficult to identify whether this choice was based on the inability to detect the signal in a different paradigm, or if it was a random answer choice when they were not sure.

For this reason, proving confidence rating in answer choice helps to reduce guessing as listeners are given more choices when they are not sure about the correct response; either by choosing the “probably” option when they believe they have heard a signal but are not 100% sure, or choose “I Don’t Know” when the discrimination task is beyond their ability (Macmillan & Creelman, 2005). This allows us to estimate listeners’ abilities as accurately as possible. 24 listeners from Group 1 and 36 listeners from Group 2 were invited back for test-retest a week later.

Figure 4.13. An example of the Melody subtest in *Limesurvey*. Listeners were told they would hear the "Reference" (standard stimulus) twice, followed by the Comparison stimulus. Their tasks were to decide whether the "Comparison" was the "same" or "different" compared to the "Reference". The rating guide was given: Definitely - If you are 100% sure about the answer, either Same or Different, please choose the "Definitely" option; Probably - If you are slightly confident about the answer, either Same or Different, please choose the "Probably" option; I Don't Know - If you have missed the question or you have no idea about the answer, please Do Not guess; choose "I Don't Know".

Figure 4.14. Senheiser HD-25-1 II are closed-back, professional monitoring headphones that offer high attenuation of background noise, suitable for the listening test.

4.2.5 Results

4.2.5.1 Scoring and descriptive statistics

A correct response chosen with maximum confidence (“definitely same” or “definitely different”) was awarded 1 point; whereas a correct response chosen with less confidence (“probably same” or “probably different”) was awarded 0.5 points. Incorrect responses (both probably and definitely) and the choice of “I don’t know” were awarded 0 points (see “5-rating” column in Table 4.5). This scoring method is calibrated to the confidence ratings and provided the best results in terms of internal consistency and test-retest reliability when compared with alternative scoring schemes[26] (see Table 4.5 and Table 4.6). Because there were 18 trials in each subtest, the maximum score listeners could obtain is 18/18, the minimum 0/18. The level of chance performance with the current scoring system is 6.75 (however if “I don’t know” is included as a response option, the level of chance is 5.4). After the raw score was calculated, the score was transformed to *d'* (pronounced as “dee-prime”*)* by using the standard *d’* model (z(H)-z(F); Macmillan & Creelman, 2005). Results are shown in Table 4.7. The discrepancy between the raw mean and *d’* are due to (1) the score distribution of the raw score and *d’* is different, (2) *d’* is coefficient that controlled for response bias (Hits & False Alarms) whereas raw scores are just Hits. A general interpretative framework for *d’* scores is *d’* = 0 (no discrimination ability), *d’* = 1 denotes 69% correct for both different and same trials. d’ values may range up to 2 (Keating, 2005). Negative *d’* was reported in some studies (i.e., Elhilali et al., 2009), the negative and near zero *d’* (-0.09) in the melody subtest suggests that the subtest was too difficult for most listeners hence they showed no discrimination skill (the false alarms and hits are almost equal).

Table 4.5. Three Types of Scoring Schemes (Study 2)

Method	Incorrect answer given			Correct answer given	
	Probably	Definitely	I Don’t Know	Probably	Definitely
5-point rating	0	0	0	0.5	1.0
3-point rating	0	0	0	1.0	1.0
Scale scoring	-0.5	-1.0	0	0.5	1.0

[26] The same analysis was conducted for the subsequent studies, and the result has shown to be the consistent, supporting the reliability of this scoring system.

Table 4.6. Reliability Coefficients of Three Scoring Schemes (Composite Score) (Study 2)

Method	Internal Consistency	Test-Retest
5-point rating	.87	.82**
3-point rating	.49	.51**
Scale scoring	.54	.54**

Note. N=39; ***p* <.01

Table 4.7. Descriptive Summaries for Subtests and Composite Score (Study 2)

Test	Raw Mean	Raw *SD*	Mean *d'* (*SD*)
Tempo	11.46	3.02	0.92 (1.16)
Timbre	10.56	2.82	0.70 (1.13)
Accent	9.72	2.76	0.26 (0.93)
Melody	8.68	2.27	-0.09 (0.79)
COMPOSITE (Group 1)	40.42 (out of 72 max)	8.78	0.35 (0.69)
Loudness	13.82	2.08	1.73 (0.89)
Standard Rhythm	11.47	2.10	0.81 (0.74)
Pitch	11.30	2.14	0.91 (0.85)
Rhythm-to-Melody	11.21	2.71	0.79 (1.07)
Tuning	9.71	1.84	0.48 (0.87)
COMPOSITE (Group 2)	57.50 (out of 90 max)	8.42	0.81(0.58)
COMPOSITE (Groups 1 & 2)	97.92 (out of 162 max)	17.20	0.58 (0.64)

Note. Group 1: *N* = 39. Group 2: *N*= 39. Performance at chance levels is 6.75 (discounting the answer option "I don't know" and assuming that choice is random).

4.2.5.2 Internal consistency and test-retest reliability

Following recent literature on internal consistency estimates, one comes to the conclusion that Cronbach's Alpha (α) may not always be the best reliability estimate (Revelle & Zinbarg, 2009; Sijtsma, 2009). Ability tests such as the present study, where homogeneity in item content is sided with heterogeneity in item difficulty, using α or greatest lowest bound (glb) might lead to an over- or under- correction. Therefore using McDonald's Omega (ω) would lead to a more accurate correction as it estimates the proportion of total common variance (general bound estimate coefficient) rather than restricting to the lower bound estimate such as α or glb. For this reason, this study follows the recommendation of these authors (Revelle & Zinbarg, 2009; Sijtsma, 2009) of reporting McDonald's ω in addition to α to gain a better estimate of internal consistency (see Table 4.8). The test-retest reliabilities, which were computed on the subset of twenty-four listeners who came back one week later, are provided in the right-hand column of Table 4.8. Overall, internal consistencies and test-retest reliabilities were moderately encouraging (> .60 for McDonald's ω), all the while pointing to the need for improvements.

Table 4.8. Cronbach's Alpha, McDonald's Omega Reliability, and Test-Retest Coefficient for Subtests and Composite Score (Study 2)

Subtest	α	ω	Test-Retest
Timbre	.74 (.70)	.86	.81**
Tempo	.78 (.89)	.84	.66**
Melody	.48 (.44)	.68	.56**
Accent	.62 (N.A)	.71	.68**
COMPOSITE (Group 1)	.87 (N.A)	.92	.82**
Rhythm-to-Melody	.67 (N.A)	.70	.79**
Loudness	.63 (.81)	.83	.64**
Standard Rhythm	.50 (.60)	.73	.67**
Tuning	.49 (.56)	.67	.62**
Pitch	.59 (.72)	.73	.63**
COMPOSITE (Group 2)	.85 (N.A)	.93	.84**

Note. $^{**}p < .01$, two-tailed. Sample size for internal consistency was $N = 39$; sample size for test-retest was $N = 24$. Values in parentheses are coefficients from Study 1; (N.A= Not Applicable) Accent and Rhythm-to-Melody subtests were not available in Study 1.

4.2.6 Discussion

This chapter has presented the use of elementary musical elements to assess individual differences in music perception ability. The advantage of this approach facilitates the assessment of a broader range of specific perceptual musical skills without being restricted to a specific culturally evolved musical system

Study 1 presented the initial stage of the test-development with the same-different paradigm, along with 2AFC and 3AFC as the answer choice. Overall, the composite score of the entire test-battery was shown to be excellent (α = .92). However, several subtests were shown to have poor and therefore unsatisfactory internal consistency (e.g., rhythm-in-melody, tuning (orchestra) and melody subtests). Because the test-battery employed a mixture of 2AFC and 3AFC as the answer choice, the estimates related to means, standard deviations, and psychometric properties are difficult to compare across subtests because of these different designs of answer choices. Confidence rating and guessing problems were also insufficiently controlled in this study.

Study 2 aimed to improve the test-battery by addressing these issues. First, the new paradigm “double reference playback with confidence rating scheme” was employed. Specifically, listeners were exposed twice to the standard stimulus for better encoding. After the comparison-stimulus was played, listeners decided among 5 answer options (Definitely Same, Probably Same, Probably Different, Definitely Different and I Don’t Know). In particular, Study 2 has selected and revised nine perception tests in total, including melody and rhythm perceptions that are considered essential in music; it has also revisited perceptual dimensions that were rarely examined, for example timbre, tempo, pitch, loudness, accent subtests; and finally a novel approach such as rhythm-to-melody and tuning perceptions.

The analysis of the study was also improved by using *d’*, which is a better scoring method that controls for response bias. In addition, McDonald’ ω was computed along with Cronbach’s α to get a better estimation of the internal consistency of the test. A test-retest session was also conducted to examine the variation of listeners’ responses over a short-period. Overall, internal consistencies and test-retest reliabilities were moderately encouraging, all the while pointing to the need for improvements.

The next chapter aims to replicate and extend the findings from Study 2, in particular to examine whether the estimates related to means, standards deviations, and psychometric properties of the test would replicate when a different sample of listeners take the entire music-test battery. The next chapter also aims to examine the validity of the current test-battery.

CHAPTER 5

Reliability and Validity Examination of the Profile Music Perception Skills (PROMS)

Chapter 4 provided two studies (Study 1 and 2) for the development of the music test-battery. Overall, internal consistencies and test-retest reliabilities were moderately encouraging, all the while pointing to the need for improvements.

This chapter (describing Study 3) has two goals. First, this study was conducted to examine whether the estimates related to means, standards deviations, and psychometric properties of the PROMS would replicate for a different sample of listeners who took the entire music-test battery at once. This provided the possibility of examining intercorrelations between all of the subtests. Second, this study set out to examine the test validity of the PROMS.

5.1 Study 3

As discussed in Chapter 2 (section 2.4.1), reliability analysis only examines whether a test gives consistent results, but it does not tell us whether a test measures what it is supposed to measure, or in other words, it does not tell us about "the validity of the test". Whilst the current study continued to examine the reliability of the PROMS score with a different sample group, it also attempted to answer the second part of the question; whether the PROMS measures music perception abilities.

Previous musical ability tests were mostly musical aptitude tests, where the main intention of the test authors was to create a music-test battery that could measure untrained musical potential in order to predict the musical achievement of the individuals. Therefore the *predictive validation* was an important aspect of previous music aptitude tests to examine whether participants who performed well on their tests also appeared to be more successful in future music achievement (see Chapter 2, Table 2.2). Several investigators also employed *criterion validation* where music teachers were recruited to rate the quality of the instrumentation performance of the music students (either their own music students or someone else's), where these ratings were compared against the students' performances on the music tests (see Chapter 2, Table 2.2). The third validation procedure commonly used by previous music ability research was convergent validation, where the participants were asked to do the music test developed by the investigator, as well as another music assessment developed by other people. The scores of the two (or more) music tests were then compared.

To choose an appropriate validation of the PROMS requires several considerations. First, using a published music piece might seem sensible, but it is likely to give an advantage for individuals who happen to be familiar with the music. Second, similar problems would arise if a musical performance were used, as it is advantageous to individuals who have had the opportunity to undertake music training (e.g., the teacher rating criterion validation method used in previous research). Therefore using a published music piece or musical performance would inevitably conflict with the aims of this research which state that the test should be equally suitable for listeners differing in their extent and type of musical background. Third, the aim of the current research was not to measure untrained musical potential (musical aptitude), and that the target sample population of the PROMS were adults who were likely to have been exposed to some form of musical experience in their life (whether involuntary or through music trainings), therefore the *predictive validation* method that employed by previous research was not the best validation procedure for the PROMS at this stage. However, the possibility of using predictive validation for the PROMS is discussed in Chapter 8, Section 8.4.3.

With these considerations in mind, and from the selection of validation procedures that are available (see Chapter 2, Table 2.1), three validation methods were chosen for the PROMS: namely criterion validation, convergent validation and content validation (see Table 5.1).

Criterion Validation: A set of music background assessments was chosen as the criterion validation for the PROMS. In particular, to deal with the problems in previous studies in which often only one type of self-reported musical background estimation was used (such as the length of musical training, as reported in Chapter 1), this study employed three types of music background assessment to strengthen the limitation of self-reported measurement: (1) a composite score of musical background aggregates that consisted of 7 items measuring objective musical training experience, critical listening experience and family influence; (2) self-rated musicianship; and (3) a self-rated music competence from Music-mindedness scale (Zentner, in progress). These music background assessments are reported in more detail in Section 5.1.1.2.5.

Convergent Validation: Knowing the limitations of only using self-reported music background as validation, this study also employed the convergent validation procedure, by using established music tests for validating the PROMS to further strengthen the validity of the test. Because of the relatively extensive work supporting the validity of Gordon's Advanced Measures of Music Audiation (AMMA) and Musical Aptitude Profile (MAP; Gordon, 1990, 1995), these test-batteries were used for a validation of the current melody, rhythm-to-melody, and tempo test as specified below. Because Gordon's AMMA rhythm perception task is embedded in a melodic context, the rhythm subtest of the Musical Ear Test (MET; Wallentin et al., 2010) was employed to validate the standard rhythm subtest. The accent subtest of this research, however, is similar to the MET's rhythm test, which uses percussive sounds. It is therefore reasonable to also expect a positive correlation between the accent subtest and MET's rhythm test.

Content Validation: Unfortunately, there are no established tests to validate the tuning and timbre subtests. Thus, content validity was examined rather than convergent validity. To this end, a series of timbre trials of a different nature (monophonic rather than polyphonic) was compiled and the performance compared with the timbre subtest.

Table 5.1. Validation Methods for the PROMS

Criterion	Convergent	Content
• Composite Music Background • Self-rated musicianship • Music-mindedness Scale	• Advanced Measures of Audiation (Tonal and Rhythm subtests; Gordon, 2008) • Musical Aptitude Profile (Tempo subtest; Gordon, 1995) • Music Ear Test (Rhythm subtest; Wallentin et al., 2010)	• Timbre (Monophonic)

5.1.1 Method

5.1.1.1 Listeners

56 listeners (15 males; 41 females) aged 18 to 38 years (mean = 22, *SD* = 4.6) participated in Study 3. They were students and staff from the University of York, U.K., recruited through PEEBS (Psychology Participants Recruitment System) who participated in exchange for either course credit or a cash reward of 15 pounds. Both types of musicians (trained and self-taught) and non-musicians were invited to take part in the study. Two listeners in Study 3 also participated in Study 2; however, this was not expected to have a significant impact on the results, as there was a gap of more than 6 months between the two studies. There were 33 listeners with an average of 12.3 years of music training and 23 listeners without a music background for the latter group, that is, no instrumental learning and/or other form of music education.

5.1.1.2 Stimuli

The number of trials and subtests was the same as in Study 2, but the trials were slightly revised based on the standard stimuli revision method that was described in Chapter 4 (section 4.2.4.2). The stimuli and test design for the validation tests were described in Chapter 2, but will be provided here for ease of reading.

5.1.1.2.1 Tempo (MAP; Gordon, 2001c)

This test consists of 40 items all played with a violin. Two short musical phrases were presented to the listeners; a "musical question" followed by a "musical answer". The ending of the musical answer could be slower, faster or exactly the same as the ending of the musical question. The listeners were asked to decide whether the musical answer (standard-stimulus) was the "same" as or "different" to the musical question (comparison-stimulus). The test takes 20-25 minutes to complete.

5.1.1.2.2 Tonal memory and Rhythm (AMMA; Gordon, 2008)

AMMA consists of 30 questions played with an electric piano. As with MAP (above), two short musical phrases are presented to the listeners, a musical question followed by the musical answer. Listeners are asked to decide whether the pairs are the same or different. If different, they are also asked to decide whether the difference was attributed to Tonal or Rhythm Change (Gordon, 2008). The scores (raw and percentile) for AMMA were generated automatically by the accompanying software. The test takes 20 minutes to complete.

5.1.1.2.3 Rhythm (MET; Wallentin et al., 2010a)

Each rhythm sequence consists of four to eleven beats, using the voice of the woodblock. All rhythm sequences were at a tempo of 100 bpm. The Different trial consists of one rhythm change, and includes triplets in 21 trials varied the complexity. The scores for MET are calculated by the percentage of the number of items correctly answered. The test takes 11 minutes to complete.

5.1.1.2.4 Timbre (Monophonic)

To examine the content validity of the PROMS, the trials for the timbre test were from the McGill University Master Samples (MUMS) and Vienna Symphonic Library (VSL), kindly supplied by Eerola, Ferrer and Alluri (2012). There were 36 pairs of trials, double the number of items of the polyphonic timbre test in the PROMS to increase the reliability of the test. The duration of each stimulus was one second. The timbres of the different trials were modified by Steinberg Nuendo 4 plugins such as *equaliser (EQ), vibrato*, *tremolo* or *flanger*. The test-design of the timbre (mono) test was the same as the PROMS design where the listeners' task was to decide whether the comparison stimulus was the same or different compared to the standard stimulus. Multiple answer options were provided, involving levels of confidence, namely, "definitely same," "probably same," "probably different," "definitely different," and "I don't know." The test takes 10 minutes to complete.

5.1.1.2.5 Music background assessment

Listeners' musical experiences were assessed using three types of questionnaire assessments:

Music Background Questionnaire (see Appendix 2). The music background questionnaire aimed to measure listeners' *objective* musical experience in three dimensions that were frequently used in previous research to assess musical ability (Ollen, 2006): (1) musical training experience, (2) superior perceptual skill or critical listening experience, and (3) family influence. Music training experience asked listeners about their abilities to play musical instruments and how long they have been playing the instruments, the amount of time that they spend practicing, their ability to read western notation, and whether they have any music qualifications. Listeners were also asked whether they have perfect (absolute) pitch ability, and whether they are involved in any activities that require critical listening skills (e.g., conducting, sound engineering, or professional performance). Finally, "family influence" was assessed by asking listeners whether they have any families who are musicians. All these items were examined as a composite music background score (average score of the 7 items), and also were analysed individually (see the Results section). Composite or aggregate measures provide more reliable and representative estimates compared with their individual components thereby protecting against Type II error[27] (e.g., Haynes & O'Brien, 2000). Thus, a composite index of music background was created (see the Results section). The *item-to-total correlation analysis* shows that all music background subcomponents contributed positively to the composite music background measurement.

Self-rating musicianship (see Appendix 2, item 15): The self-rating musicianship question was adapted from Ollen (2006), asking listeners to self-rate their musicianship from five options. The ratings were: (1) Non-Musician, (2) Music Loving Non-Musician, (3) Amateur Musician, (4) Semi-Professional Musician and (5) Professional Musician. The purpose of this questionnaire was to examine how listeners perceived their own musicianship, how it was related to their musical training experience (assessed by music background questionnaire), and how was the self-rated musicianship measurement related to perception skill.

Music-Mindedness Scale (Zentner, in progress; see Appendix 3): Music-Mindedness Scale (MM-Scale) was also a self-evaluation assessment, which consisted of 20

---

[27] Type II Error occurs when we believe there is no effect in the population, in reality, there is (Field, 2009).

questions of self-evaluation of musical ability and emotivity in the area of pitch and rhythm perception, composing, music reproducing skills, music emotion engagement, music commitment and engagement. For the purpose of the thesis, only 7 items that related to musical skills were used for analysis as a self-rated musical competence measurement (MM-Competence Scale). Listeners were asked to rate in a 5-rating scale whether the statements were “True” or “Untrue” for them.

5.1.1.2.6 Demographic questionnaires

Listeners were also asked to complete a demographic questionnaire, providing their details in gender, age, department, ethnicity etc. The full demographic questionnaire can be found in Appendix 4.

5.1.1.3 Procedure

There were two experimental sessions, and one additional session for the test-retest subgroup. In one session, listeners completed the current test battery. In a second session, they completed the validation tests. Finally, a subgroup of 20 listeners participated in the additional session as a means for obtaining new test-retest data. The procedure for the current battery was identical to that in Study 2. The validation sessions had four external tests, namely, AMMA, MAP, MET, and timbre (monophonic). During this session, the listener’s hearing ability was also examined by an audiometric examination in a sound-attenuated booth in accordance with BS EN ISO 8253-1 (BSA, 2004). Pure-tone thresholds were measured at octave frequencies from 250 to 8000 Hz inclusive for each ear. Half of the listeners did the PROMS first and the validation tests second, with the other half doing the reverse. Each session took about 1 to 1.5 hour, with a 5 to 10 minute break in the middle of the session. Twenty listeners accepted our invitation to return for the third session (test-retest).

5.1.2 Results

5.1.2.1 Descriptive statistics

All listeners in this study were found to have normal hearing (i.e., listeners who can hear all frequencies that with an intensity less than 20dB as recommended by BSA, 2004). Table 5.2 displays the means and standard deviations for each subtest. The performance levels were slightly better than in Study 2, possibly due to the slightly higher proportion of music students in this sample. So as to obtain a better sense of how a musically untrained population might perform on the PROMS, the descriptive statistics for the entire sample and for the sample after removal of either professional or semi-professional musicians (Table 5.2, in parentheses) was provided.

Table 5.2. Descriptive Summaries for PROMS ranked in order of Mean score (out of 18 max) for Subtest; and out of 162 max for the Composite Score (Study 3)

Test	Raw Mean N = 56 (N = 34)*	*Raw SD* N = 56 (N = 34)*	*Mean* *d′* (*SD*) N=56
Loudness	13.05 (13.02)	2.61 (2.68)	1.55 (1.13)
Tempo	12.88 (12.56)	2.40 (2.54)	1.45 (1.02)
Tuning	12.65 (11.59)	3.15 (3.43)	1.46 (1.29)
Standard Rhythm	12.60 (11.82)	2.45 (2.53)	1.23 (0.90)
Rhythm-to-Melody	12.31 (11.31)	3.15 (3.03)	1.33 (1.28)
Timbre	12.23 (11.79)	2.70 (3.13)	1.42 (1.11)
Pitch	12.21 (11.60)	2.40 (2.49)	1.37 (0.97)
Accent	11.28 (10.69)	2.49 (2.47)	0.80 (0.89)
Melody	10.40 (9.34)	2.56 (2.54)	0.51 (0.91)
COMPOSITE	109.60 (103.72)	17.88 (18.55)	1.02 (0.67)

Note: N = 56. *Values in brackets relate to a subsample (N=34) in which self-rated professional and semi-professional musicians were removed (see main text). A general interpretative framework for *d*' scores is *d*' = 0 (no discrimination ability), *d*' = 1 denotes 69% correct for both *different* and *same* trials.

5.1.2.2 Reliability and Validity of the PROMS

The internal consistency and test-retest reliabilities, reproduced in Table 5.3, show that the revisions to the stimulus material led to improvements in both types of reliabilities. The greatest improvements were on the level of the subtest reliabilities, which reached acceptable levels in most cases (>.60). All external validation tests also showed acceptable internal consistency reliabilities: Timbre (mono) α = .85; MET α = .69; Tempo α = .70. The internal consistency for the AMMA score could not be computed as the score for individual item was not provided in the software.

Overall, the listeners' performances on the current subtests were substantially intercorrelated with the tests selected for validation. Table 5.4 shows the validity intercorrelation matrix. In many cases, the subtests were also distinctively linked to the corresponding validation tests and the correlation strengths are within the recommended range (Kaplan & Saccuzzo, 2009, Table 2.3, Chapter 2). For example, the rhythm test taken from the MET correlated most strongly with both of our rhythm subtests (rhythm and rhythm-to-melody) and, perhaps not surprisingly, also showed moderate correlation with the accent subtest. In contrast, the correlations of the MET rhythm test with the rhythm-irrelevant tasks, such as timbre or pitch, were insubstantial. A similar pattern was also discerned for the timbre subtest. Although the highest correlation of the AMMA melody was indeed with our melody subtest, it also correlated rather strongly with other test components. This could be due to the AMMA tonal test measuring more than just melodic skills, to melodic perception skills reflecting a confluence of various musical skills, or to a combination of both.

Although the MAP tempo task was significantly correlated with the PROMS tempo task, it correlated even more strongly with *other* test components. The most likely explanation for this pattern is that the MAP tempo test also heavily taxes tonal memory. It not only uses melodic sequences, but also requires listeners to judge whether the tempo of the *ending* of the melodies is the same or different compared with the ending of the standard stimulus. As such, the MAP tempo test may be a better measure of tonal memory than of tempo perception per se.

Table 5.3. Cronbach's Alpha, McDonald's Omega, and Test-Retest Coefficient for Subtests and Composite Score (Study 3)

Subtest	α	ω	Test-Retest
Tuning	.81	.87	.68 **
Rhythm-to-Melody	.78	.83	.82**
Pitch	.73	.79	.77 **
Timbre	.77	.84	.68 **
Melody	.56	.73	.77 **
Loudness	.72	.80	.83 **
Standard Rhythm	.61	.75	.62 **
Accent	.55	.70	.71 **
Tempo	.65	.72	.81 **
COMPOSITE	.94	.95	.90 **

Note. $^{**}p < .01$. Sample size for internal consistency was $N = 56$; sample size for test-retest was N=20. The coefficient size of Alpha and Omega depending on whether the general factors are all equal: if there is lots of variability in the general factor loading, the Alpha coefficient can be less than Omega; if there is little variability in the general factor loading, the Omega coefficient can be less than Alpha (Zinbarg et al., 2005)

Table 5.4. Validity Correlation Between AMMA, MET, and Timbre (Mono) with the PROMS (Study 3)

Subtests (PROMS)	Tonal (AMMA)	Rhythm (AMMA)	Rhythm (MET)	Tempo (MAP)	Timbre (Mono)
Melody	**.68****	.60**	.46**	.60**	.23
Rhythm-to-Melody	.43**	**.42****	**.64****	.44**	.33*
Standard Rhythm	.51**	.44**	**.60****	.37**	.23
Accent	.48**	.37**	.37**	.44**	.24
Tempo	.33*	.33*	.22	**.33***	.36**
Timbre	.30*	.27	.15	.32*	**.53****
Tuning	.48**	.41**	.28*	.47**	**.41****
Pitch	.34*	.33*	.12	.37**	**.49****
Loudness	-.10	-.11	-.05	.05	**.40****

Note. AMMA = Advanced Measures of Music Audiation; MET = Musical Ear Test; MAP = Musical Aptitude Profile. $N = 52$. Targeted validity correlations are in bold. $^*p < .05$. $^{**}p < .01$, two-tailed.

5.1.2.3 PROMS correlations with musical background variables

Self-rated musicianship showed a strong significant correlation with composite music background, $r = .87$, $p < .01$; suggesting that listeners rated their level of musicianship based on their musical training experience. To examine the criterion validity of the PROMS, correlation analyses between the three music background assessments and the PROMS composite score were computed. There was a moderately strong relationship between the PROMS composite score and the composite music background[28], $r = .63$, $p < .01$ (7 dimensions, $\alpha = .74$; $\omega = .82$); self-rated musicianship, $r = .66$, $p < .01$; as well as the MM-Competence Scale[29] , $r = .66$, $p < .01$ (7 items, $\alpha = .87$, $\omega = .89$). This sizeable relationship between PROMS scores and real-life musical proficiency provides support to the test's criterion validity. However, it is also important to note that only about 30-40% of the variance in test performance could be explained in terms of musical training alone. This suggests that the current battery accounts for variance in musical skills beyond those that can be captured by self-reported musicianship status, such as music perceptual ability unaffected by music training or other factors that were not examined

[28] Test-retest reliability for the Music Background composite was $r(18) = 1.0$, suggesting that listeners were consistent when reporting their objective musical experience.
[29] Test-retest reliability for MM-competence scale was $r(18) = .96$.

in this thesis (i.e., environmental, genes, motivation).

5.1.2.4 Intercorrelations and preliminary factorial structure of test components

Overall, the correlations among subtests were substantial (see Table 5.5), pointing to a generic "musicality" factor, akin to Spearman's general intelligence or *g* factor. Unfortunately, the sample of this study was too small to examine whether a similar dual or ternary structure may subtend general musical perceptual ability by means of factor analysis. However, a preliminary inspection of factor structure is nonetheless possible if subtest scores, rather than individual items, are used as variables since the required subject-to-variable ratio was about 6:1 (Field, 2009).

To examine the factorial structure underlying the patterns of correlations, a principal component analysis (PCA) was conducted on the 9 subtests with orthogonal rotation (varimax). The Kaiser-Meyer-Olkin measure verified sampling adequacy for the analysis, KMO = .85 ('great' according to Field, 2009), and all KMO values for individual subtests were >.74, which is well above the acceptable limit of .5 (Field, 2009). Bartlett's test of sphericity X^2 (36) = 267.428, $p < .001$, indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. 2 components had eigenvalues over Kaiser's criterion of 1 and in combination explained 68.73% of the variance. Given the convergence of the scree plot (see Figure 5.1) and Kaiser's criterion on two components, this is the number of components that were retained in the final analysis. Table 5.6 shows the factor loadings after rotation. The items that cluster on the same components suggest that component 1 represents the structural processing factor, and component 2 the sensory processing factor. It should be noted that the loudness subtest was only loosely connected to the other subtests. In contrast to other subtests of the sensory factor, it had no cross-loading on the first factor (Table 5.6), and only comparatively modest correlations with the other subtests (Table 5.5).

Table 5.5. Intercorrelations of All PROMS Subtest Scores Including the Composite Score (Study 3)

Subtests	Comp	Tuning	Pitch	Tempo	Timbre	Accent	Melody	R-M	Rhythm
Tuning	.83**								
Pitch	.79**	.72**							
Tempo	.77**	.63**	.65**						
Timbre	.75**	.42**	.49**	.52**					
Accent	.75**	.62**	.65**	.52**	.51**				
Melody	.73**	.56**	.49**	.44**	.68**	.47**			
R-M	.70**	.50**	.40**	.43**	.53**	.38**	.50**		
Rhythm	.70**	.52**	.42**	.32*	.55**	.46**	.64**	.62**	
Loudness	.55**	.53**	.43**	.54**	.23	.38**	.07	.33*	.16

Note. Comp = Composite score; R-M= Rhythm-to-Melody; $N = 56$.

*$p < .05$. **$p < .01$ (two-tailed).

Table 5.6. Factor Analysis (Varimax Rotation) of the PROMS (Study 3)

Test	Structural	Sensory
Melody	**.86**	.17
Rhythm-to-Melody	**.84**	.15
Accent	**.77**	.28
Rhythm	**.68**	.31
Loudness	-.07	**.85**
Tempo	.32	**.77**
Tuning	.46	**.73**
Pitch	.42	**.73**
Timbre	.47	**.61**
Eigenvalues	3.21	2.98
% of variance	35.64	33.01
α	.87	.92

Note: N= 56; Factor loading over .50 in bold

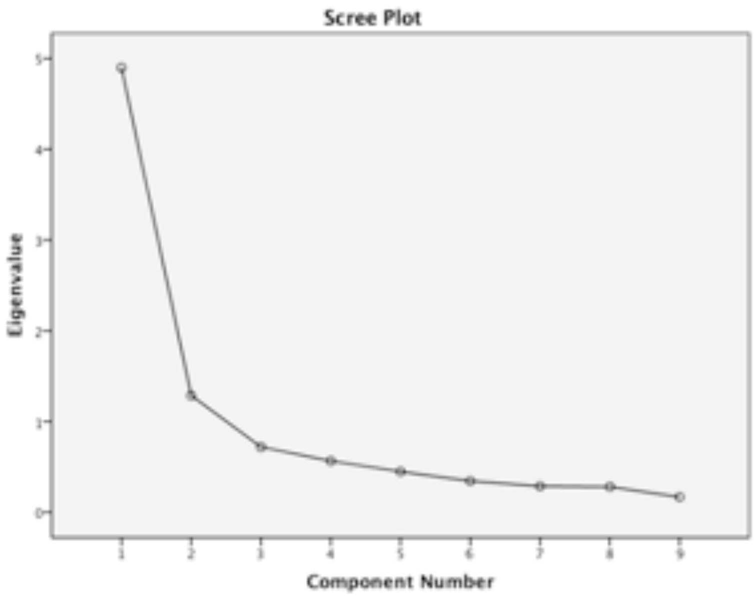


Figure 5.1. Scree plot suggested that there were two factors extracted (eigenvalue > 1)

5.1.3 Discussion

The aim of this chapter was to examine the psychometric properties (reliability and validity) of the PROMS. Three existing music batteries were selected for convergent validity, an additional timbre test was developed in monophonic (single instrument) to examine the content validity, and three music background assessments were used to examine the criterion validity of the PROMS.

Overall, the test development was successful. Specifically, both internal consistency and test-retest reliability were excellent for the composite score. The reliability coefficients for the individual subtests were modest, but nonetheless respectable given their relatively small number of trials. The convergent and content validity of the test-battery were also established, as well as the criterion validity where individuals with extensive musical training were found to perform significantly better on the test than those with no training. It was also found that the PROMS accounts for variance in musical skills beyond self-reported musicianship status and previous musical ability tests. Although the PROMS has shown good reliability and validity set by the contemporary standard which outlines a criteria for a good test (see Chapter 2, section 2.4.1 for the criteria), acquiring evidence about the meaning of the test should be an ongoing process in order to learn what the test does and what it means (Dunnette & Borman, 1979). This will be discussed further in the "Future Direction, section 8.4" of

Chapter 8.

The next chapter attempts to learn more about the PROMS by examining the correlation between musical perceptual ability with other non-musical cognitive functions. The examination of these cognitive functions has two purposes: (1) To examine musical perception ability links with other non-musical abilities, and (2) to examine the discriminant validity of the PROMS.

CHAPTER 6

The Associations between Music Perception Ability with General Mental Ability (Intelligence), Short-term Memory, Working Memory and Auditory Discrimination Skill

Chapter 4 presented the construction and development of the Profile of Music Perception Skills (PROMS) and Chapter 5 has shown that the reliability and validity of the PROMS is satisfactory. However, a question to be addressed is whether this perceptual ability is correlated with other non-musical cognitive abilities, such as general mental ability (GMA), short-term memory (STM), working memory (WM), auditory discrimination skill and, as a by-product, whether such "*transfer* effects" (transferral of knowledge from one skill to another; Barnett & Ceci, 2002) could be unique to listeners who have had music training?

On the contrary, music ability, in Gardner's (1985) view, deserves to be considered as an "autonomous intellectual realm" (p.126) that is independent from other non-musical cognitive functions and intellectual systems. This chapter serves to examine these issues: Are music abilities a reflection of more general cognitive abilities and do musical skills, aptitudes and training contribute to general cognitive abilities? Or are musical abilities a separate domain of cognitive ability? In addition to examining the associations between non-musical cognitive functions and music perception ability, the present study also examines the discriminant validity of the PROMS.

6.1 Introduction

Scholarly interest in associations between music and cognitive abilities has grown in recent years. Comparisons of musically trained and untrained participants represent natural experiments that have ramifications for issues central to cognitive science, including plasticity (Trainor, 2005), modularity (Peretz & Coltheart, 2003), talent (Howe, Davidson, & Sloboda, 1998), and transfer (Schellenberg, 2005, 2006).

The chapter first gives a general overview of the research into intelligence and memory, and subsequently their relationships with music. As there are various types of intelligence, memory and auditory discrimination tests, the names of the test-batteries and their conceptual distinctions are specified. The discussion then addresses the specific non-musical variables that are examined in this study in the following order: intelligence (general mental ability; Section 6.1.1), memory (STM and WM; Section 6.1.2, and auditory discrimination skill (Section 6.1.3). The findings and interpretation of the results are reported.

6.1.1 Individual differences in intelligence

Individual differences in intelligence are often measured in psychometric tests, and these tests can be categorised into three main types as shown in Table 6.1.

Table 6.1 Type of Intelligence Tests and Their Measurement Outcome

Test	Example of Test	Outcome
Verbal and Non-verbal	Wechsler Intelligence Scale for Children	General Intelligence
Non-Verbal	Culture Fair Intelligence Test	General Intelligence (fluid intelligence + crystallized intelligence)
Verbal and/or Non-verbal	Verbal: Raven's Progressive Matrices (RPM) Non-verbal test: Mill Hill Vocabulary (MHV)	RPM= Eductive ability (or general intelligence- see Section 6.1.1.4 for more detail) MHV = Reproductive ability RPM + MHV = General Intelligence

The term (general) intelligence as shown above, can refer to various things, and is often used interchangeably with other terms, such as cognitive ability, mental ability, and IQ (intelligence quotient) (Deary, Penke, & Johnson, 2010). In order to increase the reader's understanding of previous studies, this study will retain the original intelligence terms that were used in each study (i.e., intelligence, cognitive ability, mental ability, and IQ).

The following section reports the discussion about the relationship between intelligence and music: in particular the relationship between intelligence with music listening, music training and music reading ability. The types of intelligence tests and music tests (criteria) are described. Next, the intelligence measure used in this study is introduced (Raven's Test), the rationale of the selection of this test is reported, and the relationship with music and Raven's Test is discussed.

6.1.1.1 Music listening and intelligence

Music originally was appreciated as an art form but has received unprecedented attention in research with regards to its usefulness in understanding cognitive functions of the human brain (Thaut, 2005). One of the areas that particularly intrigued investigators was the relationship between music and intelligence: whether music capacity improves intelligence and vice versa, and this is reported in the following section.

Shuter-Dyson and Gabriel (1981) reviewed more than 50 studies that used ranges of intelligence tests with a sample size of around 160000, published from the year 1924 to 1979 and concluded that the finding "does not show many consistent trends…this may be partly because many of the subtests are much less reliable when considered separately" (p.80). But in general, Shuter-Dyson and Gabriel (1981) reported the correlation between the general intelligence tests and musical ability tests were mostly found to be around .30. Shuter-Dyson (1999) also commented that non-aural music tests such as music grades, theory of harmony or history of music were found to have higher correlations with intelligence scores, although such correlation details were not reported.

The inconsistent findings between the association of intelligence and music continued after 1980. This is probably due to the different types of intelligence measurements as well as the music ability measurements that were used across the studies (i.e., active or passive music listening tests, music reading, music theory). To give an example, the study which investigated the question of whether "Mozart's music improves spatial skill" grouped 36 college students in three conditions: listening to Mozart's Piano Sonata K.948 for 10 minutes; listening to a relaxation tape for 10 minutes; and silence for 10 minutes (Rauscher, Shaw, & Ky, 1993). Listeners' spatial reasoning skills were measured immediately after the listening tests and it was found that only listening to the music of Mozart enhanced spatial performance. This enhancing effect however, was brief (10-15 minutes). This study was later repeated and no significant connection was found between the two variables (Stough, Kerkin, Bates, & Mangan, 1994). The only difference between these two studies was the spatial tests: Rauscher and colleagues (1993) employed the spatial subtests from the *Stanford-Binet Intelligence Scale* whereas Stough and colleagues (1994) employed the *Raven's Advanced Progressive Matrices*.

Although several follow-up studies supported Stough and colleague's (1994) conclusion that Mozart's music did not improve spatial skill (McKelvie & Low, 2002; McCutcheon, 2000), other studies who were in line with Rauscher et al. (1993) suggested such enhancement was the result of listeners' arousal level and mood (Husain, Thompson, & Schellenber, 2002; Schellenberg, Nakata, Hunter, & Tamoto, 2007; Thompson, Schellenberg, & Husain, 2001). It has been suggested that low arousal and negative mood impair performance on cognitive tasks whereas moderate arousal and positive mood facilitate performance and learning (Berylne, 1967; Koester & Farley, 1982).

6.1.1.2 Music training and intelligence

Halpern and Bower (1982) proposed that long term music learning has trained musicians to encode task-relevant information in a rapid and accurate manner, and that this enhancement of cognitive processes in music is also shown in other areas of performance processing. The effects of music training are generally reported as having a positive association with intelligence components such as spatial ability (see Hetland, 2000a for a review), reading ability (Anvari et al., 2002; Deasy, 2002; Butzlaff, 2000; Hurwitz, Wolff, Bortnick, & Kokas, 1975), selective attention (Kraus & Chandrasekaran, 2010; Loui & Wessel, 2007; Hurwitz et al., 1975), and mathematical achievement (Schellenberg, 2005; Vaughn, 2000; Graziano et al., 1999; Cheek & Smith, 1999).

Several studies interpreted the relationship between intelligence and music training by categorising listeners into "musician" or "non-musicians" based on their length of music training experience. For instance, using three test-batteries (*Leistungspruefsystem*, *Cattell's Culture Free Intelligence Test* and *Berliner Intelligenzstruktur-Test*) that measure verbal comprehension, word fluency, space, closure, perceptual speed, reasoning, number and memory, and with ages and education levels controlled, it was found that musicians showed no general intellectual advantage over musically untrained listeners, except for the closure and perceptual speed skills. The authors concluded that music talent is based on intuitive rather than logical thinking (Brandler & Rammsayer, 2003; Helmbold, Rammsayer, & Altenmüller, 2005).

In contrast, in a study with children, it was found that verbal ability and non-verbal reasoning was associated with the extent and types of instrumental training (Forgeard, Winner, Norton, & Schlaug, 2008). In this study, 59 children were divided into three groups: participants with a minimum of three years traditional instrumental training (where note reading was emphasised); participants with a minimum of three years of Suzuki lessons (where listening skill was emphasised) and a "no-lessons" group. Both the musically-trained groups were found to perform better in the tonal and rhythm subtests of *Gordon's Intermediate of Music Audiation*, as well as the motor skill subtest and verbal ability subtest of the *WISC-III Vocabulary,* and the non-verbal reasoning tasks of the *Raven's Standard and Advanced Progressive Matrices*. Conversely they did not perform well in the phonemic awareness of the *Auditory Analysis test* or the spatial skills of the *WISC-III's* Block Design and Object Assembly subtests.

In order to ascertain that improvement in intelligence scores is not due to parents' background whereby children are exposed to more learning opportunities, such variables were controlled in a study by Schellenberg (2006). Even with the variables of family income, parents' education and involvement in other extracurricular activities held constant, the length of music lessons (number of months) of 147 children was shown to have significantly modest and positive association with composite intelligence score (*WISC-III Full Scale*: Verbal Comprehensive, Perceptual Organisation, Freedom from Distractibility, Processing Speed) and academic achievement *(Kaufman Test of Educational Achievement [K-TEA]*: Mathematical Application and Computation, Reading Decoding and Comprehension, Spelling), but not social adjustment measures (*Parent Rating Scale of the Behavioral Assessment System for Children [BASC]*). A similar effect was also observed in 150 undergraduates where the composite intelligence score (*Wechsler Intelligence Scale for Adult test [WAIS-III])* was found to significantly associate with the number of years of playing music (Schellenberg, 2006).

Using a more controlled experimental approach, instead of categorising listeners into musicians and non-musicians based on the extent of their training experience as in the above research, Schellenberg (2004) incorporated actual music lessons during the experiment. 144 six-year-old children were randomly divided into four groups: participants with keyboard lessons and voice lessons as the music group; participants with drama lessons and the "no-lessons" acted as the control group. The full scale of IQ using *WISC-III* and *K-TEA* was measured before and after the 36 weeks of skill training and it was found that the IQ for the music group improved significantly compared to the control group, but again, the differences were relatively modest, with the size of the effect (d = 0.35) being midway between effects considered small (0.2) and medium (0.5) by Cohen (1988). Similarly, Rauscher and Zupan (2000) also used actual music lessons in their study. After just four months of keyboard training, the music group performed significantly better at spatial tasks (Puzzle Solving subtest from the *McCarthy Scales of Children's Abilities and* Block Building subtest from the *Learning Accomplishment Profile Standardized Assessment*). The performance between the music and non-music groups was significantly greater after eight months.

More strikingly it seems that the positive correlation between music lessons and intelligence scores may vanish over time. During a three-year longitudinal study, 78 children were divided into a piano lesson group and a no-lessons group. Their cognitive and spatial skills were measured prior to the experiment and for the subsequent three years using *Developing Cognitive Ability Test (DCAT)* that measures verbal, quantitative and spatial skills. It was reported the children who received piano lessons performed statistically significantly better in the spatial and cognitive tasks after the first two years. Although both groups showed enhancement in cognitive and spatial skills after the third year, there was no significant difference between the control and music groups even with the children's sexes, the parents' incomes and employments, and their family structures (single or two parent-family) were controlled (Costa-Giomi, 1999).

6.1.1.3 Music reading and intelligence

Using a different approach to look at non-instrumental skill - music reading ability (the ability to read a musical score with respect to pitch, time, rhythm, and expression) was found to have positive relationship with intelligence score. In these experiments 128 listeners were asked to choose the correct music score that matched the piano performance played by the experimenter. It was found that poor music readers scored lower in the *Otis Self-administering Tests of Mental Ability* compared to good music readers (King, 1954). Likewise, a study by Tierney and colleagues (2008) also used music reading skill as the criterion of "musicians" and "non-musicians" as the authors were interested in the possible link between *music training* and cognitive functions such as memory and intelligence. The authors argued that, "While some of the non-musician subjects may have attained a certain level of musical skill without gaining the ability to read music, it is rather unlikely that they could have gone through any significant musical training regimen without acquiring that ability" (p.180). Therefore four groups of participants were recruited: gymnasts, video gamers (who played video games for at least 10.5 hour/week), psychology students and music students. The gymnasts, video gamers and psychology students were all categorised as non-musicians because they could not read a music score. In contrast with King's (1954) study, Tierney and colleagues (2008) reported the performance in the *Word-familiarity (FAM) Vocabulary Test* between musicians and non-musicians were equal, showing that musicians did not have better intellectual ability than non-musicians.

6.1.1.4 The current study

Overall, there have been a number of substantial studies examining the relationship between intelligence and music. The intelligence tests ranged from examining a specific intelligence skill such as verbal (reproductive) ability (e.g., Mill Hill Vocabulary; Raven et al., 1993) to a general intelligence test that consisted of a series of subtests such as memory, reasoning, processing speed, executive function, mathematic, vocabulary, reading, spelling and spatial ability (*WISC, CFIT, K-TEA*). Similar trends are presented in music ability measurement where it ranges from the number of years of musical training, actual music lessons, music reading skill, or the administration of music test-batteries such as *MBEA* or *Gordon's Intermediate of Music Audiation*. The findings between the relations of intelligence and music do not seem to follow a systematic trend, possibly due to the differences in the intelligence tests or music ability measurements.

In this study, *The Raven's Advanced Progress Matrices* (APM, Raven et al., 1993) test was selected to measure individual differences in intelligence. One reason for this is because compared to other intelligence tests that were reviewed in this chapter, the Raven's Test is "relatively uncontaminated by linguistic background" (Raven et al., 1993, p.4) which suits the purpose of this study to eliminate language issues with non-native speakers. Another advantage is that compared to other intelligence tests where the participants are required to use some prior knowledge skill to solve problems such as vocabulary or mathematic tests, using the Raven's Test only requires the abilities of reasoning and being able to solve problems when dealing with new information, rather than relying extensively on an explicit base of knowledge derived from previous experience (Raven et al., 1993).

Raven (1948, 1993, 2000) stated that APM was developed to assess one of Spearman's *g - the eductive ability*. Eductive ability (from the Latin *educere,* meaning "to draw out") refers to "the ability to make meaning out of confusion, the ability to generate high-level, usually nonverbal, schemata which makes it easy to handle complicity" (Raven, 2000, p.2). Despite Raven (1948) claiming that the Raven's Test is "not a test of general intelligence, and it is always a mistake to describe it as such" (p.13), his original intention of referring to the outcome of the Raven's Test as *eductive ability* unfortunately does not seem to be used widely. Rather it is more common for researchers to reinterpret the score of Raven's Test as general intelligence or general mental ability (Schellenberg & Moreno, 2009; Pajares & Kranzler, 1995; Deary et al., 2000; Frey & Detterman, 2004; Bódizs et al., 2005; Judge, Hurst, & Simon, 2009), analytical intelligence (Carpenter, Just, & Shell, 1990), fluid intelligence, *Gf* (Catell, 1963), reasoning ability (Marshalek et al., 1983) and spatial ability (Newman et al., 1995; Schweizer et al., 2007; Stough et al., 1994). Nevertheless, this thesis refers to the outcome of Raven's Test as the measurement of general mental ability to provide a clearer understanding of the thesis, with the original intention of the test author acknowledged.

Several psychometric tests were found to correlate highly with the Raven's Test, and it has also been proposed that amongst other reasoning tests, it was at the centre of the reasoning solution (Marshalek, Lohman, & Snow, 1983, as shown in Figure 6.1), suggesting the test is a good measure of reasoning ability. Some studies found positive association between the Raven's Test and music ability (e.g., Thompson et al., 2004; Trimmer & Cuddy, 2008; Forgeard et al., 2008) whilst others found otherwise (Schellenberg & Moreno, 2009; Franklin et al., 2008; Stough et al., 1994; Newman et al., 1995). The summaries of these studies are provided in Table 6.2.

The current study consisting of a self-reported music background measurement, as well as the individual music perception score from the PROMS, would allow us to compare this study in parallel with previous research, enabling us to examine the relationship between general mental ability and music ability more closely.

Table 6.2. Summary of Previous Studies Who Investigated the Relationship Between Music Ability and Raven's Test

Study	**Samples**	**Music Criteria**	**Result**
Trimmer & Cuddy (2008)	100 undergraduate students	1. *Montreal Battery Evaluation of Amusia (MBEA)* 2. Raw years of music training 3. Music Factor Score (a weighted combination of 28 music questionnaires)	1. Correlation between Raven's APM and MBEA Full Score, r = .31** (d = .65) 2. Correlation between Raven's APM and Raw years of music training, r = .17 (d = .35) 3. Correlation between Raven's APM and Music Factor, r = .34** (d = .72)
Thompson et al. (2004)	56 university students	Music group: 28 music students (Training experience: M =2.5 years, *SD*= 3.5 years) Control group: 28 students with no musical experience.	Music students performed better in Raven's APM, $t(54)$, 2.88, p =.006 (d = .39)
Forgeard et al. (2008)	59 children (ages 8-11)	Music group: 41 children received at least three years of instrumental music training. The mean number of years training was 4.63 (SD = 1.10). Control group: 28 children with no musical training experience.	Controlling for ages, music training duration significantly predicted *Raven's Colored Matrices* (partial r^2 = .13, p < .01; d = .54), *Standard Progressive Matrices* (partial r^2 = .10, p = .02; d = .41), and *Advanced Progressive Matrices*, (partial r^2 = .12, p = .01; d = .49)
Schellenberg & Moreno (2009)	40 undergraduates	Music group: 20 students who have at least 8 years of music training. "Each year of music lessons on two or more instruments was considered as two years, and each additional year of playing regularly (beyond lessons) was considered to be equivalent to half a year of lessons. Using these criteria, the group had 14.3 years of lessons on average	The music and the controlled groups did not differ on Raven's Test, (APM) d = .04. The correlations between Raven's Test and all music tests were not significant, as presented below.

Study	Samples	Music Criteria	Result
		(*SD* = 3.8)" (p.21). Control group: 20 students with no music training experience. Music Task: Pitch Processing Speed: Using a three-up, one-down adaptive procedure, participants heard two pure tones that differed in pitch by 2 semitones and were asked to judge the order of the tones (high-low or low-high). Frequency Discrimination: Using a three-up, one-down adaptive procedure, participant heard two pure tones and was asked to judge whether the higher tone was presented first or second. The range of the frequencies was from 3 semitones to 1 cent (1/100 of a semitone). There were two frequency tasks, one at lower frequency at 400Hz, and another one of higher frequency at 4000Hz. Relative Pitch: Listeners heard two lines of "Twinkle Twinkle Little Star" and "Happy Birthday To You". Their task was to judge whether the melody was in tune or out of tune. The tuning difference occurs on the dominant notes of the scale, ranges from 2 semitones to 5 cents. Adaptive procedure was used.	Pitch Processing Speed: $r = .03$ ($d = .06$) Frequency Discrimination (Low): $r = -.16$ ($d = -.32$) Frequency Discrimination (High): $r = .25$ ($d = .51$) Relative Pitch (Twinkle): $r = -.13$ ($d = -.26$) Relative Pitch (Happy): $r = -.20$ ($d = -.41$)
Franklin et al. (2008)	25 students	There were two experiments in this study. There were 12 musicians and 13 non-musicians in the first experiment; and 11 musicians and 9 non-musicians in the second experiment. The musician and non-musician criteria are as below:	Experiment 1: musicians: mean = 27.6, SD = 4.2; non-musicians: mean = 25.7, SD = 3.9 ($d = .25$) Experiment 2: musicians: mean = 25.7, SD = 7.9; non-musicians: mean = 28.3, SD = 5.2 ($d = .22$)

Study	Samples	Music Criteria	Result
		Music Group: • formal training in music began at age 10 or younger; • at least nine years of continuous training in music; • currently played and practiced at least 15 hours/week; • enrolled in an undergraduate or graduate music program; and • a self-rated sight-reading skill of 4 or better on a seven-point scale. Control Group: • did not currently play an instrument; • no history of playing an instrument prior to age 10; • never played an instrument for longer than one year; and • a self-rated sight-reading skill of 1 on a seven-point scale	*t*-tests revealed there was no significant difference between the Raven's scores for musicians and non-musicians in both experiments, $p > .05$.

Study	Samples	Music Criteria	Result
Stough et al. (1994) Newman et al. (1995)	30 participants 114 participants	Listeners' spatial skill was measured using Raven's APM before and after the following conditions: Condition 1: Listening to Mozart's Piano Sonata K.948 for 10 minutes Condition 2: Listening to popular or relaxation music (for 10 minutes Condition 3: Sitting quietly for 10 minutes	There were no differences in Raven's score amongst groups before or after the conditions, nor there is significant effect for type of music. (*d* <.19) .

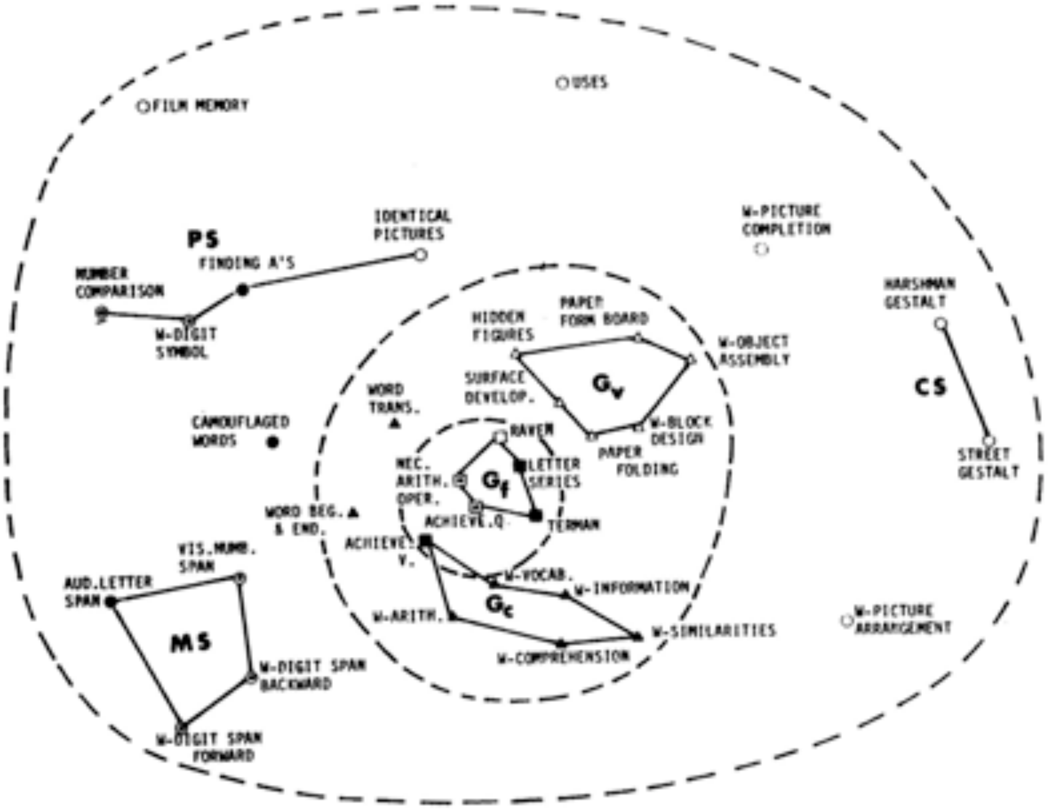


Figure 6.1. The intercorrelations amongst various tests in multidimensional scaling in three levels (N=241), showing the centrality of the Raven as reasoning solution. The shapes indicate the complexities: squares (complex); triangles (intermediate); circles (simple). The shading of the shapes indicates the content of the test: black (verbal), dotted (numerical) and white (figural-spatial). Figure from Marshalek et al. (1983, p.122).

6.1.2 Memory

The positive association between music training and other forms of memories have been described before, such as verbal memory (Brandler & Rammsayer, 2003; Chan et al., 1998, Kilgour et al., 2000; Ho et al., 2003), auditory memory (Wallentin et al., 2010a), and visual memory (Jakobson et al., 2008). Highly skilled musicians often display an amazing capacity to perform a large repertoire of complex musical pieces from memory. This skill requires the coordination of successful encoding and retrieval of musical information processed through auditory, visual and kinaesthetic senses (Williamon & Valentine, 2002; Tierney, Bergeson, & Pisoni, 2008).

The ability of memorisation can be measured by memory span, which refers to the maximum amount of sequential information that an individual can remember accurately (Gathercole, 1999). Short-term memory and working memory are examples of memory capacity that can be measured by memory span. Short-term memory is typically measured by a simple span task that only requires the passive retention of information. Working memory, in contrast, involves complex span tasks that require simultaneous storage and processing of information, such as backward digit span or operation span (Baddeley, 1997; Engle, 2002). Working memory is also influenced substantially by genetic factors, with heritability estimates around 50% (e.g., Kremen et al., 2007; Polderman et al., 2006), and is stable across time (e.g., Unsworth, Heitz, Schrock, & Engle, 2005).

Baddeley (2000) proposed a working memory model (WMM) that consists of four main components: the central executive and its storage system; the phonological loop; the visuo-spatial sketch pad and the episodic buffer, a multi-dimensional coordination store. The phonological loop which deals with verbal materials has two components: articulatory control processes and phonological store. The articulatory control processes deal with incoming *visual information* where it is assumed to be verbally encoded and rehearsed – this process is also known as subvocalisation (Baddeley, 1997). The phonological store functions as a temporary centre for *speech information* where it is assumed that auditory spoken material has automatic access to this store.

It was unclear how music is processed in the human memory system until Salamé and Baddeley (1989) examined whether musical material is processed in the phonological system originally proposed to be for speech processing only. The study found that short-term memory can be disrupted by background music, particularly vocal music rather than instrumental music. They concluded that there is a filter system within the phonological store that governs access to the store based on the acoustic characteristic of speech. This would include some music, particularly vocal music, which shares some acoustic characteristics of speech.

Further evidence of the overlapping of speech and music in the phonological system was supported in a study by Semal, Demay, Ueda and Halle (1996). Listeners were asked to recognise whether two test tones were the same or different when speech and tones were used as the intervening stimuli. The intervening stimuli were found to have little effect on the discrimination performance, however, the variation in pitch, regardless of the stimulus type, had a large disruption effect. It was proposed that the pitch of non-speech and speech is stored in the same location in the phonological system. In a more recent study to test this theory, Williamson, Baddeley, and Hitch (2006) manipulated the pitch proximity of melody sequences in parallel with phonological similarity in a recall paradigm, and it was found that performances with melodies composed of closer pitch proximity were significantly poorer compared to melodies with distant notes, as observed in the phonological similarity in letters. The study further supported previous research where music may be processed in a similar way to speech in the memory system.

6.1.2.1 Short-term verbal and non-verbal memory

Memory research in section 6.1.2 proposed that music and speech are processed in the same memory system, the phonological store. Evidence related to this issue also emerged in studies that examined the relationship between music ability and short-term verbal memory. For example, Chan and colleagues (1998) divided their participants into two groups: 30 of the participants were musically trained students (who had at least 6 years of musical training before the age of 12), and the other half were non-musicians who had no musical training experience. The participants were asked to do two short-term memory tasks: verbal memory task and visual memory task. In the verbal memory task, the participants were asked to recall a 16-word list that was presented orally to them. In the visual memory task, the participants were presented ten simple figures from the *Benton Visual-Retention Test*, and asked to draw the figures from memory. The findings found that musicians scored significantly higher in the word-recall task, but performed the same as non-musicians for the visual memory task.

In a different verbal memory study where 60 undergraduate students with a wide range of years of formal music instruction (0 to 15 years) were recruited, the participant's short-term verbal memory was assessed using the Logical Memory subtest from *Wechsler Memory Scale*. Two stories were presented orally to the participants three times, and the participants were then asked to recall the story after the presentation. A significant correlation was found between years of music training and verbal-story recall, $r = .44$, $p < .01$ (Jakobson, Cuddy, & Kilgour, 2003).

Similar findings also emerged in a lyrics memory study (Kilgour, Jakobson, & Cuddy, 2000). In this study, 78 listeners were recruited. Half of the listeners were musically trained to at least Royal Conservatory Grade VIII level, while the other half were listeners who had less than one year musical training or no musical training at all. These listeners were randomly assigned to three experimental groups (spoken, sung, or sung with piano prelude) where they were required to listen to the verses four times in each condition.

Listeners were asked to recall the material verbatim immediately following the first, second, and fourth presentations. If the listeners could not recall the material verbatim, they were asked to recall whatever they remembered. After the fourth presentation and subsequent recall, listeners were required to complete a number search tasks during a 15-minute retention interval. The purpose of this task was to prevent rehearsal without interfering with the memory because it was a non-verbal task. Immediately after the retention interval, listeners gave their final verbatim recall. The scoring was out of a maximum 25 points, which was based on the criteria of the Logical Memory subtest. Music-trained listeners were found to outperform untrained listeners in all three conditions where the authors concluded that music training leads to enhanced memory even with non-musical verbal materials (Kilgour, Jakobson, & Cuddy, 2000).

Stronger evidence with regards to the positive relationship between verbal memory and music training emerged when short-term verbal memory (participants were asked to reproduce 20 nouns that were aurally presented to them) was found to be the *only* cognitive test from an entire intelligence test-battery (*Berliner Intelligenzstruktur-Test*) that showed significant differences between music students and non-musicians (Brandler & Rammsayer, 2003)

In the interest of examining the relationship between digit-verbal memory and music, Huntsinger and Jose (1991) adopted the tests of digit recall, tone recall, digit recognition and tone recognition with 56 musically trained and untrained children aged between 6 and 10. The children were asked to reproduce and recognise (singing and speaking) comparable 3-, 4-, 5-, and 6- digital complex tone sequences and digit sequences using the numbers 1-8. Not only was it found that musically trained children performed better in both digit and music tasks, there were also moderate and strong correlations between the music and digit task. A principal components factor analysis was performed and only one factor was extracted (eigenvalue = 2.74) which accounted for 67.4% of the variance.

Using the ability of reading score as the criterion of “musician” and “non-musician”, Tierney and colleagues (2008) attempted to delve further into whether the memory capacity in trained musicians expands beyond non-verbal materials by using a parallel audio-visual test. Participants were asked to reproduce the sequence of coloured lights on a box in three conditions: visual-only (coloured panels were illuminated); audio-only (names of colours were read through headphones); audio-visual (coloured panels and names of colours were prompted simultaneously). Musicians were found to significantly outperform non-musicians in the auditory task, but not in the visual and audio-visual tasks.

6.1.2.2 Working memory and music experience

Evidence so far has suggested that short-term auditory memory may be a unitary phenomenon, and that music training can enhance it. This may be due to music training augmenting auditory temporal-order processing, where it then mediates in the association of length of musical training with memory recall (Jakobson, Cuddy, & Kilgour, 2003).

Looking beyond the association between short-term memory and music training, two recent studies suggested that working memory is associated with musical training. In the study by Lee, Lu, and Ko (2007), 40 adults (mean age = 22 years) and 40 children (mean age = 12 years) were recruited from universities and primary schools. Half of the *children* had received an average of 6.1 years of music training; half of the *adults* had received an average 14.3 years of training. All participants performed the short-term and working memory tasks, including forward and backward digit span, non-word span, operation span, and simple spatial span tasks. The findings emerged that the musically trained adults and the musically trained children performed better in short-term *and* working memory tasks than the control groups.

In the study by Meinz and Hambrick (2010), 57 pianists' sight-reading skills, working memory capacities, and "estimate of deliberate practice" were measured. All pianists performed six sight-reading piano pieces of three difficulty levels. The sight-reading performances were audio-recorded and evaluated using a scale from 1 (lowest) to 7 (highest) by two experts who held a graduate degree in music and taught piano at university level. Interrater correlation was high (average r = .86), supporting the rating consistency between the two experts.

The working memory capacities of the pianists were measured with four working-memory tasks (operation span, rotation span, reading span and matrix span tasks), and an average score of the four tasks were calculated. The "estimates of deliberate practice" were measured by interviewing participants about their piano-playing history based on the procedures used by Ericssson, Krampe, and Tesch-Römer (1993, cited in Meinz & Hambrick, 2010). In particular, the pianists were asked to indicate on a time line when lessons began and ended, when teachers changed and other events, the number of hours per week they spent on deliberate practice alone, and the number of hours they spent on sight-reading practice.

The results showed that the "estimate of deliberate practice" accounted for nearly half the variance (45.1%) in the sight-reading performance, and that the working memory capacity accounted for a significant 7.4% proportion of the variance. More strikingly, there was no evidence that the "estimate of deliberate practice" reduced the working memory effect in sight-reading performance. The authors concluded that although deliberate practice helps to develop a high level of sight-reading skill, working memory capacity might limit the ultimate level of performance that can be attained.

Evidence from Section 6.1.2 seems to suggest the possibility of using memory processing functions when engaging in music activities, however, it is not clear how or which part of the music component (s) or activities relate to the short-term memory and working memory. This question will be examined in this study.

6.1.3 Auditory Discrimination Skill

Chapter 5 provided convergent, criterion and content validities of the PROMS. This study aims to examine an additional validity of the PROMS - the discriminant validity. The discriminant validity refers to the degree to which the PROMS is *not similar* to another test that is supposedly unrelated. One way to do this is to compare the PROMS with a generic auditory discrimination test in order to examine whether the PROMS measures music-specific skill or if it only measures generic auditory skill. The discriminant validity of the PROMS would be supported if it shows non-significant correlation with the auditory discrimination test.

Amongst a selection of auditory tests such as Masking Level difference (MLD, Bocca & Antonelli), Cross-Faded Synthetic Tones (cochlear implant users, Rahne, Rasinski,& Neumann, 2010), and Spectral ripple discrimination (Won, Drennan, & Rubinstein, 2007), the Gap Detection Task was chosen for this study as it does not have a strong pitch element and it is also an established auditory discrimination test to measure individual differences in auditory ability that appeared to be a more generic auditory skill amongst healthy listeners (Kidd et al., 2007).

The Gap Detection Task is widely used to assess auditory temporal acuity and resolution [the ability to detect an auditory signal in brief duration presented at rapid rates] (Bertoli, Smurzynki, & Probst, 2002; Muluk, Yalcinkaya, & Keith, 2011), auditory nerve activity (Zeng et al., 2005) and intensity or amplitude resolution (Horwitz, Ahlstrom, & Dubno, 2011; Garadat & Pfingst, 2011; Kidd et al., 2007). The Gap Detection Task was also frequently used in investigating development of speech perception in children (Muluk, Yalcinkaya, & Keith, 2011; Wightman et al., 1989; Hall III & Grose, 1994) and hearing-impaired patients (Glasberg & Moore, 1992, 1989).

In summary, the discussion so far has presented the relationship between musical ability and non-musical variables: being that auditory memory is mostly found to have a positive association with musical ability, and that the association between general mental ability and musical ability does not seem to follow a consistent trend. The methodology to examine these associations is presented in the following section.

6.2 Methods

6.2.1 Listeners

76 listeners (18 males, 58 females), aged 18-38 (mean=21, SD=4) and with self-reported normal hearing, took part in this study. They were students and staff from the University of York, U.K. who gave their time in exchange for course credit or cash rewards of 14 Pounds Sterling. There were 57 listeners with an average of 10.4 years of music training and 19 listeners without a music background for the latter group, that is, no instrumental learning and/or other form of music education.

6.2.2 Stimuli

The music trials, music background questionnaire and MM-Competence Scale questionnaire were identical to Study 2 (Appendices 2 and 3). In order to investigate whether the non-musical cognitive tests are related with other non-musical skills, listeners in this study were also asked to report and comment in a text box if they had other outstanding non-musically related skills (e.g. High Academic Achiever, Sports, Programming, Medical, Maths and others, see Appendix 2 for details). The stimuli design for the auditory, memory and general mental ability tests are described below.

6.2.2.1 Gap Detection Task

This study employed the Gap Detection Task[30] as specified by Zeng et al. (2005), who used an adaptive two-down and one-up procedure, yielding a 70.7% performance level (Levitt, 1971). 11 independent 750-ms digital samples of white noise contained gaps of silence of 10 different durations at their temporal centre. The gap durations ranged from

[30] The program for the Gap Detection Task was developed with the assistance of Michael Cheung.

0.5 to 256ms, in log steps, whilst the total durations remained constant. An adaptive 3-alternative forced choice procedure, with visual feedback regarding the correct response, was used to determine the gap detection thresholds. The interstimulus interval was 1,250 ms and the order of the signal and standard sounds were randomised.

The test started with a medium-large signal at 32ms to facilitate listeners' understanding of the test. The level increased (or the difference was reduced) after two consecutive correct responses and the level decreased after one incorrect response (2-down, 1-up). If the listener made an incorrect response from two or more consecutive correct responses or vice versa, a reversal was recorded. Each run was terminated after 12 reversals or after a maximum of 70 trials. The average score from the last eight reversals was used to determine the gap detection threshold. A lower gap detection threshold score indicates that a listener has better auditory discrimination ability.

6.2.2.2 Raven's Test

The *Raven's Advanced Progressive Matrices (APM)* was used to assess the general mental ability of the listener. The APM was digitized and run using Limesurvey 1.91 (http://www.limesurvey.org) and there were two sets of the APM tests. 12 items from Set I acted as a practice test and 36 items from Set II constituted the actual test. Participants were given 50 minutes to complete the test. In each question for the Raven test, participants were asked to select a 'missing' piece from a choice of eight to complete an incomplete picture. The Set II questions became progressively more difficult as they required greater reasoning ability and intellectual capacity. The score was calculated from the number of correct responses. An example of Raven's Test is presented in Figure 6.2.

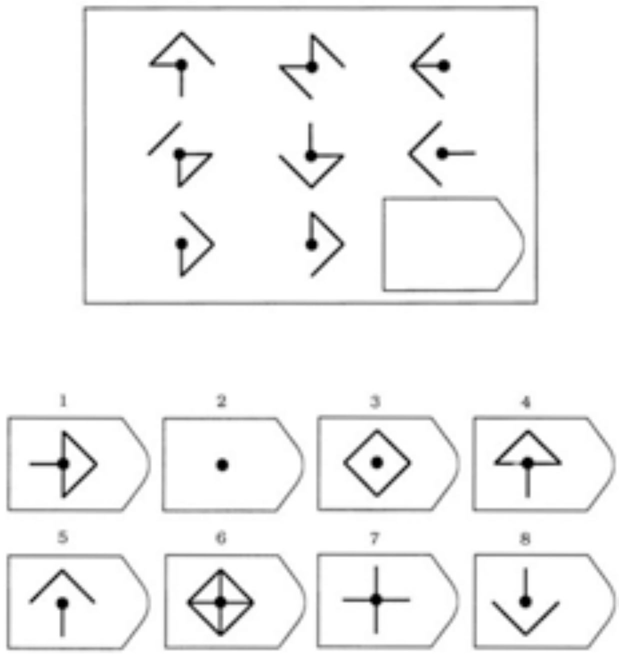


Figure 6.2. An example of Raven's Test (Raven et al., 1993).

6.2.2.3 Digit-Span Memory Test

The Digit-Span Memory Test is widely used as a short-term and working memory measurement in test-batteries and studies, for example *WASI* (Wechsler, 1999), *WISC* (Wechsler, 1991), *Automated Working Memory Assessment* (AWMA; Alloway, 2007) and other studies (i.e., Sloboda, 1985b; Wallentin et al., 2010a; Huntsinger & Jose, 1999). More specifically, the Forward Digit-Span Test is used to measure auditory *short-term memory;* and the Backward Digit-Span Test is used to measure auditory *working memory*.

The Digit-Span Memory Test from the *Automated Working Memory Assessment* (AWMA, Alloway, 2007) was computerised using Limesurvey 1.91 for group testing. The digit numbers were read to the listeners and they were asked to reproduce the number from memory by typing them in the provided answer columns in the Limesurvey. The Forward Digit-Span Test had nine levels (one to nine digits) with each level consisting of six questions. The Backward Digit-Span Test consisted of six levels (two to seven digits) where the listeners were asked to type the numbers in reverse order. The total number of correct responses calculated became the scores.

6.2.3 Procedure

There were two experimental sessions. In one session, listeners completed the PROMS and questionnaires. In a second session, they completed the non-musical cognitive tests. The procedure for the music test-battery was identical to that in section 5.2.3. The non-musical cognitive test session had four tests, namely, Forward Digit-Span Test, Backward Digit-Span Test, the Gap Detection Task and the Raven's Advanced Progressive Matrices. Half of the listeners did the PROMS first and the non-musical cognitive tests second, with the other half doing the reverse. The order of each session was also counter-balanced. Each session took about 60 to 90 minutes with a 5 to 10 minute break in the middle.

6.2.4 Data-Analytical Strategy

To answer the study's research question, "*how do non-musical variables associate with the PROMS score?*", the following strategy was used.

1. Factor analysis and Correlation: Before answering the primary question of this study, the factor structure of the PROMS, as well as the intercorrelations between the non-musical variables and the PROMS scores were first examined to provide a general overview of the data.

2. Regression analysis: To answer the primary question of this study, multivariate regression analysis with the hierarchical regression method was performed to examine how various non-musical variables predicted the PROMS score. The advantage of using the hierarchical regression method is that it allows us to test whether there are any additional variables in the subsequent block that add or reduce a significant amount of the explained variance to the outcomes. Other methods such as stepwise techniques have the limitation of over-fitting[31] and under-fitting which is not suitable for the current study (Field, 2009). Moderation analysis will also be conducted to examine whether a variable (moderator) alters the direction or strength of the relation between a predictor and an outcome (Frazier, Tix, & Barron, 2004; Baron & Kenny, 1986; Holmbeck, 1997).

[31] Fitting: Open fitting refers to "having too many variables in the model that essentially make little contribution to predicting the outcome" and under fitting refers to the converse scenario of "leaving out important predictor" (Field, 2009, p. 213)

3. Mediation analysis: Mediation analysis allows us to understand whether the relationship between the predictor (non-musical variable) and the outcome variable (PROMS scores) is mediated or influenced by a different variable (Frazier, Tix, & Barron, 2004). Mediation analysis was performed using the recent practices and new recommendations from Rucker, Preacher, Tormala, and Petty (2011, Figure 6.3). In addition, a bootstrapped analysis for estimating the indirect effect was obtained using procedures described by Preacher and Hayes (2008). This is considered a preferred method as it does not assume normality of the distribution of the indirect effects and hence provides stronger protection against Type 2 errors compared to normal procedures such as the Sobel Test (Frazier et al., 2004).

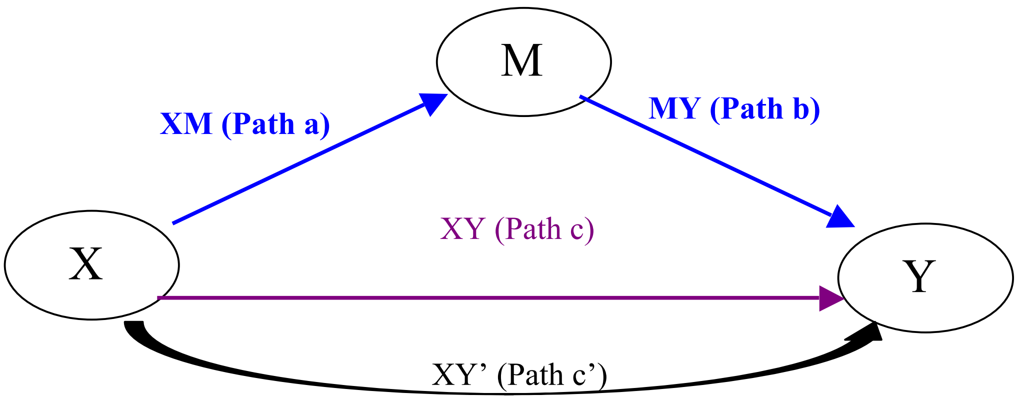


Figure 6.3. Mediation Model: X = Predictors; Y = Outcome; M= Mediator. The total effect of X is denoted by *c*. The product ab estimates the strength of the mediated or indirect effect of X on Y, that is the amount of increase in Y that occurs in X is due to M. The *c’* coefficient estimates the strength of the direct effect of X on Y, that is any effect of X on Y that is not mediated by M (Frazier et al., 2004).

6.3 Results

6.3.1 Factor Analysis and Correlation

To examine the factorial structure of the PROMS score, a factor analysis with varimax rotation was computed on the subtest scores. A two-factorial structure was replicated (Structural Processing Factor and Sensory Processing Factor) and is presented in Table 6.3. The intercorrelations between the PROMS, the two factor structures, the non-musical variables and the music background composite score are presented in Table 6.4. Scores on both factors were found to relate to listeners' musical experience, but sensory processing factor has shown lower correlation compared to structural processing factor. The effect size between the sensory processing factor and structural processing factor with music background appeared to be significantly different using Steiger's test, z = 2.416, p <. 05 (2 tailed). The negative correlation between the auditory discrimination skill and the PROMS composite score suggests that better performances on the PROMS were associated with better performances on the Gap Detection Task. The results suggest that the correlation was not significant, and that the PROMS cannot be seen as a measure of generic auditory skill, supporting the preliminary discriminant validity of the PROMS. Other details of the correlations will be discussed further in the regression analysis.

Table 6.3. Factor Analysis (Varimax Rotation) of the PROMS (Study 4)

Test	Structural Processing Factor	Sensory Processing Factor
Melody	**.77**	.10
Rhythm	**.74**	.38
Rhythm-to-Melody	**.72**	.18
Accent	**.70**	.43
Loudness	-.02	**.82**
Tuning	.36	**.70**
Tempo	.33	**.69**
Pitch	.49	**.67**
Timbre	.35	**.67**
Eigenvalues	2.74	2.89
% of variance	32.19	30.46
α	.85	.88

Note. N= 76. Factor loading over .50 appear in bold.

Table 6.4. Correlations between Composite Music Background Score (MB), PROMS Composite, General Mental Ability (GMA), Short-term memory (STM) and Working memory (WM) [Study 4]

	MB	PROMS	Struc	Sensor	GMA	STM	WM
PROMS	.51**						
Struc	.57**	-					
Sensor	.37**	-	-				
GMA	.12	.17	.28*	.04			
STM	.27*	.26*	.42**	.08	.47**		
WM	.22*	.24*	.39**	.06	.45**	.78**	
AUD	-.27*	-.21	-.19	-.16	-.21	-.07	-.03

Note. N= 76. MB- Music background composite score; PROMS – PROMS composite score; Struc-Structural Processing Factor; Sensor – Sensory Processing Factor; GMA – General mental ability (Raven's Test); STM- Short-term memory (Forward Digit-Span Test); WM - Working memory (Backward Digit-Span Test); AUD - Auditory Discrimination Skill (Gap Detection Task); The negative correlation in the Gap Detection shows *positive association* with performance of other variables as a larger gap coefficient represents poorer performance. ***p* < .01; **p* < .05 (2 tailed).

6.3.2 Regression Analysis

To understand how various non-musical variables are related with the PROMS, multivariate regression analysis with hierarchical regression method was performed to examine how various non-musical variables predicted the PROMS composite scores and the structural processing factor. The Gap Detection Task and Sensory Processing Factor are excluded from the regression analysis as Gap Detection Task shows no significant correlation with the PROMS score; and the Sensory Processing Factor shows no significant correlation with any of the external variables (see Table 6.4). Regression analysis procedure is described below:

Step 1 [Explained Variance]. The first step (Block 1) of the regression was to examine the explained variance (R^2) of the non-musical variables on the PROMS score. To do this, the non-musical variables were entered individually as the *independent variable*, and the PROMS scores were entered individually as the *dependent variable*. For example, the STM was entered as the "independent variable", and the PROMS composite score as the "dependent variable". This step was repeated with WM, GMA and Structural Processing Factor respectively (See Table 6.5). Overall, the memory capacity (short-term memory and working memory) significantly predicted the PROMS composite score and the Structural Processing Factor; and general mental ability also accounted for a significant portion of the variance in the Structural Processing Factor.

Table 6.5 Regression Model (Block 1): The Explained Variance of Non-Musical Variables on the Composite Scores and Structural Processing Factor.

	Outcome: Composite PROMS			Outcome: Structural Processing Factor		
Predictor	R^2	***F***	**β**	R^2	***F***	**β**
STM	.07	5.50*	β = .26*	.17	15.40**	β = .42**
WM	.06	4.51*	β = .24*	.15	13.11**	β = .39**
GMA	.03	2.10	β = .17	.08	6.15**	β = .28**

Note. STM= Short-term memory; WM= Working memory; GMA= General Mental Ability; $^{**}p < .01$, $^{*}p < .05$ (2 tailed).

Step 2 [Explained Variance with Music Background Score]. Table 6.4 shows that listeners' music background composite score showed to have significant correlations with most of the non-musical variables. In order to examine whether this experience would reduce or increase the explained variance in *Block 1*, the music background composite score was entered as the independent variable in *Block 2*. Statistically significant increments in variance (ΔR^2) in *Block 2* would indicate contributions of music background to the PROMS scores above and beyond the performance of the individual non-musical variables (see Table 6.6).

Overall, there was a medium to large effect (Cohen, 1988) of music background that explained the variance of the PROMS scores, above and beyond the non-musical cognitive performance (ΔRs^2 = .21-.29, ΔFs = 21.36-33.97). The beta coefficients of the non-musical variables were also reduced compared to Table 6.5, in some cases, from significant to non-significant. For example, the beta coefficient of STM on PROMS score was original significant in Table 6.5, β = .26*, but has dropped to .14 (n.s.) when music background composite score was included in the model. This result suggests that listeners' music background has a significant role on the influence of the association between the non-musical variables and the PROMS score (Table 6.6). There are two ways to examine this influence: (1) Moderation analysis allows us to examine whether "the strength of the association between the external variables and the PROMS score" are dependant on listeners' *level of music background*; (2) Mediation analysis examines *how significantly* the role of listeners' music background influences the association between the external variables and the PROMS score.

Table 6.6 Regression Model (Block 2): Music Background Score is added to the Regression Model to Predict the PROMS Composite score and Structural Processing Factor.

	Outcome: Composite PROMS			Outcome: Structural Processing Factor		
Predictor	**Δ*R*²**	**Δ*F***	**β**	**Δ*R*²**	**Δ*F***	**β**
STM & MB	.21**	21.36**	β (STM) = .14 β (MB) = .48**	.23**	27.62**	β (STM) = .28** β (MB) = .49**
WM & MB	.22 **	22.44**	β (WM) = .13 β (MB) = .48**	.25**	29.72**	β (WM) = .28** β (MB) = .51**
GMA & MB	.25 **	24.79**	β (GMA) = .11 β (MB) = .50**	.29 **	33.97**	β (GMA) = .21* β (MB) = .55**

Note. The R Square change (ΔR^2) is significant, suggesting music background score explained the variance of the PROMS scores, above and beyond the non-musical variables. STM = Short-term Memory; WM= Working Memory; GMA= General Mental Ability; MB = Music Background Composite Score **$p < .01$, *$p < .05$.

Step 3 [Moderation Effect]. To examine whether the effect of the non-musical variables on the PROMS scores is dependant on the levels of the music background, the score of the non-musical variables, as well as the music background composite score, were centralised and converted to an *interaction score* to examine the interaction (moderator) effect (Frazier et al, 2004). The interaction scores were then entered as the independent variable in the third block of the regression analysis and the result is presented in Table 6.7. The interaction effect was not significant (as shown in the ΔR^2 *and* ΔF columns), suggesting the association strengths between the non-musical variables and the PROMS scores were not affected by the *levels* of listeners' musical experience.

Table 6.7 Regression Model (Block 3): Interaction Effect of the Non-Musical Variables and Music Background score in predicting the PROMS Composite score and Structural Processing Factor.

Predictor	**Outcome: Composite PROMS**		**Outcome: Structural Processing Factor**	
Predictor	**Δ*R*²**	**Δ*F***	**Δ*R*²**	**Δ*F***
STM x MB	.02	2.40	.02	2.33
WM x MB	.02	2.26	.02	1.90
GMA x MB	.01	0.64	.00	0.07

Note: The R Square change (ΔR^2) is not significant, suggesting the level of music background score did not affect the association between the predictor and the outcome. STM = Short-term Memory; WM= Working Memory; GMA= General Mental Ability; MB = Music Background Composite Score

Step 4 [Mediation effect]. To gain a deeper understanding of how the relationships between the PROMS score and the non-musical variables are influenced by the Composite Music Background score, mediation analysis was conducted. The "initial causal variable/predictor" was the non-musical variables; the "outcome variable" was the PROMS scores; and the proposed mediating variable was the music background composite score. Both unstandardised and standardised coefficients are reported.

Mediation Analysis (Step 1, Path a). As recommended by (Rucker et al., 2011), the first step of mediation analysis is to examine the relationship between the mediator (music background composite score, MB) and the predictors (external variables), presented in Table 6.8. Both of the STM and WM show significant relationships with the music background score, except for the GMA.

Table 6.8 Mediation Analysis – The Association Between Predictor and Mediator (Path a)

Testing steps in mediation model	*B*	SE *B*	95% CL	β
[Path a]				
Outcome: Music Background (MB)				
Predictor: STM	.138	.058	.022 to .255	.265*
Predictor: WM	.142	.072	-.002 to .285	.220*
Predictor: GMA	.062	.062	-.062 to .187	.115

Note. STM= Short-term memory; WM= Working memory; GMA= General Mental Ability; $^{**}p < .01$, $^{*}p < .05$ (2 tailed).

Mediation Analysis (Step 2, Path b). The second step of mediation analysis is to examine the relationship between the mediator (i.e. MB) and the outcome (i.e. PROMS score), shown in Table 6.9. Both of the PROMS composite score and Structural Processing Factor showed significant relationships with the MB (mediator).

Table 6.9 Mediation Analysis – The Association Between Predictor and Outcomes (Path b)

	Outcome: PROMS				Outcome: Structural Processing Factor			
	Coef	SE	95% CL	β	Coef	SE	95% CL	β
Path b								
Predictor: MB	.089	.017	.055 to .124	.512**	.110	.019	.074 to .147	.570**

Note: MB = Music background composite score; ***p* < .01, **p* < .05 (2 tailed).

Mediation Analysis (Step 3, Path *c* and *c'*). The next step of mediation analysis is to examine the association between the predictors and the outcomes *including* the mediator (path c, Total Effect), and the association between the predictors and the outcomes *excluding* the mediator (path c', Direct Effect). As seen in Table 6.10, the coefficient of the Direct Effect is reduced when the music background is excluded from the model (i.e. compare Total Effect with Direct Effect).

Table 6.10 Mediation Analysis for testing Total Effect and Direct Effect

	Outcome: PROMS				Outcome: Structural			
	Coef	SE	95% CL	β	Coef	SE	95% CL	β
Path c and c'								
Predictor: STM								
Total Effect (c)	.024	.010	.004 to .044	.263*	.041	.011	.021 to .063	.415**
Direct Effect (c')	.012	.010	-007 to .032	.137	.029	.010	.009 to .048	.284*
Predictor: WM								
Total Effect (c)	.027	.011	.002 to .051	.240*	.048	.012	.023 to .072	.388**
Direct Effect (c')	.015	.012	-.010 to .039	.132	.033	.012	.011 to .057	.275*
Predictor: GMA								
Total Effect(c)	.016	.011	-.006 to .037	.166	.029	.012	.006 to .052	.277*
Direct Effect(c')	.010	.001	-.009 to .029	.109	.023	.010	.003 to .042	.214

Note. The Direct effect is reduced when the music background score (mediator) is excluded from the model. ***p* < .01, **p* < .05 (2 tailed).

Mediation Analysis (Step 4, *ab*). To examine whether the coefficient reduction in Table Table 6.10 is significant, the Indirect Effect (ab) and the effect ratio is reported in Table 6.11. The results showed that the reductions of these coefficients were significant (*indirect effect*) as shown in the bootstrapping analysis (i.e., the bootstrap test is statistically significant if both the lower and upper limit of confidence level have the same sign, either both positive or both negative), except for GMA. The effect ratios were also provided in Table 6.11 to express the amount of the total effect that is explained by the indirect effects via the mediators. For example, an effect ratio of 0.48 was observed for the significant indirect effect of the STM on the PROMS composite score, indicating that about 48% of the total effect was explained by the music background composite score. Overall, the results suggested that music background composite score significantly mediated the relationship between the non-musical variables and the PROMS scores, except for the GMA which also showed to have no significant relationship with the mediator in Table 6.8. An example of the mediation effect between STM and the PROMS is illustrated in Figure 6.4.

Table 6.11 Testing Indirect Effect and Effect Ratio using Bootstrapping Technique

	Outcome: PROMS			Outcome: Structural		
	Coef	SE	95% CL	Coef	SE	95% CL
Path ab						
Predictor: STM						
Indirect Effects (ab)	.012	.005	.003 to .023	.013	.006	.004 to .026
Effect Ratio	.480			.316		
Predictor: WM						
Indirect Effects (ab)	.012	.007	.001 to .029	.014	.008	.000 to .032
Effect Ratio	.449			.292		
Predictor: GMA						
Indirect Effect (ab)	.005	.006	-.006 to .017	.007	.010	-.008 to .021
Effect Ratio	.346			.227		

Note. The Indirect Effect for STM and WM are significant (the bootstrap test is statistically significant if both the lower and upper limit of confidence level have the same sign, either both positive and both negative). STM= Short-term memory; WM= Working memory; GMA= General Mental Ability

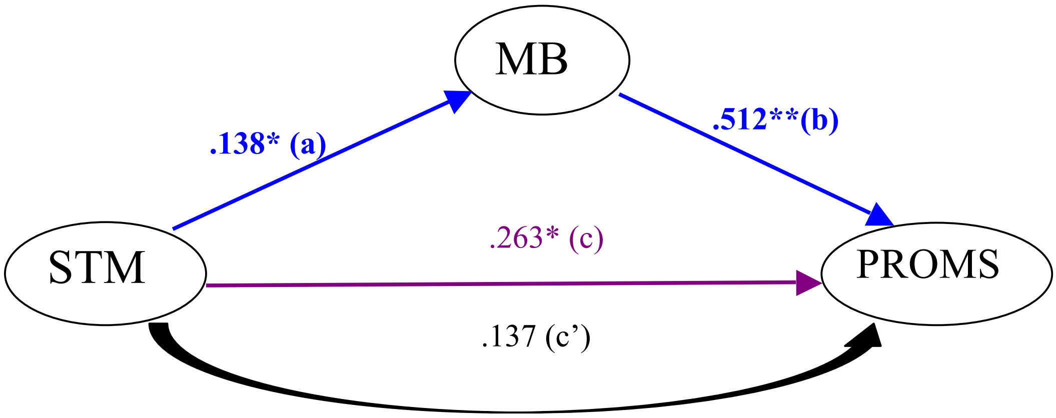


Figure 6.4. An example of mediation effect presented using standardised coefficients. The total effect of STM (denoted by *c*) reduced significantly (denoted by *c'*) when the music background composite score is controlled in the mediation model. The drop of the coefficient was significant (see Table 6.11), supporting music background composite score as the mediator. STM- Short-term memory; MB- Music Background Composite Score; PROMS- Total PROMS score.

6.3.3 The Relationships between the PROMS subtests and Non-Musical variables

To examine the relationship between the PROMS subtests and the non-musical variables, correlation and mediation analyses were conducted. It was found GMA was correlated with the rhythm-to-melody subtest (r = .38, p <.01) and the rhythm subtest (r = .23, p <.05); and auditory discrimination skill was correlated with the accent subtest, r = .24, p < .05. Mediation analyses of these correlations were examined. The results showed that the relationship between GMA and the rhythm-to-melody and rhythm subtest were significantly mediated by STM (Figure 6.5; effect ratio [rhythm-to-melody] = .537; effect ratio [rhythm subtest] = .512); and the relationship between auditory discrimination skill and the accent subtest was mediated by the music background composite score (effect ratio = .363).

Listeners in this study were also asked to report if they had any other outstanding non-musically related skills. Amongst various answers, participants who answered "logic or mathematical skill" as their outstanding non-musical skill were found to have significant correlation with the GMA, r = .52, p <.01, this effect was not found to be mediated or moderated by music background or memory capacity.

There was no significant correlation between the age and gender with the PROMS scores or the non-musical variables, $p > .05$, nor were they found to have any moderation or mediation effect on the relationship between the non-musical variables and the PROMS scores.

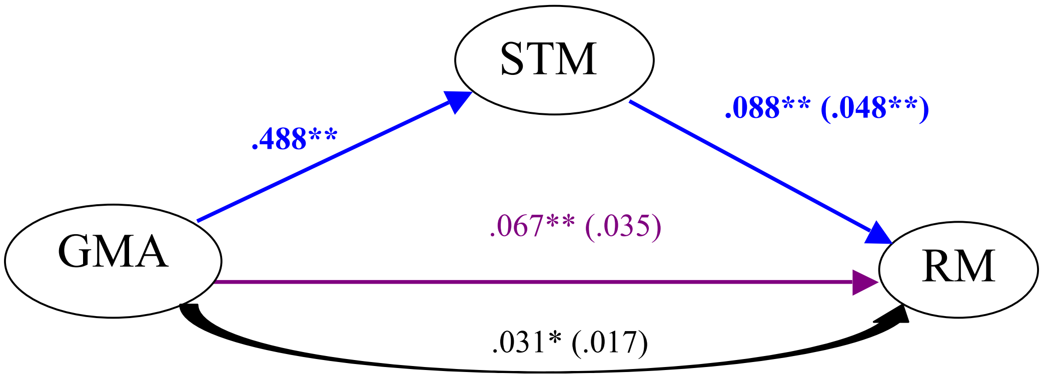


Figure 6.5. Mediation analysis between general mental ability (GMA),Rhythm-to-Melody (RM) and short-term memory (STM). Mediation analysis for rhythm subtest is provided in the parentheses. Coefficients reported are unstandardized.

6.3.4 Correlations between non-musical cognitive tests and self-reported music background

In order to compare the present study with previous studies which have used self-reported music background measurement, the correlation between the Raven's Test, the Gap Detection Task, Forward and Background Digit-Span Test and self-reported music background such as composite music background, years of musical training, music qualification and music reading skills are reported in Table 6.12. Overall, the findings show that the composite music background and years of musical training correlated significantly with short-term memory, working memory and auditory discrimination skill but not with the general mental ability. Music reading ability only showed significant correlation with the Gap Detection Task; and music qualification was shown to relate significantly with short-term and working memory only.

Table 6.12. Correlations between Raven Matrices, Gap Detection, Digit Recall Tests and Music Background Questionnaire

	Raven	Digit FWD	Digit BCK	GAP
Year	.14	.22*	.20	-.31**
Qualification	.06	.29*	.24*	-.16
Music Reading	.11	.03	.13	-.37**
Composite Music Background	.12	.27*	.25*	-.27*

Note. Raven= Raven Matrices Test; Gap = Gap Detection Task; Digit FWD = Forward Digit Recall; DigitBCK- Backward Digit Recall. The negative value in the Gap Detection shows *positive correlation* with other variables as a larger gap coefficient represents poorer performance. **$p < .01$; *$p < .05$

6.4 Discussion

The role of intelligence, short-term and working memory in musical perception ability has long been debated. This chapter has taken a fresh look at the issue by relating established measures of non-musical variables to the newly devised objective music battery. This chapter has also examined the discriminant validity of the PROMS. Four additional non-musical tests were employed in this study. The first - the Raven's Test - measures general mental ability to reason and solve problems in new information. The Forward Digit-Span Test measures auditory short-term memory and the Backward Forward Digit-Span Test measures auditory working memory, it is worth noting that working memory also involves some form of short-term memory hence the two tests are not entirely separable. The final test, The Gap Detection Task, measures generic auditory discrimination skill, particularly dealing with the function of auditory temporal capacity. The present study was also carried out to investigate whether self-reported music background score is related to the non-musical cognitive functions as it has often been used as the measure of musical ability by previous research.

Overall, the present main findings suggest the PROMS cannot be seen as a measure of generic auditory skills, memory skill or general mental ability, supporting the discriminant validity of the PROMS. However, these non-musical skills are not completely independent from listeners' musical experience, and are reported in the following section.

6.4.1 General Mental Ability (Raven's Test)

The present study shows that rhythm and rhythm-to-melody perceptions are associated with Raven's Test, a test to measure general mental ability, as well as spatial ability

(Newman et al., 1995; Schweizer et al., 2007; Stough et al., 1994). Although there was evidence that the relationship between the rhythm-to-melody subtest and the Raven's Test was mediated by short-term memory capacity, the relationship between spatial ability and rhythm skills was interpreted as being the overlapping function of the cerebellum in processing spatial properties and musical rhythm tasks (Thaut, 2005; Parsons & Fox, 1995). In particular, Parson and colleagues' reported that visual rhythms have the same effect as auditory rhythms on spatial ability enhancement, and that such enhancement is greater on mental rotation tasks (Cube Comparison) than spatial visualisation tasks (Paper Folding) (Hetland, 2000b).

The correlation between spatial skill and music ability was also found to be positive where children who were given 6 months of keyboard lessons performed significantly better compared to a control group (who did not have keyboard lessons) in the "Spatial-Temporal Math Video Games" which teach both proportional math and fractions (Graziano, Peterson, & Shaw, 1999). Although the exact details of the music lesson was not reported, it is reasonable to believe that this spatial-temporal mathematical ability may be related to rhythm perception as that is generally believed to be foundational in music learning.

When compared to previous studies who have examined the relationship between music ability and the Raven's Test (see Table 6.2), this study is in line with Schellenberg and Moreno (2009) as well as Franklin and colleagues (2008) where (1) no significant correlation was found between the PROMS's tonal-related subtests (melody, pitch, melody, tuning, timbre) and the Raven's Test, (2) and no significant relationships between the self-reported music background and the Raven's Test.

In addition, despite the present study being divided into separate sessions for the music test and the Raven's test on different days - whether to do the PROMS or Raven's Test first was also counterbalanced - the correlation between the rhythm-to-melody and rhythm perceptions and the Raven's Test were still significantly related, suggesting this is not just a short-term enhancement effect as reported by previous research study (e.g., Rauscher, Shaw, & Ky, 1993; Costa-Giomi, 1999). Rather it presents a long-term enhancement effect on listeners who have shown superior rhythmic perceptual skill.

The present findings suggest the application of using reasoning skill (Raven's Test) in music is apparent when dealing with rhythm properties. The reasons for the positive or null relationship between Raven's Test and musical ability in previous research is probably due to the specific rhythmic properties of the music stimuli presented in their studies, or the rhythm perception ability of their recruited listeners.

6.4.2 Short-term and working memory

The present study aimed to examine how short-term memory and working memory are engaged when listening to musical structures and how memory capacity associated with musical background.

The overall correlation between "short-term memory with the PROMS" appeared to be very similar with "working memory with the PROMS". The reason for this pattern is probably due to the PROMS only requiring a simple span task (short-term memory) rather than complex manipulation of the original information that is required in the case of working memory. Therefore only the "short-term memory processing" in the working memory task was activated (Cowan, 2008), hence the correlation of PROMS with short-term and working memory were equivalent.

The rhythm-to-melody, rhythm, accent and melody subtests showed positive associations with short-term memory and working memory capacity. Similarly, these four subtests also showed to have the highest loading solution on the structural processing factor. To further examine the structures of the structural processing and sensory processing factors, the correlations of these two factors were examined with short-term and working memory, general mental ability and auditory discrimination capacity, and further analyses with moderation and mediation effect were also conducted. The structural processing factor was found to relate to memory capacity (both STM and WM), even with music background controlled as the mediator, $\beta = .28$, $p < .05$. In addition, although short-term and working memory showed significant correlation with the PROMS composite score, mediation analysis revealed that around 45% to 50% of this relationship was mediated by listeners' musical experience. This suggests that if the listeners did not have any musical training or experience, the relationship between memory capacity and the PROMS composite score would be insignificant, a finding that further supports the discriminant validity of the PROMS not being a strong memory test.

These findings suggest that the various representations of musical stimuli in the present study have yielded two distinct listening behaviours: (1) using memory to store and recall auditory elements, (2) using auditory sensory mechanisms to process auditory elements. These two behaviours in this study are associated with musical representation rather than the length of the music. For instance, the tempo stimuli that had a similar length as other structural processing factor stimuli appeared to share the same factor loading with the sensory materials that had shorter lengths. A possible explanation is that listeners successfully extracted the tempo information during the first few seconds of the stimuli presentation, just like the sensory processing factor stimuli that lasted only two seconds. The structural processing factor stimuli, on the other hand, required attention throughout the stimuli presentations as the music elements were not presented in a uniform manner.

Overall, these findings accord with previous research where music training was found to associate with memory capacity (Chan et al., 1998; Jakobson, Cuddy, & Kilgour, 2003; Brandler & Rammsayer, 2003; Franklin et al., 2008; Tierney et al., 2008; Wallentin et al., 2010a) and that music memory was correlated with auditory memory (Huntsinger & Jose, 1991; Kilgour, Jakobson, & Cuddy, 2000). Indeed, some musical activities such as musical practice, remembering songs, and music theory have trained listeners to develop a strategic memory to process incoming information into chunks of meaningful musical elements. Studies have reported musicians displayed greater ability of indexing and categorising musical information into meaningful chunks rapidly, retrieve it in memory and then use it to organise the practise sessions in order to produce quality musical performances (Halpern & Bower, 1982; Williamon & Valentine, 2002). This memory ability, however, worsens as the musical notes are presented randomly further supporting the notion that trained musicians use a specific memory structure strategy to remember music information rather than having a good memory capacity in general (Halpern & Bower, 1982).

In addition to the findings of the positive relationship of using memory strategy in music, the present study is consistent with previous research, where it has been found that both the short-term and working memory capacities correlate with the Raven's Test. Ackerman, Beier, and Boyle (2005) performed a meta-analysis of 86 correlational studies to evaluate the claims of the close relationship between working memory, general intelligence and fluid intelligence (Raven's Test). The authors concluded that there was a correlation between the fluid intelligence (Raven's Test) and memory capacity and that they are not isomorphic to each other.

In summary, this study is consistent with the idea that music training is associated with superior performance in multiple domains of memory functioning such as music memory and auditory memory. The present study has also provided direct evidence of the use of two perceptual strategies when processing musical information: memory (structural processing factor) and immediate judgement (sensory processing factor).

These two perceptual skills are distinct from each other and are used to process different type of musical information.

6.4.3 Auditory ability

Music can be seen as the organisation of auditory events based on musical rules, therefore the relationship between music perception and auditory discrimination skill is interdependent because of their shared core properties. Indeed music learning and performance requires the mapping and organisation of auditory patterns to produce meaningful musical output based on musical rules. The aim of the PROMS is to measure auditory skills that are *related to musical rules*, rather than just generic auditory skills. To examine whether the PROMS measures musically-specific auditory skills, the association between the PROMS and generic auditory skill such as the Gap Detection Task was examined. Overall, the PROMS shows non-significant correlation with the Gap Detection Task, thus supporting the discriminant validity of the PROMS.

Despite the PROMS having shown itself to be a non-generic auditory test, the Gap Detection task has shown low-moderate correlation with the composite music background (r = .27), but this correlation is considered as modest since the PROMS has shown moderate-high correlation with the music background composite scores (r = .51). This finding is inline with several studies that have reported long-term musical practice strengthens generic auditory skills (Kraus et al., 2009; Rammsayer & Altenmullaer, 2006; Parbery-Clark et al., 2009b). The positive correlation between auditory skill and music could be the influence of musical practice on the neural enhancement of trained listeners (i.e., neuroplastic adaptation, or mismatch negativity - MMN), and years of musical practice (Gaser & Schlaugh, 2003; Hutchinson et al., 2003; Musacchia et al., 2007).

6.4.4 General summary

Several researchers have examined and interpreted the relationship between cognitive function and music based on self-reported musicianship as a simple binary classification, for example comparing the performance of musicians versus non-musicians on variables such as general IQ and mental abilities (e.g., Helmbold, Rammsayer, & Altenmüller, 2005); memory (e.g., Williamson, Baddeley, & Hitch, 2010); and auditory skills (Strait et al., 2010). The present study improves this investigation by using music perception tasks rather than just using a self-reported musical background measurement.

In particular, rhythm and rhythm-to-melody perceptions were found to be related to general mental ability; melody, rhythm, rhythm-to-melody and accent perception abilities were related to short-term and working memory; whilst loudness, pitch, tuning and tempo, in contrast, seem to be sensory processing skills that deal with judgment in brief music listening analysis.

In summary, the limitation of using binary categorisation of musician and non-musician using self-reported music background was revealed in this study, showing the importance of using a standardised music perceptual tool, in addition to a music background questionnaire when interpreting the relationship between music ability and other non-musical cognitive functions.

The chapter concludes the controlled studies conducted in this thesis. The next chapter investigates the PROMS using more diverse samples gleaned from the Internet to examine whether the findings from the controlled studies can be replicated using a more diverse sample, as well as to examine the preliminary norm of the PROMS.

CHAPTER 7

An Online Study using the PROMS

Chapter 6 (Study 4) demonstrated the relationship between music perception ability and non-musical abilities (general mental ability, short-term memory, working memory and auditory discrimination skill), as well as supported the discriminant validity of the PROMS. This chapter explores the potential of the PROMS beyond a controlled laboratory setting. Results and their implications are reported.

7.1 Study 5

All listeners from previous studies were recruited from the academic sector; they were either students or staff from the University of York, U.K. The problem with such a population is that they are unlikely to be representative of humankind in general and therefore the result cannot be generalised with the normal population (Henrich, Heine, & Norenzayan, 2010). To address this issue, Gosling and colleagues (2010) proposed that the Internet holds great promise for broadening the participant base of research in behavioural science. They further suggested that "Internet methods offer researchers many advantages over traditional methods in terms of improved efficiency, accuracy, cost effectiveness and reach" and that "Internet samples are generally more diverse than the 'traditional samples' in the top psychology journals with respect to gender, socioeconomic status, geographic region, and age" (p.94).

Recently, Amazon Mechanical Turk (MTurk, www.mturk.com), an online marketplace, has been increasingly employed by researchers to harvest Internet samples because of its function as an online resource, connecting people and tools to enable task creation, labour recruitment, compensation, and data collection (Gosling et al., 2010). The MTurk consists of over 100,000 users from over 100 countries who complete tens of thousands of tasks daily (Pontin, 2007). Experimenters are registered as “requesters” and participants are registered as “workers” as defined in the MTurk platform. Requesters can create any computer task using the MTurk default template, or link workers to other online tools and resources (e.g., Limesurvey). Workers browse for available tasks in the search bar and are paid upon successful completion of each task. A requester can refuse payment to workers for unsatisfactory work, and as a negative consequence, workers with higher rejection rates may not be allowed to further participate in tasks should their reputation be deemed unacceptable. Likewise, if a requester refuses a worker’s payment without a valid reason, workers can raise the issue to the MTurk support for further investigation.

The aim of the present study was to examine the factor structure of the music test-battery and also to provide a preliminary norm of the current test. Another purpose of this study was to see whether the result found in controlled studies can be replicated with a more diverse population, and how an uncontrolled testing environment and equipment may have affected the result. This research employed MTurk to recruit participants, as well as invitations via email, social media, word of mouth and direct visits to the online PROMS website that collects Internet data. The reason for including non-MTurk participants was to collect a more diverse population including participants who volunteered to take part in the study without seeking a compensational reward like the MTurk participants.

7.1.1 Method

7.1.1.1 Stimuli

Music stimuli were identical with Studies 3 and 4 for parallel comparisons. Music background and demographic questionnaires, as well as their scorings are reported in Appendices 2, 3 and 4.

7.1.1.2 Procedure

As in Study 4, the online PROMS test (including the questionnaires) was computerised and delivered via LimeSurvey software (version 1.91). The order of the test was presented slightly differently to make it more user-friendly and attractive for the public population. In particular, the test was divided into two main sections, “tonal test” and “temporal test”. The tonal test consisted of the timbre, melody, tuning, pitch and loudness subtests; the temporal test consisted of the rhythm, rhythm-to-melody, tempo and accent subtests. The order of the test was counter-balanced where half of the listeners did the tonal test first and the other half did the temporal test first.

A sound volume calibration test and two practice trials preceded the first music listening test, enabling the listeners to familiarise themselves with the test design and stimuli materials. Subsequently, one practice trial preceded each subtest. Listeners received their scores and text feedbacks[32] after they completed the first section, as well as at the end of the second section. Listeners were given the opportunity to save and resume the test at any point, allowing them to complete the test at their own convenience. The entire test took about 60 to 90 minutes (if completed in one sitting), depending on the listener's Internet connection speed and computer processor power. Data collection was conducted over 6 weeks.

7.1.1.3 Data screening

547 listeners completed the online PROMS. Data were screened using the following strategies as recommended by Lee (2010), Mason and Suri (2011) and Downs et al. (2010).

Time stamp. Time duration for the test was recorded; this included the start and completion time, as well as the time that listeners spent on each question. The durations of the whole test and individual questions were recorded during the pilot test as a reference guide for the actual test. If a listener was constantly found to be spending a much shorter time on each question than the reference time, the data of that listener was removed. For example, each rhythm trial should have taken about 20 seconds to load, play and to make a decision. Several listeners were found to be spending less than 5 seconds in total, hence their data were removed. When listeners spent less time than they expected, it indicated they had given their answer before the music finished playing (N=80).

d'. Listeners who spent less time than expected on each question were also found to have negative *d'* on all subtests. This suggests that *d'* is sensitive with data that contains

[32] The feedback function of the online PROMS was developed by Melissa Saviste.

randomly guessed responses. Therefore, additional listeners who scored negative *d'* for *more than* 6 out of the 9 subtests were also removed. Negative *d'* is sometimes removed in study (e.g., Elhilali et al., 2009) as it indicates either that participants are lacking in motivation to do the test, misunderstand the test instructions, or are simply unable to do the test (N=24).

*Unusual response***.** Listeners who gave their answers in an unusual or recognisably consistent manner (i.e., choosing *Definitely Same* or *Probably Same* answer for the entire test) were also removed (N= 9).

Maths. Listeners were also asked to answer a simple math question (i.e., 5 + 2) in order to make sure they were not 'spammers' (those who attempt to complete the test to receive compensation without regard to the instructions of the requesters) or "bots" (an autonomous computer program that is used to complete the test) (N=4).

Outlier procedures. Once the above procedure was conducted, an outlier detecting procedure was conducted to screen potential outliers such as univariate outliers (listeners who scored 2.5 lower or higher than the standard score) and multivariate outliers (Mahalanobis D^2 that is less than *p* value of .05) [33]. There was no significant difference when these outliers were removed, hence all data remained.

7.1.1.4 Listeners

In total, 117 listeners of 547 listeners were removed (21%). This is considered a normal range as previous research have removed up to 44.3% of the original data when screening for Internet samples (Lee, 2010). Also note that there were higher unsatisfactory data in the beginning of the data collection, where it was found more than 85% of the workers from India violated the screening rules as stated above, hence samples from India were eventually removed from the MTurk data collection. The removal quantity would have been higher if India was included throughout the data collection in MTurk.

Therefore, data for 430 listeners (187 male, 243 females), aged 15-74 (mean= 29,

[33] Mahalanobis D^2 refers to the distance of cases from the means (s) of the predictor variables, cases with $p <.05$ suggest a significant influence of the case to the overall data

SD =11) was used in this study. Listeners came from over 50 countries, with the majority from the United States (37.4%), followed by United Kingdom (16.3%) and Canada (4.7%). 65.8% of the listeners were native English speakers. Education and ethnicity details are provided below in Table 7.1 and Table 7.2. From the 430 listeners, 344 listeners were recruited from MTurk and 86 listeners were invited to do the test via email, social media, word of mouth and directly visiting the website. Only listeners from MTurk received cash compensation for completing the survey ($0.50 to $0.70). 70% of the listeners used earphones or headphones for the test. There were 274 listeners with an average of 12.6 years of music training, and 156 listeners without a music background for the latter group, that is, no instrumental learning and/or other form of music education.

Table 7.1. Education background of the Internet Samples (Study 5)

Education	**%**	% [Age >24] (UNESCO)
College/University Degree (e.g. BA, MA= 41.4%; PhD = 3.7%)	45.11	34.65 (15.87)
Professional School Diploma/Certificate/Similar	33.70	1.23 (8.07)
Completed high school	13.51	5.58 (26.93)
Attended some university/college, but not graduated	6.02	4.65 (-)
Did not finish high school	1.65	0.69 (-)

Note. N=430. Data from UNESCO (2012) is provided in parentheses, because UNESCO only provided data for people age 25 and above, therefore parallel data for the Internet study (listeners age >24) is also provided in the same column with the UNESCO's data. (-) refers to data was not provided in UNESCO (2012).

Table 7.2. Ethnicity background of the Internet Samples (Study 5)

Ethnicity	%
White	75.3
Asian	13.7
Black	3.5
Hispanic	3.0
Middle-Eastern	1.2
Others	3.3

Note. N=430.

7.1.2 Results and Discussion

7.1.2.1 Descriptive statistics and test-reliability of the PROMS

Table 7.3 displays the means and standard deviations for each subtest. The performance levels of the Internet samples were no different compared to Study 4 (d < .20; Cohen, 1988). The internal consistency reliability, reproduced in Table 7.4, shows that the reliability coefficients were lower compared to the previous studies. The drop in reliability is as expected for an Internet test as the testing environment was different across listeners, as well as the equipment that they chose to use (sound card, headphones, speakers, computer models) (Honing & Ladinig, 2008).

For instance, the internal consistencies of pitch and timbre subtests have dropped substantially. An explanation for this is that pitch and timbre perception can vary quite largely depending on the quality of the audio equipment. The Internet samples in this study reported using a wide range of speakers and headphones of different qualities (from "$3 amazon earphones" to high end headphones such as *Sennheiser HD280 Pro*). This variation has inevitably affected the consistencies of perception judgments. In addition, the homogeneity of the samples (age range, education background etc.) is more diverse compared to the controlled study, which may also have affected the reliability coefficients as also observed in several studies (Wing, 1968; Drake, 1957; Stamou, Schmidt, & Humphreys, 2010; Caruso, 2000; Henson, Kogan, & Vacha-Haase, 2001).

Another reason for the lower reliability coefficient may be due to there being fewer professional musicians in the Internet study. As can be seen in Table 7.5, the listeners in Study 3 had higher musical training experience compared to other studies, and the overall internal consistency of Study 3 was also shown to be higher than Study 4 and Study 5. A similar pattern is observed between Study 4 and 5 where the internal consistency of Study 4 is higher than Study 5, but lower than Study 3.

The descriptive mean of the Internet study shows no difference with previous studies and is probably due to the contribution of 'musical sleepers' (see Figure 7.1). Musical sleepers are defined as listeners whose music background is in the lowest quintile of music background composite score *and* who scored above average in the PROMS. The number of musical sleepers in the Internet study (16%) was significantly higher than the number of musical sleepers in the controlled studies (3-4%).

The variations in internal consistency reliability estimates with different sample groups has been described before: "The same measure, when administered to more heterogeneous or more homogeneous sets of subjects, will yield scores with differing reliability" (Thompson, 2003, p.93). This occurs as reliability estimates are heavily affected by total score variability. Indeed, a reliability estimate "is a property of the score of a test *for a particular population of examinees*" (by Wilkinson and Task Force on Statistical Inference (1999), p.569, with italics for emphasis), rather than the reliability of a *test*. Incorrect assumptions of using reliability to define *the quality of a tool or instrument* have been noted by several authors (e.g., Thompson, 2003; Kaplan & Saccuzzo, 2009; Meyer et al., 2001; Henson et al., 2001) who have gone on to suggest that reliability should be reexamined in every sample group.

Nevertheless, the composite score of the Internet study has shown to have excellent internal consistency. The scores for most subtests showed an acceptable internal consistency range, particularly in McDonald's omega reliability (.60 to .70), and the internal consistency for other subtests' scores showed to have fluctuated more when the test was administered to the public population (< .60). The norm distribution (histogram) of the subtests and composite scores can be found in Appendix 5.

Table 7.3. Descriptive Summaries for Subtest and Composite Scores (Study 5)

Test	Raw Mean	Raw SD	*d'* Mean	*d'* SD
Loudness	13.32 (13.08)	2.77(2.55)	1.68 (1.48)	1.20 (1.03)
Tempo	12.87 (12.42)	2.30 (2.37)	1.54 (1.24)	1.00 (0.98)
Tuning	12.05 (12.41)	2.58 (2.53)	1.29 (1.31)	1.10 (1.04)
Rhythm	12.86 (12.23)	2.37 (2.60)	1.38 (1.10)	0.99 (0.95)
Rhythm-to-Melody	11.58 (11.36)	3.00 (2.71)	1.08 (0.91)	1.24 (1.09)
Timbre	12.64 (12.16)	1.87 (2.31)	1.56 (1.33)	0.86 (0.99)
Pitch	12.09 (11.79)	1.82 (2.25)	1.46 (1.24)	0.85 (0.94)
Melody	10.40 (10.79)	2.43 (2.31)	0.47 (0.64)	0.87 (0.91)
Accent	11.31 (10.65)	2.56 (2.53)	0.81(0.56)	0.94 (0.86)
Composite	109.12 (106.9)	14.98 (16.03)	1.02 (0.91)	0.58 (0.59)

Note. N=430. Scores of Study 4 (Chapter 6) are given in brackets. Performance at chance levels is 6.75 (assuming total randomness in choice). A general interpretative framework for *d*' scores is *d'*= 0, no discrimination ability. A *d'* score of 1 denotes 69% correct for both *Different* and *Same* (Keating, 2005)

Table 7.4. Cronbach's Alpha and McDonald's Omega Reliability (Study 5)

Test	α (Study 4)	ω (Study 4)
Loudness	.75 (.71)	.72 (.69)
Tuning	.70 (.70)	.61 (.72)

Rhythm-to-Melody	.75 (.70)	.79 (.71)
Timbre	.48 (.67)	.67 (.75)
Pitch	.58 (.66)	.61 (.86)
Rhythm	.59 (.66)	.60 (.74)
Tempo	.63 (.60)	.68 (.57)
Accent	.57 (.58)	.51 (.68)
Melody	.52 (.57)	.71 (.63)
COMPOSITE	.91 (.88)	.89 (.90)

N=430. Scores of Study 4 (Chapter 6) are given in brackets.

Table 7.5. Mean and Standard Deviations of Composite Music Background (Study 5)

Studies	Music Background Composite Score (Mean)	SD
3	5.86	4.06
4	5.21	3.38
5	3.51	2.90

Note. N= 430

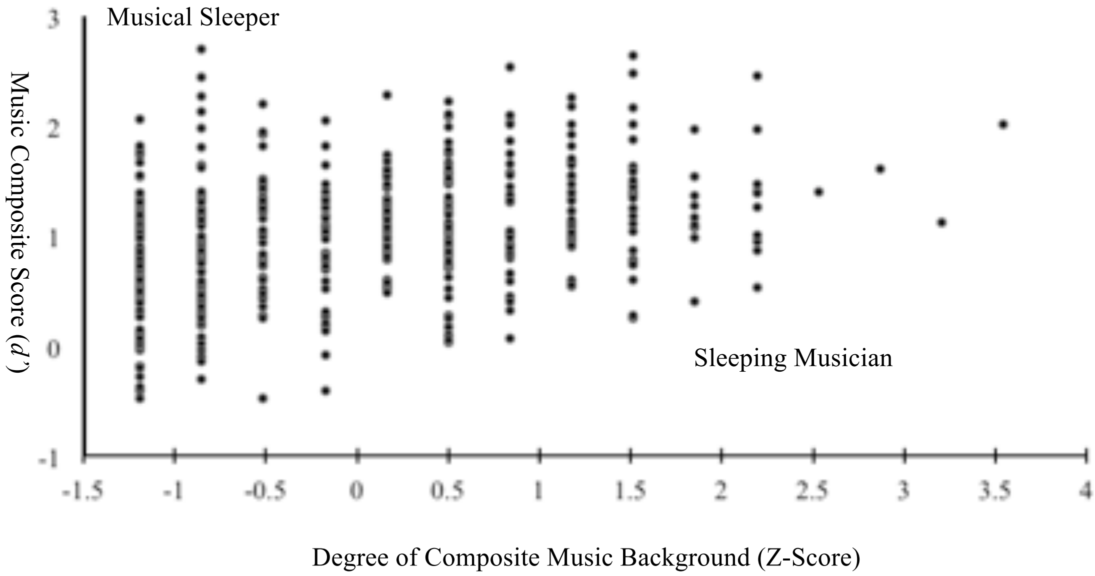


Figure 7.1 N= 430. Scattergram plotting PROMS performance (*d'*) against an aggregate index of musical training including self-reported years of musical training, involvement in critical listening activities, music degrees and qualifications, hours of practice time, perfect pitch, music reading abilities and family influence (see main text and Appendix 2). Range for music background (x-axis: 1-4); range for raw and *d'* score for the PROMS (y-axis): 0-162 (-1 to 3 for *d'*). Extent of training predicts PROMS performance substantially but imperfectly (r = .34, p <. 01). *Upper left corner:* Example of a "musical sleeper" performing well despite minimal musical training. *Lower right corner:* Example of a "sleeping musician" posting a lesser performance despite extensive musical training.

7.1.2.2 Intercorrelations of subtests, factorial structure of test components

As in Study 4 in this thesis, this study shows that the correlations amongst subtests were substantial (see Table 7.6), and that all subtests show moderately strong correlations with the composite score (see Table 7.7). This suggests that musical perception abilities are not completely independent from each other, pointing to the notion of the unitary "musical ability".

To examine the factorial structure underlying the patterns of correlations, a principal component analysis (PCA) was conducted on the 9 subtests with orthogonal rotation (varimax). The Kaiser-Meyer-Olkin measure verified sampling adequacy for the analysis, KMO = .90 ('superb' according to Field, 2009), and all KMO values for individual subtests were >.87, which is well above the acceptable limit of .5 (Field, 2009). Bartlett's test of sphericity X^2 (36)= 1277.90, p < .001, indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. 2 components had eigenvalues over Kaiser's criterion of 1 and in combination explained 58.08% of the variance. Given the large sample size, and the convergence of the scree plot (see Figure 7.2), and Kaiser's criterion on two components, this is the number of components that were retained in the final analysis. Table 7.8 shows the factor loadings after rotation. The items that cluster on the same components lead to a suggestion that component 1 represents "structural processing factor", and component 2 is the "sensory processing factor".

This result is quite similar to the previous study (Study 4), except that the tempo subtest had higher loading in the structural processing factor together with the melody, rhythm, rhythm-to-melody and accent subtests; whilst the timbre, pitch, tuning, and loudness subtests scores loaded highly on Factor 2, the sensory processing factor.

Table 7.6. Correlation Between PROMS Subtests and PROMS Composite Score (Internet Study)

	Composite score including autocorrelation	Composite score excluding autocorrelation	Composite score including autocorrelation	
Subtest	*Pearson r*	*Pearson r*	*beta*	*t*
Rhythm-to-Melody	.75**	.64**	.20	18.76
Accent	.72**	.61**	.20	20.14
Tuning	.70**	.60**	.15	15.36
Rhythm	.69**	.58**	.16	16.04
Tempo	.66**	.57**	.16	16.64

Melody	.65**	.52**	.16	17.35
Loudness	.63**	.49**	.19	20.71
Pitch	.63**	.55**	.10	10.12
Timbre	.62**	.54**	.12	12.54

Note. Autocorrelation refers to the correlation between a subtest and “a composite score that includes this subtest as well as others”. For example, the correlation between “melody subtest” and “Composite Score including autocorrelation” refers to the total score that includes the melody subtest (and other subtests). N= 430; $p < .05$. **$p < .01$ (two-tailed).

Table 7.7. Intercorrelation Between PROMS Subtests (Internet Study)

	R-to-M	Accent	Tuning	Rhythm	Tempo	Melody	Loudness	Pitch
Accent	.53**							
Tuning	.43**	.37**						
Rhythm	.59**	.50**	.40**					
Tempo	.43**	.46**	.41**	.42**				
Melody	.46**	.42**	.45**	.38**	.31**			
Loudness	.30**	.31**	.46**	.28**	.38**	.28**		
Pitch	.38**	.34**	.50**	.30**	.36**	.38**	.50**	
Timbre	.39**	.41**	.43**	.34**	.36**	.35**	.37**	.39**

Note. N= 430. Comp = Composite score; R-to-M = Rhythm-to-Melody; *$p < .05$. **$p < .01$ (two-tailed).

Table 7.8. Factor Analysis (Varimax Rotation) of the PROMS (Internet Study)

Test	Structural Factor	Sensory Factor
Rhythm	**.81**	.13
Rhythm-to-Melody	**.78**	.24
Accent	**.76**	.21
Rhythm	**.74**	.13
Melody	**.56**	.36
Tempo	**.54**	.40
Loudness	.11	**.81**
Pitch	.21	**.78**
Tuning	.37	**.67**
Timbre	.38	**.54**
Eigenvalues	2.57	2.32
% of variance	28.59	25.78
α	.87	.84

Note. N=430. Factor loading over .50 appear in bold

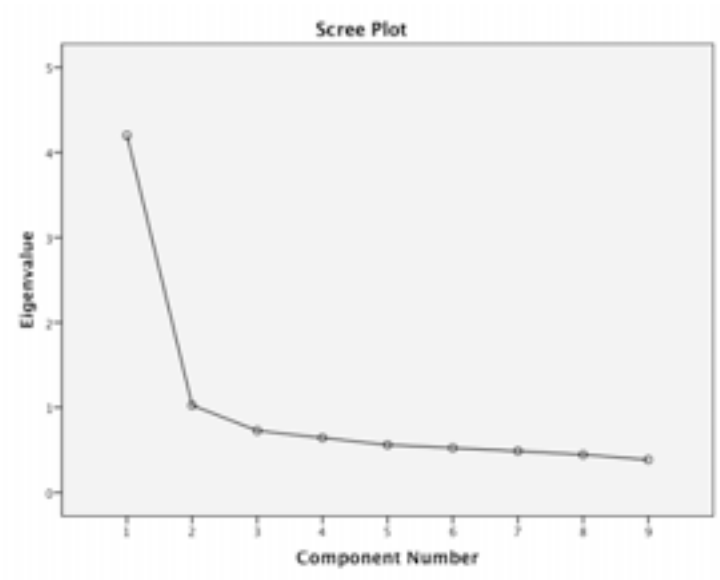


Figure 7.2. Scree plot of factor analysis of nine components. 2 factors were extracted (eigenvalue >1).

7.1.2.3 Correlations of the online PROMS with the musical background variables

Not surprisingly, the lack of professional musicians and the higher proportion of musical sleepers (Figure 7.1) in this study also affected the correlation between the online PROMS with the composite music background ($\alpha = .66$; $\omega = .75$; $r = .34$, $p < .01$), MM-Competence Scale ($\alpha = .89$; $r = .41$, $p < .01$) and self-rated musicianship ($r = .30$, $p < .01$).

One item from the music questionnaire asked listeners at what age they began their musical training (see Appendix 2, item 2). Listeners who had received music training at age 7 or earlier are considered to be individuals who have early music training experience as that age relates to exceptional musical abilities such as perfect pitch ability (Brown, Sachs, Cammuso, & Folstein, 2002; Chin, 2003). This study found most of the PROMS subtests showed positive significant correlation with the *early music training experience* and *music background composite score*, mostly around $r = .25$ to $.34$, $p < .01$; except for the loudness and tempo subtests. Consistent with previous studies in this thesis (i.e., Studies 2, 3 and 4), loudness perception, which is also a generic auditory skill, shows no significant correlations with listeners' musical experience (Zimmerman, 2011; Riley & Mckee, 1963; Williams, Sievers, & Hattwick, 1932). In addition to loudness perception, this study revealed that individual differences in tempo perception, like loudness perception, also showed to be less affected by musical training. This result is consistent with studies that have found infants were able to match the tempo or beat of music pieces (Zentner & Eerola, 2010; Winkler et al., 2009), suggesting that individual differences in tempo perception may be less influenced by musical training. In order to ascertain how much loudness and tempo subtests have influenced the correlation between music background and the composite score, the PROMS score without tempo and loudness subtests was computed. It was found that the correlation of the *composite music background* and the *new PROMS score* without tempo and loudness subtests was higher, $r = .43$, $p < .01$. However the Steiger's test shows that this increase showed no significant difference with the composite scores that included the tempo and loudness subtests, $z = -6.013$, $p > .05$

Another music background item - family influence[34], did not show significant correlation with the composite perceptual score, in line with the findings from previous research (Shuter-Dyson, 1981). However family influence has shown low significant correlation with early music training, $r(428) = .10$ to $.13$, $p < .05$. Similar patterns also emerged in Study 4, $r(74) = .25$ to $.34$, $p < .01$.

In an attempt to examine how musical sleepers perceive the same or higher level of perceptual ability compared to trained musicians, the criteria for a musical sleeper were defined in section 7.1.2.1. The study found that musical sleepers' timbre perception was positively correlated with the amount of music listening (see Appendix 2, item 11), $r(68) = .38$, $p < .01$. This finding however, is preliminary and more work needs to be done in order to verify this result.

This study also found there is no significant correlation between the PROMS score with age and education, $r = .04$, $p > .05$.

7.1.3 General Discussion

Study 5 was conducted to investigate the PROMS with a more diverse population and to examine whether the result found in controlled studies could be replicated with the Internet population, and how an uncontrolled testing environment and equipment may have affected the result.

The data of this study was collected via MTurk as well as via direct invitation to

[34] Family influence means either influence from parents, siblings or extended family.

the test. Overall, the Internet data replicated previous controlled studies with several exceptions. First, the internal consistency reliabilities of several subtests were lower compared to controlled studies. Although this is as expected with a more diverse population who completed the test in various environments and with differing equipment, this test can be improved upon so that it is more practical to use with public population. This will be discussed in more detail in the next chapter.

Secondly, the result of the factor analysis has shown to be similar with the controlled studies, however, the tempo test has shown to load higher in the structural processing factor rather than in the sensory processing factor. This might be due to the order effect where the tempo test was always presented in the temporal test together with other structural processing subtests such as rhythm, rhythm-to-melody and accent subtests. Thus listeners were “encouraged” subliminally to use the same processing strategy to process the tempo subtest as they had used for other structural subtests. Compared to the controlled studies, the tempo subtest was presented *before* the sensory processing subtests, hence it has shown the tempo test uses the same strategy as sensory processing.

Several links between the PROMS score and music background emerged. First, as in previous research (Shuter-Dyson, 1981), family (either parents or extended family) did not seem to have significant influence upon perceptual ability, but contributed to the opportunity for early music training. This study also revealed that tempo perception, in addition to loudness perception, was found to be less affected by music training.

The Internet study also attracted a substantial portion of musical sleepers to take part in the experiment and the preliminary finding shows that timbre perception of musical sleepers was moderately correlated with the amount of music listening.

Overall, the Internet study has been successful, although some of the findings deviated slightly compared to the controlled study. But this was to be expected given the variations of participants, environment and other uncontrolled issues. Using an Internet method nevertheless has allowed investigators to collect data in a shorter period (and/or more cheaply), however, this study is subject to limitations. First, despite participants having the flexibility to pause and resume the test at any time, the total length of time to complete the test was about 1 to 1 ½ hours. This length may have led to a bias in participants, for example participants who have a higher attention span or persistence. Future research to improve this aspect will be discussed in the next chapter.

Second, the online PROMS was constructed in English only, and indeed more than 50% of the samples came from English speaking countries or they were native English speakers. This study might have neglected important findings about tonal speakers and perceptual ability that was found in previous research (Klein, Zatorre, Milner, & Zhao, 2001; Delogu, Lampis, & Olivetti Belardinelli, 2006).

Third, the samples of the Internet study were not generally representative of humankind because participants required a computer or some form of Internet accessible device, as well as requiring a relatively good Internet connection speed in order to load the audio files properly. Most importantly, participants needed to know how to use a web browser or have a basic computer literacy level.

This chapter concludes the empirical studies of this thesis. The next chapter presents the summaries of the thesis, as well as providing an integrative view of findings of all studies in this thesis. The usefulness of the test-battery, limitation and future work are also presented.

CHAPTER 8

Summary and General Discussion

8.1 Summary of Thesis

Musical ability is an abstract concept. Using psychometric approaches, several researchers have developed a series of musical test-batteries to measure ability objectively (e.g., Seashore, 1919a; Wing, 1948; Bentley, 1966; Gordon, 1965; Karma, 1980). Chapters 1, 2 and 3 reviewed these test-batteries and provided the evidence of their limited use in contemporary research.

The thesis reviewed this issue and narrowed down the concept of musical ability to one of its components, music perception ability, which can be seen as the fundamental cognitive skill in music. The importance of perceptual ability in music is emphasised by Sloboda (2000, p.397): "…the musical experience of the listener is at the heart of all musical activity". Indeed, most people's experience with music is through listening rather than playing musical instruments.

There were two main goals for this thesis. The first was to develop a validated music test-battery that could be used to objectively measure music perception ability. This test-battery is called the PROMS, which stands for the Profile of Music Perception Skills. The second goal was to examine substantive issues in music perception with the application of the PROMS. For this reason, the thesis was organised into four main sections: the development of the PROMS (Study 1 and Study 2); the validation of the PROMS (Study 3); the examination of the PROMS's associations with general mental ability, short-term and working memory, and auditory discrimination skill (Study 4); and lastly the examination of the PROMS structure using Internet samples (Study 5). The main results of these studies are summarised in Table 8.1 and Table 8.2.

Table 8.1. Summary of All Empirical Studies in the Present Thesis

Study	Aims	Test-Design	Summary of findings
1	Pilot Study N=24	PROMS Stimuli Presentation: Standard & Comparison-Stimulus 18 stimuli/subtest (9 *Same*, 9 *Different*) 2AFC (Same/Different) 3AFC: Tempo (Slower/Faster/Same) 3AFC: Loudness (Softer/Louder/Same) 3AFC: Pitch (Lower/Higher/Same)	Study 1 was a pilot study to examine the psychometric properties of the novel music test. The internal consistency reliability for the composite score was excellent, however, the ranges of the reliability coefficients across subtests' scores were very wide, from very poor internal consistency reliability to a good reliability coefficient (see Table 8.2). Because the test-battery employed a mixture of 2AFC and 3AFC, the estimates relating to means, standard deviations, and psychometric properties are hard to compare across subtests due to the different designs of answer choices. Confidence rating and guessing problems were also insufficiently controlled in this study.
2	Pilot test *Double Reference Playback with Confidence Rating Scheme* N=39	PROMS Stimuli Presentation: Standard- stimulus plays twice and followed by the Comparison-Stimulus* 18 stimuli/subtest (9 *Same*, 9 *Different*) 5 Rating scale: Definitely Same, Probably Same, I Don't Know, Probably Different, Definitely Different*	Study 2 aimed to improve the test-battery by pilot testing a new test paradigm "double reference playback with confidence rating scheme". The balance of the test-difficulty was also revised. The analysis of the study was also improved by using *d'*, which is a better scoring method that controlled for response bias. In addition, McDonald' ω was computed along with Cronbach's α to get a better estimation of the internal consistency of the test. A test-retest session was also conducted to examine the variation of listeners' response over a short period (1 week). Overall, the internal consistencies and the test-retest reliabilities of the composite

Study	Aims	Test-Design	Summary of findings
			and subtest scores were moderately encouraging, all the while pointing to the need for improvements (see Table 8.2)
3	Validation Study N=56	The PROMS' stimuli presentation was the same as Study 2 Music Background Questionnaires* MM-Competence Scale* External Music Tests*: AMMA, MAP, MET, Timbre (Mono)	Study 3 replicated the test-design in Study 2 (with minor revisions of the music stimuli) and the purpose of this study was to validate the music test-battery. Three existing music batteries were selected for convergent validity; and an additional timbre test was developed in monophonic (single instrument) format with double the numbers of stimuli to examine the content validity of the PROMS. Additionally Study 3 was also expanded by the addition of musical experience questionnaires, which include self-reported measurement such as music background questionnaires, self-evaluation measurement such as musicianship rating and the MM-Competence scale to examine the criterion validity of the PROMS scores. Overall, the PROMS showed moderate to moderate-strong correlation with the existing music tests and all the self-reported music background questionnaires, supporting the convergent, content and criterion validities of the PROMS.
4	External Variables Study N=76	The PROMS' stimuli presentation was the same as Study 2 External variables*: Forward Digit Recall, Backward Digit Recall, Raven's Matrices Test, Gap	Study 4 aimed to investigate whether the PROMS correlated with general mental ability, short-term memory, working memory and auditory discrimination skill. Study 4 also served to examine the discriminant validity of the PROMS. Overall, the finding supported the discriminant validity of the PROMS, the details of the relationship between PROMS

Study	Aims	Test-Design	Summary of findings
		Detection Task	and general mental ability, short-term memory, working memory and auditory discrimination skill are discussed in later sections.
5	Internet Study N= 430	The PROMS' stimuli presentation was the same as Study 2, however the subtests were divided into two sections (Tonal and Temporal). Score and text feedback was also given after each section.	Study 5 aimed to examine whether the result found in controlled studies could be replicated with a more diverse population, and how an uncontrolled testing environment and equipment may have affected the result. Study 5 also examined the preliminary norm of the PROMS. Overall, the Internet data replicated previous controlled studies with several exceptions. First, the internal consistency reliabilities of several subtests were lower compared to controlled studies, however this was to be expected with a more diverse population who completed the test in various environments and with differing equipment. Secondly, the tempo subtest of the Internet study had higher loading in the structural processing factor rather than in the sensory processing factor. Overall, the Internet study was successful but also pointed to the need for improvements. This is discussed further in later sections in this chapter.

Note: * (Test-Design and Stimuli) refers to the main change or addition from the preceding study.

Table 8.2. Summary of Cronbach's Alpha, McDonald's Omega and Test-Retest Reliabilities for All Studies

	Study 1	Study 2			Study 3			Study 4		Study 5	
	(N=24)	(N=39)	(N=39)	(N=39)	(N=56)	(N=56)	(N=20)	(N=76)	(N=76)	(N=430)	(N=430)
Test	α	α	ω	Test-Retest	α	ω	Test-Retest	α	ω	α	ω
Tempo	.89	.78	.84	.66	.65	.72	.81	.60	.57	.63	.68
Loudness	.81	.63	.83	.64	.72	.80	.83	.71	.69	.75	.72
Pitch	.72	.59	.73	.63	.73	.79	.77	.66	.86	.58	.61
Timbre	.70	.74	.86	.81	.77	.84	.68	.67	.75	.48	.67
Rhythm	.60	.50	.73	.67	.61	.75	.62	.66	.74	.59	.60
Tuning (Chord)	.56	.49	.67	.62	.81	.87	.68	.70	.72	.70	.61
Melody	.44	.48	.68	.56	.56	.73	.77	.57	.63	.52	.71
Tuning (Orchestra)	.33	-	-	-	-	-	-	-	-	-	-
Rhythm-in-Melody	.11	-	-	-	-	-	-	-	-	-	-
Rhythm-to-Melody	-	.67	.70	.79	.78	.83	.82	.70	.71	.75	.79
Accent	-	.62	.71	.68	.55	.70	.71	.58	.68	.57	.51
COMPOSITE	.92	.86	.93	.83	.94	.95	.90	.88	.90	.91	.89

8.2 Understanding Music Perception Ability via the PROMS

The implementation of the PROMS not only allows us to measure music perception ability objectively, it also helps us to examine more closely the interrelationships of various music perception domains. The following section describes the findings that emerged from the application of the PROMS, in particular what PROMS tells us about music perception ability. These interpretations, however, are only based on the preliminary findings of the work conducted in the current thesis therefore more work is required to corroborate these findings. The potential usefulness, limitations and future direction of the PROMS will also be addressed.

8.2.1 Musical Training Effects on the PROMS Scores

The origins of musical ability have been widely debated and even today it remains an open question (Wallin, Merker, & Brown, 1999; Peretz & Zatorre, 2003; Bigand & Poulin-Charronnat, 2006). It has been proposed that one way to examine the nature of musical ability more accurately is through implicit music tasks such as asking listeners to discriminate whether two sounds are the "Same or Different". Explicit music tasks, such as asking listeners to judge whether a pitch is lower or higher, have limitations because they give an advantage to trained listeners in recognising such a change, and this would inevitably distract the researchers from understanding the nature of musical ability (Bigand & Poulin-Charronnat, 2006; Underwood, 1996). For this reason, this thesis employed an implicit music task design ("Same or Different" with a confidence rating) and examined several issues of musical ability, in particular evaluating the influence of intensive musical training on music perception abilities.

The PROMS is a music-test battery that was developed based on musical rules. Unsurprisingly the PROMS composite score was positively correlated with the self-reported music background questionnaires. Music training often requires a substantial number of hours of practice time over a very long period (around 10 to 15 years; Bigand et al., 2006). During this period, music apprentices learn specific perceptual skills in ear training, music analysis consisting of theoretical and practical knowledge of understanding the construction of musical structures, and motor skill development for playing musical instruments. All of these skills are likely to have deeply influenced the way that music apprentices process musical stimuli. For instance, musical stimuli processing during music listening can be influenced by involuntary motor activities such as moving fingers along with music (Haueisen & Knösche, 2001; Janata & Graftton, 2003).

Across the nine PROMS subtests, all studies in this thesis have shown that loudness perception skill did not significantly correlate with music background, suggesting that individual difference in loudness perception appears to be comparatively less affected by musical training. In addition, despite tempo perception showing influence from musical training in the controlled studies, Study 5 (the Internet study with the largest sample) revealed that tempo perception showed no significant correlation with music background. This suggests that individual differences in tempo perception may be quite similar to loudness perception which is less affected by musical training. However, more research is needed to corroborate this finding.

Nevertheless, previous research has shown both loudness and tempo perception abilities to be skills found in infant or children without the aid of explicit musical training, and that these skills are part of the initial perception development prior to other perception abilities such as melody or rhythm perceptions (Geringer, 1995; Riley & Mckee, 1963; Williams, Sievers & Hattwick, 1932; Winkler et al., 2009; Zimmerman, 2011; Zentner & Eurola, 2010). This finding is perhaps not surprising; the ability to perceive intensity change and speed change is often used in daily life even in a non-musical context such as speech or during physical movement such as walking. Whilst other perception abilities such as melody or rhythm perceptions also occur in daily life, these abilities improve with increasing attention span and the improvement of the memory function that is associated with musical training (Zimmerman, 2011; Brandler & Rammsayer, 2003; Chan et al., 1998, Kilgour et al., 2000; Ho et al., 2003; Wallentin et al., 2010a; Jakobson et al., 2008). Therefore, the enhancement of melody and rhythm perception skills from musical training shows a stronger correlation with musical background compared to loudness perception. This implies that human perception of music follows a developmental sequence: from daily life experience to improvement through musical training. This pattern was evident in children where their loudness discrimination ability developed first, with pitch and rhythm discrimination developing only somewhat concurrently (Zimmerman, 2011).

While music training has shown a positive relationship with most of the PROMS subtests, listeners with stronger musical influence from their family (either genetically from their parents or musical exposure from other musician family members), did not seem to perform statistically better in their PROMS score than those listeners who did not have such family influence.

However, family influence correlated significantly with “early musical training (age 7 or/and before)” which has shown a positive contribution to the PROMS scores. This suggests that being born in or surrounded by a musical family does not necessary give genetic advantage in musical perceptual skill, but it does provide an environmental opportunity for early music training which contributes substantially on the later growth of musical perceptual ability. This result is in line with previous genetic studies with twins where the researchers commented, “While genes are not unimportant, they often play a role secondary to family environment… the differences between pairs who did and pairs who did not take private lessons suggest that parental influence might be one critical factor in training” (Coon & Carey, 1989, p.190 & p.191). However, it is possible that the listeners used in this thesis had constrained musical abilities ranging only from relatively mild to high only [based on Gagné (2003)’s metric model of levels of giftedness: mildly (1 in 10), moderately (1 in 100), highly (1 in 1,000), exceptionally (1, in 10,000) and extremely gifted (1 in 100,000)], which has attenuated the correlation strength between family influence and the PROMS scores. As Vandenberg suggests, "it may be that only the exceptional talent of great composers and musicians has an hereditary factor (1962, p.233).

Bigand and Poulin-Charronnat (2006) also reported a series of other musical characteristics that do not depend on formal music training: perceiving musical tensions and relaxations in both melodies and harmonic sequences; anticipating musical events on the basis of subtle syntactic-like features of the prime sequence; integrating local structures in large scale structures without an explicit context, learning new compositional systems and finally performance in cognitive and emotional tasks.

In the attempt to examine how musical sleepers perceive the same or higher level of perceptual ability compared to trained musicians, the criteria for ‘musical’ were defined. Listeners whose music background is in the lowest quintile of music background composite score *and* who scored above average in the PROMS are categorised as musical sleepers in this study. Study 5 revealed that musical sleepers’ timbre perception was positively correlated with their amount of music listening. This finding however is only preliminary, and more research needs to be done to corroborate it.

8.2.2 Music Perception Structures in the PROMS

One of the aims of this thesis was to investigate the intercorrelations within the PROMS' subtests, for example, whether perception A is highly correlated with perception B but not others, or whether they point to a general musical ability factor akin to Spearman's *g*. In general, the intercorrelations between most subtests were equally substantial, generally around .30 to .40. There was no significant difference even when memory (STM and WM) and general mental ability (Raven's Test) was controlled using partial correlation analysis.

The perception dimensions that tap similar abilities have shown higher correlations with each other (≥.50), demonstrating the common variance between them, for example rhythm and rhythm-to-melody subtests, r = .60; pitch and tuning subtests, r = .50; pitch and loudness; r = .50; rhythm and accent, r = .50. This suggests that individual perception abilities contribute to each other to a substantial level, pointing to the presence of a generic "musical perception ability" factor. In addition, studies 3, 4, and 5 showed that the PROMS subtests can be extracted into two perceptual factors: "structural processing factor" and "sensory processing factor". The structural processing factor consists of melody, rhythm, rhythm-to-melody and accent subtests and the sensory processing factor consists of pitch, tuning, timbre and loudness subtests.

A mini meta-analysis of studies 3, 4, and 5 was conducted with reliability controlled to examine the effect size of the two factor structures with music background, and further analysis with Steiger's Test (Steiger, 1980) to investigate whether there was a significant difference between the two factors with musical training experience. The result showed that there was no significant difference between the two factors, z = -4.93, p >. 05 (2 tailed).

Overall, the data suggest that that the structural processing factor is explained by a strategy of using attention and memory to store and recall auditory elements as shown by the significant correlation between short-term and working memory with structural processing factor in Study 4. In contrast, the sensory processing factor applies less analytical strategy and represents the ability to extract auditory information during the initial seconds of audio listening. Individual differences in tempo perception seem to vary depending on the strategy adopted in processing the preceding musical stimuli.

For instance, studies 3 and 4 have shown that tempo perception loaded higher on the sensory processing factor together with pitch, tuning, timbre and loudness; however, the internet study (Study 5) has shown that tempo perception loaded higher alongside the structural processing factor that correlates with memory capacity. The reason for this may be due to the order effect of the stimuli presentation. The tempo subtest was preceded by other structural processing subtests during the Internet study, as opposed to the controlled studies where the tempo subtest was preceded by other sensory processing subtests. This might suggest the change of tempo perception, and the way tempo is perceived, can be influenced by the preceding use of another perception strategy.

8.2.3 The PROMS and the Non-Musical Cognitive Abilities

Chapter 6 presented the evidence from previous research of investigating the relationship between music training and non-musical abilities such as short-term and working memory, auditory discrimination ability and general mental ability. But few studies have examined the relationship between the *specific* music perception ability and the non-musical abilities. This issue was examined in this thesis.

It was found that listeners who have more extensive musical experience also showed to have better memory capacity, and this musical experience mediated the positive relationship between the PROMS and the memory capacity. On the other hand, Raven's Test, a reasoning test that is commonly used to measure general mental ability (e.g., Judge, Hurst, & Simon, 2009), as well as spatial ability (e.g., Schweizer et at al., 2007), did not show significant correlation with the listeners' musical experience or the PROMS composite score. However, two of the PROMS subtests, rhythm and rhythm-to-melody subtests significantly correlated with the Raven's Test. Explaining from the spatial ability point of view, this finding is in line with previous study which suggests that rhythm and spatial ability are processed in the same brain area (i.e., cerebellum) and thus, are related (Shaw, 2000; Parson & Fox, 1995). The Gap Detection Task was employed to measure auditory discrimination skill, and although it correlated significantly with the listeners' musical background, it did not show significant correlation with the PROMS composite and subtests.

In summary, the PROMS composite score did not show significant correlation with the general mental ability and the auditory discrimination skill, and that the relationship between memory capacity and the PROMS composite was mediated by the listeners' musical experience. These findings supported the discriminant validity of the PROMS, as well as further strengthening the notion of the close relationship between auditory memory capacity and music training (Jakobson, Cuddy, & Kilgour, 2003; Brandler & Rammsayer, 2003; Tierney and colleagues, 2008).

8.2.4 PROMS Scores According to Gender, Ethnicity, Nationality, Age, Education, and Handedness

Studies 3, 4 and 5 showed that there was no significant correlation between the PROMS scores with gender, handedness, age, education, ethnicity and nationality. This suggests that the PROMS is consistent with the rationale behind the test construction, which was to create a test that should be equally applicable across people differing in gender, ethnicity and education status. Also note that despite handedness in this thesis not being measured using a validated handedness questionnaire such as the Edinburgh Handedness Inventory (Oldfield, 1971), this result is in line with previous studies that have employed validated handedness measurement to examine its relationship with musical ability (Byrne, 1974; Aggleton, Kentridge, & Good, 1994).

8.3 The Usefulness of the PROMS

This thesis has demonstrated the usefulness of the PROMS in several ways. In particular it has revealed the inter-relationship within the nine dimensions of music perception skills, and how they are influenced by musical training and non-musical cognitive functions. The following section proposes additional uses of the PROMS that were not examined in this thesis.

8.3.1 The PROMS, Musical Aptitude and Musical Achievement

To be able to objectively measure innate musical talent has always been one of the core motives of previous music education researchers (e.g., Seashore, 1919; Wing. 1948; Bentley, 1960; Gordon, 1965). The PROMS is likely to attract this same question: does it function as an innate music talent measurement tool?

The quick answer to this would be “no”; the PROMS does not claim to measure innate talent, rather it aims to measure individual differences in music perception abilities. Although previous music test authors expressed the view that the measurement of innate musical ability can be undertaken successfully with children before they are exposed to formal music training or music-related environmental factor (Seashore, 1919; Wing, 1948; Bentley, 1960; Gordon, 1965), more recent research suggests that prenatal influence may play an even more important role in the development of musical ability (Arabin, 2002; Whitwell, 1999).

Despite the PROMS having limited usage as a measure of innate ability, Study 5 has shown that it has detected musical sleepers who possess latent musical perceptual ability despite their lack of musical training. It is worth noting that this superior perceptual ability that lacks musical training is quite distinct from innate musical talent. This ability only represents an exceptional perceptual skill that untrained listeners share with musically-trained listeners; how it is developed and whether there are other factors that might have contributed was not intensively examined in this thesis.

For the same reason, the test score of the present battery is not used to categorise listeners into a “good” or “bad” group, rather it examines listeners’ perception behaviour. The positive correlation between music background and the perceptual test-score, for example, suggests that musically-trained listeners may have searched for music cues in the music subtests that required training to succeed, such as the interval or mode of melody or the number of rhythmic elements to determine the differences in total sequence duration.

Unlike musically-trained listeners, non-trained listeners chose to apply implicit judgement rather than explicit music judgement (Bigand & Poulin-Charronnat, 2006; Thaut, 2005; Zendel & Alain, 2008). Therefore musically-untrained listeners may have chosen a less analytical strategy and, therefore, appeared to judge the stimuli by using a more simplistic and holistic gestalt strategy. For example, when a listener chose a “same” answer when listening to a “different-pair”, particularly for a difficult trial that consisted of only subtle changes between the pairs, it implied that the listener applied an holistic judgement. In fact, non-musicians posses other areas of music proficiency for example perceiving music emotion (an holistic judgement) to an as complex and highly developed degree as trained musicians (Bigand & Poulin-Charronnat, 2006).

Likewise, when an individual obtains a high perceptual score in this test, this may imply that the individual possesses a perceptual ability or potential that would promote musical learning or music-related activities. But it does not affirm or predict that the individual will be successful in musical achievement (Révész, 1953). Take the analogy of a computer that has been programmed to 'understand' pitch. The computer will always score highly in the pitch test but this does not mean that the computer will be successful in musical achievement.

Being successful in music is not merely due to superior perceptual ability; it is often a combination of perceiving, memorising, reproducing, motivation, determination, deliberate practise, opportunities, parental encouragement, attitude, and other environmental factors (Shuter-Dyson, 1969; Meinz & Hambrick, 2010; Gordon, 1995; Ericsson, Nandagopal, & Roring, 2005). Just as perfect pitch ability is an exceptional perceptual ability that gives an advantage in musical activity such as transcription, it is not an essential skill for individual success in music (Parncutt & Levitin, 1999; Slominsky, 1930). Bigand and Poulin-Charronnat (2006) also noted that although trained musicians (who may have higher achievement in music qualifications) seem to have the advantage of being able to interpret music on a higher order such as being able to accurately transform a musical score into sound, this does not imply that trained musicians have more creative musical skills than untrained individuals. For example, many musicians who have received Western music training are often unable to improvise or compose music, whereas many famous self-taught musicians such as John Coltrane and Django Reinhart have remarkable improvisational abilities.

Nevertheless, the PROMS has presented initial evidence of measuring musical potential, in particular to discover listeners who have not had intensive music training and yet possess superior music perception skill (known as 'musical sleepers' in this thesis). To corroborate this, more research needs to be conducted and this is discussed in the "future direction", Section 8.4.3.

8.3.2 Added Values of the PROMS Relative to Self-Reported Musical Background

The PROMS helps to attenuate errors of categorisation that are the result of reliance on the self-reported extent of musical training alone. For instance, a considerable number of listeners in the present thesis scored far better (or worse) on the test than was to be expected based on their extent of their musical training. Specifically, among the listeners who lacked music training (lowest quintile), PROMS test scores ranged from a low of 63 (d = -.65) up to an impressive 128 (d = 1.69). In turn, the scores of the most highly trained participants (highest quintile) ranged from a high of 143 (d = 2.46) down to a below-average 103 (d = .70). This finding adds substance to the notion of *musical sleepers* and *sleeping musicians,* that is, musically untrained but capable individuals, and, vice versa, highly trained individuals of limited musical ability. It is easy to see how routinely allocating musically skilled and unskilled individuals to the labels “musicians” or “non-musicians” based only on the extent of musical training can lead to distorted estimates and interpretations of the effects of musical ability on any outcome, be this language processing, autism spectrum disorder, or brain anatomy.

In fact, several studies have shown the problems of using just the length of musical training experience to determine music ability, finding that so called “musicians” are not as good as “non-musicians” in music perception skills (Eagleson & Eagleson, 1947; Butler & Brown, 1984; Panion, 1989; Krumhansl, 1996). By improving the sensitivity in the assessments of musical ability, the PROMS should be helpful in attenuating such biases, especially when used in combination with musical training indicators. Thus, when a high PROMS score coincides with advanced musical qualifications, depending on the context of the research, one might infer musical proficiency with maximum confidence. Such confidence is lessened when the same qualifications are paired with a modest PROMS performance. Musically untrained individuals who score highly on the PROMS, in turn, represent a special group of musically gifted individuals, who may exhibit a very different response pattern in outcome measures compared to untrained individuals with low PROMS scores.

8.3.3 Examining the Meaning of Music Perception Ability and Its Usefulness with Other Non-Musical Abilities

The PROMS may be useful for addressing substantive questions about the nature of music perception itself. For example, an important but unresolved question relates to the interrelationships between various musical skills. Custom has it that music is composed of distinct elements such as rhythm, meter, tempo, melody, harmony, timbre, and so on. However, as pointed out by Serafine (1988), these elements are often the by-product of musical writing and analysis, and it is not clear whether these distinctions are sensible from a perceptual point of view. The current work suggests that performance on the nine subtests may be subtended by two higher order perceptual abilities; a structural-processing and a sensory-processing music perception ability.

Furthermore, categorisation based on musicianship usually only allows linking an outcome to musical ability or expertise generically, but not to any specific musical skill. Yet, this is what researchers are often interested in. For example, the transfer effects from music experience to language processing are well established in behavioral and neurophysiological studies (Bidelman, Gandour, & Krishnan, 2011). However, there are limited studies to examine whether or not linguistic expertise, for example speaking a tonal language, enhances music-related processing and its perception, such as tonally related perception properties (e.g., pitch, timbre).

In addition, the PROMS may provide a pathway towards potential remedies for language-music related disorders. For instance, what seems to be driving the link between musical ability and dyslexia are timing skills, rather than skills in pitch or timbre discrimination (Overy, Nicolson, Fawcett, & Clarke, 2003; Overy, 2000). This is not only important for understanding the disorder, but it could also have a role in devising treatment plans for a specific disorder using music materials (Thompson & Goswami, 2008).

Finally, the PROMS may have some limited uses in special populations. For example, hearing aids or cochlear implants enhance speech perception, but do little to improve the quality of music perception. Research about the underlying problems has only just begun, so exactly why the corrective devices do so little to restore music perception is far from clear (e.g., Won, Drennan, Kang, & Rubinstein, 2010). Comparing population norms of normal hearing adults to the performance of populations with hearing impairments on standardised batteries such as the PROMS may help to diagnose the type and extent of their musical perceptual deficits.

8.4 Limitation and Future Direction

Although the PROMS has comparatively more strengths than previous musical ability batteries, it is by no means a perfect or exhaustive test of musical perceptual ability. A salient point is raised by Bentley (1966),

> No test, no examination, no measurement of human abilities, is perfect. The most we can hope to achieve

> are the best tests that human fallibility, both on the part of the person making the tests and of those who are tested, will allow. We must accept this limitation. (p.79)

The following section describes the limitation of the PROMS, as well as discussing how these limitations can be overcome in the future research.

8.4.1 The Limitation of the PROMS and the Necessity of Modification for Special Populations, the Elderly and Children

Although this thesis included a comparatively large number of music components in the test, it may not have accounted for all the perceptual competences that have a role in musical ability. For example, individual differences in the perception of expressive perceptual musical qualities such as phrasing, balance and musical expression (Gordon, 1965); whether the skill was inherited or acquired (Gagné, 1999; Sloboda & Howe, 1991); motivation and determination of the listeners to learn music (McAuley, Henry, & Tuft, 2011) or environmental factors (Ericsson et al., 2005); are all excluded from measurement in the present battery.

Likewise, the relationship between music perception, musical performance, music creation and improvisation, and music analysis, cannot be taken for granted (Boyle, 1992). Creating suitable stimuli for measuring such advanced skills is arduous because, as soon as one moves away from basic units or patterns of music, for example by using musical excerpts instead, one is bound to measure familiarity with culturally evolved musical systems as well as, or even instead of, any kind of basic musical ability.

Nevertheless, the PROMS is a potential tool for examining basic music perception skill if the researcher is that way inclined. The thesis has only examined the PROMS with an adult normal population. Modification of the PROMS would be necessary if it were to be used amongst special populations, such as the elderly or children. For example, a child-centric PROMS should be more interactive, fun and attractive, and the general difficulty of the test should be easier than the current PROMS. The overall length of the PROMS for children should also be shorter as children's concentration span is shorter than adults'.

8.4.2 The Practical Length of the PROMS

Despite best efforts at keeping the test short, the full battery takes about an hour to complete. This practical disadvantage is attenuated if the researcher is only interested in using a selection of individual subtests. Also, if the user is merely interested in an overall assessment of the level of perceptual musical skill, three to four subtests will provide an acceptable approximation to the score that would be obtained from all nine subtests.

For example, a short-comprehensive test could be created by selectively choosing the four best items from each subtest according to their item-total-correlations from studies 3, 4 and 5 reducing 162 items to 36 items. The second possible solution is to select two sensory-processing subtests and two structural-processing subtests with high loadings on their respective factors: i.e., Rhythm-to-Melody, Accent, Tuning, and Loudness, reducing 162 items to 72 items. An added benefit of this second selection is that pitch-, timing- and dynamic-related skills remain. A preliminary reliability analysis was conducted based on the second solution using studies 3, 4, and 5. It has been satisfactory with respect to the most important quality criteria: it correlated highly with the full PROMS, $r(536) = .90$, $p < .001$, exhibited good internal consistency ($\alpha = .86$), and posted a correlation with external indicators of musical proficiency comparable to that of the full PROMS, $r(536) = .37$, $p <.001$. Furthermore, all of the subtests of the brief PROMS exhibited satisfactory test-retest reliabilities (Study 3). Taking about 20 to 30 minutes to complete, the brief version commends itself as an acceptable, time-efficient alternative to the full PROMS. Nevertheless, the reliability and validity of the shorter version would need to be re-examined.

Another alternative way of making the full test shorter is by devising an adaptive version, i.e., one where parts are skipped by automatically adapting to the participant's ability level. However, this procedure is not optimal for structural-processing subtests as adaptive procedures involve a reversal, i.e., where participants are returned to the trials of the previous level if an incorrect answer is given. As stimuli from structural processing factor are "memory based", participants' responses to the reversal stimuli will be based on memory training rather than a true response. Therefore it is recommended that adaptive is the optimum technique to be used with trials that that are stationary with time (Levitt, 1971; Leek, 2001), unless a different stimuli of equal difficulty level are prepared during each reversal stage.

8.4.3 Examining The PROMS as a Musical Potential Measurement Tool in Longitudinal Study

The PROMS has provided initial evidence of measuring musical potential, however it is not clear to what extent this musical potential facilitates music learning and achievement. It is proposed that a longitudinal study be conducted to examine this issue. In particular, a future study could compare whether musical sleepers learn faster than non-musicians.

This study could be conducted in two sessions. A group of non-musicians who have less than one year of musical training will be recruited. The first session would be a screening session where listeners are asked to complete the PROMS as well as a music background questionnaire to confirm their musical experiences. Listeners who score above average in the PROMS would be categorised in a “musical sleeper” group; listeners scoring below average would be categorised in a “control” group. Both groups are provided with identical musical instrument lessons that would prepare them for a basic music examination after a period of 6 months or 1 year. Different music teachers (who were not involved with the music lessons) would be recruited to assess participants’ performance and the scores of the two groups would be compared. If the music sleeper group performed significantly better than the non-musician group, it would further corroborate the notion of using the PROMS as a ‘musical potential’ measurement tool.

8.4.4 PROMS Examination in a Normal Population

Despite Study 5 having examined the PROMS with a more diverse population (N = 430), the collected data cannot be viewed as representative of the general population. For example, the sample in this thesis consists of a majority of participants from English speaking countries, such as the United States and the United Kingdom, in contrast with world population statistics which state that the majority of the world’s population are in Asian countries such as China and India (UN, 2011). The current samples might have led to overestimates of average performance scores on the one hand, and attenuated the size of correlations as a result of restriction of range on the other. The collection of data from larger and more diverse samples is an important step toward understanding the distribution of musical skills in the general population, for example, whether they conform to a normal or positively skewed distribution and whether distributions vary according to parameters such as age, gender, and socioeconomic status, or to the presence of strong musical institutions.

8.4.5 Understanding Musical Ability and Conclusions

Finally, the question of “what is musical ability?” should continue to be explored in the future. Everyday life experience has intensively trained humans to understand the basic perception of sound and music. Adding supplementary music training enables the acquisition of specific skills (perceptual, cognitive and motor that are required when learning musical symbols and playing musical instruments) , which are indispensable in becoming a professional musician. However, this thesis has shown that musical training and everyday life experience are not the only factors that determine the musical ability of human beings. Therefore the questions surrounding “what is musical ability?” should continue to be investigated.

In summary, after an absence of suitable music perception ability test batteries stretching over thirty years, it is hoped that the current test-battery can reignite interest in the scientific study of musical ability, its measurement, and its relationship with other human abilities.

APPENDIX 1

Matlab Code for LTware ™

Study 1 was developed using Matlab program. A sample of the code is provided here.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% This is a Listening Test Software (LTware)implemented in GUI %
%            by Lily Law, University of York, May 2010         %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


function varargout = rhythm(varargin)
% RHYTHM M-file for rhythm.fig
%      RHYTHM, by itself, creates a new RHYTHM or raises the existing
%      singleton*.
%
%      H = RHYTHM returns the handle to a new RHYTHM or the handle to
%      the existing singleton*.
%
%      RHYTHM('CALLBACK',hObject,eventData,handles,...) calls the local
%      function named CALLBACK in RHYTHM.M with the given input arguments.
%
%      RHYTHM('Property','Value',...) creates a new RHYTHM or raises the
%      existing singleton*.  Starting from the left, property value pairs are
%      applied to the GUI before listening1_OpeningFunction gets called.  An
%      unrecognized property name or invalid value makes property application
%      stop.  All inputs are passed to rhythm_OpeningFcn via varargin.
%
%      *See GUI Options on GUIDE's Tools menu.  Choose "GUI allows only one
%      instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help rhythm

% Last Modified by GUIDE v2.5 21-Jul-2009 10:20:32

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                   'gui_Singleton',  gui_Singleton, ...
                   'gui_OpeningFcn', @rhythm_OpeningFcn, ...
                   'gui_OutputFcn',  @rhythm_OutputFcn, ...
                   'gui_LayoutFcn',  [] , ...
                   'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if nargout
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%IMPLEMENTATION STARTS HERE
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%This is the function where all the variable and default values are defined
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


% --- Executes just before rhythm is made visible.
function rhythm_OpeningFcn(hObject, eventdata, handles, varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
% varargin   command line arguments to rhythm (see VARARGIN)

% Choose default command line output for rhythm
handles.output = hObject; %handle to figure
handles.count= 0; % counter for the Play Order
handles.scount= 1; % counter for the array
handles.f= zeros(100,5);% default array value and size

% Update handles structure
guidata(hObject, handles); %update handles structure


%end here
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% UIWAIT makes rhythm wait for user response (see UIRESUME)
% uiwait(handles.figure1);

% --- Outputs from this function are returned to the command line.
function varargout = rhythm_OutputFcn(hObject, eventdata, handles)
% varargout  cell array for returning output args (see VARARGOUT);
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Get default command line output from handles structure
varargout{1} = handles.output;



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%This is the function where the audio are loaded and played.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% --- Executes on button press in select_audio.
function select_audio_Callback(hObject, eventdata, handles)
% hObject    handle to select_audio (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

tic;

handles = guidata(gcbo); % returns the handle of the graphics object whose callback is
executing.

if(handles.count<=0) % set number of audio playing 18 times
handles.count= handles.count+1;

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R1.wav');
samples{200} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R2.wav');
samples{201} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R3.wav');
samples{202} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R4.wav');
samples{203} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R5.wav');
samples{204} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R6.wav');
samples{205} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R7.wav');
samples{206} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R8.wav');
samples{207} = {audio, fs};
```

```
[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R9.wav');
samples{208} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R10.wav');
samples{209} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R11.wav');
samples{210} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R12.wav');
samples{211} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R13.wav');
samples{212} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R14.wav');
samples{213} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R15.wav');
samples{214} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R16.wav);
samples{215} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R17.wav);
samples{216} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R18.wav);
samples{217} = {audio, fs};




%**********************************************************************

% RANDOMISE and PLAY AUDIO

handles.randValue=199+handles.count;

PLAY= handles.randValue;
PLAY

sound(samples{handles.randValue}{:}) ; % then play the value from the cell



else
set(hObject, 'enable', 'off') % switch off button when reach the limit
end

guidata(gcbo, handles);%update

%**********************************************************************
% Load AUDIO


guidata(hObject, handles); %update


%end here

%**********************************************************************
%% ANSWER FUNCTIONS START HERE
%**********************************************************************

%%%%%%%%%%%%%%%%%%%%%
% "SECOND" answer   %
%%%%%%%%%%%%%%%%%%%%%

% --- Executes on button press in answer_Different.
function answer_Different_Callback(hObject, eventdata, handles)
% hObject    handle to answer_Different (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

%------------------------------------------------------------

tic;
```

```
handles = guidata(gcbo); % returns the handle of the graphics object whose callback is
executing.

if(handles.count<=18) % set number of audio playing 18 times
handles.count= handles.count+1;

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R1.wav');
samples{200} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R2.wav');
samples{201} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R3.wav');
samples{202} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R4.wav');
samples{203} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R5.wav');
samples{204} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R6.wav');
samples{205} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R7.wav');
samples{206} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R8.wav');
samples{207} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R9.wav');
samples{208} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R10.wav');
samples{209} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R11.wav');
samples{210} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R12.wav');
samples{211} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R13.wav');
samples{212} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R14.wav');
samples{213} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R15.wav');
samples{214} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R16.wav);
samples{215} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R17.wav);
samples{216} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R18.wav);
samples{217} = {audio, fs};


%**********************************************************************

% RANDOMISE and PLAY AUDIO

handles.randValue=199+handles.count;

PLAY= handles.randValue;
PLAY
handles.ans= (handles.randValue-1);
handles.no= (handles.count-1);

sound(samples{handles.randValue}{:}) ; % then play the value from the cell


else

    set(hObject, 'enable', 'off') % switch off button when reach the limit
    msgbox ('Press the "Submit" button now', 'Next Step')

end
```

```
guidata(gcbo, handles);%update

%**********************************************************************
% Load AUDIO


%------------------------------------------------------------

toc;
t= toc;
handles.fc{handles.scount}= [handles.no, 99, (handles.ans), 1, t];

% create matrix with counter
handles.f(handles.scount,:)= handles.fc{handles.scount};

% counter increase
handles.scount= handles.scount+1;


guidata(hObject, handles);%update


%**********************************************************************

%%%%%%%%%%%%%%%%%%%%%
% "FIRST" answer  %
%%%%%%%%%%%%%%%%%%%%%

% --- Executes on button press in answer_Same.
function answer_Same_Callback(hObject, eventdata, handles)
% hObject    handle to answer_Same (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)


%-------------------------------------------------------------------

tic;

handles = guidata(gcbo); % returns the handle of the graphics object whose callback is
executing.

if(handles.count<=18) % set number of audio playing 18 times
handles.count= handles.count+1;

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R1.wav');
samples{200} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R2.wav');
samples{201} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R3.wav');
samples{202} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R4.wav');
samples{203} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R5.wav');
samples{204} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R6.wav');
samples{205} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R7.wav');
samples{206} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R8.wav');
samples{207} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R9.wav');
samples{208} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R10.wav');
samples{209} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R11.wav');
samples{210} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R12.wav');
samples{211} = {audio, fs};
```

```
[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R13.wav');
samples{212} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R14.wav');
samples{213} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R15.wav');
samples{214} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R16.wav);
samples{215} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R17.wav);
samples{216} = {audio, fs};

[audio, fs] = wavread('/Volumes/STORAGE/Research/MQ RESEARCH/Matlab_MAC/R18.wav);
samples{217} = {audio, fs};



%**********************************************************************

% RANDOMISE and PLAY AUDIO

handles.randValue=199+handles.count;

PLAY= handles.randValue;
PLAY
handles.ans= (handles.randValue-1);
handles.no= (handles.count-1);

sound(samples{handles.randValue}{:}) ; % then play the value from the cell


else

    set(hObject, 'enable', 'off') % switch off button when reach the limit
    msgbox ('Press the "Submit" button now', 'Next Step')

end

guidata(gcbo, handles);%update

%*********************************************************************

%------------------------------------------------------------

toc;
t= toc;
handles.fc{handles.scount}= [handles.no, 99, (handles.ans), 0, t];

% create matrix with counter
handles.f(handles.scount,:)= handles.fc{handles.scount};

% counter increase
handles.scount= handles.scount+1;


guidata(hObject, handles);%update


%-------------------------------------------------------------------


%**********************************************************************

%**********************************************************************
%**********************************************************************

%%%%%%%%%%%%%%%%%%%%%%%%%%%
% WRITE THE DATA TO FILE  %
%%%%%%%%%%%%%%%%%%%%%%%%%%%


% --- Executes on button press in Submit.
function Submit_Callback(hObject, eventdata, handles)
% hObject    handle to Submit (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
```

```
%c= xlswrite(handles.name, [handles.f]);
c= xlswrite('rhythm.xls', [handles.f]);
msgbox('Press "Next" to proceed to the next section', 'Next Step')
%trial;
%close(listening1);

guidata(hObject, handles);


%**********************************************************************
%**********************************************************************

% GET THE ID FROM THE EDIT BOX




% --- Executes on button press in NEXT.
function NEXT_Callback(hObject, eventdata, handles)
% hObject    handle to NEXT (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
rhythmmelody; % open this test
close (rhythm); % close this test
```

APPENDIX 2

Music Background Questionnaire

1. Do you play musical instrument(s) or sing?

   Yes [Go to Question 2]

   No [Go to Question 8]

2. At what age did you start learning a musical instrument/sing*?

   [Age 8 and above] – 1 point

   [Age 7] – 2 points

   [Age 6] – 3 points

   [Age 5] – 4 points

   * This question was used in studies 4 and 5 only.

3. Which instrument(s) do you play? Please state how many years* you have been playing the instrument(s). [text]

   [Instrument, 1-5 years] - 1 point

   [Instrument, 6-9 years] - 2 points

   [Instrument, 10 years and above] – 3 points

   *If the listener played more than one musical instrument, the longest experience was used as the criteria

4. Do you still practise/play instrument(s)?

   Yes [Go to Question 5]

   No [Go to Question 6]

5. How often do you practise weekly?

   [1-5 hours/week] - 1 point

   [6-10 hours/week] - 2 points

[11-15 hours/week] - 3 points

[16 hours/week and above] - 4 points

6. Do you have any formal music qualification(s) or music award(s)?

Yes [Go to Question 7]

No [Go to Question 8]

7. Which music qualification(s)* or award(s)? Please state the Grade and other additional comments

[Qualification, Grade 1-5] - 1 point

[Qualification, Grade 6-8] - 2 points

[Qualification, Bachelor] - 3 points

[Qualification, Masters] - 4 points

[Qualification, PhD] – 5 points

*Practical qualification was used as the main criteria

8. Do you have perfect/absolute pitch? (an ability of a person to identify or recreate a musical note without the benefit of an external reference)

Yes [Go to Question 9]

No [Go to Question 10]

9. What note is this? [Music note playing]

[Correct] – 1 point

[Incorrect] – 0 point

10. Can you read western musical notation?

[Yes] – 1 point

[No] – 0 point

11. How often do you listen to music*?

[Never] - 0 point

[Occasionally] - 1 point

[1-2 days/week] - 2 points

[3-4 days/week] - 3 points

[5-6 days/week] - 4 points

[Everyday] - 5 points

*This question was used in study 5 only.

12. Are you involved actively in professional listening activities? (e.g. Conducting, Sound Engineering, Piano Tuning, Performing, DJ-ing, Music Perception Research and others)

    Yes [Go to Question 13] - 1 point

    No [Go to Question 14] - 0 point

13. What are the activities?

    [Text]

14. Do people who know you consider you to have outstanding skills NOT RELATED TO MUSIC? (e.g. High Academic Achiever, Sports, Programming, Medical, Maths and others).

    [Yes, “comment”] – 1 point

    [No] – 0 point

15. Would you consider yourself as

    [Non-musician] – 1 point

    [Music Loving Non-musician] – 2 point

    [Amateur Musician] – 3 points

    [Semi-Professional Musician] – 4 points

    [Professional Musician] – 5 points

16. Are any members of your family musicians?

    [Yes, “comment”] – 1 point

    [No] – 0 point

APPENDIX 3

MusicMindedness Scale (MM-Scale, Zentner, in progress)

Musical competence section consists 20 questions of self-rated musicality in perception of pitch and rhythm, composing and music reproducing skills, music emotion engagement, music commitment and engagement, musical influence of daily life [Rating Scale: 1 = Very Untrue; 2 = Somewhat Untrue; 3 = Neither True Nor Untrue; 4 = Somewhat True; 5 = Very True].

1. Even when a musical instrument is just slightly out of tune, I notice it instantly.*
2. When listening to music, I experience bodily sensations (temperature, heart beat, shivers).
3. Overall, I consider myself unmusical.*
4. I am ready to travel more than 100 miles to hear one of my favourite artists.
5. Musical experiences belong to the most precious experiences in my life
6. My sense of rhythm is not the best.*
7. Without music, my life would be meaningless.
8. The emotions I feel when listening to my favorite music can be as intense as those I experience when in love.
9. Music is a recurrent topic in my daily conversations.
10. Playing a musical instrument or singing is very important to me.*
11. I have a constant yearning for music.
12. I can easily reproduce (e.g., sing, play) a song that I have only heard once.*
13. I find making and/or listening to music more important than other leisure activities (e.g.,TV, books, sports).
14. I have composed music.*
15. I could easily do without music for a month.

16. After a moving concert, I often find myself in a state of trance.
17. When I listen to music, I experience chills down my spine.
18. I am able to look at musical notes for a moment, and then reproduce them from memory (e.g., sing, play).*
19. Music touches me unlike anything else.
20. Among art forms, I tend to prefer literature and/or the visual arts to music.

* MM-Competence Scale - Only these items were used in the analysis in the thesis as they are related to musical skill rather than emotional.

APPENDIX 4

Demographic Questionnaire

1. Gender

   [ ] Male

   [ ] Female

2. Age

   [Comment Box]

3. Are you Right- or Left- Handed?

   [ ] Left

   [ ] Right

   [ ] Both

4. Education *

   [ ] Did not finish high school – 1 point

   [ ] Completed high school – 2 points

   [ ] Attended some university/college, but not graduated– 3 points

   [ ] Professional School Diploma/Certificate/Similar – 4 points

   [ ] College/University Degree (e.g. BA, MA) – 4 points

   [ ] Doctoral Degree (PhD) – 5 points

   * this question was only used in Study 5 only (Internet Study)

5. Department *

   [Comment Box]

   * this question was used in studies 3 and 4 only

6. Nationality*

[Drop down of country list]

* this question was only used in Study 5 only (Internet Study)

7. Ethnicity background

[ ] Asian

[ ] White

[ ] Black

[ ] Hispanic

[ ] Others

8. Are you multilingual (speak more than one language fluently)?

[ ] Yes

[ ] No

9. I am a native English speaker

[ ] Yes

[ ] No

10. Did you use headphones or speakers for the listening test? Please provide the headphones or speakers model if known *

[ ] Headphones

[ ] Speakers

[Comment Box]

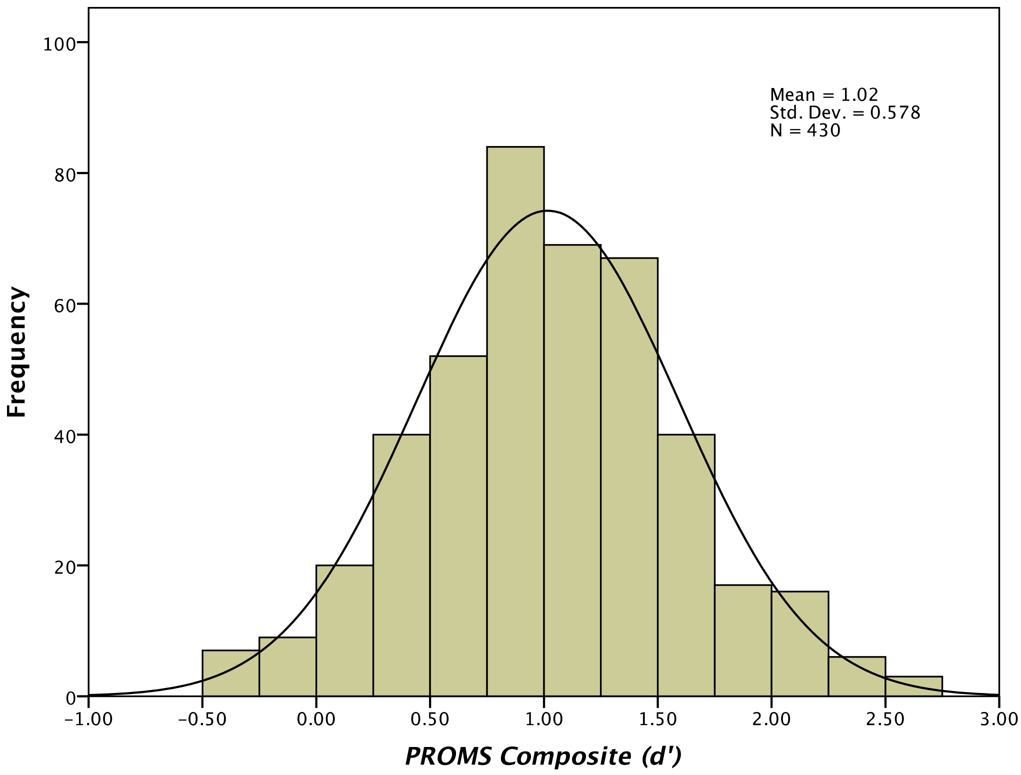
* this question was only used in Study 5 only (Internet Study)

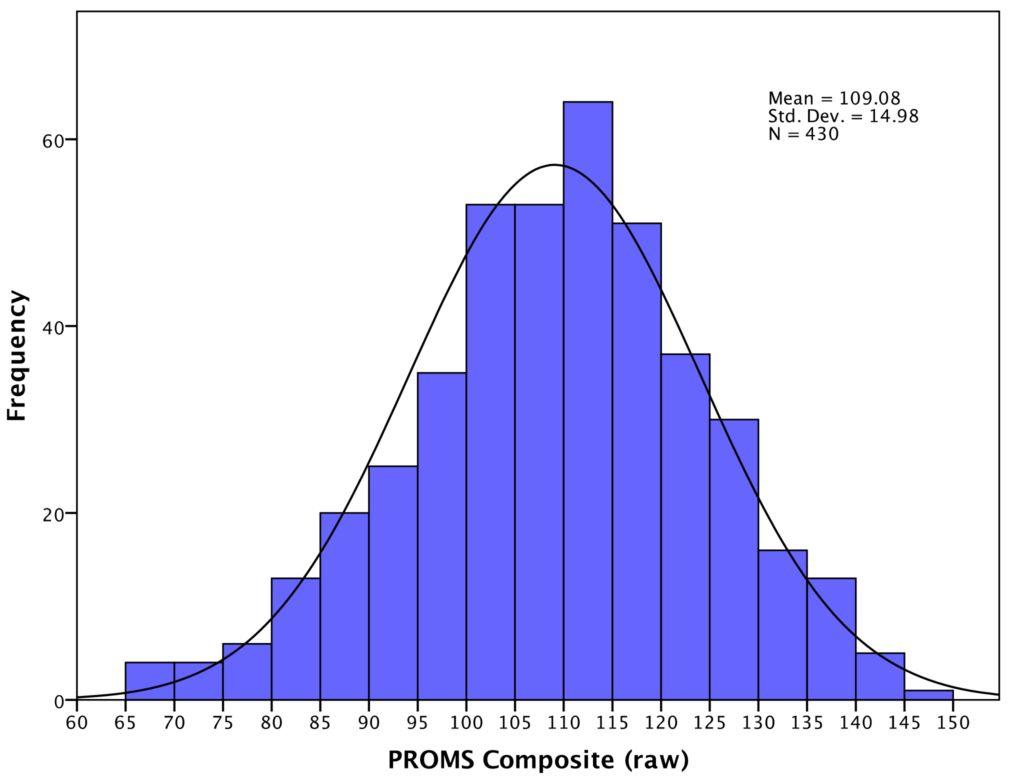
APPENDIX 5

Histogram (Norm Distribution) of the Internet Study

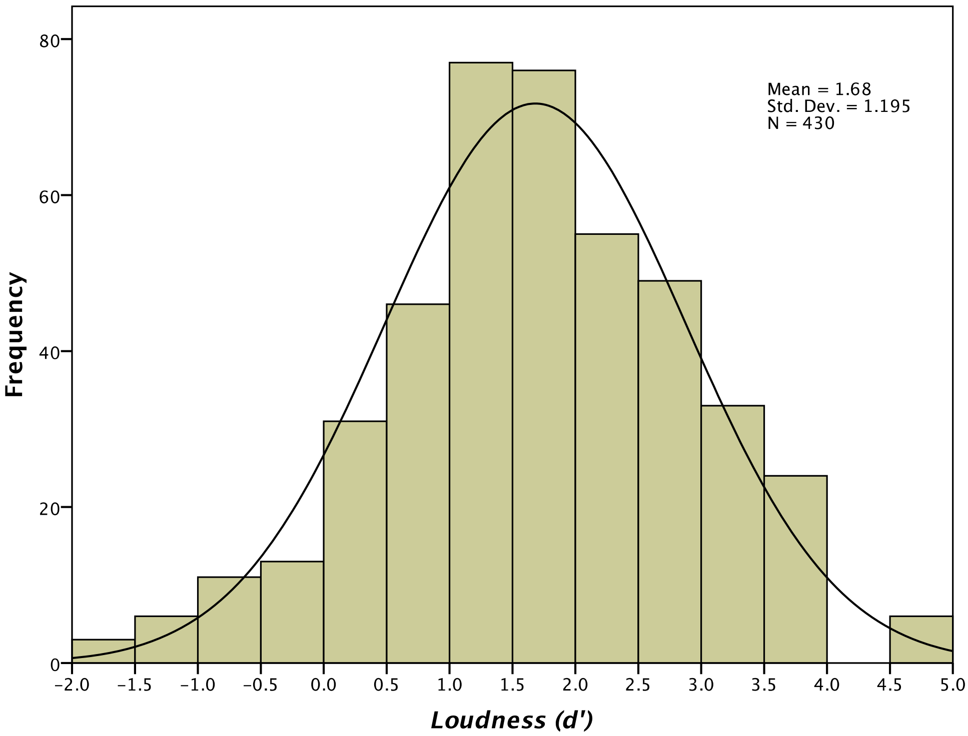
Preliminary normality of the PROMS was examined in Study 5 (Internet Study). Due to the large samples of the study (N=430), Kolmogorov-Smirnov tests (K-S) and Shapiro-Wilk tests are likely to show significant difference even when the scores are only slightly different from a normal distribution, therefore the data of the PROMS was also checked against with the visual inspection of Q-Q plots and histograms, as well as the values of skew and kurtosis analysis (z scores < 3.29; Field, 2009). The histograms with raw score were also computed in addition to the *d'* score (Macmillan & Creelman, 2005) to provide a comparison with the *d'* distribution. However, please note that the *d'* score was the main analysis used in this thesis. Overall, the normality distribution of the PROMS composite score (*d'*) showed to be normally distributed; several subtests were slightly skewed but were still within the acceptable range of normal distribution, except the skewness (3.86) of the Rhythm subtest (*d'*) that was slightly above the recommended threshold (z scores < 3.29; Field, 2009). The raw score of the subtests on the other hand, showed small to moderate skew. Norm distribution histograms are provided in this appendix and arranged according to the level of means from high to low.

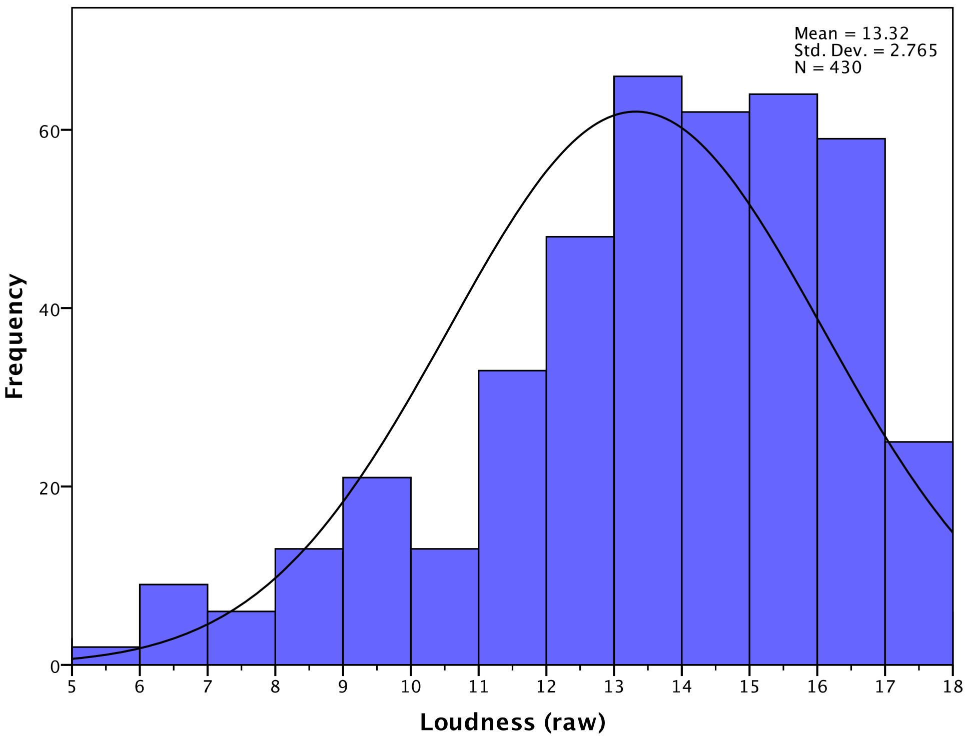
PROMS Composite



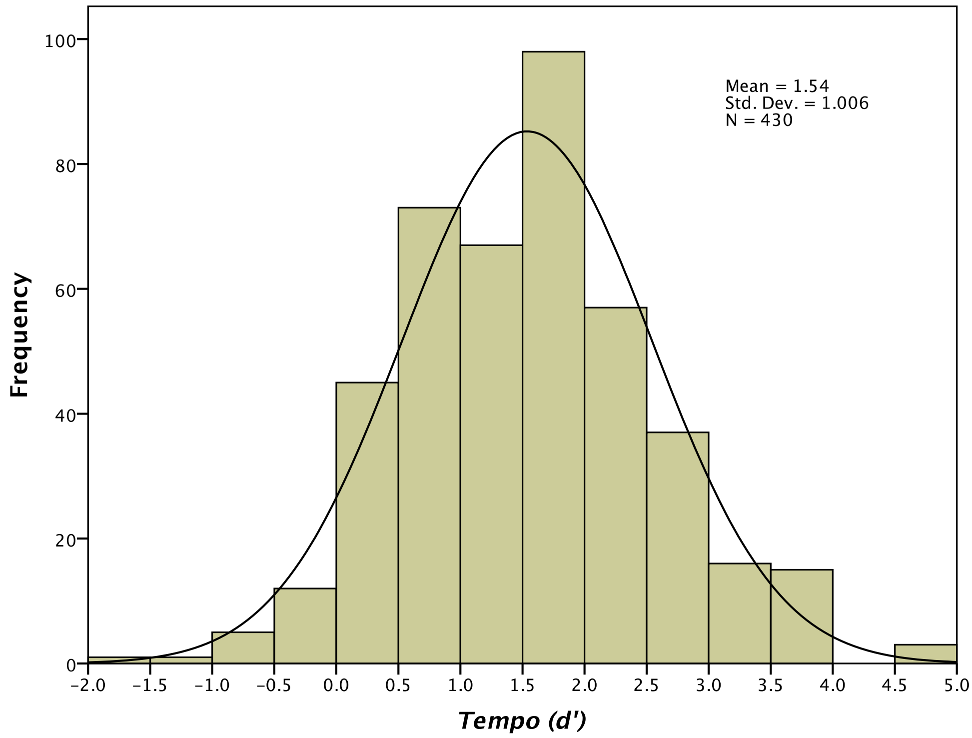


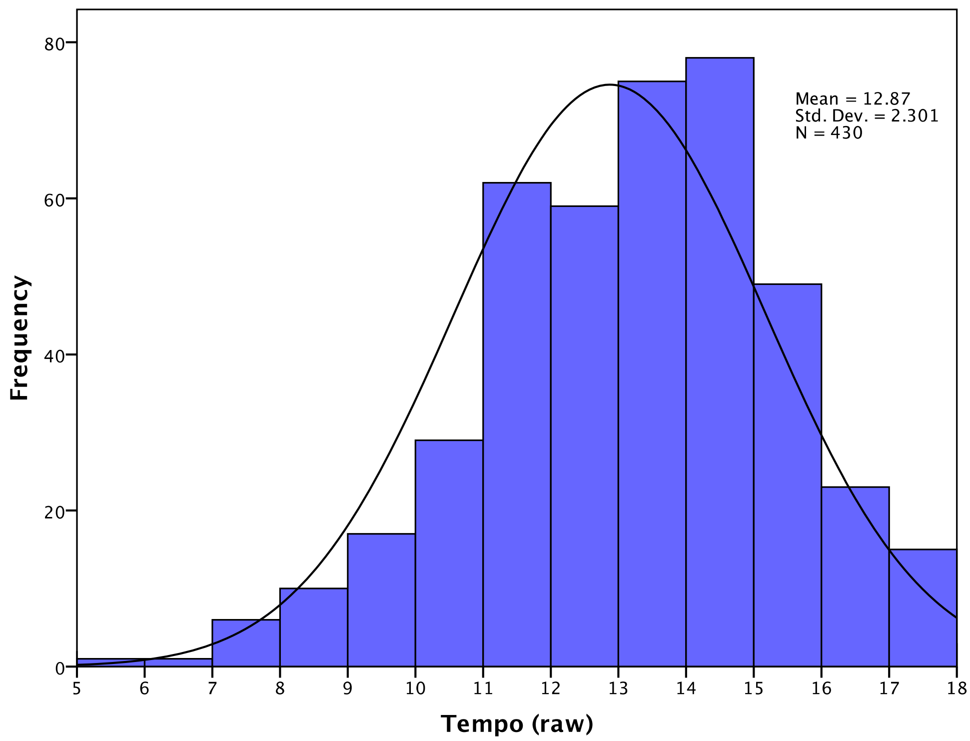
Loudness


Mean = 1.68
Std. Dev. = 1.195
N = 430
Frequency
0
20
40
60
80
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Loudness (d')

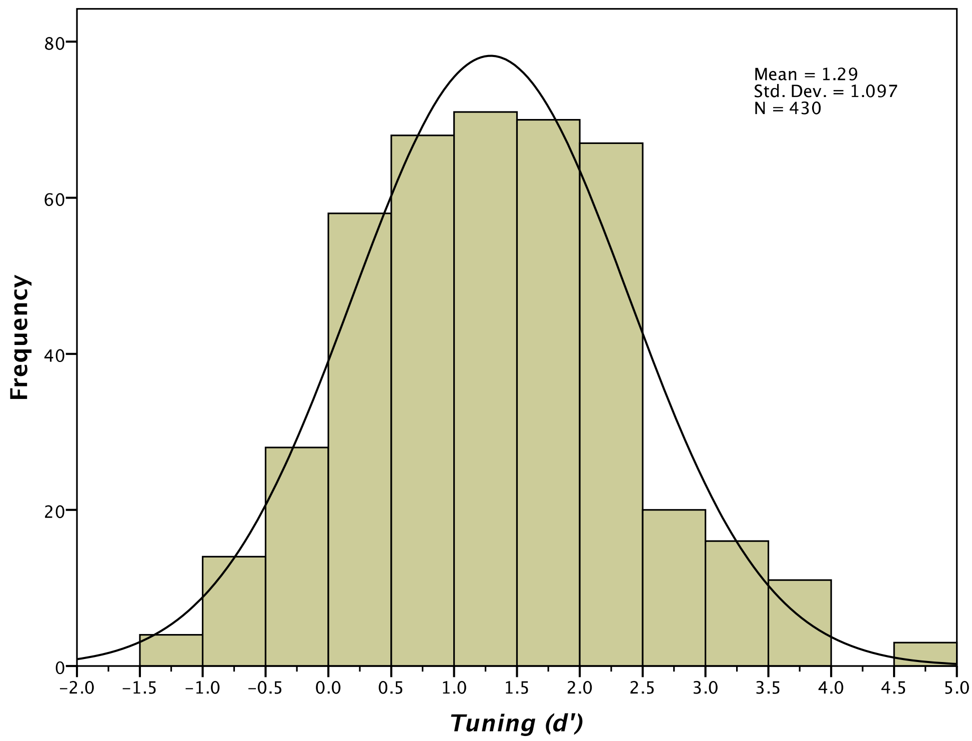

Mean = 13.32
Std. Dev. = 2.765
N = 430
Frequency
0
20
40
60
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Loudness (raw)

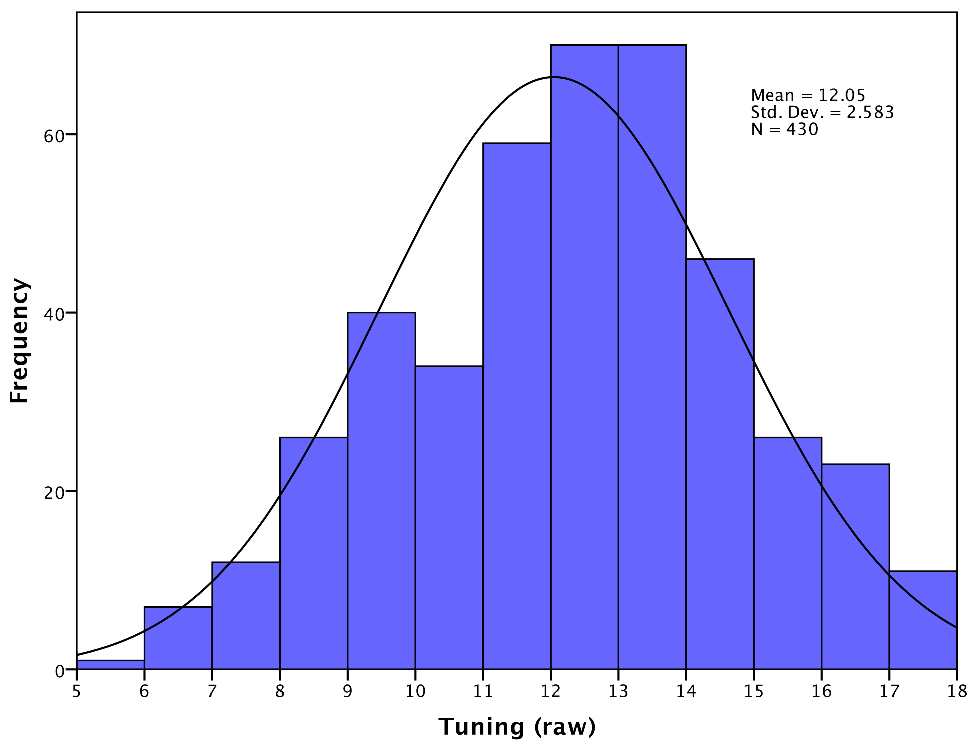
Tempo


Mean = 1.54
Std. Dev. = 1.006
N = 430
Frequency
100
80
60
40
20
0
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Tempo (d')

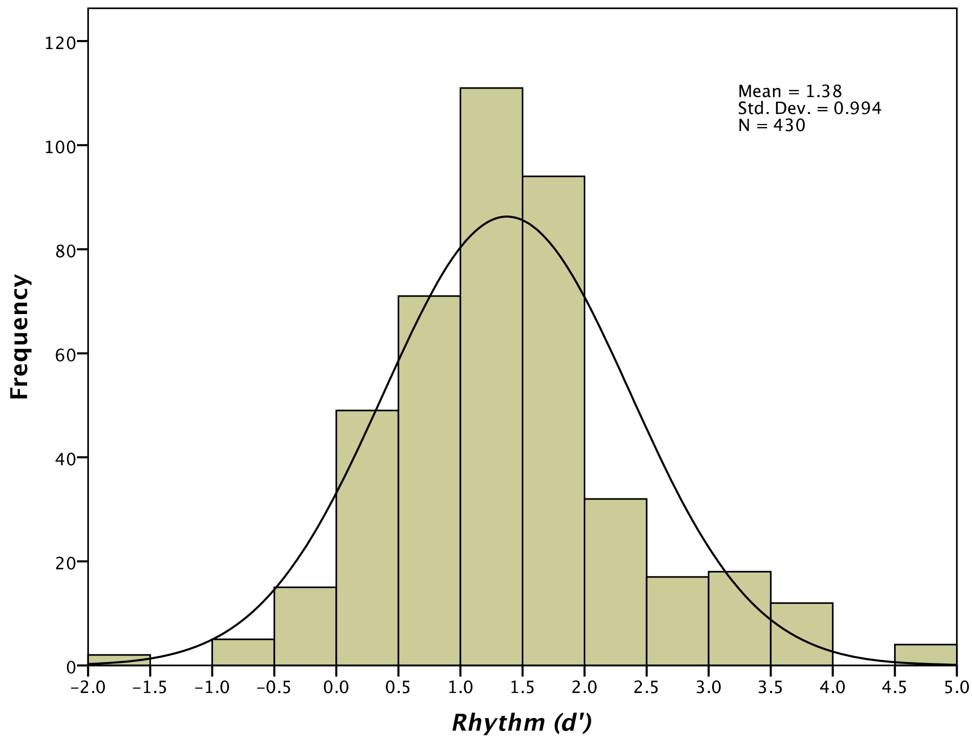

Mean = 12.87
Std. Dev. = 2.301
N = 430
Frequency
80
60
40
20
0
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Tempo (raw)

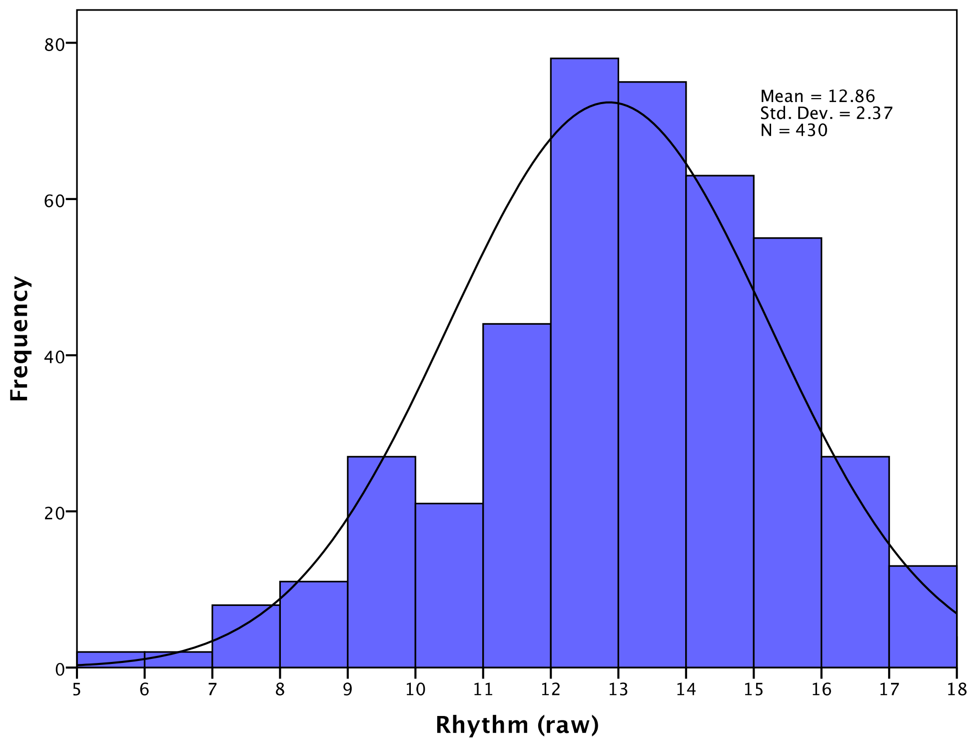
Tuning


Mean = 1.29
Std. Dev. = 1.097
N = 430
Frequency
80
60
40
20
0
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Tuning (d')

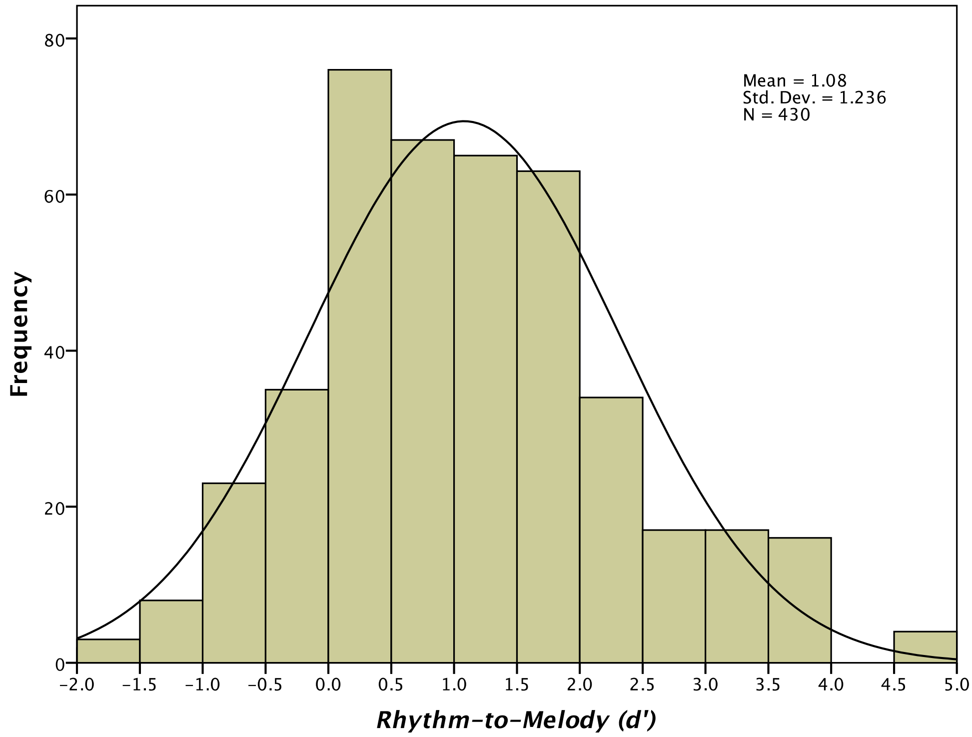

Mean = 12.05
Std. Dev. = 2.583
N = 430
Frequency
60
40
20
0
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Tuning (raw)

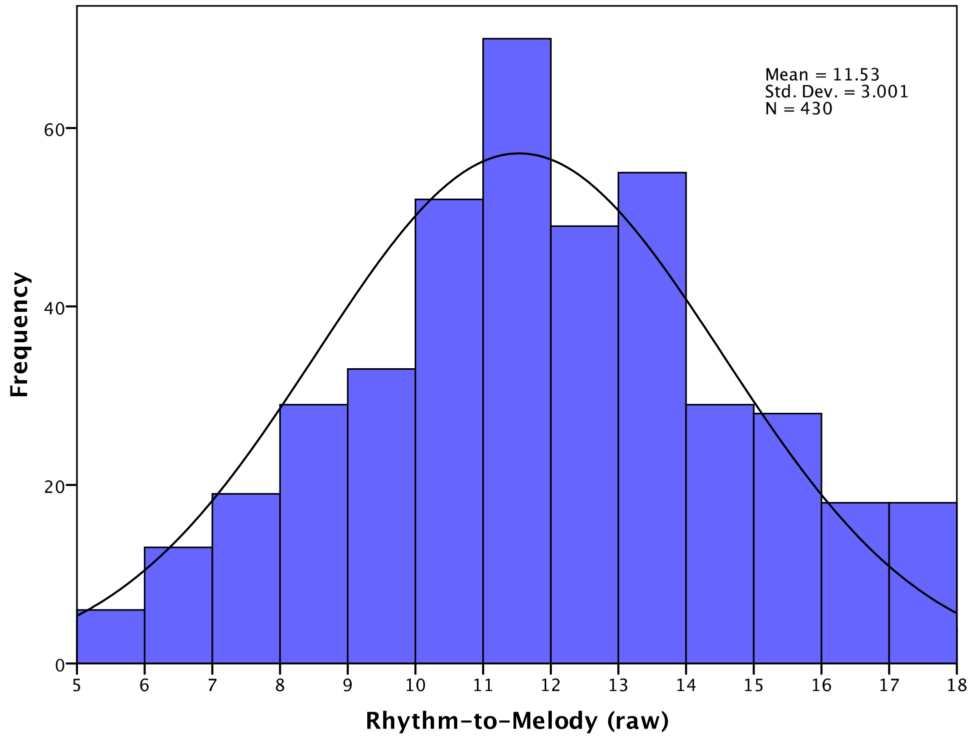
Rhythm


Mean = 1.38
Std. Dev. = 0.994
N = 430
Frequency
120
100
80
60
40
20
0
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Rhythm (d')

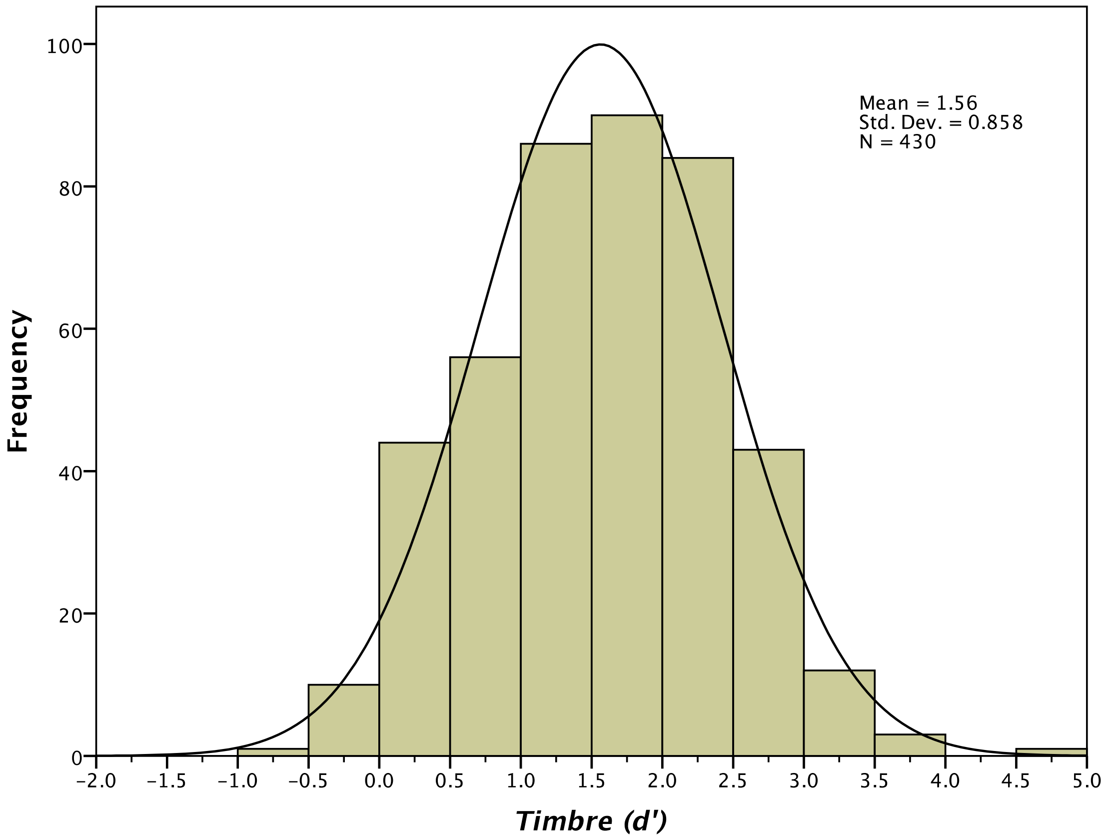

Mean = 12.86
Std. Dev. = 2.37
N = 430
Frequency
80
60
40
20
0
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Rhythm (raw)

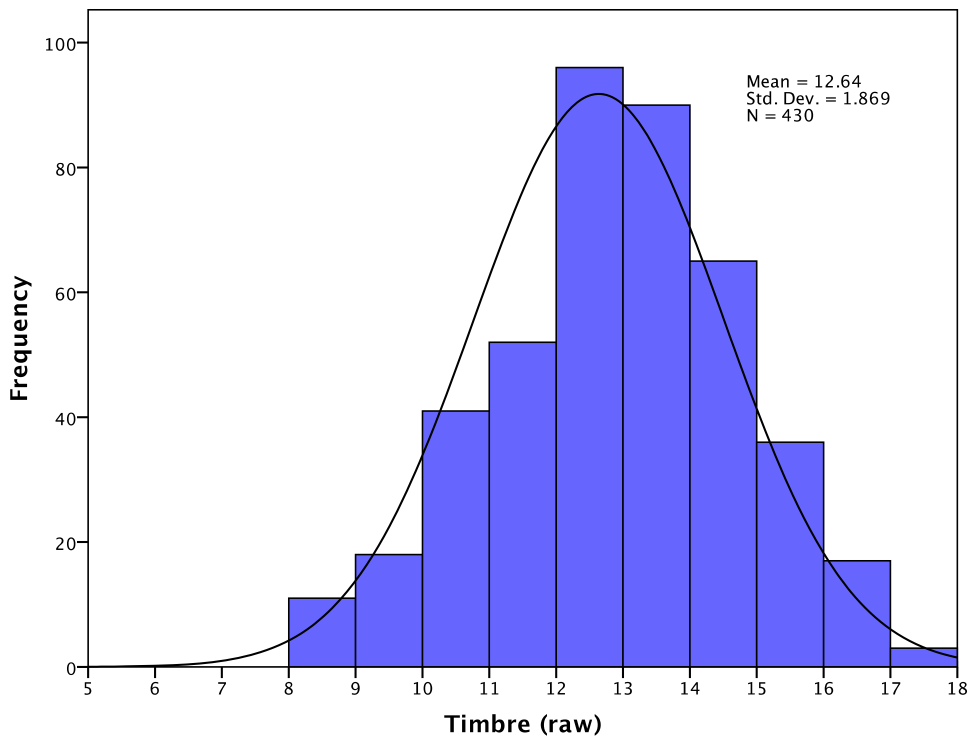
Rhythm-to-Melody



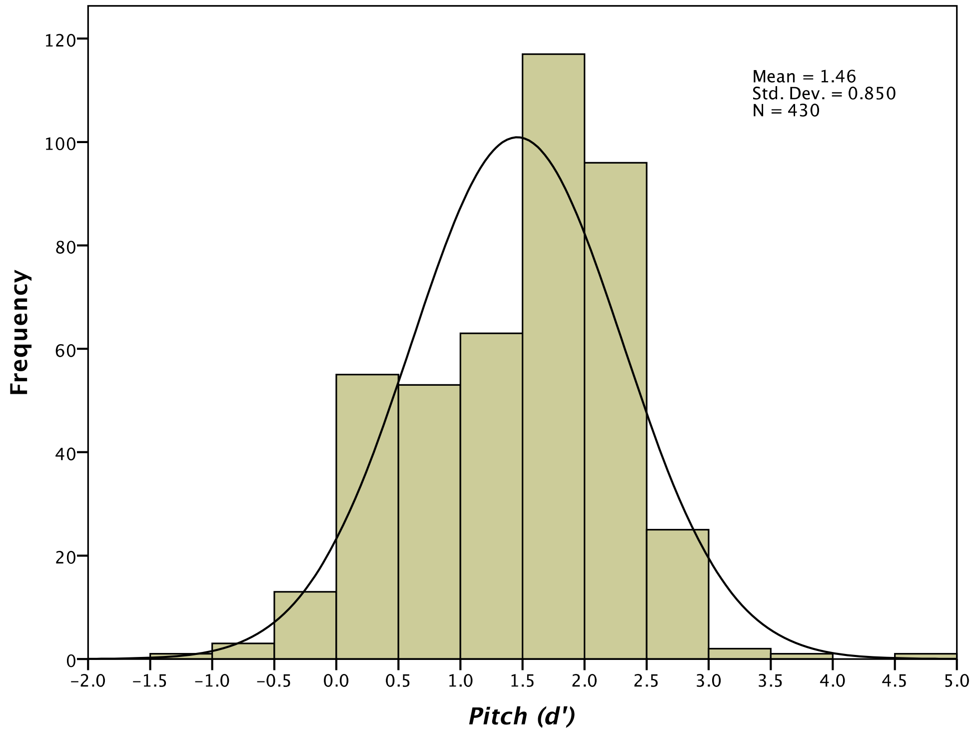


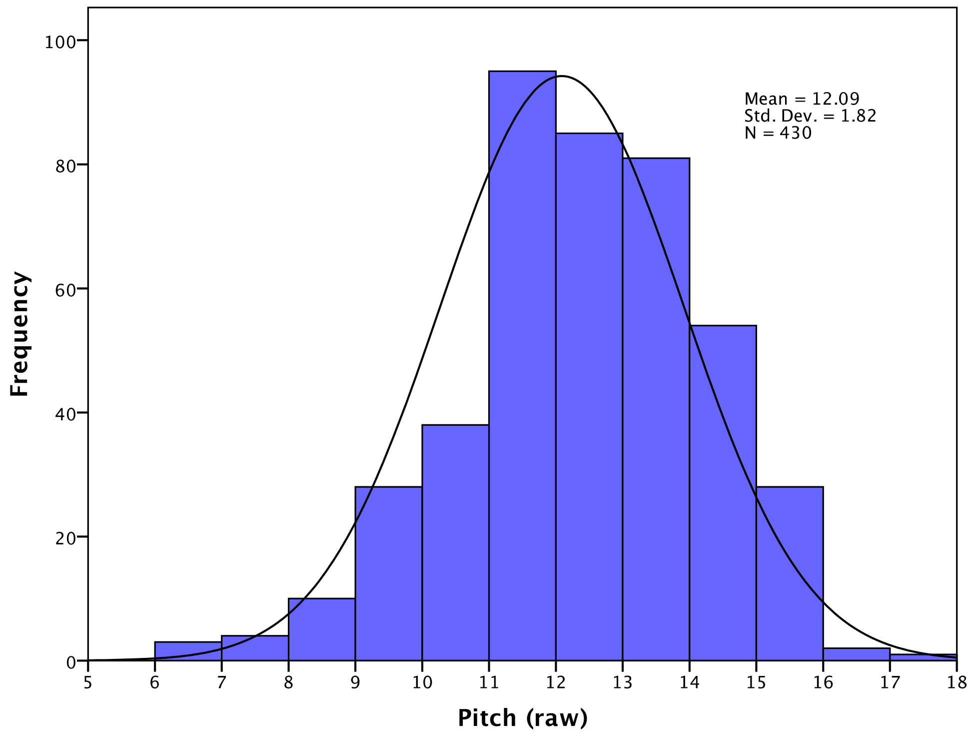
Timbre


Mean = 1.56
Std. Dev. = 0.858
N = 430
Frequency
Timbre (d')
100
80
60
40
20
0
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

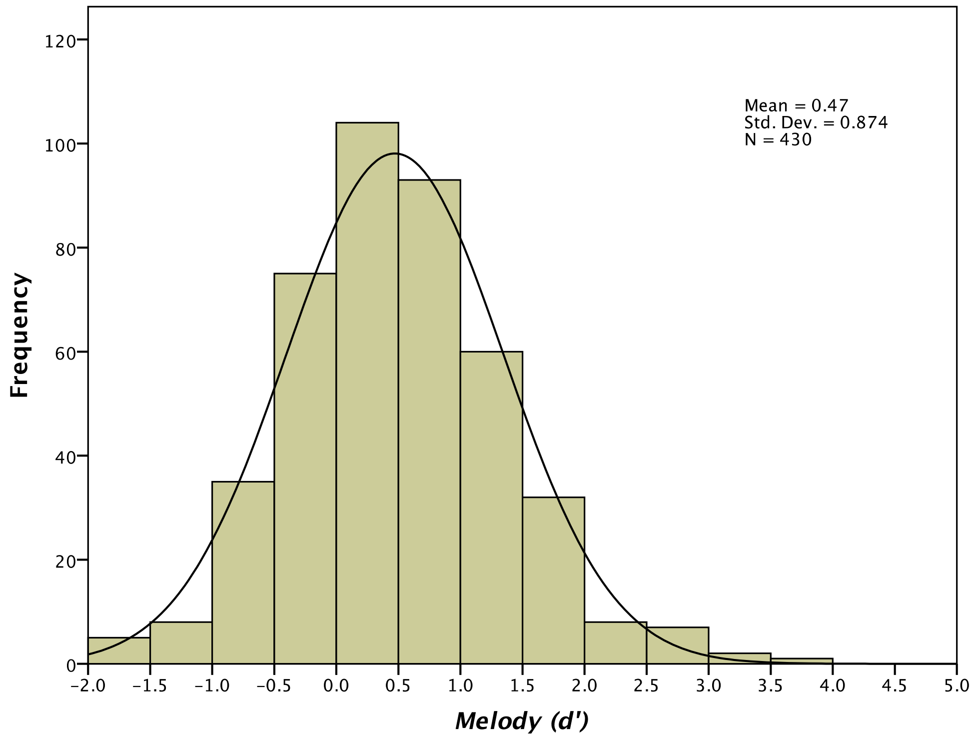

Mean = 12.64
Std. Dev. = 1.869
N = 430
Frequency
Timbre (raw)
100
80
60
40
20
0
5
6
7
8
9
10
11
12
13
14
15
16
17
18

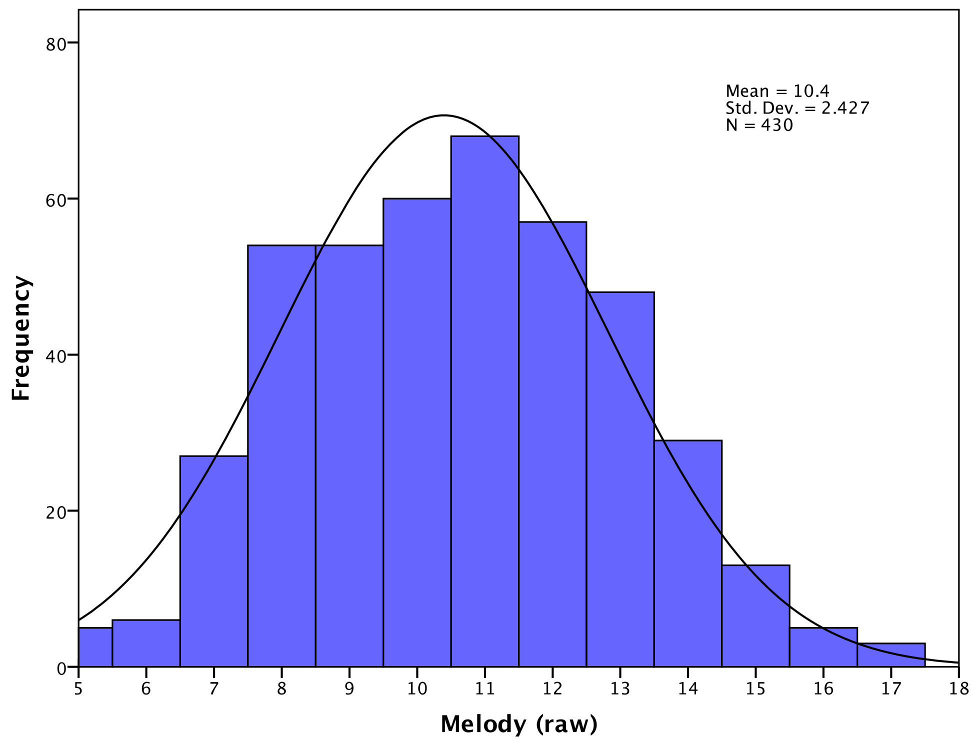
Pitch


Mean = 1.46
Std. Dev. = 0.850
N = 430
Frequency
Pitch (d')
120
100
80
60
40
20
0
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

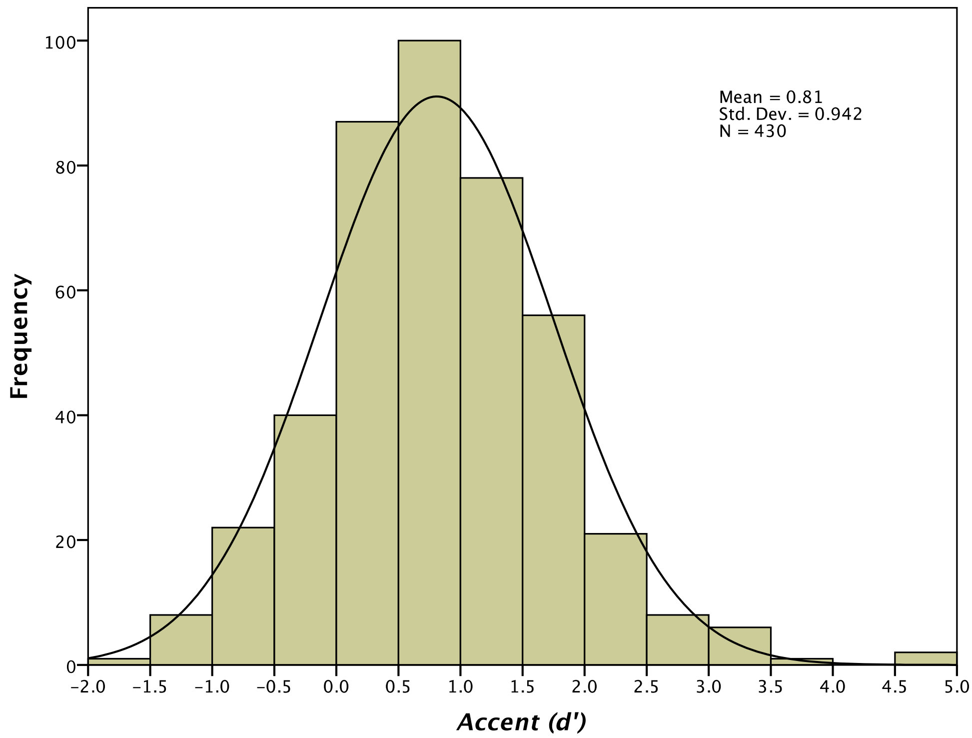

Mean = 12.09
Std. Dev. = 1.82
N = 430
Frequency
Pitch (raw)
100
80
60
40
20
0
5
6
7
8
9
10
11
12
13
14
15
16
17
18

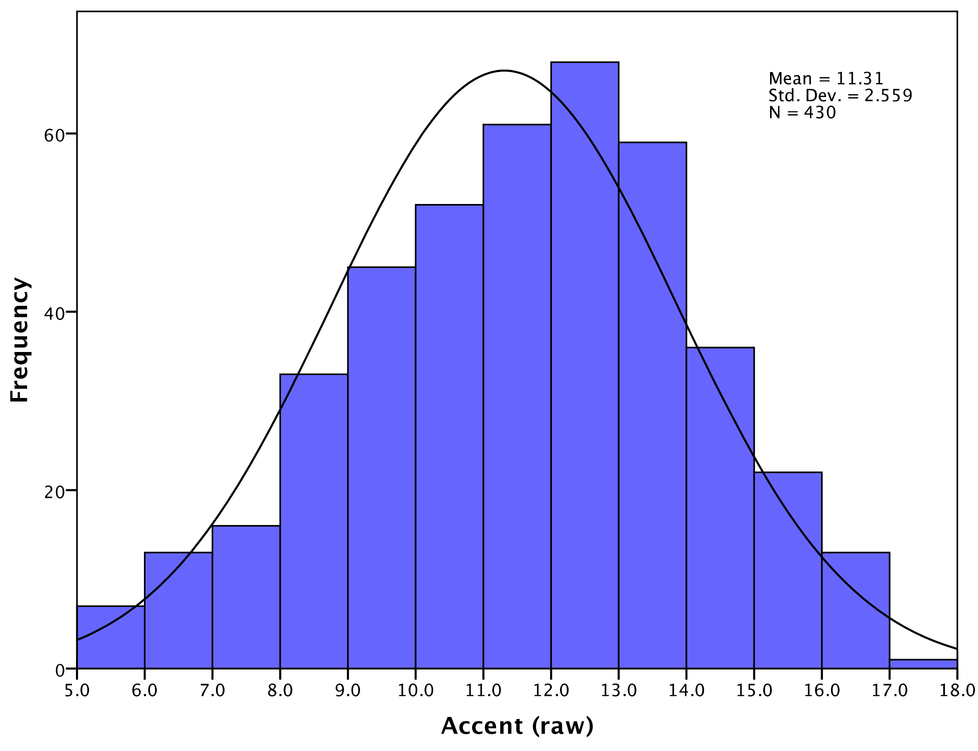
Melody


Mean = 0.47
Std. Dev. = 0.874
N = 430
Frequency
0
20
40
60
80
100
120
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Melody (d')


Mean = 10.4
Std. Dev. = 2.427
N = 430
Frequency
0
20
40
60
80
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Melody (raw)

Accent


Mean = 0.81
Std. Dev. = 0.942
N = 430
Frequency
Accent (d')


Mean = 11.31
Std. Dev. = 2.559
N = 430
Frequency
Accent (raw)

APPENDIX 6

PROMS Stimuli Transcription

The stimuli transcriptions of the PROMS (studies 3, 4, 5) are provided in this appendix. Please note that some of the subtest transcriptions are provided in Western score format (i.e., melody, rhythm, rhythm-to-melody, accent), but this does not imply that the stimuli were created strictly based on western musical rules, therefore aspects such as key signature and time signature were omitted in the scores. Also note that the PROMS is a perceptual test, therefore these transcriptions are only provided for record purposes. To fully experience the PROMS stimuli individuals should always refer to the audio stimuli (Appendix 7) rather than the visual representation provided in this appendix.

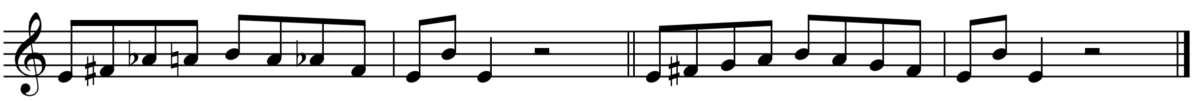
Reminder:

Same - Standard-stimulus and the comparison stimulus are identical
Different - Standard-stimulus and the comparison stimulus are identical

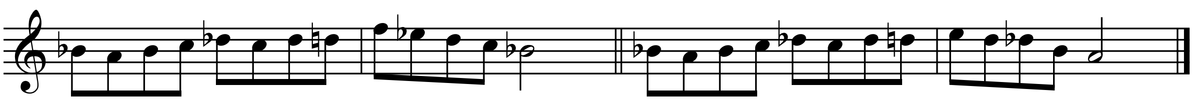
Melody

Difficulty Level: Easy

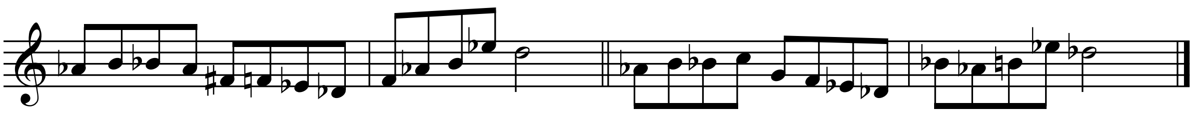
Different 1



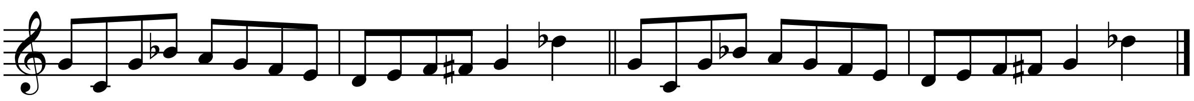
Different 2



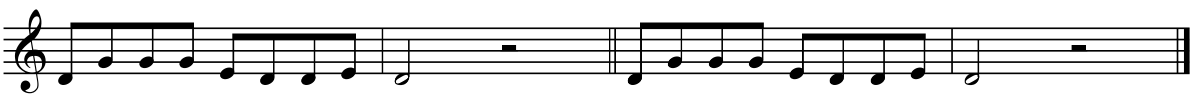
Different 3



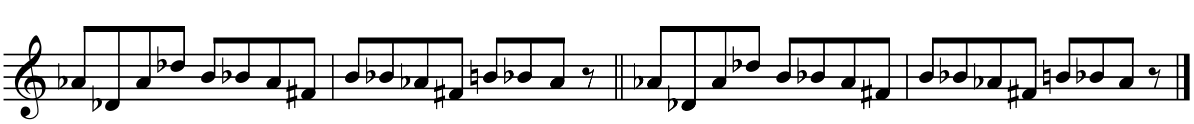
Same 1



Same 2



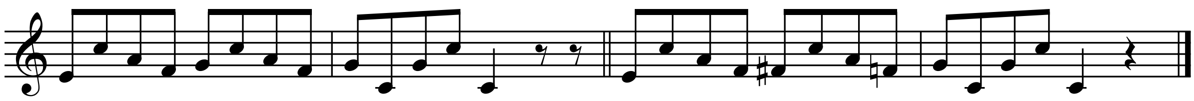
Same 3



Melody

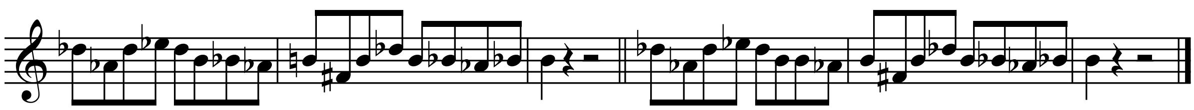
Difficulty Level: Moderate

Different 1

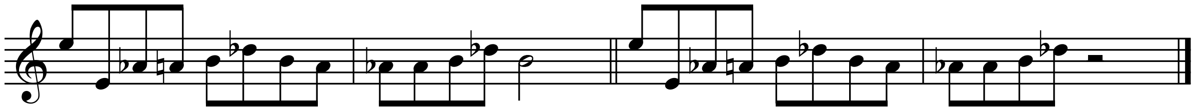


Different 2

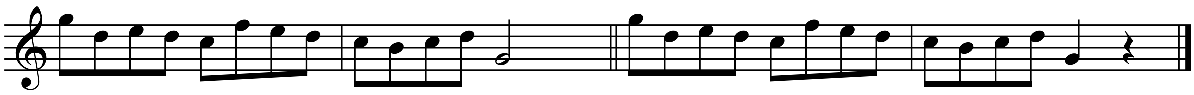
Different 3



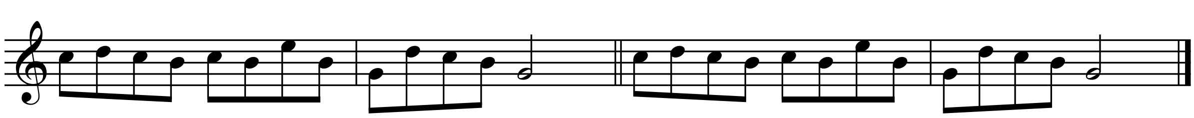
Same 1



Same 2



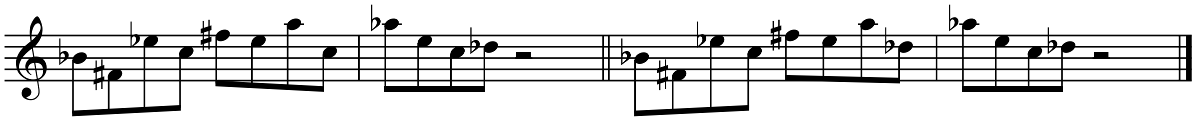
Same 3



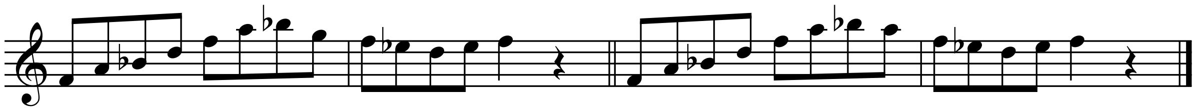
Melody

Difficulty Level: Complex

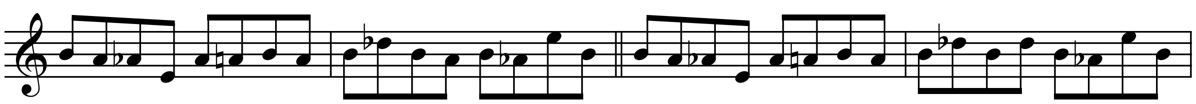
Different 1



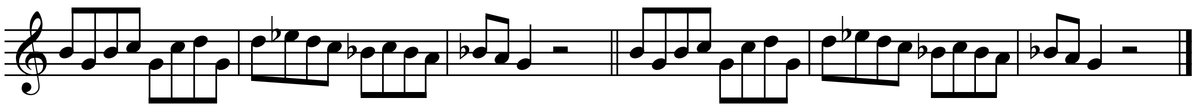
Different 2



Different 3

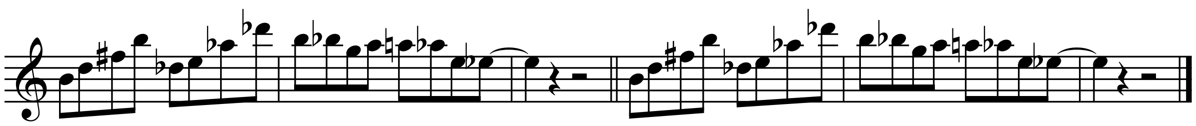


Same 1



Same 2

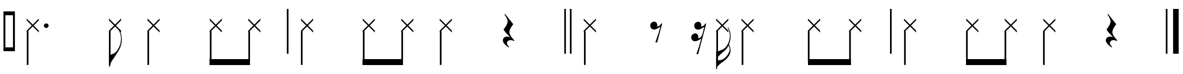
Same 3



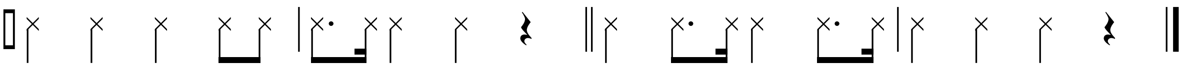
Rhythm

Difficulty Level: Easy

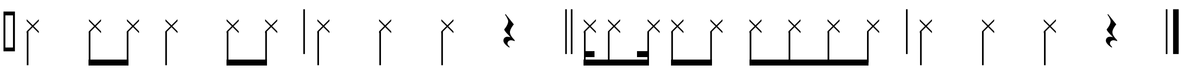
Different 1



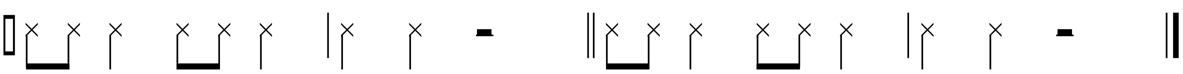
Different 2



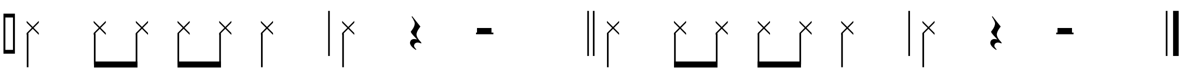
Different 3



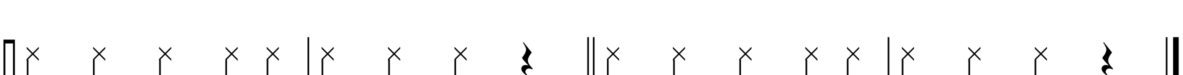
Same 1



Same 2



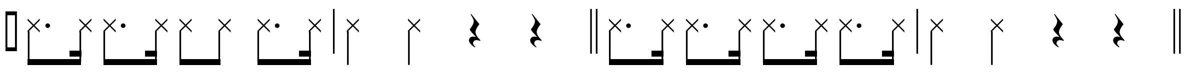
Same 3



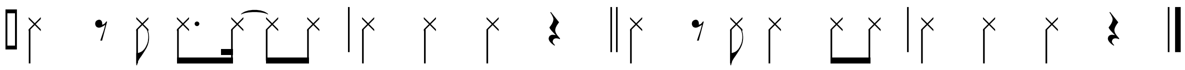
Rhythm

Difficulty Level: Moderate

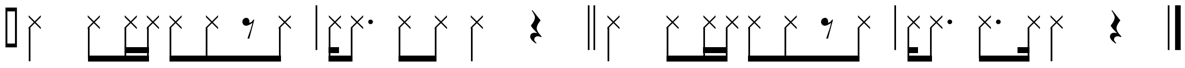
Different 1



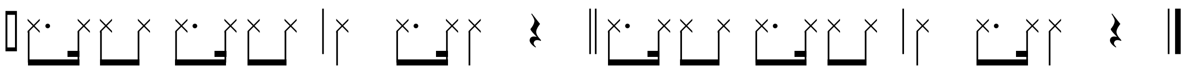
Different 2



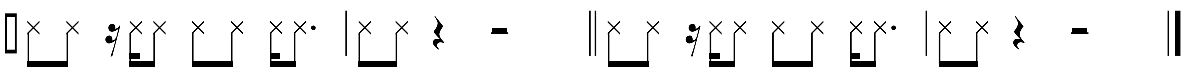
Different 3



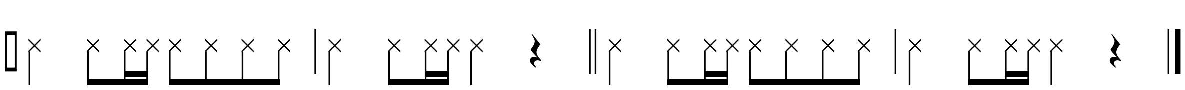
Same 1



Same 2



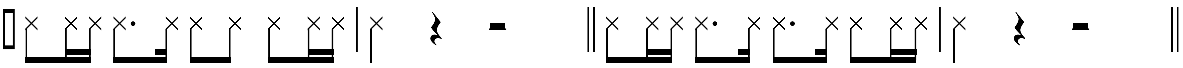
Same 3



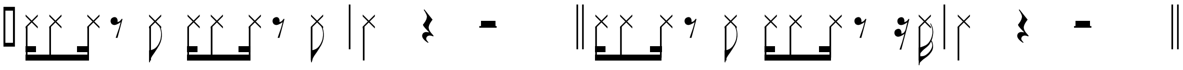
Rhythm

Difficulty Level: Complex

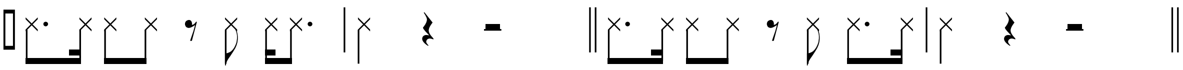
Different 1



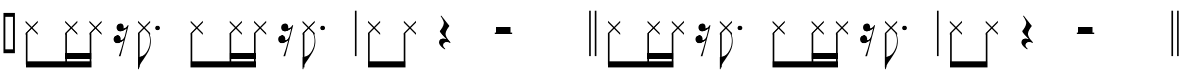
Different 2



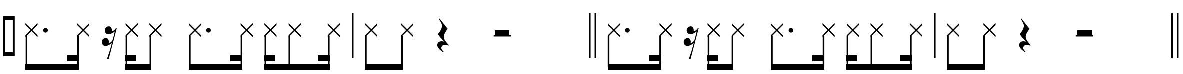
Different 3



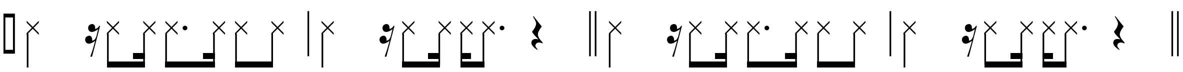
Same 1



Same 2



Same 3



Rhythm-to-Melody

Difficulty Level: Easy

Different 1

Different 2

Different 3

Same 1

Same 2

Same 3

Rhythm-to-Melody

Difficulty Level: Moderate

Different 1

Different 2

Different 3

Same 1

Same 2

Same 3

Rhythm-to-Melody

Difficulty Level: Complex

Different 1

Different 2

Different 3

Same 1

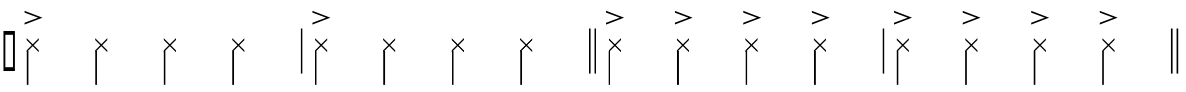
Same 2

Same 3

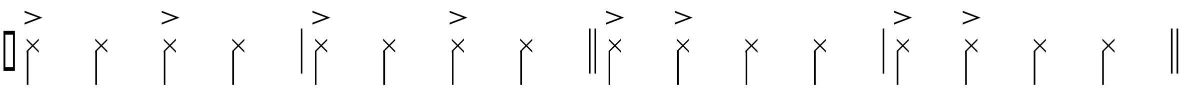
Accent

Difficulty Level: Easy

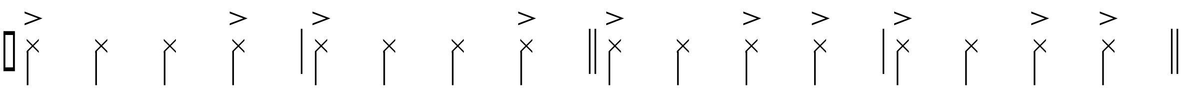
Different 1



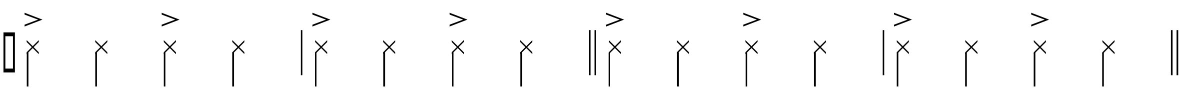
Different 2



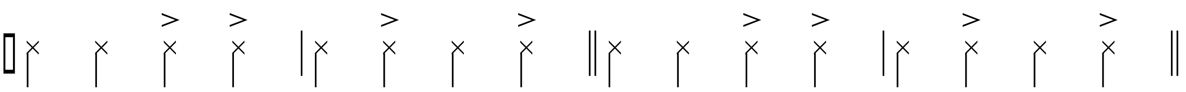
Different 3



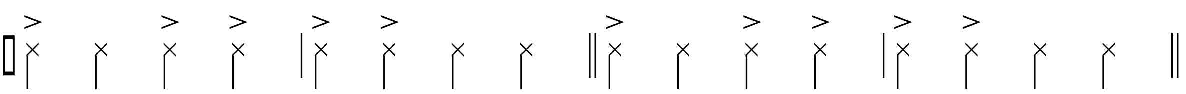
Same 1



Same 2



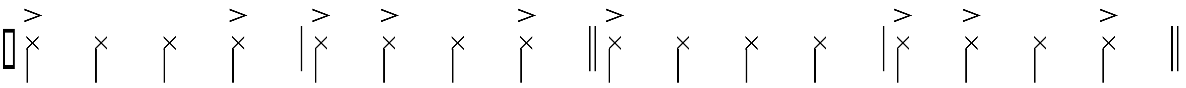
Same 3



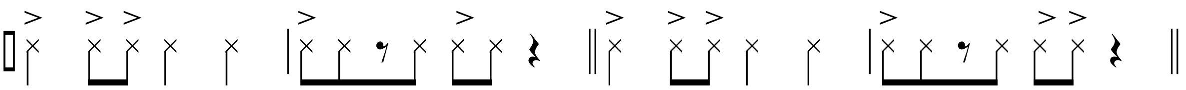
Accent

Difficulty Level: Moderate

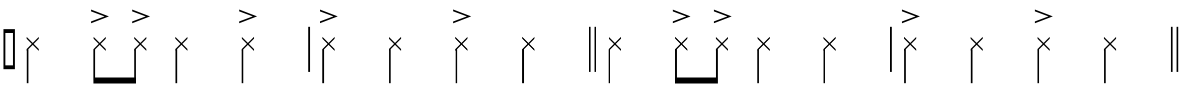
Different 1



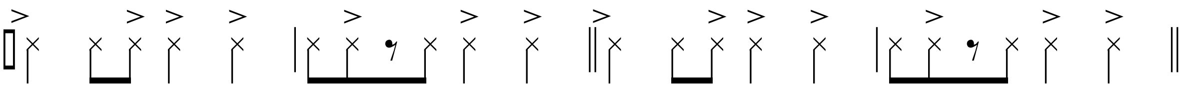
Different 2



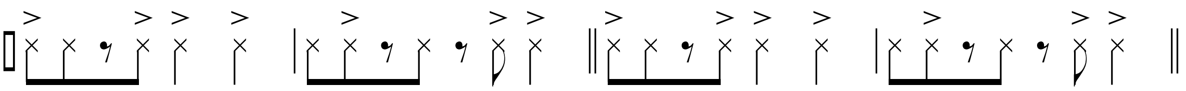
Different 3



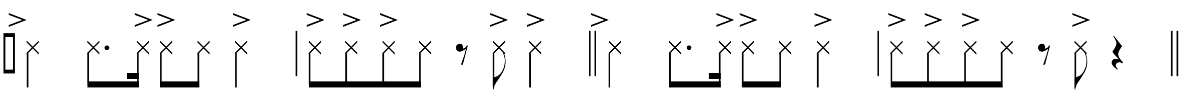
Same 1



Same 2



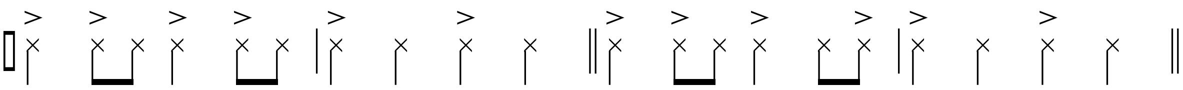
Same 3



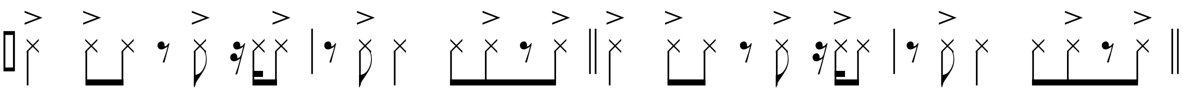
Accent

Difficulty Level: Complex

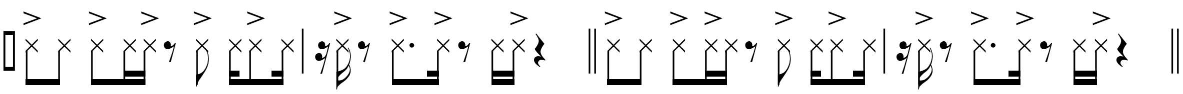
Different 1



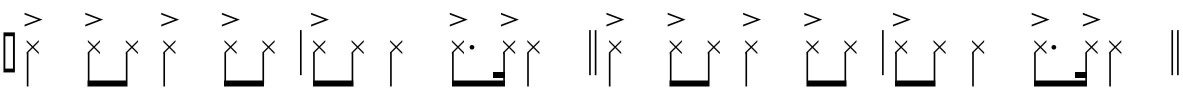
Different 2



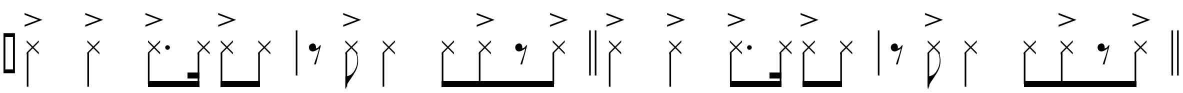
Different 3



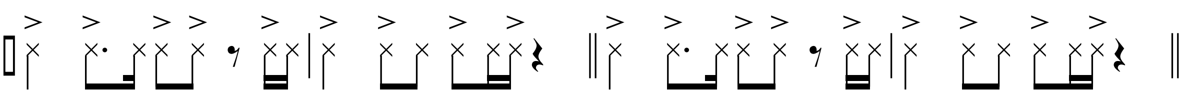
Same 1



Same 2



Same 3



Tempo

Trial Type/Difficulty Level	BPM
Easy 1	110 : 130
Easy 2	100 : 120
Easy 3	120 : 112
Moderate 1	125 : 113
Moderate 2	107 : 101
Moderate 3	113: 110
Complex 1	112 : 120
Complex 2	112: 109
Complex 3	117 : 120
Same 1	130
Same 2	114
Same 3	113
Same 4	115
Same 5	102
Same 6	120
Same 7	111
Same 8	110
Same 9	128

Pitch

Trial Type/Difficulty Level	Hz
Easy 1	447 : 435
Easy 2	445 : 485
Easy 3	440 : 466
Moderate 1	440 : 432
Moderate 2	432 : 437
Moderate 3	440 : 427
Complex 1	443 : 440
Complex 2	446 : 448
Complex 3	432 : 435
Same 1	466
Same 2	415
Same 3	434
Same 4	436
Same 5	443
Same 6	437
Same 7	440
Same 8	485
Same 9	427

Tuning

Trial Type/Difficulty Level	Cents
Easy 1	0 : 40
Easy 2	0 : 50
Easy 3	50 : 10
Moderate 1	0 : 30
Moderate 2	25 : 0
Moderate 3	10 : 40
Complex 1	30 : 10
Complex 2	10 : 25
Complex 3	20 : 0
Same 1	0
Same 2	20
Same 3	10
Same 4	50
Same 5	10
Same 6	40
Same 7	20
Same 8	30
Same 9	0

Timbre

Trial Type/ Difficulty Level	Instruments
Easy 1	Horn section vs. Woodwind section
Easy 2	Woodwind section vs. Horn section + CELLO
Easy 3	Woodwind section + VIOLIN vs. Horn section
Moderate 1	Woodwind section + VIOLIN vs. Woodwind section
Moderate 2	Viola section + CLARINET vs. Viola section
Moderate 3	Clarinet vs. Clarinet section + ENGLISH HORN
Complex 1	Bassoon vs. Bassoon +CLARINET
Complex 2	Viola vs. Viola section + VIOLIN
Complex 3	Clarinet section vs. Clarinet section + HORN
Same 1	Woodwind Section + TRUMPET
Same 2	Horn section
Same 3	Woodwind section
Same 4	Horn section + FLUTE
Same 5	Woodwind section + PIANO
Same 6	Horn section + CLARINET
Same 7	Horn Section + CELLO
Same 8	Woodwind Section + VIOLIN
Same 9	Horn Section + PIANO

Note. “Section” refers to a chord that consists of C_4, E_4, G_4, C_5. Capitalised instruments denote that the E note of the chord is replaced with said capitalised instrument. For example, Horn Section+ PIANO refers to Horn C_4, Piano E_4, Horn G_4 and Horn C_5.

Loudness

Trial Type/Difficulty Level	dB
Easy 1	-5 : 2
Easy 2	0 : -6
Easy 3	-1 : -6
Moderate 1	0 : -3
Moderate 2	0 : 3
Moderate 3	-6 : 0
Complex 1	0 : 2.5
Complex 2	2 : -1
Complex 3	-1.5 : 0
Same 1	-3
Same 2	0
Same 3	-1
Same 4	-2
Same 5	-6
Same 6	-4
Same 7	0
Same 8	1
Same 9	2

APPENDIX 7

Thesis Support Website (Audio & Video Demo)

This documentation provides information on how to access the content of this CD.

Option 1: Via Internet

Use one of these links:

1. http://tinyurl.com/appendix7
2. http://php.york.ac.uk/fs/psyc523/Thesis_Appendix/Welcome.html

Option 2: Local Browsing

Click on Welcome.html file in the folder, it will open in your default Internet browser. If not, you can ***right click*** on Welcome.html, and ***open with*** your favourite Internet browser.

Option 3: Manual Browsing

1. PROMS stimuli: Go to " Audio" ➔ PROMS
2. Supporting materials of Literature Review: Go to " Audio" ➔ Literature Review
3. PROMS video demo: Go to "Video"

Abbreviations and symbols

2AFC	Two-alternative Forced Choice
3AFC	Three-alternative Forced Choice
AMMA	Advanced Measures of Music Audiation
AWMA	Automated Working Memory Assessment
d’	Index of Sensitivity (pronounced ‘dee-prime’)
dB	Decibel
dBA	Decibel using “ ‘A’ weighting filter ”
EEG	Electroencephalography
fMRI	Functional Magnetic Resonance Imaging
KMO	Kaiser-Meyer-Olkin measures of sampling adequacy
MAP	Musical aptitude profile
MBEA	Montreal Battery of Evaluation of Amusia
MMN	Mismatch Negativity
MTurk	Amazon Mechanical Turk
PCA	Principal Component Analysis
SDT	Signal Detection Theory
SPL	Sound Pressure Level
STM	Short-term memory
TBAC	Tests of Basic Auditory Capabilities
WM	Working memory
Crotchet	Quarter note
Quaver	Eighth note
Semiquaver	Sixteenth note

Abbreviations and symbols for statistics, for units of measurement under the International System of Units, those appearing in references, and others used very commonly in publications from the American Psychological Association are omitted from the list above.

REFERENCES

Abel, S. M. (1972). Discrimination of temporal gaps. *The Journal of the Acoustical Society of America*, *52*, 519-524.

ABRSM (2011). Associated Board of the Royal Schools of Music: History. *History of ABRSM*. Retrieved May 11, 2011, from http://www.abrsm.org/en/regions/usa/united-states/about-abrsm/history/

Ackerman, P. L., Beier, M. G., & Boyle, M. O. (2005). Working Memory and Intelligence: The same or different constructs. *Psychological Bulletin*, *131*(1), 30-60.

Aggleton, J. P., Kentridge, R. W., & Good, J. M. M. (1994). Handedness and musical ability: A study of professional orchestral players, composers, and choir members. *Psychology of Music*, *22*(2), 148-156.

Alcock, K. J., Passingham, R. E., Watkins, K., & Vargha-Khadem, F. (2000). Pitch and timing abilities in inherited speech and language impairment. *Brain and Language*, *75*(1), 34–46.

Allan, L. G. (1979). The perception of time. *Perception & Psychophysics*, *26*, 340-354.

Allen, G. (1878). Note-Deafness. *Mind*, *3*(10), 157-167.

Alloway, T. P. (2007). *Automated Working Memory Assessment Manual*. London: Harcourt Assessment.

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice Hall.

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of experimental child psychology*, *83*(2), 111–130.

Apple (2011). Apple - Logic Studio - Logic Pro 9. Retrieved October 28, 2011, from http://www.apple.com/logicstudio/logicpro/

Arabin, B. (2002). Music during pregnancy. *Ultrasound Obstet Gynecol*, *20*(5), 425-430.

References

Auer, P., Couper-Kühlen, E., & Müller, F. (1999). *Language in time: The rhythm and tempo of spoken interaction (Oxford Studies in Sociolinguistics)*. New York: Oxford University Press.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia A group study of adults afflicted with a music-specific disorder. *Brain*, *125*(2), 238–251.

Baddeley, A., D. (1997). *Human memory: Theory and practice (Revised edition).* Hove: Psychology Press.

Baddeley, A., D. (2000). The episodic buffer: A new component of working memory? *Trends in cognitive sciences*, *4*(11), 417–423.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological bulletin*, *128*(4), 612637.

Baron, R. M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Bartlett, J.C. & Dowling, W.J. (1980). The recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance,6*, 501-515

Becker, J. (1979). *Time and tune in Java. In A.L. Becker & A.A. Yengoyam (Eds.).* Norwood: Ablex Publishing Corporation.

Bentley, A. (1966) Musical ability in Children and its measurement. London: Harrap.

Berlyne, D. E. (1967). Arousal and reinforcement. In D. Levine (Ed.). *Nebraska symposium on motivation* (pp. 1–110). Lincoln: University of Nebraska Press.

Bertoli, S., Smurzynski, J., & Probst, R. (2002). Temporal resolution in young and elderly subjects as measured by mismatch negativity and a psychoacoustic gap detection task. *Clinical neurophysiology*, *113*(3), 396–406.

Besson, M., & Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(6), 1278–1296.

BGT. (2007, 2009). Britain's Got Talent. Retrieved August 23, 2012, from http://talent.itv.com/

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition*, *77*(1), 1-10.

References

Bienstock, S. F. (1937). A review of recent studies on musical aptitudes. *Journal of educational psychology*, *33*(6), 427–442.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*, 100-130.

Bisesi, E., Parncutt, R., & Friberg, A. (2011). An accent-based approach to performance rendering: Music theory meets music psychology. In A. Williamon, D. Edwards, & L. Bartel (Eds.). Presented at the International Symposium on Performance Science (ISPS 2011) (pp. 27-32). Utrech: The European Association of Conservatoires (AEC).

Bocca, E., & Antonelli, A.R. (1976). Masking Level Difference. *Audiology, 15*, 480–487.

Bódizs, R., Kis, T., Lázár, A. S., Havrán, L., Rigó, P., Clemens, Z., & Halász, P. (2005). Prediction of general mental ability based on neural oscillation measures of sleep. *Journal of sleep research*, *14*(3), 285–292.

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology*, *6*, 145-238.

Boyle, D. J. (1992). *Evaluation of musical ability. In R. Colwell (Ed.)*. New York: Macmillan: Handbook for Research in Music Teaching and Learning.

Brandler, S., & Rammsayer, T. H. (2003). Differences in Mental Abilities between Musicians and Non-Musicians. *Psychology of Music*, *31*(2), 123-138.

Bregman, A. S. (1990). Auditory Scene Analysis: The Perceptual Organization Of Sound. Cambridge, MA: Bradford Books, MIT Press.

Brezinka, C., Lechner, T., & Stephan, K. (1997). The fetus and noise. *Gynakol Geburtshilfliche Rundsch*, *37*(3), 119–129.

British Society of Audiology. (2004). Pure tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels.

Brockmann, C. L. (2012). Anatomy of the Human Ear. Retrieved August 21, 2012, from http://commons.wikimedia.org/wiki/File:Anatomy_of_the_Human_Ear.svg

Brown, J. (1999). Beat of a different drummer. Retrieved May 10, 2011 from http://www.pbs.org/newshour/bb/entertainment/jan-june99/drummer_6-14.html

Brown, S., & Jordania, J. (2011). Universals in the world's musics. *Psychology of Music*.

Brown, W. A., Sachs, H., Cammuso, K., & Folstein, S. E. (2002). Early music training and absolute pitch. *Music Perception*, *19*(4), 595-597.

BSA. (2004). Pure tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels. British Society of Audiology.

Bugos, J. A., Perlstein, W. M., McCrae, C. S., Brophy, T. S., & Bedenbaugh, P. H. (2007). Individualized piano instruction enhances executive functioning and working memory in older adults. *Aging & Mental Health, 11*(4), 464–471.

Burns, E. D. (1999). Intervals, scales and tuning. *The Psychology of Music. In D. Deutsch (Ed.).* (pp. 215-264). California, USA: Academic Press.

Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal for the Educational Psychology*, *91*(3), 403–313.

Butler, D., & Brown, H. (1984). Tonal structure versus function: Studies of the recognition of harmonic motion. *Music Perception*, *2*(1), 6-24.

Butzlaff, R. (2000). Can music be used to teach reading? *Journal of Aesthetic Education*, *34*(3/4), 167–178.

Byrne, B. (1974). Handedness and musical ability. *British Journal of Psychology*, *65*(2), 279-281.

Cage, J. (1993). *4'33"*. Hungrary: Hungaroton.

Cambridge Dictionaries Online (CDO). (2011). Music- Definition. Retrieved February 14, 2012, from http://dictionary.cambridge.org/dictionary/

Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicology: Music, Mind and Brain*, *1*(1), 12–29.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological review*, *97*(3), 404-431.

Carol, G. (2007). Decibel (Loudness) Comparison Chart. Retrieved February 14, 2012, from http://www.gcaudio.com/resources/howtos/loudness.html

Carson, A. D. (1998). Why Has Musical Aptitude Assessment Fallen Flat? And What Can We Do About It? *Journal of Career Assessment*, *6*(3), 311–327.

Carterette, E. C., & Kendall, R. (1999). Comparative music perception and cognition. *The psychology of music* (2nd ed., pp. 725–791). New York, NY: Academic Press.

Caruso, J. C. (2000). Reliability generalization of the NEO Personality Scales. *Educational and Psychological Measurement*, *60*(2), 236-254.

Cary, H. (1922). Are you a musician? Professor Seashore's specific psychological tests for specific musical abilities. *Scientific American*, *12*, 326–327.

Castagne, N., & Cadoz, C. (2003). 10 criteria for evaluating physical modelling schemes for music creation. *Proc. 6th Int. Conf. Digital Audio Effects (DAFX-03)* (pp. 270-277). London, UK.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General; Journal of Experimental Psychology: General*, *113*(3), 394.

Cattell, R. B. (1940). A culture-free intelligence test. *The Journal of Educational Psychology*, *31*(3), 161–179.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1-22.

Chan, S., Ho, Y., & Cheung, M. (1998). Music training improves verbal memory. *Nature*, *396*, 128.

Chang, H. W., & Trehub, S. E. (1977). Infants' perception of temporal grouping in auditory patterns. *Blackwell Publishing on behalf of the Society for Research in Child Development*, *48*(4), 1666–1670.

Cheek, J. M., & Smith, L. R. (1999). Music training and mathematics achievement. *Adolescence*, *34*, 759–762.

Chin, C. S. (2003). The development of absolute pitch: A theory concerning the roles of music training at an early developmental age and individual cognitive style. *Psychology of Music*, *31*(2), 155-171.

Chowning, J. M., & Bristow, D. (1987). *FM theory and applications: By musicians for musicians*. Tokyo, Japan: Yamaha Foundation.

Christopherson, L. A., & Humes, L. E. (1992). Some psychometric properties of the Test of Basic Auditory Capabilities (TBAC). *Journal of Speech and Hearing research*, *35*(4), 929–935.

Clarke, E. F. (1987). Categorical rhythm perception: An ecological perspective. In A. Gabrielsson (Ed.). *Action and perception in rhythm and music* (pp. 19–34). Stockholm, Sweden: Royal Swedish Academy of Music.

Clarke, E. F. (1999). Rhythm and timing in music. *The Psychology of Music. In D. Deutsch (Ed.).* (2nd ed.). California: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum.

Cook, N. D. (2009). Harmony perception: Harmoniousness is more than the sum of interval consonance. *Music Perception*, *27*(1), 25–42.

Coon, H., & Carey, G. (1989). Genetic and environmental determinantes of musical ability in twins. *Behavior Genetics*, *19*(2), 183-193.

Conard, N. J., Malina, M., & Münzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. Nature.

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: tapping to the beat. *Cortex*, *45*(1), 119–130.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, *78*(1), 98-104.

Cosi, P., De Poli, G., & Lauzzana, G. (1994). Auditory modelling and self-organizing neural networks for timbre classification. *Journal of New Music Research*, *23*(1), 71–98.

Costa-Giomi, E. (1999). The Effects of Three Years of Piano Instruction on Children's Cognitive Development. *Journal of Research in Music Education*, *47*(3), 198-212.

Coutinho, E., & Cangelosi, A. (2011). Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion*, *11*(4), 921–937.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research*, *169*, 323–338.

Creelman, C. D. (1962). Human discrimination of auditory duration. *Journal of the Acoustical Society of America*, *34*, 582-593.

Crocker, R.L. (2000). *An introduction to Gregorian Chant*. New Haven: Yales University Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, *52*(4), 281-302.

Cuddy, L. L., & Cohen, A. J. (1976). Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology, 28,* 255-270.

Cuddy, L. L, Balkwill, L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare: A study of 'tone deafness' among university students. Annals of the New York Academy of Sciences, The Neurosciences and Music II: From perception to performance, 1060, 311-324.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of Structure in Short Melodic Sequences. *Journal of Experimental Psychology, Human Perception and Performance*, *7*(4), 869–883.

Cuddy, L. L., Cohen, A. J., & Miller, J. (1979). Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology*, *33*(3), 148-157.

Cutler, A. (1996). Prosody and the word boundary problem. In J.L. Morgan and K. Demuth (Eds.). *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 87-99). Mahwah: Lawrence Erlbaum.

Dahl, S., & Granqvist, S. (2003). Looking at perception of continuous tempo drift- a new method for estimating internal drift and just noticeable difference. In R. Bresin (Ed.). *Proceedings of the Stockholm Music Acoustics Conference (SMAC 03)* (Vol. II, pp. 595–598). Stockholm, Sweden.

Darwin, C. J., & Ciocca, V. (1992). Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component. *The Journal of the Acoustical Society of America*, *91*(6), 3381-3190.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51–62.

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, *11*, 201-211.

Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, *28*(1), 49–55.

Deasy, R. J. (2002). *Learning in the Arts and Student Academic and Social Development*. Washington: Arts Education Partnership.

Delogu, F., Lampis, G., & Olivetti Belardinelli, M. (2006). Music-to-language transfer effect: may melodic ability improve learning of tonal languages by native nontonal speakers? *Cognitive Processing*, *7*(3), 203-207.

DePoli, G., Roda, A., & Vidolin, A. (1998). Note-by-note analysis of the influence of expressive intentions and musical structure in violin performances. *Journal of New Music Research*, *27*, 293-321.

Deutsch, D. (1969). Music recognition. *Psychological Review*, *76*(3), 300–307.

Dewar, K. M., Cuddy, L. L., & Mewhort, D. J. (1977). Recognition memory for single tones with and without context. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(1), 60–67.

Dewitt, L.A., & Crowder, R.G. (1986). Recognition of novel melodies after brief delays. *Music Perception*, *3*(3), 259–274.

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, *11*(3), 189–212.

Dowling, W. J. (1978). Scale and contour: Two components of memory for melodies. *Psychology Review,* 1978, 85(4), 341-354.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Attention, Perception, & Psychophysics*, *50*(4), 305–313.

Dowling, W.J. & Bartlett, J.C. (1981). The importance of interval information in long term memory for melodies. *Psychomusicology, 1*, 30-49.

Dowling, W.J. & Fujitani D.S (1971). Contour, interval and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America,* 1971, *49,* 524-531.

Dowling, W. J., Kwak, S., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Attention, Perception, & Psychophysics*, *57*(2), 136–149.

Downs, J. S., Holbrook, M. B., Sheng, S., & Cranor, L. F. (2010). Are your participants gaming the system? : Screening mechanical turk workers. *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 2399–2402).

Drake, C. (1993). Perceptual and performed accents in musical sequence. *Bulletin of Psychonomic Society*, *31*, 107-100.

Drake, C., & Botte, M. C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Attention, Perception, & Psychophysics*, *54*(3), 277–286.

Drake, C., Botte, M. C., & Gérard, C. (1989). A perceptual distortion in simple musical rhythms. Presented at the Proceedings of the International Society of Psychophysics: Fifth Annual Meeting, Cassis, France.

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception*, *10*, 343-378.

Drake, C., Gros, L., & Penel, L. (1999). *How Fast Is That Music? The Relation Between Physical and Perceived Tempo* (In S. Yi, ed.). Seoul: Seoul National University Press.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, *77*(3), 251–288.

Drake, R. M. (1954). Musical Aptitude Test: Manual (Chicago: Science Research Associates).

Drake, R. M. (1957). *Manual for the Drake Musical Aptitude Tests* (Second Edition). Chicago: Science Research Associates.

Drennan, W., and Watson, C.S. (2001). Sources of variation in profile analysis. Individual Differences and Extended Training. *Journal of the Acoustical Society of America*, *110*, 2491-2503.

Dubno, J. R., & Levitt, H. (1981). Predicting consonant confusions from acoustic analysis. *Journal of the Acoustical Society of America*, *69*(1), 249–261.

Dumesnil, R. (1979). *Le rythme musical: Essai historique et critique*. Paris: H: Champion.

Dunnette, M. D., & Borman, W. C. (1979). Personnel selection and classification systems. *Annual Review of Psychology*, *30*, 477-525.

Dyson, M. C., & Watkins, A. J. (1984). A figural approach to the role of melodic contour in melody recognition. *Attention, Perception, & Psychophysics*, *35*(5), 477–488.

Eagleson, H. V., & Eagleson, O. W. (1947). Identification of musical instruments when heard directly and over a public-address system. *The Journal of the Acoustical Society of America*, *19*(1), 291–291.

Echols, C. H. (1966). A role for stress in early speech segmentation. In J.L Morgan and K. Demuth (Eds.). *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 151-170). Mahwah: Lawrence Erlbaum Associate.

Eck, D. (2001). Finding downbeats with a relaxation oscillator. *Psychological Research, 66,* 18–25.

Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception*, *2*(3), 375–368.

Eerola, T. (2012). Modeling listeners' emotional response to music. *Topics in Cognitive Science, 4*(4), 607–624.

Eerola, T., Alluri, V., & Ferrer, R. (2008). Emotional connotations of isolated instruments sounds. *In Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC10)* (pp. 483–489). Sapporo: University of Hokkaido.

Eerola, T., Ferrer, R., & Alluri, V. (2012). Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception*, *30*(1), 49–70.

Efron, B., & Tibshirani, R.J. (1993). *An introduction to the bootstrap*. Boca Raton, B.L.: CRC Press.

Eisler, H. (1976). Experiments on subjective duration 1868-1975: A collection of power function exponents. *Psychological Bulletin*, *83*(6), 1154–1171.

Elhilali, M., Xiang, J., Shamma, S. A., & Simon, J. Z. (2009). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. (T. D. Griffiths, Ed.). *PLoS Biology*, *7*(6), e1000129.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, *11*(1), 19–23.

Ericsson, K. A., Nandagopal, K., & Roring, R. W. (2005). Giftedness viewed from the expert-performance perspective. *Journal for the Education of the Gifted*, *28*(3-4), 287–311.

d' Errico, F., Henshilwood, C., Lawson, G., Vanhaeren, M., Tillier, A. M., Soressi, M., Bresson, F., et al. (2003). Archaeological evidence for the emergence of language, symbolism, and music–an alternative multidisciplinary perspective. *Journal of World Prehistory, 17*(1), 1–70.

Fabiani, M., & Friberg, A. (2011). Influence of pitch, loudness, and timbre on the perception of instrument dynamics. *Journal of the Acoustical Society of America*.

Farnsworth, P. R. (1969). *The social psychology of music (2nd ed.)*. Ames: Iowa State University Press.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: Sage.

Fitch, W. T., & Rosenfeld, A. J. (2007). Perception and production of syncopated rhythms. *Music Perception*, *25*(1), 43–58.

Flowers, P. J. (1984). Attention to Elements of Music and Effect of Instruction in Vocabulary on Written Descriptions of Music by Children and Undergraduates. *Psychology of Music*, *12*(1), 17-24.

Forgeard, M., Winner, E., Norton, A., & Schlaug, G. (2008). Practicing a musical instrument in childhood is associated with enhanced verbal ability and nonverbal reasoning. *PLoS ONE*, *3*(10), 1-8.

Fox, P. T., & Parsons, L. M. (1996). Neural basis of the comprehension of musical harmony, melody , and rhythm. *Music Education Research*, 21-24.

Foxton, J. M., Brown, A. C. B., Chambers, S., & Griffiths, T. D. (2004). Training Improves Acoustic Pattern Perception. *Current Biology*, *14*(4), 322–325.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, D. (2004b). Characterization of deficits in pitch perception underlying “tone deafness.” *Brain*, *127*(4), 801–810.

Franklin, M. S., Sledge Moore, K., Yip, C.Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, *36*(3), 353-365.

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing Moderator and Mediator Effects in Counseling Psychology Research. *Journal of Counseling Psychology*, *51*(1), 115–134.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological science*, *15*(6), 373–378.

Friberg, A., & Battel, G. U. (2002). Structural communication. In R. Parncutt & G.E. McPherson (Eds.). *The science and psychology of music performance: Creative strategies for teaching and learning* (pp. 199–218). New York: Oxford University Press.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of cognitive neuroscience*, *16*(6), 1010–1021.

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*(1), 69–91.

Gagné, F. (1999). Nature or Nurture? A Re-examination of Sloboda and Howe's (1991) Interview Study on Talent Development in Music. *Psychology of Music*, *27*(1), 38-51.

Gagné, F. (2003). Transforming gifts into talents: The DMGT as a developmental theory. In N. Colangelo & G.A. David (Eds.). *Handbook of gifted education* (3rd ed., pp. 60-74). Boston: Pearson Education.

Garadat, S. N., & Pfingst, B. E. (2011). Relationship between gap detection thresholds and loudness in cochlear-implant users. *Hearing research*, *275*(1), 130–138.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution.* New York: Basic Book.

Gardner, H. (1999). Intelligence reframed: Multiple intelligences for the 21st century. New York: Basic Books.

Garner, W. R. (1974). *The processing of information and structure.* Toronto: Wiley.

Garner, W. R. & Clement, D. E. (1963). Goodness of pattern and pattern uncertainty. *Journal of Verbal Learning and Verbal Behavior, 2*, 446-452. Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience*, *23*(27), 9240-9245.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience*, *23*(27), 9240–9245.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410-419.

Geem, Z. W. (2009). *Music-inspired harmony search algorithm: Theory and applications*. New York, NY: Springer.

Gembris, H. (1997). Historical phases in the definition of musicality. *Psychomusicology: Music, Mind and Brain*, *16*(1-2), 17–25.

Gérard, C., Drake, C., & Botte, M. C. (1993). Rhythm perception: Interactions between time and intensity. *Contemporary Music Review*, *9*(1-2), 179-189.

Geringer, J. M. (1991). Temporal Discrimination of modulated intensity in music excerpts and tones. *Journal of Research in Music Education*, *39*, 113–120.

Geringer, J. M. (1993). Loudness estimations of noise, synthesizer, and music excerpts by musicians and nonmusicians. *Psychomusicology: Music, Mind and Brain*, *12*(1), 22–30.

Geringer, J. M. (1995). Continuous loudness judgments of dynamics in recorded music excerpts. *Journal of Research in Music Education*, *43*(1), 22.

Geringer, J. M., & Johnson, C.M. (2007). Effects of excerpt duration, tempo, and performance level on musicians' ratings of wind band performances. *Journal of Research in Music Education*, *55*(4), 289–301.

Geringer, J. M., Madsen, C. K., MacLeod, R. B., & Droe, K. (2006). The effect of articulation style on perception of modulated tempo. *Journal of Research in Music Education*, *54*(4), 324–336.

Geschwind, N. (1984). The brain of a learning-disabled individual. *Annals of Dyslexia*, *34*(1), 319–327.

Getty, D. J. (1975). Discrimination of short temporal intervals: A comparison of two models. *Perception & Psychophysics*, *18*, 1-8.

Getty, D. J. (1976). Counting processes in human timing: A comparison of two models. *Perception & Psychophysics*, *20*, 191-197.

Gibson, C., Folley, B. S., & Park, S. (2009). Enhanced divergent thinking and creativity in musicians: A behavioral and near-infrared spectroscopy study. *Brain and Cognition*, *69*(1), 162–169.

Glasberg, B. R., & Moore, B.C.J. (1989). Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. *Scandinavian Audiology*, *32*, 1-25.

Glasberg, B. R., & Moore, B. C. J. (1992). Effects of envelope fluctuations on gap detection. *Hearing research*, *64*(1), 81–92.

Glennie, D. E. (1993). Hearing essay. Retrieved from http://www.evelyn.co.uk/

Gordon, A., & Martin, P. J. (1993/1994). A study of the rhythmic skills of musically unsophisticated secondary school students when playing the electronic keyboard with a drum-machine. *Bulletin of the Council for Research in Music Education*, *119*, 59-64.

Gordon, E. E. (1965). The musical aptitude profile: A new and unique musical aptitude test battery. *Bulletin of the Council for Research in Music Education*, 12–16.

Gordon, E. E. (1967). The Musical Aptitude Profile. *Music Educators Journal*, *53*(6), 52-54.

Gordon, E. E. (1989). *Advance Measures of Music Audiation*. Chicago: Riverside Publishing Company.

Gordon, E. E. (1990). *Predictive Validity Study of AMMA- A One-Year Longitudinal Predictive Validity Study of the Advanced Measures of Music Audiation*. Chicago: GIA Publication.

Gordon, E. E. (1995). *Manual Musical Aptitude Profile* (Third Edition.). Chicago: GIA Publication.

Gordon, E. E. (2001a). *Music Aptitude and Related Tests- An Introduction*. Chicago: GIA Publication.

Gordon, E. E. (2001b). *A three-year study of the Musical Aptitude Profile*. Chicago: GIA Publication.

Gordon, E. E. (2001c). *Musical Aptitude Profile (CD)*. GIA Publication.

Gordon, E. E. (2008). *Advanced Measures of Music Audiation (AMMA)-Computer Version 1.0 Help Manual*. Chicago: GIA Publication.

Gosling, S. D., Sandy, C. J., John, O. P., & Potter, J. (2010). Wired but not WEIRD: The promise of the internet in reaching more diverse samples. *Behavioral and Brain Sciences*, *33*(2-3), 94–95.

Goswami, U. (in press). Language, music, and children's brains: a rhythmic timing perspective on language and music as cognitive systems. In P. Rebuschat, M. Rohmeier, J.A.Hawkins, & I.Cross (Eds.). *Language and music as cognitive systems*. Oxford: Oxford University Press.

Gott, P. S. (1973). Language after dominant hemispherectomy. *Journal of Neurology, Neurosurgery & Psychiatry*, *36*(6), 1082–1088.

Grahn, J. A., & Brett, M. (2007). Rhythm and Beat Perception in Motor Areas of the Brain. *Journal of Cognitive Neuroscience*, *19*(5), 893-906.

Graziano, A. B., Peterson, M., & Shaw, G. L. (1999). Enhanced learning of proportional math through music training and spatial-temporal training. *Neurological Research*, *21*(2), 139-152.

Gregory, D. (2010). The continuous response digital interface: An analysis of reliability measures. *Psychomusicology: Music, Mind and Brain*, *14*(1-2), 197–208.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, *61*(5), 1270–1277.

Gromko, J.E. (2005). The effect of music instruction on phonemic awareness in beginning readers. *Journal of Research in Music Education*, *53*(3), 199–209.

Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). New York: Wiley.

Hailstone, J. C., Omar, R., Henley, S. M., Frost, C., Kenward, M. G., & Warren, J. D. (2009). It's not what you play, it's how you play it: timbre affects perception of emotion in music. *The Quarterly Journal of Experimental Psychology*, *62*(11), 2141-2155.

Hall, H. H. (1937). Sound analysis. *The Journal of the Acoustical Society of America*, *8*(4), 257-262.

Hall III, J. W., & Grose, J. H. (1996). Development of temporal resolution in children as measured by the temporal modulation transfer function. *The Journal of the Acoustical Society of America*, *96*(1), 150-154.

Hallam, S. (2010). 21st century conceptions of musical ability. *Psychology of Music*, *38*(3), 308–330.

Halpern, A. R., & Bower, G. H. (1982). Musical expertise and melodic structure in memory for musical notation. *The American Journal of Psychology*, 31–50.

Halpern, A. R., & Darwin, C. J. (1982). Duration discrimination in a series of rhythmic events. *Perception & Psychophysics*, *31*, 86-89.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge: MIT Press.

Handel, S., & Oshinsky, J. S. (1981). The meter of syncopated auditory polyrhythms. *Perception & Psychophysics*, *30*(1), 1–9.

Hannon, E. E., & Trehub, S. E. (2005). Metrical Categories in Infancy and Adulthood. *Psychological Science*, *16*(1), 48–55.

Hartmann, W. M., McAdams, S., & Benneth, K. S. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *The Journal of the Acoustical Society of America*, *88*(4), 1712-1724.

Haueisen, J., & Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, *13*(6), 786-1792.

Hayes, A. F. (in-press). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling.

Hayes, J. R. (1981). *The Complete Problem Solver*. Philadelphia: Franklin Institute Press.

Haynes, S. N., & O'Brien, W. H. (2000). *Principles and practice of behavioral assessment.* New York: Springer.

Hayward, C. M., & Eastlund Gromko, J. (2009). Relationships among music sight-reading and technical proficiency, spatial visualization, and aural discrimination. *Journal of Research in Music Education, 57*(1), 26–36.

Helmbold, N., Rammsayer, T., & Altenmüller, E. (2005). Differences in primary mental abilities between musicians and nonmusicians. *Journal of Individual Differences*, *26*(2), 74-85.

von Helmholtz, H. (1877). *On the sensations of tone (English translation A.J. Ellis, 1885, reprinted 1954)*. New York, Dover: Dover Publications.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world. *Behavioral and Brain Sciences*, *33*(2-3), 61–83.

Henson, R. A., & Wyke, M. A. (1982). The performance of professional musicians on the Seashore Measures of Musical Talent: An unexpected finding. *Cortex*, *18*(1), 153-157.

Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization study of the teacher efficacy scale and related instruments. *Educational and Psychological Measurement*, *61*(3), 404–420.

Herndon, M. (1974). Analysis: The herding of scared cows? Ethnomusicology, *18,* 219-262.

Hetland, L. (2000a). Learning to make music enhances spatial reasoning. *Journal of Aesthetic Education*, *34*(3/4), 179–238.

Hetland, L. (2000b). Listening to Music Enhances Spatial-Temporal Reasoning: Evidence for the" Mozart Effect." *Journal of Aesthetic Education*, *34*(3/4), 105–148.

Hiller, L., & Ruiz, P. (1971). Synthesizing musical sounds by solving the wave equation for vibrating objects: Part I. *Journal of the Audio Engineering Society*, *19*, 463-470.

Hirsh, I. J. (1959). Auditory perception of temporal order. *The Journal of the Acoustical Society of America, 31*, 759-767.

Hirsh, I. J., Monahan, C.B., Grant, K. W., & Singh, P. G. (1990). Studies in auditory timing: 1. Simple patterns, *47*, 215-226.

Ho, Y. C., Cheung, M. C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology*, *17*(3), 439.

Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: examples from the child-clinical and pediatric psychology literatures. *Journal of consulting and clinical psychology*, *65*(4), 599–610.

Horwitz, A. R., Ahlstrom, J. B., & Dubno, J. R. (2011). Level-dependent changes in detection of temporal gaps in noise markers by adults with normal and impaired hearing. *The Journal of the Acoustical Society of America*, *130*(5), 2928–2938.

Honing, H. (2003). The final retard: On music, motion and kinematic models. *Computer Music Journal, 27*(3), 66–72.

Honing, H. (2006). Computational modeling of music cognition: A case study on model selection. *Music Perception, 23*(5), 365–376.

Honing, H., & Ladinig, O. (2008). The potential of the internet for music perception research: A comment on lab-based versus web-based studies. *Empirical Musicology Review, 3*(1), 1–7.

Howard, D. M., & Angus, J. A. S. (2006). *Acoustics and psychoacoustics* (3rd ed.). London: Focal Press 2006.

Howe, M. J. A., Davidson, J. W., & Sloboda, J. A. (1998). Innate talents: Reality or myth? *Behavioral and brain sciences*, *21*(03), 399–407.

Hulford, D. P. (1990). Training phonemic segmentation ability with a phonemic discrimination intervention in second- and third-grade children with reading disabilities. *Journal of Learning Disabilities*, *23*(9), 564–569.

Huntsinger, C. S., & Jose, P. E. (1991). A test of Gardner's modularity theory: A comparison of short-term memory for digits and tones. *Psychomusicology: Music, Mind and Brain*, *10*(1), 3–18.

Hurwitz, I., Wolff, P. H., Bortnick, B. D., & Kokas, K. (1975). Nonmusical effects of the Kodaly Music Curriculum in primary grade children. *Journal of Learning Disabilities*, *8*(3), 167-174.

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception*, *20*, 151-171.

Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2010). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, *47*, 674-689.

Hutchinson, S., Lee, L. H. L., Gaab, N., & Schlaug, G. (2003). Cerebellar volume of musicians. *Cerebral Cortex*, *13*(9), 943–949.

Ikeda, R. (2005). *Dataplex (CD).* Raster-Noton.

Irwin, R., & Francis, M. (1995). Perception of simple and complex visual stimuli: Decision strategies and hemispheric differences in same-different judgments. *Perception*, *24*, 787-809.

ISO. (1975). Acoustics-Standard tuning frequency (Standard musical pitch). *International Organization for Standardization*.

Iversen, J. R., & Patel, A. D. (2008). The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population. *10th International Conference on Music Perception & Cognition (ICMPC10). Sapporo, Japan, August* (Vol. 26).

Jakobson, L. S., Cuddy, L. L., and Kilgour, A. R. (2003). Time-tagging: A key to musicians' superior memory. *Music Perception*. *20*, 307-313.

Jakobson, L. S., Lewycky, S. T., Kilgour, A. R., & Stoesz, B. M. (2008). Memory for verbal and visual material in highly trained musicians. *Music Perception*, *26*(1), 41–55.

Janata, P., & Grafton, S. T. (2003). Swinging in the brain: Shared neural substrates for behaviors related to sequencing and music. *Nature Neuroscience*, *6*(7), 682-687.

Jensen, A. R. (1998). The g factor: The science of mental ability. In R. J. Sternberg & W. M. Williams (Eds.). *Intelligence, instruction, and assessment: Theory into practice* (pp. 111–131). Mahwah, NJ: Lawrence Erlbaum.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia*, *48*(9), 2630–2639.

Johnson, D. M., Jensen, J. K., & Watson, C. S. (1980). Individual differences in auditory capabilities: A preliminary test battery (A). *The Journal of the Acoustical Society of America*, *67*(S1), S58–S58.

Johnson, K. L., Nicol, T. G., Zecker, S. G., & Kraus, N. (2007). Auditory brainstem correlates of perceptual timing deficits. *Journal of Cognitive Neuroscience, 19*(3), 376–385.

Jones, M. R. (1978). Auditory patterns: Studies in the perception of structure. In E.C.

Carterette and M.P. Friedman (Eds.), *Handbook of perception VIII: Perceptual coding.* New York: Academic Press.

Jones, M. R., Fay, R. R., & Popper, A. N. (2010). *Music Perception*. New York, NY: Springer.

Joseph, R. (1988). The right cerebral hemisphere: Emotion, music, visual-spatial skills, body-image, dreams, and awareness. *Journal of Clinical Psychology*, *4*(5), 630–673.

Judge, T. A., Hurst, C., & Simon, L. S. (2009). Does it pay to be smart, attractive, or confident (or all three)? Relationships among general mental ability, physical attractiveness, core self-evaluations, and income. *Journal of Applied Psychology*, *94*(3), 742–755.

Juslin, P. N., & Sloboda, J. A. (2010). *Handbook of Music and Emotion: Theory, Research and Application*. Oxford: Oxford University Press.

Justus, T. C., & Bharucha, J. J. (2002). Music perception and cognition. In H. Pashler & S. Yantis (Eds.). *Steven's handbook of experimental psychology.* (3rd ed., Vols. 1-1, Vol. Sensation and Perception, pp. 453-492). New York: John Wiley & Sons.

Kang, R., Nimmons, G. L., Drennan, W., Longnion, J., Ruffin, C., Nie, K., Won, J. H., et al. (2009). Development and validation of the University of Washington Clinical Assessment of Music Perception Test. *Ear & Hearing*, *30*(4), 411-8.

Kaplan, R. M., & Saccuzzo, D. P. (2009). *Psychological testing: Principles, application, and issues* (7th ed.). Belmont: Wadsworth.

Karlin, J. E. (1942). A factorial study of auditory function. *Pscyhometrika*, *7*(4), 251–279.

Karma, K. (1973). The ability to structure acoustic material as a Measure of Musical Aptitude. 1. Background Theory and Pilot Studies. *Research Bulletin No. 38, Institute of Education, University of Helsinki*.

Karma, K. (1975). The ability to structure acoustic material as a Measure of Musical Aptitude. 2.Test construction and results. *Research Bulletin No. 43*, Institute of Education, University of Helsinki.

Karma, K. (1980). The ability to structure acousitc material as a Measure of Musical aptitude. 5. Summary and Conclusions. *Research Bulletin No. 52, Institute of Education, University of Helsinki*.

Karma, K. (1984). Musical Aptitude as the Ability to Structure Acoustic Material. *International Journal of Music Education*, *3*(1), 27-30.

Karma, K. (2007). Musical aptitude definition and measure validation: ecological validity can endanger the construct validity of musical aptitude tests. *Psychomusicology: Music, Mind and Brain*, *19*(2), 79–90.

Keating, P. (2005). D-prime (signal detection) analysis. Retrieved August 24, 2012, from http://www.linguistics.ucla.edu/faciliti/facilities/statistics/dprime.htm

Kemp, A. E. (1996). *The musical temperament*. Oxford: Oxford University Press.

Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *The Journal of the Acoustical Society of America*, *122*(1), 418.

Kilgour, A. R., Jakobson, L. S., & Cuddy, L. L. (2000). Music training and rate of presentation as mediators of text and song recall. *Memory & cognition*, *28*(5), 700–710.

Kim, Y. E., Chai, W., Ricardo, G., & Barry, V. (2000). Analysis of a contour-based representation for melody. *In Proceedings of International Symposium on Music Information Retrieval* (Vol. 120, p. 1579). Plymouth, Massachusetts.

King, H. A. (1954). A study of the relationship of music reading and I.Q. scores. *Journal of Research in Music Education*, *2*(1), 35-37.

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology*, *95*(3), 453–464.

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of basic and clinical physiology and pharmacology*, *12*(2 Suppl), 125-143.

Kitterick, P. T. (2008). *The bases of difficulties in spatial hearing for speech: Investigations using psychoacoustic techniques and magneto-encephalography* (Unpublished Doctoral Dissertation). University of York, York, UK.

Klein, D., Zatorre, R. J., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage*, *13*, 646-653.

Kline, P. (1993). *Personality: The psychometric view*. New York, NY: Routledge.

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing:“nonmusicians” are musical. *Journal of Cognitive Neuroscience*, *12*(3), 520–541.

Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *NeuroReport, 10*(6), 1309.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A.D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, *7*, 302 – 307.

Koester, L. S., & Farley, F. H. (1982). Psychophysiological characteristics and school performance of children in open and traditional classrooms. *Journal of Educational Psychology*, *74*(2), 254-263.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605.

Kraus, N., Skoe, E., Parbery-Clark, A., & Ashley, R. (2009). Experience-induced Malleability in Neural Encoding of Pitch, Timbre, and Timing. *Annals of the New York Academy of Sciences*, *1169*(1), 543–557.

Kremen, W. S., Jacobsen, K. C., Xian, H., Eisen, S. A., Eaves, L. J., Ysuang, M. T., & Lyons, M. J. (2007). Genetics of verbal working memory processes: A twin study of middle-aged men. *Neuropsychology*, *21*(5), 569-580.

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception*, *13*(3), 401-432.

Krumhansl, C.L & Kessler, E.J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review,* 1982, *89,* 334-368.

Kubik, G. (1962). The phenomenon of inherent rhythms in East and Central African instrumental music. *African Music*, *31*(1), 31–42.

Kubik, G. (1979). Pattern perception and recognition in African music. In J. Blacking & J.W. Kealiinohonoku (ed.). *The performing arts* (pp. 221–249). The Hague, Netherlands: Mouton Publishers.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160.

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, *26*(1), 1–37.

Lee, C.Y., & Hung, T.H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *124*(5), 3235.

Lee, J. H. (2010). Crowdsourcing music similarity judgments using mechanical turk. *Proceedings of the 11th ISMIR (International Society for Music Information Retrieval) conference* (pp. 183–188).

Lee, Y., Lu, M., & Ko, H. (2007). Effects of skill training on working memory capacity. *Learning and Instruction*, *17*(3), 336-344.

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, *63*(8), 1279-1292.

Lenhoff, H. M., Perales, O., & Hickok, G. (2001). Absolute pitch in Williams Syndrome. *Music Perception*, *18*(4), 491-503.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge: MIT Press.

Levine, M. (1995). *The Jazz theory book*. Petaluma, CA: Sher Music.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*, 467-477.

Licklider, J. C. R. (1954). "Periodicity" pitch and "Place" pitch. *The Journal of the Acoustical Society of America*, *26*(5), 945-945.

Liégeois, F., Connelly, A., Cross, J. H., Boyd, S. G., Gadian, D. G., Vargha-Khadem, F., & Baldeweg, T. (2004). Language reorganization in children with early-onset lesions of the left hemisphere: an fMRI study. *Brain*, *127*(6), 1229–1236.

Lienhard, J. H. (2011). No. 1736: The Seashore Test. *The Seashore Test*. Retrieved February 25, 2011, from http://www.uh.edu/engines/epi1736.htm

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, *11*(5), 1021–1031.

Lindström, E. (2003). The contribution of immanent and performed accents to emotional expression in short tone sequences. *Journal of New Music Research*, *32*(3), 269-280.

Lochy, A., Hyde, K.L., Parisel, S., Van Hyfte, S., & Peretz, I. (2004). *Discrimination of speech prosody in congenital amusia.* Poster presented at the 2004 meeting of the Cognitive Neuroscience Society, San Francisco, California.

Loewenthal, K. M. (1996). *An introduction to psychological tests and scales*. London, UK: UCL Press.

Loui, P., & Wessel, D. (2007). Harmonic expectation and affect in Western music: Effects of attention and training. *Attention, Perception, & Psychophysics*, *69*(7), 1084–1092.

Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide* Mahwah: Lawrence Erlbaum Associates.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge, UK: Cambridge University Press.

Madsen, C. K. (1979). Modulated beat discrimination among musicians and nonmusicians. *Journal of Research in Music Education*, *27*(2), 57-67.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, *18*(2), 199–211.

Margulis, E. H., & Levine, W. H. (2004). Melodic expectation: A priming study. *Proceedings of the Eighth International Conference on Music Perception and Cognition* (pp. 364–366). Evanston, IL.

Marin, R. (1982). Neurological aspects of music perception. In D. Deutch (Ed.). *The Psychology of Music* (pp. 453–477). New York: Academic Press.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*(2), 107–127.

Martin, G. (2008). The Opera Companion. *Hal Leonard Corporation*.

Mason, W., & Suri, S. (2011). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1-23.

Mathias, S. (2010). *Individual differences in pitch perception* (Unpublished Doctoral Dissertation). University of York, York, UK.

MathWorks. (2009). *Matlab 2009b*. UK. Retrieved from http://www.mathworks.co.uk/products/matlab/

McAuley, J. D. (2010). Tempo and rhythm. *Music Perception. In Jone M.R., Fay, R.R. and Popper A.N. (Eds).* (pp. 165–200). London: Springer.

McAuley, J. D., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(6), 1102–1125

McAuley, J. D., & Kidd, G. R. (1998). Effect of deviations from temporal expectations on tempo discrimination of isochronous tone sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1786–1800.

McAuley, J. D., Henry, M. J., & Tuft, S. (2011). Musician advantages in music perception: An issue of motivation, not just ability. *Music Perception: An Interdisciplinary Journal*, *28*(5), 505–518.

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General*, *135*(3), 348–367.

McClelland, D. C. (1980). Motive dispositions: The merits of operant and respondent measures. In LWheeler (Ed.). *Review of personality and social psychology* (Vol. 1, pp. 10-41). Beverly Hill: Sage.

McCutcheon, L. E. (2000). Another failure to generalize the Mozart effect. *Psychological Reports*, *87*, 325-330.

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, *18*(4), 452–463.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Hillsdale: Lawrence Erlbaum Associate.

McDonnell, M. (2007). Visual music. Presented at the Visual Music Marathon Event, Boston, USA.

McKelvie, P., & Low, J. (2002). Listening to Mozart does not improve children's spatial ability: Final curtains for the Mozart effect. *British Journal of Developmental Psychology*, *20*(2), 241–258.

McKinney, M.F., & Moelants, D. (2006). Ambiguity in tempo perception: What draws listeners to different metrical levels? *Music Perception*, *24*(2), 155–166.

McLaurin, M. C., Jenkins, J. F., William, E. F., & Rumore, M. C. (1973). Correlations of IQs on verbal and nonverbal tests of intelligence. *Psychological reports*, *33*(3), 821–822.

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate Practice Is Necessary but Not Sufficient to Explain Individual Differences in Piano Sight-Reading Skill: The Role of Working Memory Capacity. *Psychological Science*, *21*(7), 914-919.

Meister, I., Krings, T., Foltys, H., Boroojerdi, B., Müller, M., Töpper, R., & Thron, A. (2005). Effects of long-term practice and task complexity in musicians and nonmusicians performing simple and complex motor tasks: Implications for cortical motor organization. *Human Brain Mapping*, *25*(3), 345–352.

Melson, W. H., & McCall, R. B. (1970). Attentional responses of five-months girls to discrepant auditory stimuli. *Child Development*, *41*(4), 1159-1171.

Merriam, A.P. (1981). *African musical rhythm anc concepts of time-reckoning. In T. Nobbliee (ed)*. New York: Pendragon Press.

Merzbow. (1996). *Pulse Demon*. US: Release Entertainment.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., & Dies, R. R. (2001). Psychological Testing and psychological Assessment. *American Psychologist*, *56*(2), 128–165.

Micheyl, C., Kaernbach, C., & Demany, L. (2008). An evaluation of psychophysical models of auditory change perception. *Psychological review*, *115*(4), 1069.

Michon, J. A. (1964). Studies on subjective duration: I. Differential sensitivity in the perception of repeated temporal intervals. *Acta Psychologica*, *22*, 441-450.

Minati, L., Rosazza, C., D'Incerti, L., Pietrocini, E., Valentini, L., Scaioli, V., Loveday, C., et al. (2009). Functional MRI/Event-related potential study of sensory consonance and dissonance in musicians and nonmusicians. *NeuroReport*, *20*(1), 87–92.

Milovanov, R., & Tervaniemi, M. (2011). The interplay between musical and linguistic aptitudes: A review. *Frontiers in Psychology*, *2*, 321–326.

Moelants, D. (2002). Preferred tempo considered. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, J. Renwick (Eds.). *Proceedings of the 7th International Conference on Music Perception and Cognition* (pp. 580-583). Adelaide: Causal Productions.

Monahan, C. B., & Hirsh, I. J. (1990). Studies in auditory timing: 2. Rhythm patterns. *Perception & Psychophysics*, *47*, 227-242.

Montgomery, C. R., Morris, R. D., Sevcik, R. A., & Clarkson, M. G. (2005). Auditory backward masking deficits in children with reading disabilities. *Brain and language*, *95*(3), 450–456.

Moore, B. C. J. (1982). *An introdction of the psychology of hearing*. London: Academic Press.

Moore, B. C. J., & Glasberg, B. R. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *The Journal of the Acoustical Society of America*, *80*(2), 479-483.

Moore, B. C. J., & Glasberg, B. R. (2004). A revised model of loudness perception applied to cochlear hearing loss. *Hearing research*, *188*(1-2), 70–88.

Moore, B. C. J., Peters, R. W., & Glasberg, B. R. (1985). Thresholds for the detection of inharmonicity in complex tones. *The Journal of the Acoustical Society of America*, *77*(5), 1861-1867.

Moore, R. S. (1981). Modulated intensity discrimination among musicians and nonmusicians. Presented at the Music Educators National Conference, Minneapolis, Minnesota.

Muluk, N. B., YalÁInkaya, F., & Keith, R. W. (2011). Random gap detection test and random gap detection test-expanded: Results in children with previous language delay in early childhood. *Auris Nasus Larynx*, *38*(1), 6–13.

Murphy, C. (1999). How far do tests of musical ability shed light on the nature of musical intelligence? *British Journal of Music Education*, *16*(01), 39–50.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, *104*(40), 15894-15898.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, cocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Developmental Psychology*, *40*(5), 665–681.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The Implication-Realization Model*. Chicago: University of Chicago Press.

Nettl, B. (2005). *The study of ethnomusicology. Thirty-one issues and concepts*. Champaign: University of Illinois Press.

Nettles, B., & Graft, R. (2002). *The chord scale theory and Jazz harmony*. Rottenburg N., Germany: Advance Music.

Newman, J., Rosenbach, J., Burns, K. L., Latimer, B. C., Matocha, H. R., & Rosenthalvogt, E. (1995). An experimental test of “The Mozart Effect”: Does listening to his music improve spatial ability? *Perceptual and Motor Skills*, *81*, 1379-1387.

van Noorden, L. (1975). *Temporal coherence in the perception of tone sequences*. Institute of Perception Research, Eindhoven.

Noto, A. (2008). *Unitxt (CD)*. Germany: Raster-Noton.

Ohm, G. S. (1843). On the definition of a tone with the associated theory of the siren and similar sound producing devices (translated by Lindsay, reprinted in Lindsay, 1973, pp 242-247). *Poggendorf’s Annalen der Physik und Chemie 59, 497ff.*

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., & Nishikawa, M. (2001). Functional anatomy of musical perception in musicians. *Cerebral Cortex*, *11*(8), 754-760.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 97–113.

Olivetti Belardinelli, M. (1995). Capacità specifiche nella elaborazione cognitiva di materiale musicale rilevate mediante i Wing Standardised Tests of Musical Intelligence [English translation]. *Comunicazioni Scientifiche di Psicologia Generale*, *14*, 77–93.

Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings* (Doctoral dissertation). The Ohio State University.

Olsen, K. N., Stevens, C. J., & Tardieu, J. (2010). Loudness change in response to dynamic acoustic intensity. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1631–1644.

Overy, K. (2000). Dyslexia, temporal processing and music: The potential of music as an early learning aid for dyslexic children. *Psychology of Music*, *28*(2), 218-229.

Overy, K. (2003). Dyslexia and music: From timing deficits to musical intervention. *Annals of the New York Academy of Sciences*, *999*, 497–505.

Overy, K., Nicolson, R. I., Fawcett, A. J., & Clarke, E. F. (2003). Dyslexia and music: Measuring musical timing skills. *Dyslexia*, *9*(1), 18-36.

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology*, *20*(4), 426-443.

Panion, H. (1989). *An experimental study of relationships among music listening abilities in formally and informally trained musicians* (Doctoral dissertation). Ohio State University, United States.

Papoušek, M., & Papoušek, H. (1981). Musical elements in the infant's vocalization: Their significance for communication, cognition, and creativity. *Advances in Infancy Research*, *1*, 163-224.

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience*, *29*(45), 14100–14107.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, *11*, 409-464.

Parncutt, R. (2003). Accents and expression in piano performance. *Perspektiven und Methoden einer Systemischen Musikwissenschaft*, 163–185.

Parncutt, R., & Levitin, D. J. (1999). Absolute pitch. *Grove Dictionary: Absolute Pitch*.

Parsons, L. M., & Fox, P. T. (1995). Neural basis of mental rotation. *Society for Neurosciences Abstracts*, *116*, 172-191.

Patel, A. (2003). Language, music, syntac and the brain. *Nature Neuroscience*, *6*(7), 674–680.

Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford, UK: Oxford University Press.

Patel, A. D., & Iversen, J. R. (2003). Acoustic and perceptual comparison of speech and drum sounds in the north indian tabla tradition: An empirical study of sound symbolism. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona* (pp. 925–928).

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*(6), 717–733.

Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception, 23*(5), 377–405.

Pedhazur, E. J., & Schemelkin, L. P. (1991). *Measurement, design, and analysis: An Integrated approach.* Hillsdale: Lawrence Erlbaum Associates.

Peretz, I. (2003). Brain specilisation for music: New evidence from congenital amusia. *The cognitive neuroscience of music*. In I. Peretz & R. Zatorre (Eds). (pp. 192–203). Oxford, UK: Oxford University Press.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. [Review]. *Trends in Cognitive Sciences*, 7(8), 362-367.

Peretz, I., & Zatorre, R. J. (2005). Brain Organization for Music Processing. *Annual Review of Psychology*, *56*(1), 89-114.

Peretz, I., Champod, A. S., & Hyde, K. L. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*(1), 58-75.

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., Paquette, S., et al. (2008). On-line identification of congenital amusia. *Music Perception*, *25*(4), 331–343.

Perron, M. (1994). *Checking tempo stability of MIDI sequencers*. Presented at 97th Convention of the Audio Engineering Society, San Francisco.

Pfordresher, P. Q. (2005). Auditory feedback in music performance: The role of melodic structure and musical skill. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1331–1345.

Pickles, J. (1988). *An introduction to the physiology of hearing*. London: Academic Press.

Plack, C. J. (2005). *The sense of hearing*. Mahwah: Lawrence Erlbaum Associations.

Plomin, R., & Thompson, L. A. (1993). Genetics and high cognitive ability. In K. Ackrill (Ed.). *The origins and development of high ability: Proceedings of Ciba symposium 178* (pp. 67-84). London: Wiley.

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *The Journal of the Acoustical Society of America*, *38*(4), 548–560.

Polderman, T. J. C., Stins, J. F., Posthuma, D., Gosso, M. F., Verhulst, F. C., & Boomsma, D. I. (2006). The phenotypic and genotypic relation between working memory speed and capacity. *Intelligence*, *34*(6), 549-560.

Pontin, J. (2007). Artificial intelligence, with help from the humans. *The New York Times*. Retrieved from http://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html

Povel, D. J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(1), 3-18.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891.

Prinz, W. (1990). A common-coding approach to perception and action. *Relationships between perception and action: Current approaches* (pp. 167-201). Berlin: Springer.

Purwins, H., Herrera, P., Grachten, M., Hazan, A., Marxer, R., & Serra, X. (2008a). Computational models of music perception and cognition I: The perceptual and cognitive processing chain. *Physics of Life Reviews, 5*(3), 151–168.

Purwins, H., Grachten, M., Herrera, P., Hazan, A., Marxer, R., & Serra, X. (2008bb). Computational models of music perception and cognition II: Domain-specific music processing. *Physics of Life Reviews, 5*(3), 169–182.

Rahne, T., Rasinski, C., & Neumann, K. (2010). Measuring timbre discrimination with cross-faded synthetic tones. *Journal of Neuroscience Methods, 189*(2), 176–179.

Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception*, *24*(1), 37–48.

Rasch, R., & Plomp, R. (1999). The perception of musical tones. In D. Deutsch (ed.). *The Psychology of Music* (2nd ed., pp. 89–112). New York, NY: Academic Press.

Rauscher, F. H., & Zupan, M. A. (2000). Classroom keyboard instruction improves kindergarten children's spatial-temporal performance: A field experiment. *Early Childhood Research Quarterly*, *15*(2), 215-218.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature*, *365*, 611.

Raven, J. C. (1948). The comparative assessment of intellectual ability. *British Journal of Psychology*, *39*(1), 12-19.

Raven, J. C. (2000). The Raven's Progressive Matrices- Change and Stability over Culture and Time. *Cognitive Psychology*, *41*, 1-48.

Raven, J. C., Court, J. H., & Raven, J. (1993). *Advanced Progressive Matrices (1994 Edition)*. Oxford, UK: Oxford Psychologists Press Ltd.

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, *1*(3), 86–89.

Repp, B. H. (2000). Compensation for subliminal timing perturbations in perceptual-motor synchronization. *Psychological Research*, *63*(2), 106–128.

Repp, B. H., & Doggett, R. (2007). Tapping to a very slow beat: a comparison of musicians and nonmusicians. *Music Perception*, *24*(4), 367–376.

Restle, F. (1973). Coding of nonsense vs. the detection of patterns. *Memory & Cognition, 1*, 499-502.

Revelle, W., & Zinbarg, R. (2009). Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145-154.

Révész, G. (1953). *Introduction to the psychology of music. Translated by G.I.C. de Courcy*. London, UK: Longmans, Green and Co.

Riley, D. A., McKee, J. P., Bell, D. D., & Schwartz, C. R. (1967). Auditory discrimination in children: The effect of relative and absolute instructions on retention and transfer. *Journal of Experimental Psychology, 73,* 581-588.

Risset, J. C., & Wessel, D. L. (1999). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.). *The Psychology of Music* (pp. 113-169). California: Academic Press.

Ropshkow, O. (2012). Cochlea Cross-section. Retrieved August 21, 2012, from http://commons.wikimedia.org/wiki/File:Cochlea-crosssection.png

Ross, D., Choi, J., & Purves, D. (2007). Musical intervals in speech. *Proceedings of the National Academy of Sciences*, *104*(23), 9852.

Rousseau, L., & Rousseau, R. (1996). Stop-reaction time and the internal clock. *Attention, Perception, & Psychophysics*, *58*(3), 434–448.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation Analysis in Social Psychology: Current Practices and New Recommendations. *Social and Personality Psychology Compass*, *5*(6), 359–371.

Sacks, O *(2007)*. *"*Musicophilia – Amusia. *" Retrieved March 24, 2011, from http://www.goodreads.com/videos/show/2921-oliver-sacks-musicophilia-amusia, Mar 24, 2011*

Salamé, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology Section A*, *41*(1), 107-122.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology*, *88*(6), 1068–1081.

Scheirer, E. D. (1998). *Music perception systems* (Unpublished Doctoral Dissertation). USA: Massachusetts Institute of Technology.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*(8), 511-514.

Schellenberg, E. G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science*, *14*(6), 317-320.

Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology*, *98*(2), 457.

Schellenberg, E. G., & Moreno, S. (2009). Music lessons, pitch processing, and *g*. *Psychology of Music*, *38*(2), 209-221.

Schellenberg, E. G., & Peretz, I. (2008). Music, language and cognition: unresolved issues. *Trends in Cognitive Sciences*, *12*(2), 44–45.

Schellenberg, E. G., Adachi, M., Purdy, K. T., & McKinnon, M. C. (2002). Expectancy in melody: Tests of children and adults. *Journal of Experimental Psychology: General*, *131*(4), 511–537.

Schellenberg, E. G., Nakata, T., Hunter, P. G., & Tamoto, S. (2007). Exposure to music and cognitive performance: Tests of children and audlts. *Psychology of Music*, *35*(1), 5-19.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, *5*(7), 688–694.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349

Schouten, J. F. (1938). The perception of subjective tones. *Nederlandse Akademie van Wetenschappen Procs.*, *41*, 1086–1093.

Schouten, J. F., Ritsma, R. J., & Lopes Cardozo, B. (1962). Pitch of the Residue. *The Journal of the Acoustical Society of America*, *34*(8), 1418–1424.

Schulze, H. H. (1989). Categorical perception of rhythmic patterns. *Psychological Research*, *51*(1), 10–15.

Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Maintenance and manipulation of tonal and atonal sequences in non-musician and musician listeners. *Music Perception*, *29*(3), 255–267.

Schweizer, K., Goldhammer, F., Rauch, W., & Moosbrugger, H. (2007). On the validity of Raven's matrices test: Does spatial ability contribute to performance? *Personality and Individual Differences*, *43*(8), 1998–2010.

Seargent, D. (1968). Experimental investigation of absolute pitch. *Journal of Research in Music Education*, *17*(1), 135-143.

Seashore, C., E. (1919a). *Manual of instructions and interpretations for measures of musical talent*. New York City: Educational Department, Columbia Graphophone Company.

Seashore, C. E. (1919b). *The psychology of musical talent*. Boston: Silver, Burdett, and Company.

Seashore, C. E. (1938). *Psychology of Music*. New York: McGraw-Hill Book Company.

Seashore, C., E. (1967). *Psychology of Music*. New York: Dover Publications.

Seashore, C. E., Lewis, D., & Saetveit, J. G. (1960). *A second revision of the Manual of Instructions and Interpretations for the Seashore Measures of Musical Talents (1939 Revision)*. The Psychological Corporation New York.

Semal, C., Demany, L., Ueda, K., & Hallé, P. A. (1996). Speech versus nonspeech in pitch memory. *The Journal of the Acoustical Society of America*, *100*(2), 1132–1140.

Serafine, M. L. (1988). *Music as cognition: The development of thought in sound*. New York: Columbia University Press.

Shaw, G. (2000). *Keeping Mozart in mind*. San Diego: Academic Press.

Shepard, R., N. (1964). Circularity in judgements of relative pitch. *Journal of the Acoustical Society of America*, *36*(12), 2346-53.

Shepard, R. N. (1982). Structural representations of musical pitch. In D.Deutsch (Ed.). *The Psychology of Music* (1st ed., pp. 343-390). New York: Academic Press.

Shibata, D. (2001). *Brains of deaf people "hear" music*. Presented at the 87th Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA), Chicago.

Shuter-Dyson, R. (1969). Some problems in psychology of musical ability. *Journal of Research in Music Education*, *17*(1), 90-93.

Shuter-Dyson, R. (1982). *Musical Ability. The Psychology of Music edited by Diana Deutsch* (United Kingdom.). London: Academic Press.

Shuter-Dyson, R. (1999). Musical Ability. In D. Deutsch (Ed.). *The Psychology of Music* (2nd ed., pp. 627-651). California, USA: Academic Press.

Shuter-Dyson, R., & Gabriel, C. (1981). *The psychology of musical ability* (2nd ed.). London, UK: Methuen.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's Alpha. *Psychometrika*, *74*(1), 107–120.

Sims, W. L., & Nolker, D. B. (2002). Individual differences in music listening responses of kindergarten children. *Journal of Research in Music Education*, *50*(4), 292–300.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*(8), 675–681.

Sloboda, J. A. (1985a). *The musical mind. The cognitive psychology of music*. Oxford: Oxford University Press.

Sloboda, J. A. (1985b). An exceptional musical memory. *Music Perception*, *3*(2), 155-170.

Sloboda, J. A. (2000). Individual differences in music performance. *Trends in cognitive sciences*, *4*(10), 397–403.

Sloboda, J. A., & Howe, M. J. A. (1991). Biographical precursors of musical excellence: An interview study. *Psychology of Music*, *19*(1), 3-21.

Sloboda, J. A., Davidson, J.W., & Howe, M. J. A. (1999). Is everyone musical? In P. Murphy (Ed.). *Learners, learning & assessment* (pp. 46-57). London: Paul Chapman Publishing.

Slominsky, N. (1930). Absolute pitch. *American Mercury*, *21*, 244-247.

Small, A. M., Jr., & Campbell, R. A. (1962). Temporal differential sensitivity for auditory stimuli. *American Journal of Psychology*, *75*, 401-410.

Smoliar, S. W. (1991). Review of the analysis and cognition of basic melodic Structures: The Implication- Realization model by Eugene Narmour. *In Theory Only*, *12*(1), 43-56.

Snell, K.B., Mapes, F.M., Hickman, E.D., and Frisina, D.R. (2002). Word recognition in competing babble and the effects of ages, temporal processing, and absolute sensitivity. *The Journal of the Acoustical Society of America*, *112*, 720-727.

SNS. (1994, July 6). Music synthesis approaches sound quality of real instruments. *Stanford News Service*. Retrieved from http://news.stanford.edu/pr/94/940607Arc4222.html

Stevens,C., & Tolbert, E. (2009). Universals in music processing. *The Oxford handbook of music psychology*. In S. Hallam, I. Corss, and M. Thaut (Ed.). (pp. 14–23). Oxford, New York: Oxford University Press.

Snyder, J. S., Hannon, E. E., Large, E. W., & Christiansen, M. H. (2006). Synchronization and continuation tapping to complex meters. *Music Perception*, *24*(2), 135–146.

Snyder, J. S. , & Krumhansl, C. L. (2001). Tapping to ragtime: Cues to pulse finding. *Music Perception*, *18*(4), 455–489.

Spearman, C. (1904). "General Intelligence," Objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292.

Spearman, C. (1923). *The nature of "intelligence" and the principles of cognition*. London: Macmillan.

Stamou, L., Schmidt, C. P., & Humphreys, J. T. (2010). Standardization of the Gordon Primary Measures of Music Audiation in Greece. *Journal of Research in Music Education*, *58*(1), 75-89.

Steiger, J.H. (1980). Tests of comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245–251.

Steinberg. (2009). Steinberg Products. Retrieved August 24, 2009, from http://www.steinberg.net/

Stern, M. K., & Johnson, J. H. (2010). Just Noticeable Difference (JND). In I. Weiner and W.E. Craighead (Eds.). In I. B. Weiner & W. E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology* (pp. 886 – 889). Hoboken, NJ: John Wiley & Sons.

Sternberg, S., Knoll, R. L., & Zukofsky, P. (1982). Timing by skilled musicians. In D. Deutsch (Ed.). *The psychology of music* (pp. 181-239). New York: Academic Press.

Stough, C., Kerkin, B., Bates, T., & Mangan, G. (1994). Music and spatial IQ. *Personality and Individual Differences*, *17*(5), 695–695.

Strait, D., L, Chan, K., Ashley, R., & Kraus, N. (in press). Specialization among the specialized: Auditory brainstem function is tuned in to timbre. *Cortex (2011)*. doi:10.1016/j.cortex.2011.03.015

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, *261*(1-2), 22-29.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, *80*(1), 99–103.

Stumpf, C. (1883). *Tonpsychologie*. Leipzig: Hirzel.

Stumpf, C. (2012). *The origins of music. Translated and edited by D. Trippett*. Oxford: Oxford University Press.

Surprenant, A. M., & Watson, C. S. (2001). Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *The Journal of the Acoustical Society of America*, *110*(4), 2085.

Temperley, D. (2008). A Probabilistic Model of Melody Perception. Cognitive Science: *A Multidisciplinary Journal, 32*(2), 418–444.

Tervaniemi, M., Ilvonen, T., Karma, K., Alho, K., & Näätänen, R. (1997). The musical brain: brain waves reveal the neurophysiological basis of musicality in human subjects. *Neuroscience letters*, *226*(1), 1–4.

Thaut, M. H. (2005). *Rhythm, music and the brain*. New York: Taylor & Francis Group.

Thompson, B. (2003). *Score reliability: Contemporary thinking on reliability issues*. London, UK: Sage Publication.

Thompson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal of Physiology-Paris*, *102*(1-3), 120–129.

Thompson, W. F., Schellenberg, E. G., & Hussain, G. (2001). Arousal, mood and the Mozart effect. *Psychological Science*, *12*, 248-251.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, *4*(1), 46-64.

Tierney, A. T., Bergeson-Dana, T. R., & Pisoni, D. B. (2008). Effects of early musical experience on auditory sequence memory. *Empirical musicology review: EMR*, *3*(4), 178.

Torgeson, J. K. (1991). Learning disabilities: An historical and conceptual overview. In B.K. Wong. (Ed.). *Learning About Learning Disabilities* (pp. 3-37). San Diego: Academic Press.

Toulson, E. R., Cuny-Cringy, C., Robinson, P., & Richardson, P. G. M. (2008). The perception and importance of drum tuning in live performance and music production. *Proceedings of The Art of Record Production Conference, Lowell, Massachusetts*.

Trainor, L. J. (2005). Are there critical periods for musical development? *Developmental Psychobiology*, *46*(3), 262–278.

Trehub, S. E. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences*, *930*(1), 1–16.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Blackwell Publishing on behalf of the Society for Research in Child Development*, *55*(3), 821–830.

Trimmer, C. G., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion*, *8*(6), 838-849.

Tyler, R.S., Summerfield, Q., Wood, E.J., and Fernandes, M.A. (1982). Psychoacoustic and phonetic temporal processing in normal and hearing impaired listeners. *The Journal of the Acoustical Society of America, 72*, 740-752.

Umemoto, T. (1990). The psychological structure of music. *Music Perception*, *8*(2), 115-128.

UN. (2011). United Nations demographic year book, 2009-2010. United Nations.

UNESCO. (2012). United Nations Educational, Scientific and Cultural Organization: 2011 Demographic and economic data (Education). Retrieved September 24, 2012, from http://www.unesco.org/

Underwood, G. (1996). *Implicit Cognition*. Oxford: Oxford University Press.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498-505.

Vaughn, K. (2000). Music and mathematics: Modest support for the oft-claimed relationship. *Journal of Aesthetic Education*, *34*(3/4), 149–166.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *The Journal of the Acoustical Society of America*, *66*(5), 1364–1380.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., et al. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30*(4), 1414–1432.

Vink, A. (2001). Music and emotion-Living apart together: A relationship between music psychology and music therapy. *Nordic Journal of Music Therapy*, *10*(2), 144–158.

Vitz, P.C. & Todd, T.C. (1969). A coded element model of the perceptual processing of sequential stimuli. *Psychological Review, 76*, 433-449.

VSL. (2002). Vienna Symphonic Library. Retrieved October 28, 2011, from http://vsl.co.at/

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010a). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, *20*(3), 188–196.

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010b). Corrigendum to "The Musical Ear Test, a new reliable test for measuring musical competence." *Learning and Individual Differences*, *20*, 705-705.

Wallin, N. L., Merker, B., & Brown, S. (1999). *The origins of music*. Cambridge: MIT Press.

Watson, B. U., & Miller, T. K. (1993). Auditory perception, phonological processing, and reading ability/disability. *Journal of Speech and Hearing Research*, *36*(4), 850.

Watson, C. S., Jensen, J. K., Foyle, D. C., Leek, M. R., & Goldgar, D. E. (1982b). Performance of 146 normal adult listeners on a battery of auditory discrimination tests. *The Journal of the Acoustical Society of America*, *71*(S1), S73–S73.

Watson, C. S., Johnson, D. M., Lehman, J. R., Kelly, W. J., & Jensen, J. K. (1982a). An auditory discrimination test battery. *The Journal of the Acoustical Society of America*, *71*(S1), S73–S73.

Wechsler, D. L. (1991). *Wechsler Intelligence Scale for Children (WISC)* (3rd ed.). San Antonio: The Psychological Corporation.

Wechsler, D. L. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI) Manual*. The Psychological Corporation New York.

Wells, J. J. (2006). *Real-Time Spectral Modelling of Audio for Creative Sound Transformation*. University of York, York, UK.

Wertheim, N., & Botez, M. I. (1961). Receptive amusia: A clinical analysis. *Brain*, *84*(1), 19–30.

Wessel, D. L. (1991). *Wechsler Intelligence Scale for Children (WISC)* (3rd ed.). San Antonio: The Psychological Corporation.

Whitwell, G. (1999). The importance of prenatal sound & music. *Journal of Prenatal and Perinatal Psychology and Health*, *13*(3-4), 255-262.

Wightman, F., Allen, P., Dolan, T., Kister, D., & Jamiescon, D. (1989). Temporal resolution in children. *Child Development*, *60*, 611-624.

Wigram, T. (1995). The psychological and physiological effects of low frequency sound and music. *Music Therapy Perspectives*, *13*(1), 16–23.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American psychologist*, *54*(8), 594-604.

Williamon, A., & Valentine, E. (2002). The role of retrieval structures in memorizing music. *Cognitive Psychology*, *44*, 1-32.

Williams, D.B. (1980). A study of tonal strength and its influence on melodic memory. In W.V. May (Ed.), *Proceedings of the Research Symposium on the Psychology and Acoustics of Music.* University of Kansas, Dept. of Music Education and Music Therapy, 45-61.

Williams, H. M., Sievers, C. H., & Hattwick, M. S. (1932). The measurement of musical development. *University of Iowa Studies in Child Welfare*, *7*(1), 187–191.

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2006). Music in working memory? Examining the effect of pitch proximity on the recall performance of nonmusicians. *9th International Conference on Music Perception and Cognition* (pp. 1581–1590).

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Memory & Cognition*, *38*(2), 163–175.

Wing, H. D. (1948). *Tests of musical ability and appreciation. An investigation into the measurement, distribution, and development of musical capacity.* The British Journal of Psychology. Monograph Supplements. London: Cambridge University Press.

Wing, H. D. (1962). A Revision of the "Wing Musical Aptitude Test". *Journal of Research in Music Education*, *10*(1), 39-46.

Wing, H. D. (1968). *Tests of Musical Ability and Appreciation: An investigation into the measurement, distribution, and development of musical capacity*. British Journal of Psychology. Monograph supplements (Second.). London: Cambridge University Press.

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *PNAS*, *106*(7), 2468-2471.

Wolpert, R., S. (1990). Recognition of melody, harmonic accompaniment, and instrumentation: Musicians vs. nonmusicians. *Music Perception*, *8*(1), 95–106.

Wolpert, R., S. (2000). Attention to key in a non-directed music listening task: Musicians vs. Nonmusicians. *Music Perception*, *18*(2), 225–230.

Won, J. H., Drennan, W. R., & Rubinstein, J. T. (2007). Spectral-Ripple Resolution correlates with speech reception in noise in cochlear implant users. *Journal of the Association for Research in Otolaryngology, 8(*3), 384–392.

Won, J. H., Drennan, W. R., Kang, R. S., & Rubinstein, J. T. (2010). Psychoacoustic abilities associated with music perception in cochlear implant users. *Ear and Hearing*, *31*(6), 796–805.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*.

Woodrow, H. (1951). Time perception. In S.S. Stevens (Ed.). *Handbook of Experimental Psychology* (pp. 1224-1236). Oxford, UK: John Wiley & Sons.

Wyatt, R. (1939). A note on the use of “omnibus” training to validate Seashore’s “Capacity” hypothesis. *The American Journal of Psychology*, *52*(4), 638-640.

Xenakis, I. (1992). *Formalized music: Thought and mathematics in composition* (2nd ed.). Hillsdale, New Jersey: Pendragon Press.

Young, R. W. (1939). Terminology for logarithmic frequency units. *The Journal of the Acoustical Society of America*, *11*(1), 134–139.

Young, W.T. (1979). The Bentley “Measures of Musical Abilities”: A congruent validity report.

Yost, W. A., Hill, R., & Perez-Falcon, T. (1978). Pitch and pitch discrimination of broadband signals with rippled power spectra. *The Journal of the Acoustical Society of America*, *63*, 1166–1173.

Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of cognitive neuroscience*, *21*(8), 1488–1498.

Zeng, F. G., Kong, Y. Y., Michalewski, H. J., & Starr, A. (2005). Perceptual consequences of disrupted auditory nerve activity. *Journal of Neurophysiology*, *93*(6), 3050–3063.

Zentner, M. (in progress). *Music-Mindedness: The Musical Personality Dimension*. Unpublished Manuscript, Department of Psychology, University of York.

Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences*, *107*(13), 5768.

Zimmer, E. Z., Fifer, W. P., Kim, Y., Rey, H. R., Chao, C. R., & Michael, M. (1993). Response of the premature fetus to stimulation by speech sounds. *Early Human Development*, *33*(3), 207–215.

Zimmerman, M. P. (2011). Musical characteristics of children. *Visions of Research in Music Education*, *17*.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and Mcdonald's ωH: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123–133.

Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2004). Grouping of sequential sounds-an event-related potential study comparing musicians and nonmusicians. *Journal of cognitive Neuroscience*, *16*(2), 331–338.