

Nature, nurture, and music –
A population-based twin study
of music perception

Erik Juhana Seesjärvi
Master's Thesis
Psychology
Institute of Behavioural Sciences
August, 2014
Supervisors:
Teppo Särkämö and
Eero Vuoksima

HELSINGIN YLIOPISTO - HELSINGFORS UNIVERSITET - UNIVERSITY OF HELSINKI

Tiedekunta - Fakultet - Faculty Faculty of Behavioral Sciences	Laitos - Institution - Department Institute of Behavioral Sciences	
Tekijä - Författare - Author Erik Juhana Seesjärvi		
Työn nimi - Arbetets titel - Title Nature, nurture, and music – A population-based twin study of music perception		
Oppiaine - Läroämne - Subject Psychology		
Työn laji ja ohjaaja(t) - Arbetets art och handledare - Level and instructor Master's Thesis, instructors Teppo Särkämö and Eero Vuoksima	Aika - Datum - Month and year August 2014	Sivumäärä - Sidoantal - Number of pages 46 + 1
<p>Tiivistelmä - Referat - Abstract</p> <p>The aim of the study was to estimate the proportional effects of genetic and environmental factors on individual differences in music perception. Previous research has demonstrated that genetic effects explain a substantial amount of these differences, and that common environmental effects have been low or have not been examined because of methodological constraints. However, in defining the accuracy of music perceptual skills, most previous studies have used somewhat simple auditory stimuli this choice limiting their generalizability. In order to reliably determine the proportional effects of genetic and environmental factors on perception of more complex musical stimuli, the present study utilized unfamiliar melodies, the classical twin design, and structural equation modelling.</p> <p>The participants of the current study were 384 twins from the longitudinal <i>FinnTwin16</i> study, aged 32–38 years. They performed an online music perception test that consists of two subtests with pitch-related tasks (Scale and Out-of-key subtests) and one subtest with a time-related task (Off-beat subtest). The test includes 30 unfamiliar melodies. The participants also answered a short questionnaire regarding their educational background and self-assessed musical ability.</p> <p>The results showed that genetic factors explained about half of the interindividual variance in ability to detect pitch changes in repeated melodies (Scale subtest), whereas common environmental factors had only marginal effect on this ability. Furthermore, genetic factors explained about a quarter or less of the interindividual variance in ability to detect a timing delay that was disrupting the meter or rhythm of melody (Off-beat subtest). There were no common environmental effects. In contrast, a different pattern of results was obtained for ability to detect a tone that violated the established scale expectations (Out-of-key subtest) in which there were only marginal genetic effects whereas common environmental factors explained over half of the interindividual variance. Together with the previous research, these results show that genetic factors explain a significant portion of the individual differences in music perception tasks especially when the task is sufficiently demanding cognitively. However, when the task demands explicit or implicit knowledge of musical scales, genetic effects disappear and the individual differences can mostly be explained by common environmental factors.</p>		
Avainsanat - Nyckelord - Keywords Music perception, Twin study, Quantitative genetics, Structural equation modelling		
Säilytyspaikka - Förvaringsställe - Where deposited Helsinki University Library, City centre campus library, Behavioural sciences / Minerva		
Muita tietoja - Övriga uppgifter - Additional information		

HELSINGIN YLIOPISTO - HELSINGFORS UNIVERSITET - UNIVERSITY OF HELSINKI

Tiedekunta - Fakultet - Faculty Käyttäytymistieteellinen		Laitos - Institution - Department Käyttäytymistieteiden laitos	
Tekijä - Författare - Author Erik Juhana Seesjärvi			
Työn nimi - Arbetets titel - Title Geenit, ympäristö ja musiikki – Väestöpohjainen musiikin havaitsemisen kaksostutkimus			
Oppiaine - Läroämne - Subject Psykologia			
Työn laji ja ohjaaja(t) - Arbetets art och handledare - Level and instructor Pro gradu -tutkielma, ohjaajat Teppo Särkämö ja Eero Vuoksimaa		Aika - Datum - Month and year Elokuu 2014	Sivumäärä - Sidoantal - Number of pages 46 + 1
<p>Tiivistelmä - Referat - Abstract</p> <p>Tutkimuksen tarkoituksena oli selvittää, kuinka suurina ovat perintötekijöiden ja ympäristötekijöiden suhteelliset selitysosuudet yksilöiden välisistä eroista musiikin havaitsemisessa. Aiempien tutkimusten mukaan perintötekijät selittävät merkittävän osan näistä eroista, kun taas jaetun ympäristön selitysosuus on joko ollut pieni tai sitä ei ole voitu metodologisista syistä tutkia. Useimmissa musiikin havaitsemisen tutkimuksissa on käytetty varsin yksinkertaisia ärsykeitä kuten äänipareja, mikä rajoittaa tulosten yleistettävyyttä. Tässä tutkimuksessa käytettiin kokonaisia musiikkimelodioita, kaksosasetelmaa ja rakenneyhtälömallinnusta, mikä mahdollistaa monimutkaisempien musiikillisten ärsykkeiden havaitsemiseen vaikuttavien perintö- ja ympäristötekijöiden laajan ja luotettavan tutkimisen.</p> <p>Tutkimukseen osallistui 384 kaksosta, jotka olivat mukana <i>Nuorten Kaksosten Terveystutkimuksessa</i> ja jotka olivat tutkimuksen suoritusvaiheella 32–38 vuoden ikäisiä. Koehenkilöt suorittivat internetin välityksellä musiikkitestin, joka sisältää kaksi sävelkorkeuspoikkeamien tunnistamista mittaavaa osatehtävää (Scale ja Off-beat -osatehtävät) ja yhden ajallisten poikkeamien tunnistamista mittaavan osatehtävän (Out-of-key -osatehtävä). Testi sisältää 30 melodiaa, jotka ovat kuulijalle entuudestaan tuntemattomia. Lisäksi osallistujat täyttivät lyhyen kyselyn koskien heidän koulutustaustansa ja itsearvioitaan musiikin havaitsemisesta.</p> <p>Tulosten perusteella perintötekijät selittivät noin puolet yksilöiden välisistä eroista kyvyssä havaita sävelkorkeusmuutoksia toistuvissa melodioissa (Scale-osatehtävä). Yhteisten ympäristötekijöiden vaikutus oli sen sijaan erittäin vähäinen. Perintötekijät selittivät arviolta alle neljäsosan yksilöiden välisistä eroista kyvyssä havaita melodian rytmiä tai tahtilajia rikkova viive (Off-beat -osatehtävä). Yhteiset ympäristötekijät eivät vaikuttaneet tähän kykyyn. Sen sijaan perintötekijät selittivät hyvin vähän yksilöiden välisistä eroista kyvyssä havaita melodian sävellajioletuksia rikkova ääni (Out-of-key -osatehtävä). Tästä kyvystä yhteinen ympäristö selitti yli puolet. Näiden tulosten ja aikaisemman tutkimuksen perusteella vaikuttaa siltä, että perintötekijät selittävät merkittävän osan yksilöiden välisistä eroista musiikin havaitsemisessa etenkin silloin, kun tehtävä on kognitiivisesti riittävän kuormittava. Kuitenkin silloin, kun tehtävä vaatii sävellajioletusten eksplisiittistä tai implisiittistä hallintaa, perintötekijät eivät selitä yksilöiden väliset eroja, vaan ympäristötekijöiden rooli näiden erojen selittäjinä nousee keskeiseksi.</p>			
Avainsanat - Nyckelord - Keywords Musiikin havaitseminen, kaksostutkimus, kvantitatiivinen genetiikka, rakenneyhtälömallinnus			
Säilytyspaikka - Förvaringsställe - Where deposited Helsingin yliopiston kirjasto, Keskustakampuksen kirjasto, Käyttäytymistieteet / Minerva			
Muita tietoja - Övriga uppgifter - Additional information			

Foreword

This study was completed in a multidisciplinary collaboration project between the *Cognitive Brain Research Unit* (CBRU), Institute of Behavioural Sciences, University of Helsinki; the *Twin Research Group*, Hjelt Institute (HI), University of Helsinki; the *Finnish Centre of Excellence in Interdisciplinary Music Research* (FCoEIMR), Universities of Jyväskylä and Helsinki; and the *International Laboratory for Brain, Music and Sound Research* (BRAMS), University of Montreal and McGill University, Canada. The project was planned and supervised by Prof. Mari Tervaniemi (CBRU & FCoEIMR), Acad. Prof. Jaakko Kaprio (HI), Dr. Teppo Särkämö (CBRU & FCoEIMR), Dr. Eero Vuoksimaa (HI), and Prof. Isabelle Peretz (BRAMS). Other researchers involved included MHS Anja Häppölä (Research coordinator, HI) and Dr. Kauko Heikkilä (Project data manager, HI). The project was financially supported by Paulo Foundation. Participants of this study were recruited from the longitudinal *FinnTwin16* study, which is conducted in collaboration with Hjelt Institute, University of Helsinki, and Indiana University, Bloomington, Indiana, USA. Data collection of the FinnTwin16 study has been funded by the Academy of Finland and the National Institutes of Health, USA.

First of all, I would like to thank Professor Mari Tervaniemi for allowing me to gain valuable experience as a research assistant in this project during the last two years. The project has been personally very rewarding, as it has allowed me to combine my background as a professional musician with my psychology studies. I am also grateful to Academy Professor Jaakko Kaprio for allowing me to do my Master's Thesis with the FinnTwin16 participants. I would also like to thank my brilliant supervisors, Dr. Teppo Särkämö and Dr. Eero Vuoksimaa, for their excellent guidance. Thank you Teppo for your comprehensive, mercilessly accurate but encouraging comments, which have been invaluable. Thank you Eero for sharing ridiculous amounts of your time and expertise when introducing me to statistically advanced method of structural equation modelling. I would also like to thank Dr. Kauko Heikkilä and MHS Anja Häppölä for their assistance with the twin data.

I want to thank Professors Mari Tervaniemi, Jaakko Kaprio and Isabelle Peretz for taking their precious time to read the manuscript and to provide many important comments. Also the members of the thesis seminar group of Prof. Mari Tervaniemi gave insightful comments on the work. Furthermore, I would like to thank Dr. Mícheál Briody for proofreading parts of the manuscript and for providing fruitful linguistic comments.

I express my gratitude to my parents, my musical sisters Anna and Iris, their companions Shane and Demian, and other members in my extended family for their support and encouragement during my life. Also thanks to all my friends, study mates and fellow musicians about inspirational conversations about music and musicality.

Last but not least, I want to thank my beloved wife Maria. Thank you for your constant support during my studies and invaluable advice. Like one famous psychologist once stated, wives are an invaluable resource for their husbands.

Erik Seesjärvi

Contents

1. Introduction.....	1
1.1. Acoustic structure of music	2
1.2. Evolutionary and developmental background of music	4
1.3. Music perception in the adult human brain	7
1.4. Quantitative genetics	9
1.5. The classical twin design	11
1.6. Twin and family studies of music perception	13
1.7. Research questions	18
2. Methods	19
2.1. Subjects	19
2.2. Assessment procedure	20
2.3. Data analysis	22
3. Results	25
3.1. Descriptive results	25
3.2. Model fitting results	28
4. Discussion.....	31
4.1. Genetic and environmental effects on pitch perception	31
4.2. Genetic and environmental effects on rhythm perception	34
4.3. Limitations of the study	36
4.4. Summary	37
References	39
Appendix I. Additional model-comparison test statistics	47

1. Introduction

Music is an integral part of human life and is present in all cultures throughout the world. It occurs in almost all social gatherings, has the capacity to unite people in emotions varying from joy to grief, and can enhance group cohesion, as when fighting a common enemy (Trainor & Hannon, 2012). Some researchers see music as a product of human culture or as a human invention (Patel, 2010; Pinker, 1999) while others (Mithen, 2005) regard it as an evolutionary adaptation. In any case, we seem to be adept at perceiving music from an early age: before the age of twelve months, infants are able to recognise a familiar tune performed at different pitches (Chang & Trehub, 1977), and the ability to recognise a constant pulse, an essential element in music, seems to be innate (Winkler, Haden, Ladinig, Sziller, & Honing, 2009).

One of the classical questions in psychology has always been whether individual differences result from genetic or environmental factors, which has come to be known as so called *nature or nurture* debate. According to Turkheimer (2000), all behavioural traits are inheritable and the effect of being raised in a particular family is smaller than the effect of genes. These claims seem to be supported by numerous twin and family studies, but for some more abstract and complex traits, such as religiousness and conservatism, environmental factors seem to play a larger role than genetic factors (Abrahamson, Baker, & Caspi, 2002; Boomsma, Busjahn, & Peltonen, 2002; McGue & Bouchard Jr, 1998; Vinkhuyzen, Van der Sluis, Posthuma, & Boomsma, 2009). So far there have been only a few population-based studies of the genetic basis of music perception (Drayna, Manichaikul, de Lange, Snieder, & Spector, 2001; Mosing, Madison, Pedersen, Kuja-Halkola, & Ullén, 2014; Ullén, Mosing, Holm, Eriksson, & Madison, 2014), which have suggested that genetic differences seem to explain majority of individual differences. Given the pivotal role of music in society, there is a clear demand for further research in this interesting area.

To my knowledge, this thesis is the first population-based study of genetic and environmental effects on individual differences in pitch and rhythm perception in unfamiliar melodies. Thus, it provides new information on music perception. Drayna et al. (2001) used familiar melodies but studied only pitch-related tasks, whereas Ullén et al.

(2014) and Mosing et al. (2014) focused on more musically and acoustically simple stimuli, such as isochronous tone sequences, rhythm sequences of constant pitch, and sinusoidal tone pairs. It has been argued that speech and music perception are closely related, especially via rhythm (Hausen, Torppa, Salmela, Vainio, & Särkämö, 2013). Therefore, the results of this study are also likely to have implications beyond music perception. The *classical twin design* is used, where groups of genetically identical and non-identical twins are compared.

1.1. Acoustic structure of music

Music can be defined as a communication system that develops under a complex set of genetic predispositions and environmental input (Trainor & Hannon, 2012). Based on their physical qualities, *rhythm* and *pitch* can be defined as the two primary dimensions of music (Krumhansl, 2000). For each dimension, there is an underlying structure that affects how music is perceived. For rhythm, this structure is the regular *meter* or *periodic pulse*. For pitch, it is the *musical key* or *scale*. These structures are hierarchical and rest on general capacities and constraints of the human nervous system (Trainor & Hannon, 2012). Although they appear to be essentially universal, there are some culture-dependent differences, such as the existence of different musical keys in different cultures (Patel & Demorest, 2012).

Rhythm consists of events (sounds) in time. If the time between events is less than 100 ms, listeners tend to hear a single, continuous event, whereas events separated by more than 1500 ms tend to be perceived as disconnected (Fraisse, 1982). Rhythm consists of several components, such as rhythmic pattern, tempo, and meter (Honing, 2012). Rhythmic pattern is a pattern of durations that can be represented discretely, for example with note values. Tempo can be described as the impression of the speed of the sounding pattern. Meter refers to the constant level of rhythm expressed by regularly occurring beats. Most music includes a constant meter, although sometimes it is more obvious and sometimes less noticeable. Different meters are expressed by time signatures. If a tone disrupts the regular beat that has been established it is referred as being *off-beat*.

Pitch can be defined as the perceived frequency of a sound. Most musical sounds are periodic in nature, involving regular repeats of sound pressure waves over time. The frequency of repetition is usually measured by the number of cycles per second (hertz, Hz). Most of the periodic sounds contain several frequency components: a fundamental frequency (F0) and higher frequencies (often integer multiples of F0) known as overtones. For example, note A4 has a fundamental frequency of 440 Hz and overtones of 880 Hz, 1,320 Hz, 1,760 Hz and so on. Each musical instrument creates its own combination of overtones of various strengths, commonly referred to as the *timbre* or sound colour of the instrument. The relative pitch difference between two tones is called the interval. Semitone is the smallest interval usually employed in Western music and refers to a pitch difference of about six per cents.

A musical scale consists of intervals and defines a set of pitches that are typically used together. The most common scales include major, natural minor, harmonic minor and melodic minor scales. A musical key can be major or minor and consists of a major or a minor scale starting on a certain note. For example, if a piece of music is said to be in C major, it mostly employs major scale that starts from C. If a tone that is foreign to the predominant key is introduced, it is can be said to be *out of key*. If the tone fails to be in a chromatic scale¹ in relation to other tones, it is also referred as being *out of tune*. Thus, tones that are out of tune are always also out of key, but a tone can be out of key but in tune.

Melodies, sets of consecutive tones that group together in a meaningful way, are often the most distinguishable elements of a musical work. A melody usually follows a musical key and meter. If a tone of a melody violates the expectations set by predominant key or meter, listeners often perceive it as being strange or "wrong", even though they might not be able to describe why it appeared to be abnormal. Indeed, empirical evidence demonstrates that listeners tend to utilize the scale structure when perceiving and recalling melodies (Tillmann, Bharucha, & Bigand, 2000).

¹ Chromatic scale consists of consecutive semitones. The keys of common keyboard instruments, such as piano or cembalo, form a chromatic scale.

1.2. Evolutionary and developmental background of music

Music might often be thought as an artefact of human culture, but comparative biology shows that a song-like communication system has developed at least three times in mammals and three times in birds (Fitch, 2006). There are also interesting animal homologues for instrumental music: African great apes are known to drum bimanually and with sticks, and woodpeckers seek out resonant trees for display drumming (Fitch, 2006). Thus, cross-species research provides valuable insights into the evolution of music-related capacities.

Music cognition involves many distinct capacities, some of which seem to be specific to music and some not. These abilities have most likely developed at different times in the evolution (Patel & Demorest, 2012). The ability to perceive the pitch of a complex harmonic sound seems to be shared among several species of mammals, birds, and even fish (Shofner, 2005). Hence it seems that this ability, which is not specific to music, has developed very early in the evolution. In spite of this basic pitch perception ability, animals seem to lack the capacity to perceive some of the more abstract structural properties related to pitch. For example, no species apart from ours is known to be able to recognize a familiar tune at different pitches (D'Amato, 1988; Patel & Demorest, 2012). Therefore, this ability, which requires implicit understanding of relational pitch differences and is specific to music, seems to be developed very late in the evolution process. This is particularly interesting given the fact that humans rely more on relative pitch (identifying a given musical note by comparing it to a reference note) than absolute pitch (identifying a given musical note without the benefit of a reference tone) when recognizing musical sequences (Lee, Janata, Frost, Hanke, & Granger, 2011; Plantinga & Trainor, 2005).

Cognitive capacities related to rhythm perception have interesting similarities and dissimilarities between species. It has been shown that pigeons can discriminate between tone sequences on the basis of differences in the tempo of the sequences and generalize this discrimination to different tempi (Hagmann & Cook, 2010). Furthermore, monkeys can discriminate between slow and fast auditory click trains (McDermott & Hauser, 2007). These results show that basic tempo perception is widespread among vertebrates and has developed early in the vertebrate evolution for identifying a variety of biological and environmental sounds (Patel & Demorest, 2012). The capability to synchronize movement

to sounds seems to be more rare, although some particular species, such as cockatoos, have been found to spontaneously synchronize their movements to the beat of human music (Patel, Iversen, Bregman, & Schulz, 2009). Interestingly, studies have shown that rhesus monkeys fail to move in synchrony with a metronome even after more than a year of training (Zarco, Merchant, Prado, & Mendez, 2009), whereas parrots succeed in doing so (Hasegawa, Okanoya, Hasegawa, & Seki, 2011). Overall, moving spontaneously to the beat of music seems to be a typically human behaviour but not common across other species.

Cross-species research also provides insights into the meaning of music. In the animal kingdom, the activities that resemble human singing and drumming are often closely related to communication. For example, chimps and gorillas use manual drumming to indicate dominance, aggression, or an invitation to play (Fitch, 2006). Humans, on the other hand, use language for the majority of everyday communicational needs, while music is more often used for pleasure, self-expression, and communication of emotions. Thus, language and music have distinctive qualities and different functions, even though these functions also overlap. The links between music and language have led some researchers to argue that both of them have developed during evolution from a *proto-language*, a song-like communication system, but over time have separated into different systems with their own specific features (Mithen, 2005). This view claims that music evolved in order to serve specific functions, such as attracting the opposite sex, communicating with others, and social bonding between mother and child. In this view, music could be seen as having survival value, which would define it as a biological adaptation in evolution (i.e., a product of natural selection). Other views about the function of music exist, too. Pinker (1999) claims that music is purely a vehicle of pleasure and has developed from the brain mechanisms responsible for speech, hearing, movement, and perception of emotions. Also Patel (2010) presents a view that music is not an evolutionary adaptation but a human invention. Since there is no clear adaptive survival function that could be attributed to music (Hauser & McDermott, 2003), this debate remains open.

Developmental psychology provides another viewpoint on the biological basis of music. Our abilities to make pitch and rhythm discriminations in music develop early in childhood, which suggests that we have a biological disposition for music. Before age of twelve months, infants are able to detect mistuned sounds in melodies (Trehub,

Schellenberg, & Kamenetsky, 1999), recognise a familiar tune performed at different pitch levels (Chang & Trehub, 1977) and detect changes in rhythm (Trehub & Thorpe, 1989) and meter (Hannon & Johnson, 2005). Beat induction, the detection of regular pulse, seems to be innate (Winkler et al., 2009) and infants above six months of age have been shown to engage in significantly more rhythmic movement to music and other rhythmically regular sounds than to speech (Zentner & Eerola, 2010).

In addition to the innate predispositions, also environmental factors affect the development of music cognition. One environmental factor common across cultures is singing to infants (Trehub & Trainor, 1998; Trehub, 2001). When infants listen to singing and other music of their culture they become influenced by the particular musical system of that culture. This influence will modify the brain by developing neural circuits specialized for processing the pitch-related and rhythm-related structure of that musical system (Hannon & Trainor, 2007; Trainor, 2005; Trainor & Corrigall, 2010; Trainor, Marie, Gerry, Whiskin, & Unrau, 2012; Trehub, 2003). An example of this influence is the acquisition of an implicit knowledge about culture-specific musical keys. Western infants of eight months of age can detect changes in a Western musical melody that are within the predominant key or that go outside it, whereas musically untrained Western adults are much better at detecting the out-of-key than within-key changes (Trainor & Trehub, 1992). Much like adults, children of four and five years of age are better at detecting the out-of-key changes than within-key changes in melodies (Corrigall & Trainor, 2010; Trainor & Trehub, 1994). This implicit knowledge about musical keys has been detected in both behavioural and event-related potential studies of musically untrained Western adults (Bigand & Poulin-Charronnat, 2006; Bischoff Renninger, Wilson, & Donchin, 2006; Brattico, Tervaniemi, Näätänen, Peretz, 2006; Koelsch, Gunter, Wittfoth, & Sammler, 2005). Presumably, this enculturation takes place automatically through everyday exposure to Western music and its musical system (Trainor & Hannon, 2012).

The above-mentioned results of comparative and developmental psychology show that our inclination to perceive, appreciate, and produce music is biologically driven. However, the environments where we live in affect our ways of perceiving music via implicit learning of musical systems. Hence, it seems that both genetic and environmental factors play a role in musical development, but their relative contribution to the individual differences in different music perception skills remains largely unknown.

1.3. Music perception in the adult human brain

Musical sounds and all other sounds set into motion a complex cascade of mechanical, chemical, and neural events in the cochlea, brain stem, midbrain nuclei, and cortex, which rapidly gives rise to our perception of the sounds (Peretz & Zatorre, 2005). The brain areas related to the processing of musical sounds have been examined in brain lesion studies and neuroimaging studies of healthy participants. Lesion studies suggest that rhythm-based and pitch-based structures are processed somewhat independently in the brain (Peretz & Zatorre, 2005). Brain damage can cause deficits of pitch perception while rhythm perception remains accurate (Ayotte, Peretz, Rousseau, Bard, & Bojanowski, 2000; Liegeois-Chauvel, Peretz, Babai, Laguitton, & Chauvel, 1998; Peretz, 1990; Peretz & Kolinsky, 1993; Piccirilli, Sciarra, & Luzzi, 2000; Vignolo, 2003). Also opposite cases with impaired rhythm perception and intact pitch perception have been documented (Di Pietro, Laganaro, Leemann, & Schneider, 2004; Peretz, 1990), suggesting that there are separable neural subsystems for pitch and rhythm. However, this distinction is yet to be confirmed in healthy participants using neuroimaging methods (Peretz & Zatorre, 2005).

Rhythm perception involves two different abilities, namely the ability to segment an ongoing sequence into temporal groups and to extract an underlying periodic pulse or meter (Fraisse, 1982). Studies utilizing tapping tasks (Fries & Swihart, 1990; Ibbotson & Morton, 1981; Wilson, Pressing, & Wales, 2002) have suggested that the right hemisphere is more involved in meter perception, while the left hemisphere could be involved in the more fine-grained task of temporal grouping. Studies of rhythm perception have supported this division by pointing to the role of left hemisphere in temporal grouping (Di Pietro et al., 2004; Vignolo, 2003) and the right temporal auditory cortex in meter perception (Liegeois-Chauvel et al., 1998; Penhune, Zatorre, & Feindel, 1999; Wilson et al., 2002). Both lesion studies (Ivry, Keele, & Diener, 1988; Ivry & Keele, 1989) and neuroimaging studies (Penhune, Zatorre, & Evans, 1998; Penhune & Doyon, 2002; Schubotz, Friederici, & Yves von Cramon, 2000) have suggested that the cerebellum is closely involved in controlling motor and perceptual timing. Other studies (Harrington, Haaland, & Hermanowitz, 1998; Rao, Mayer, & Harrington, 2001; Rao et al., 1997) have implicated that the basal ganglia have a role in these functions, particularly in beat perception (Grahn

& Rowe, 2009; Grahn & Rowe, 2013). Several studies also suggest the involvement of premotor and motor cortical areas in rhythm perception and production (Grahn & Rowe, 2009; Halsband, Ito, Tanji, & Freund, 1993; Zatorre, Chen, & Penhune, 2007).

Pitch perception has been associated with the right temporal neocortex in several brain lesion studies (Liegeois-Chauvel et al., 1998; Milner, 1962; Zatorre, 1985). Also damage to the right anterolateral part of Heschl's gyrus in the temporal lobe has been found to cause problems with pitch-related tasks (Johnsrude, Penhune, & Zatorre, 2000; Zatorre, 1988). Several neuroimaging studies (Hall et al., 2002; Hart, Palmer, & Hall, 2003; Thivard, Belin, Zilbovicius, Poline, & Samson, 2000) suggest that the analysis of pitch changes may involve areas of posterior secondary auditory cortex. Fine-grained manipulations of pitch and the perception of musical melodies have been shown to produce greater responses in the right auditory regions (Hyde, Peretz, & Zatorre, 2008; Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Zatorre, Belin, & Penhune, 2002; Zatorre & Belin, 2001). In summary, especially the right secondary auditory cortex and other regions in the right temporal lobe seems to be involved in processing pitch elements in sequentially presented tones (Peretz & Zatorre, 2005).

Music perception requires also memory. Lesion studies of working memory for pitch materials have suggested the involvement of the right auditory cortex (Zatorre & Samson, 1991). Also frontal cortical areas, such as dorsolateral and inferior frontal areas, and inferior parietal areas, seem to be involved (Griffiths, Büchel, Frackowiak, & Patterson, 1998; Schulze, Zysset, Mueller, Friederici, & Koelsch, 2011; Zatorre, Evans, & Meyer, 1994). Marin and Perry (1999) suggest that pitch memory could be seen as a specialized subsystem in general working memory. The right temporal structures seem to be less critical in recognition of highly familiar tunes than in tasks involving new tunes (Peretz & Zatorre, 2005).

In summary, music perception involves multiple brain regions that seem to be related to different acoustic, structural, and functional aspects of music. Therefore, it seems reasonable to assume that there are also multiple genetic factors that affect different aspects of music perception, such as rhythm perception, pitch perception and music-related working memory.

1.4. Quantitative genetics

Human genome is normally expressed in 23 chromosome pairs including approximately 24,000 genes (Plomin, DeFries, Knopik, & Neiderhiser, 2008). We all have a slightly different genome with the exception of (genetically) identical twins. The alternative forms of a gene are called *alleles*. Each allele is located in a specific location (i.e. *locus*) in a specific chromosome. There are two copies of each gene, one inherited from each parent. Since we cannot select our genes, the question how much they regulate our lives has a natural appeal. For complex behavioural traits, such as music perception, there is no straightforward answer to this question at the individual level. However, at the level of a population, the proportional effects of genes and environment can be estimated using *quantitative genetics* (Boomsma et al., 2002; Plomin et al., 2008).

Quantitative genetics can be defined as the study of *heritability* of different traits using the known degrees of genetic similarity between relatives. Heritability refers to the proportion of the population variance explained by genetic factors. All methods of quantitative genetics are based on the work of Fisher (1919), who derived the expected values of genetic similarity of relatives and pointed out that these can be used to estimate, how much the genetic differences explain about the total variation of a trait in a given population. The rest of the population variance is explained by environmental differences. The biological foundations of a trait are called its *genotype* and the expression of a trait is known as its *phenotype*.

Genetic effects can be divided to two parts, namely *additive genetic effects* and *dominant genetic effects*. Additive genetic effects refer to the summed effects of all the alleles that influence the given phenotype. There can also be interaction between two alleles sharing the same locus. This interaction can cause dominant genetic effects, which means that the allele that has stronger influence on the given phenotype dominates the other allele in the same locus. Dominance causes the total effect of a given allele pair to deviate from the sum of the individual alleles' effects. *Narrow-sense heritability* includes only additive effects while *broad-sense heritability* includes both additive and dominant genetic effects (Boomsma et al., 2002). In this thesis, heritability refers to narrow-sense heritability unless otherwise stated.

Environmental effects can be divided to those caused by *common environment* and *unique environment*. Common (or shared) environment refers to all those environmental factors that make individuals similar. Unique environment refers to all those environmental factors that make individuals different. This means that any given environmental factor can be part of the common environment or the unique environment depending on how it affects the given phenotype. For example, parenting style can be a part of common environment if it makes children more similar, or a part of unique environment if it makes them more different. In quantitative genetics, measurement error is included in unique environmental effects. This makes the interpretation of these effects ambiguous, since they can be a result of real environmental factors or measurement error (Turkheimer, 2000).

Estimates of relative genetic and environmental effects on different complex behavioural and personality traits vary noticeably. As each effect explains a proportion of the population variance, it has to be on range from .00 to 1.00. For personality traits of the NEO Five-Factor Inventory², estimated genetic effects of .51–.58 with no common environmental effects have been obtained (Loehlin, McCrae, Costa Jr, & John, 1998). Abrahamson et al. (2002) studied religiousness and conservatism in adolescents and obtained estimates in range of .00–.24 for genetic effects and .11–.46 for common environmental effects. In some traits the heritability changes as a function of age, especially during childhood and adolescence. In a longitudinal twin study of intelligence, genetic and common environmental effects were estimated at .30 and .60 respectively at the age of five, but gradually the proportion of genetic effects increased while the proportion of common environmental effect decreased, and at the age of 27, genetic effects accounted for .85 of the variation while common environmental effects dropped to zero (Boomsma, de Geus, van Baal, & Koopmans, 1999).

The examples mentioned above illustrate the fact that depending on the phenotype and age group very different estimates can be obtained. It should be kept in mind that because genetic and environmental effects account for a certain *proportion* of the population variance, an increase in either group lowers the estimate in the other. This means that any given estimate tells only how strong an effect is related to other sources of variance, not how strong the effect is *per se*. Thus if environment would be held constant for everyone,

² Major personality traits in NEO-FFI include neuroticism, extraversion, openness, conscientiousness, and agreeableness.

all individual differences would become heritable (Johnson, Turkheimer, Gottesman, & Bouchard, 2010). It should also be noted that each estimate is applicable only to the population it was derived from. For example, even though height seems to be strongly heritable, the environmental differences between North and South Korea cause the North Korean children to be on average eight centimeter shorter than the South Korean children (Schwekendiek & Pak, 2009). Thus, if either of the two nations is examined alone, genetic effects for height are strong (but likely greater in South Korea), but if their populations, which are genetically very close to each other, are compared, environmental effects explain almost all of the difference between the nations.

Relative risk or *relative odds* is another way to assess the genetic influence behind a trait. If a person has a certain trait or disease, relative risk tells how many times more probable it is for his or her siblings to have the same trait compared to a person whose sibling does not possess the trait. The interpretation of a relative risk is dependent on the prevalence of the trait. A common trait cannot be very much more common in relatives even when highly familial, while a rare trait can be even orders of magnitude more common in close relatives.

1.5. The classical twin design

The classical twin design, where groups of identical and non-identical twins are combined is the most common method in quantitative genetics. Identical twins are a result of one fertilized egg (zygote) dividing to two embryos early in the development and are therefore called *monozygotic* (MZ) twins. Non-identical twins are born after two different eggs fertilize separately and are called *dizygotic* (DZ) twins. Since MZ twins separate after chromosomal crossover they have an identical genome. DZ twins fertilize separately and share, like non-twin siblings, on average only 50 % of their segregating genes. First derived by Fisher (1919), these expected values of genetic similarity have later been confirmed by Visscher et al. (2006) using genome scans. Therefore, the only difference between groups of MZ and DZ twins is their different degree of genetic relatedness, if we can assume that there is no systematic difference in the environmental effects between the two groups. This is referred as the *equal environmental similarity assumption* (Bouchard & McGue, 2003).

In modern quantitative twin research, structural equation modelling is the most commonly used statistical method. The central idea is to decompose population variance of a given phenotype between variance caused by the different genetic and environmental effects (Purcell, 2008). Genetic and environmental effects are referred to as latent, unknown variables whose relations to known, measured traits are then estimated. As explained in the previous chapter, these effects include additive genetic effects, dominant genetic effects, common environmental effects, and unique environmental effects, which are often abbreviated A, D, C and E, respectively. A model where only one measured trait is included is called a *univariate model* (see Figure 1). The correlations of the genetic latent variables between twins are derived from the known genetic relatedness of MZ and DZ twins (Fisher, 1919; Visscher et al., 2006). The twin correlation of the common environmental latent variables is fixed at one, because they consist of all the environmental effects that are *the same* for both twins. The unique environmental latent variables are uncorrelated within a twin pair because they consist of all the environmental effects that are *different* between the twins.

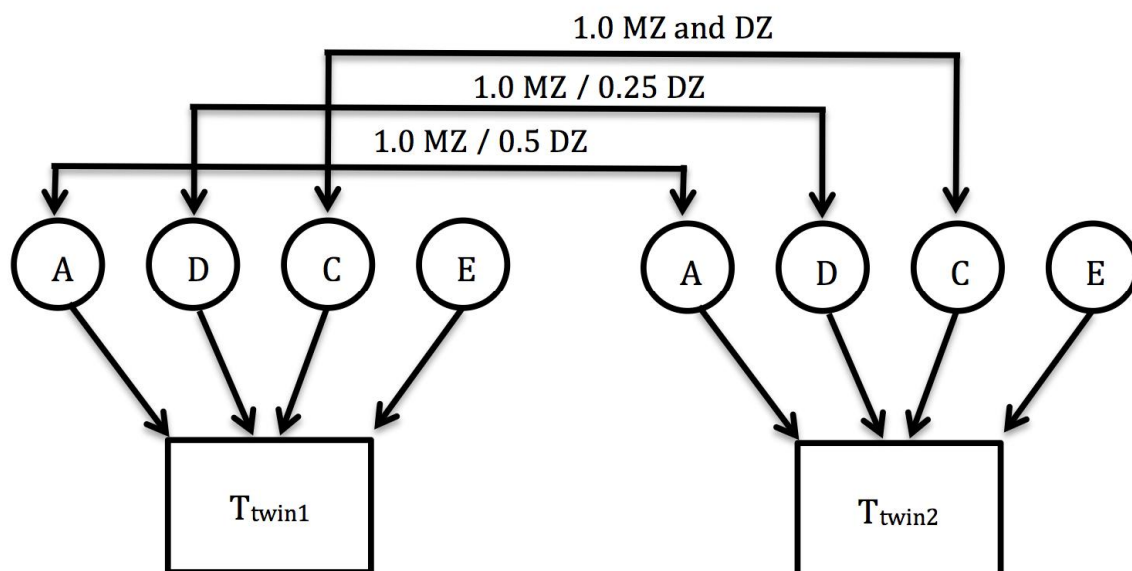


Figure 1. The univariate model and the effects of the latent components to one measured trait and the correlations of the latent components between twins as presented by Silventoinen and Kaprio (2008). T = measured trait; A = Additive genetic effects; D = Dominant genetic effects; C = Common environmental effects; E = Unique environmental effects; MZ = Monozygotic twins; DZ = Dizygotic twins. An one-headed arrow represents a direct influence of a latent component to the measured trait. A two-headed arrow represents a correlation between two latent components.

It should be noted that a full ACDE model (a model that would include all four latent variables) cannot be estimated using the classical twin design where each twin pair has been reared together. This is a consequence of the fact that the only difference between the groups of MZ and DZ twins is their different degree of genetic relatedness, and thus only three latent parameters can be estimated at time (For a mathematical proof see Neale & Cardon, 1992). In the model described above, the measurement error is included in unique environmental effects, which prevents omission of that latent variable. CDE and DE models are rarely used with complex behavioural traits, since it is unlikely that such a trait would have dominant genetic effects without additive genetic effects, i.e. effects due to dominance (interactions between alleles at a locus) in the absence of additive effects at a locus. Hence, practically possible models include ACE model, ADE model, and models with fewer latent variables (AE, CE, and E models). An ACE or ADE model is first fitted to the data and is referred as *full model*. If the twin correlation of MZ pairs is over twice that of the DZ pairs, it suggests the existence of dominant genetic effects and an ADE model should be used, otherwise an ACE model should be used (Plomin et al., 2008). After fitting an ACE or ADE model to the data, models with fewer latent variables (AE, CE, and E models) are compared against it in order to find the most parsimonious model that fits the data.

1.6. Twin and family studies of music perception

Music perception has been measured with a variety of tests that have several differences regarding the tasks and stimuli. For example, the acoustic or musical complexity of the stimuli varies from test to test. Some tests, such as the *Seashore Measures of Musical Talents* (Seashore, Lewis, & Saetveit, 1960) and the *Swedish Music Discrimination Test* (Ullén et al., 2014), use relatively simple stimuli, such as isochronous tone sequences, rhythm sequences of constant pitch, or sinusoidal tone pairs. The rationale is that by using very primitive stimuli it is easier to measure certain isolated aspect of music perception at a time, such as rhythm or pitch perception. Other tests, such as the *Distorted Tunes Test* (Fry, 1948; Kalmus & Fry, 1980), the *Montreal Battery of Evaluation of Amusia* (Peretz, Champod, & Hyde, 2003), and the *Online Test of Amusia* (Peretz et al., 2008) use musical melodies, which tend to be slightly more complex in nature. This type of material is closer to genuine real life music while still being simple enough to control easily. Crucially, when

comparing the results of different studies it should be always kept in mind that different tests measure different aspects of music perception and cognition. For example, from a music cognition perspective, comparing two isolated sinusoidal tones for a fine-grained pitch difference is a much simpler task than listening to a popular melody for any larger-than-semitone pitch alterations, even though pitch perception is clearly required in both.

Vandenberg (1962) studied 33 MZ and 43 DZ high school-aged twin pairs using five subtests from the Seashore and Wing music tests (Seashore et al., 1960; Wing, 1970). In the Seashore Pitch subtest the participant hears sinusoidal tone pairs and has to decide whether the second tone is higher or lower in pitch than the first. The difference in pitch is about half semitone for the first five pairs and decreases gradually until it is not discriminable by the human ear. In the Seashore Loudness subtest, the participant hears sinusoidal tone pairs and has to indicate whether the second tone is stronger or weaker in intensity than the first. The difference in loudness is four decibels for the first five pairs and gets smaller as the test proceeds. In the Seashore Rhythm subtest, the participant hears two rhythmic patterns of five to seven tones of equal pitch and is asked to indicate if the patterns are identical or not. The alterations are metrically congruent, i.e. they do not disturb the constant pulse that has been established. In the Wing Pitch change subtest, the participant hears two consecutive chords that differ in one tone, and the task is to recognize whether there is a tone change in the second chord. In the Wing Memory subtest, the participant hears two sequences of three to ten tones and needs to detect which tone has been altered. Heritability estimates did not statistically differ from zero for tests that involve pitch or loudness discrimination. In the Seashore Rhythm test and the Wing Memory test, the heritability estimates were .52 and .42, respectively. However, due to the small sample size and several methodological shortcomings, these results should be considered as more unreliable than the results obtained in more recent studies.

Drayna, Manichaikul, de Lange, Snieder, and Spector (2001) studied pitch perception with a sample of 284 twin pairs aged 18–74 years using the Distorted Tunes Test (DTT) originally developed by Fry (1948) and later used by Kalmus and Fry (1980). In DTT, 26 unaccompanied popular melodies are used as stimuli. Seventeen of the melodies are distorted by altering the pitch of one or several tones and the subjects are asked to judge whether each melody is correct (original) or incorrect (distorted). The authors estimated the heritability of performance in DTT to be .71 (95 % Confidence Interval [CI] .61–.78).

The data did not support an effect of common environment. All the subjects in the sample were female, so no sex differences could be studied. Another limitation of the study has to do with the DDT test itself: since it relies on well-known songs (e.g., the national anthem of USA), the detection of distortion in the melodies is based both on an implicit sense of musical key and on a memory-based comparison of the heard melody with the original one. Therefore, the test is not a pure measure of music perception but requires long-term memory as well.

Pulli et al. (2008) studied the genetic background of music perception with 15 multigenerational families (with a total of 234 family members). The subjects were recruited among the families whose members had studied or were studying music in at Sibelius Academy or other music institutes in Finland. Music perception was evaluated with three tests. The *Karma Music Test* (KMT) measures auditory structuring ability with short, abstract sound patterns that form hierarchical structures, and the participant is asked to detect structural changes in these patterns (Karma, 1984). The Seashore Pitch (SP) subtest is identical to that used by Vandenberg (1962) and was described earlier. In the Seashore Time (ST) subtest, the participant hears sinusoidal tone pairs and is asked to answer whether the second tone is longer or shorter in duration than the first. The difference in duration is initially 300 ms and gets gradually smaller. Heritability was estimated to be .42 for KMT, .57 for SP, and .21 for ST. This study was the first to provide heritability estimates on tasks that require the ability to make discriminations in the temporal domain. Importantly, the results show that there seems to be also genetic influence on test requiring the ability to make discriminations in musical time. However, it should be noted that the main aim was to study musical talent and, therefore, the families in the study all included professional musicians and serious amateurs, which clearly makes the sample biased and precludes conclusions about the heritability of music in the general population. The same research group also presented slightly different heritability estimates with a larger sample of 19 families (Ukkola, Onkamo, Raijas, Karma, & Järvelä, 2009). In this study, estimates were .39 for KMT, .52 for SP, and .10 for ST.

In a recent study, Ullén et al. (2014) studied music perception with population-based sample of 1362 twin pairs aged 27–54 years with the Swedish Music Discrimination Test (SMDT) that includes three subtests. In the SMDT Rhythm subtest, the stimuli consist of short rhythmic sequences of five to seven sinusoidal tones of constant pitch. Two

sequences separated by a brief pause are heard and the participant has to decide, whether they are identical or not. In the SMDT Pitch subtest, a pair of sinusoidal tones is heard. The participants' task is to tell, which one of the tones had higher pitch. In the SMDT Melody subtest, isochronous sequences of four to nine piano tones are heard. As in the rhythm subtest, two sequences separated by a brief pause are compared. The task is to decide which tone was different between the melodies. In the Rhythm subtest, the heritability was estimated to be .50 (95 % CI .33–.57) and the data showed no significant effect of common environment. In the Pitch subtest, the results were different for males and females. For females, the heritability was .30 (95 % CI .09–.52) and the effect of common environment was .19 (95 % CI .00–.38). For males, the respective estimates were .12 (95 % CI .00–.35) and .38 (95 % CI .17–.53). In the Melody subtest, the heritability was estimated at .59 (95 % CI .43–.64) and the data showed no significant effect of common environment. The same research group studied also the heritability of music practice and whether MZ pairs that were discordant regarding the amount of music practice performed differently in the SMDT tasks (Mosing et al., 2014). They reported that the amount of music practice was also heritable (.70 and .40 for males and females, respectively), and that the within-pair difference in music practice and the within-pair difference in the SMDT tasks did not correlate in discordant MZ pairs. The latter result supports a hypothesis that the differences in music perception ability, as operationalized by the SMDT test, are due to genetic differences, not differences caused by amount of practice (an environmental factor).

Heritability has also been studied in some special cases of music perception, namely *congenital amusia* and *absolute pitch*. Congenital amusia refers to an inability to acquire normal musical skills despite normal hearing and exposure to music (Peretz, 2008). In an aggregation study of amusic and control families by Peretz, Cummings, and Dubé (2007), the authors discovered that the siblings' relative risk for congenital amusia was estimated to be 10.8 (95 % CI 8–13.5), thus indicating that congenital amusia has an inheritable component. The amusic individuals had difficulties in their ability to detect off-key pitches, but their ability to detect incongruities in time was normal. These results suggest that the inheritable component for congenital amusia might be more related to the pitch perception than to musical time perception. Interestingly, there is also a specific rhythmic form of amusia known as *beat deafness* (Phillips-Silver et al., 2011), but its heritability has not been studied. Absolute pitch refers to the ability to identify the pitch of tones in the

absence of a reference pitch. Even though early musical training is considered essential for developing absolute pitch (Zatorre, 2003), the genetic predisposition is essential, too. Baharloo et al. (2000) estimated siblings' relative odds for absolute pitch after controlling for early musical training to be in a range of 7.8–15.1, thus indicating that also this special phenomenon has an inheritable component.

The above-mentioned studies have obtained hereditary effects in different music perception tasks in different populations. In summary, in studies of persons with “normal” musical skills (meaning no amusia or absolute pitch) and background (no special musical training), the heritability is usually found to be around .50 or higher in tasks where stimuli *longer and more complex* than tone or chord pairs are used (Drayna et al., 2001; Ullén et al., 2014; Vandenberg, 1962). This fairly high heritability seems to be found in both pitch- and rhythm-related tasks. For tasks that require the comparison of *two* tones or chords for a pitch change, Vandenberg (1962) found only very small heritability estimates (.12 or less), while Ullén et al. (2014) found the heritability to be different for male and female participants (.12 and .30, respectively). Thus, the heritability seems to be higher in tasks where longer stimuli are processed, perhaps because of the increased demands for memory. Studies with professional and amateur musicians (Pulli et al., 2008; Ukkola et al., 2009) suggest that within these special groups the heritability seems to be higher in pitch-related than in time-related tasks (.52–.57 and .10–.21, respectively). However, no direct conclusions can be made about the heritability in the general population. Studies with individuals with congenital amusia (Peretz et al., 2007) and those with absolute pitch (Baharloo et al., 2000) have demonstrated that these pitch-related special conditions have a heritable component.

Many questions about the heritability of music perception in the general population remain to be answered. The study by Drayna et al. (2001) was representative of the general population, but did not include any discrimination tasks related to musical time or rhythm. The studies by Ullén et al. (2014) and Mosing et al. (2014) were also population-based and included a rhythm discrimination task, but concentrated on perception of more musically and acoustically simple stimuli. Thus, there is a clear demand for a population-based study with more complex and ecologically valid musical stimuli (i.e. musical melodies), which include tasks related to both rhythm and pitch perception.

1.7. Research questions

The goal of the thesis was to determine genetic and environmental effects of music perception ability in the normal population using the classical twin design. Specifically, the aim was to estimate these effects on three subtests and total score of the updated version of the Online Test of Amusia (OTA; Peretz et al., 2008), one of the most widely used and comprehensive music perception tests, which involves different musical features and task constraints: 1) comparing two consecutive melodies for out-of-key (but in tune) pitch alterations in piano timbre (the Scale subtest), 2) detecting pitch incongruences (tones that are out of key *or also* out of tune) in single melodies with varying instrumental timbres (the Out-of-key subtest), and 3) detecting rhythmic incongruences in single melodies with varying instrumental timbres (the Off-beat subtest). Given the differences in the acoustic structure, development, and neural basis of pitch and rhythm in music, it was hypothesized that the contribution of genetic and environmental factors would differ between pitch and rhythm perception ability.

2. Methods

2.1. Subjects

The sample consisted of individuals from a population-based longitudinal *FinnTwin16* study³, which includes altogether 2733 Finnish twin pairs born in 1975–1979 (Kaprio, Pulkkinen, & Rose, 2002; Kaprio, 2013). The Finntwin16 study was initiated in 1991 and is headed by Academy Professor Jaakko Kaprio, University of Helsinki. The FinnTwin16 study protocol is approved by the Ethical Committee of the Helsinki and Uusimaa Hospital District, Finland, and by the Institutional Review Board of Indiana University, Bloomington, Indiana, USA. The Ethical Committee of the Hjelt Institute of the University of Helsinki approved the study, and informed consent was obtained from all participants.

The data of the current study was collected in two waves. In the first wave, the invitation to participate was mailed in September 2012 to 105 MZ and 107 DZ twin pairs (212 pairs in total). These pairs were randomly selected from those FinnTwin16 female-female pairs where both members of a pair lived in Finland, were Finnish speaking and had participated in earlier FinnTwin16 survey⁴ in 2010–2012. In the second wave, the invitation was mailed in February 2013 to 148 MZ twin pairs and 147 DZ twin pairs (295 pairs in total). These pairs were randomly selected of those male-male and female-female pairs who lived in Southern Finland and were Finnish speaking. Overall, the 507 twin pairs who received the invitation included 164 male-male and 343 female-female pairs. The sample represented all educational levels and all regions of Finland. The zygosity of pairs was determined by a validated questionnaire method which makes it possible to accurately classify over 91 % of the twins (Sarna, Kaprio, Sistonen, & Koskenvuo, 1978), while for a fraction of the pairs, zygosity has been confirmed using DNA-based tests of polymorphic genetic markers.

Of the 1014 persons invited to participate in the study 386 completed the online test (response rate of 38.1 %). Two individuals whose responses indicated that they had not actually performed the task were excluded from the data (performance was well below chance level or the task was completed in less time than stimulus presentation takes). Thus,

³ The homepage of FinnTwin16 study can be found at <http://wiki.helsinki.fi/display/twineng/Finntwin16>.

⁴ This earlier survey was the fifth wave data collection of the FinnTwin16 cohort and was done principally as an online survey.

the final data consisted of 69 MZ twin pairs with both twins participating, 44 DZ twin pairs with both twins participating, 70 MZ twins with no co-twin, and 88 DZ twins with no co-twin.

2.2. Assessment procedure

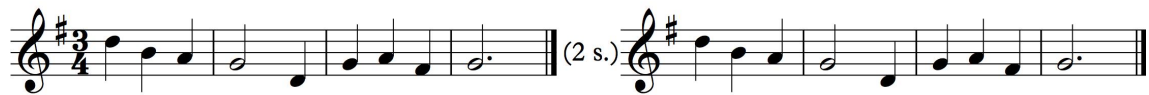
The participants received an invitation letter to participate in an internet-based study of genetic basis of music perception. In the invitation, they received a code, which they used to identify themselves when logging to a server maintained at the Hjelt Institute, University of Helsinki. After logging in, they performed the Online Test of Amusia (Peretz et al., 2008), which is based on the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003). The updated version⁵ of the OTA used here includes the Off-beat and Out-of-key subtests from the first version (Peretz et al., 2008), while the Mistuned subtest has been replaced with Scale subtest from the MBEA (Peretz, Champod, & Hyde, 2003). A total score was calculated, which is simply the sum of the three subtests. In conjunction with the test, they also answered four self-evaluation questions about their capacity to perceive music. The questions and the test instructions were presented in Finnish.

The laboratory-based MBEA was developed as a diagnostic tool for evaluation of amusia (Peretz et al., 2003). It can be used also as a general music perception test, with a high retest reliability ($r = .75$), normal distribution of the scores, and absence of floor or ceiling effects (Peretz et al., 2003). The OTA correlates fairly well with the MBEA ($r = .71$; Peretz et al., 2008). The Finnish version of the OTA has been tested with 61 healthy participants and the distributions of the scores are similar to those obtained with the English version (Hausen et al., 2013).

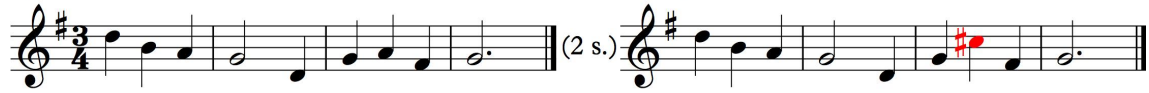
The current version of the OTA includes 30 new melodies that are all constructed in major key according to Western tonal-harmonic conventions (Peretz et al., 2008). They are 9.6 tones long on average and played by computer at a tempo of 120 beats per minute. The three subtests were presented in fixed order (Scale, Off-beat, Out-of-key). Within all subtests, the trials were presented in randomized but fixed order. In the Scale subtest, the participant is presented with two practice trials and 31 experimental trials. Each trial

⁵ The updated version of the OTA can be found at <http://www.brams.umontreal.ca/amusia-new/>.

A) Scale subtest: Two identical melodies separated by a pause of 2 seconds.



B) Scale subtest: Two non-identical melodies separated by a pause of 2 seconds.



C) Off-beat and Out-of-key subtests: An example melody with no irregularities of pulse or violations of musical key.



D) Off-beat subtest: An example melody with an additional pause of 357 ms creating irregularity in pulse.



E) Out-of-key subtest: An example melody with a tone which violates the musical key.



Figure 2. Examples of melodies used in Scale subtest (A and B), Off-beat subtest (C and E) and Out-of-key subtest (D and E) of MBEA. Violation of scale (A), incongruity in rhythm (D) and tone that is out of tune and out of key (E) are marked in red.

consists of two melodies separated by an interval of two seconds (Figure 2, examples A and B). The task of the participant is to judge whether the melodies are *identical* or not. In 15 trials, the second melody is identical to the first and in 15 trials it has been altered by changing the pitch of one tone to be out of key (but not out of tune). The subtest also included a markedly different “catch” trial, which was used in order to ensure that the participant was paying attention and was not included in the statistical analysis. In the Off-beat subtest, the participant is presented with two example trials and 24 experimental trials. In each trial, the participant hears a melody and the task is to indicate whether the melody contained an *unusual delay* or not (Figure 2, examples C and E). In half of the trials, the pulse of the melody is constant and in half of the trials an additional silence of 5/7 of the beat duration (357 ms) is introduced to the melody, thereby creating irregularity in the pulse. The Out-of-key subtest was similar to the Off-beat subtest. The difference is that the pulse of the melody is constant in all trials, but in half of the trials one tone is modified by raising or lowering its *pitch* by one, 1/2, 1/4 or 1/8 semitone, thereby creating a tone that is

out of key, i.e. irregularity in the pitch (Figure 2, examples D and E). The task of the participant is then to decide whether the melody contained a tone that is out of key or not. In the Out-of-key and Off-beat subtests the incongruent tone always occurred on the first beat of third bar of the four-bar melody, which facilitates its detection (Jones, Boltz, & Kidd, 1982). In the Scale subtest, the melodies were presented in piano timbre and in the Off-beat and Out-of-key subtests 10 different timbres were used (e.g., piano, saxophone, clarinet, recorder, harp, strings, and guitar).

Before starting the first subtest, the participants were asked to adjust the volume of the audio system to a comfortable level using three sample tones. Thus, the volume was not controlled but the participants had the opportunity to adjust it to an optimum level. It was advised to perform the test in a quiet environment and with headphones if possible. The participants received their results at the end of the test.

2.3. Data analysis

Initial descriptive data analyses were performed with IBM SPSS version 22.0 (IBM Corp., 2013) and Stata 11 software (Stata Corp., 2009). Further statistical analyses and structural equation modelling (i.e. genetic univariate modelling) was performed with MX software version 1.68e (Neale, 2004).

Initial data analyses were conducted as follows. First, the data were reviewed for any missing entries, outliers and persons who might have answered by guessing. There were no missing entries in the data. One participant was removed from the data because his/her test result was well below random chance and was creating an outlier. Another participant was removed because he/she had used markedly less time than is needed even to listen to all the melodies. Second, the score distributions for full pairs and twins with no co-twin were compared within each zygosity from histograms and with t-tests. Within both MZ and DZ twins, the distributions were very similar and there were no statistically significant differences between the distribution means. This indicates that there seems to be no selection effect in the participants. Third, the normality of distributions was evaluated from histograms. The distributions were normal or near normal and similar for both MZ and DZ twins, which is assumed in structural equation modelling. Fourth, Pearson's product-

moment correlations between subtests were calculated in order to examine their relations⁶. Fifth, Pearson's product-moment correlations within twin pairs (*twin correlations*) were calculated within each subtest, separately for both zygosity groups, and used as a basis for model selection (see below).

In genetic modelling, the scores of the three subtests and the total score were all analysed separately by using univariate models (see Chapter 1.5). The analysis was carried out by a method outlined by Neale and Cordon (1992), and is described below.

First, a *saturated model* was estimated, where the estimated variances and means of the measured trait (score) were allowed to vary between MZ and DZ twins and between first and second twins (note that the order of twins within a pair was random). The saturated model and its *minus two log likelihood* (-2LL) value⁷ were used as a baseline to which other models were compared. The difference in the -2LL values between two models (Δ -2LL) follows a χ^2 distribution with as many degrees of freedom as is the difference in the number of estimated parameters between the two models. The null hypothesis is that the models fit the data equally well, while the alternative hypothesis is that the model, which is compared to the baseline model, has a poorer fit to the data. A risk level of 0.05 was used as decision criterion.

After estimating the saturated model, further models were estimated in order to test the assumptions and constraints of genetic modelling. In order to test the *homogeneity of variances assumption*, two models were estimated. In the first model, the variances between the groups of MZ and DZ were constrained to be equal, and in the second model, the variances between the groups of first twins from each pair and second twins from each pair were constrained to be equal. Next, further two models were estimated in order to test the *equality of means assumption*. In the first model, the means between the groups of MZ and DZ were constrained to be equal, and in the second model, the means between the

⁶ In the correlations between the subtests the twin-clustered structure of the data was not taken into account, and correlations were calculated only by using full pairs. Therefore, these correlations are slightly distorted. However, the correlations calculated by using this method are close to phenotypic correlations that could be acquired using bivariate modelling. Since the correlations between subtests were used only to examine their relations and are not used in the actual univariate modelling, this more straightforward method was used.

⁷ Likelihood value of a given model is a Pearson's goodness-of-fit χ^2 statistic between the sample covariance matrix and the estimated population covariance matrix (Neale & Cardon, 1992). -2LL value is calculated by multiplying the natural logarithm of likelihood value by -2.

groups of first twins from each pair and second twins from each pair were constrained to be equal. All of these models were tested against the corresponding saturated model.

After testing the assumptions of genetic modelling, models with latent variance components affecting the measured trait were estimated using the *maximum likelihood method* (Neale & Cordon, 1992). First a full model, i.e. either an ACE or ADE model depending on twin correlations in the corresponding score, was estimated. Next, simpler models with one or two of the latent components fixed to zero were estimated and compared to the full model in order to find the most parsimonious model that still provided a good fit to the data. For these comparisons, two probability values were obtained, *naive p-value* and *corrected p-value*. The naive p-values were obtained by simply using the χ^2 distribution with as many degrees of freedom as is the difference in the number of estimated parameters between the models. As pointed out by Dominicus, Skrondal, Gjessing, Pedersen, and Palmgren (2006), the test statistic (Δ -2LL) in these comparisons actually follows a mixture of different χ^2 distributions with various degrees of freedom⁸. For the current study, corrected p-values were calculated using a method described by Dominicus et al. (2006). Furthermore, *Bayesian Information Criterion* (BIC), where a smaller value indicates better fit, was calculated for each model in order to provide information about how well it fits the data when taking the number of estimated parameters into account. The model comparison statistics and the estimated variance components for each model are provided in the next chapter.

All the models included sex and age as covariates, i.e. the variation caused by them was excluded from the analysis. Furthermore, models with educational level as third covariate were also estimated. These results were very similar to those with two covariates and therefore are not shown.

⁸ For ACE vs. AE, ACE vs. CE, and ADE vs. AE comparisons, this is a 50:50 mixture of null and $\chi^2(1)$ distributions, i.e. the corrected p-value is half of the naive p-value. Also for ACE vs. E comparisons, a halved naive p-value was used as the corrected p-value, as suggested by Dominicus et al (2006). For ADE vs. E comparison and data of 69 MZ twins and 44 DZ twins, corrected p-value is approximately a 21:50:29 mixture of null, $\chi^2(1)$ and $\chi^2(2)$ distributions (Dominicus et al., 2006).

3. Results

3.1. Descriptive results

Table 1 shows the distributions of sex, age, educational level, and answers of four self-evaluation questions for all participants and separately for both zygosity groups. The zygosity groups were compared regarding these background variables while taking the twin-clustered structure of the data into account. There were no statistically significant differences (see Table 1 for test statistics).

Table 1. Demographic variables and distributions of four self-evaluation questions for all participants and separately for each zygosity group.

	All	MZ	DZ	F-value, p-value
Sex				F(1, 270) = 2.64, p = .105
Male	109	50	59	
Female	275	158	117	
Age				F(1, 270) = 0.16, p = .685
mean (years)	35.00	34.96	35.04	
The educational level				F(3.94, 1062.64) = 0.87, p = .477
Primary level	10	3	7	
Secondary level	93	51	42	
Lowest level tertiary	47	24	23	
Bachelor level	110	66	44	
Master level or higher	124	64	60	
Do you recognise a familiar melody?				F(1, 270) = 1.17, p = .280
mean*	4.52	4.55	4.47	
Do you lack sense of music?				F(1, 270) = 1.58, p = .210
yes	53	24	29	
no	331	184	147	
Do you notice if someone is singing off-key?				F(1, 270) = 0.12, p = .734
yes	340	183	157	
no	44	25	19	
Do you notice if someone plays a wrong note?				F(1, 270) = 0.40, p = .527
yes	290	154	136	
no	94	54	40	

All, All participants together; MZ, MZ participants; DZ, DZ participants; *, the question was answered using five level Likert-scale with following level descriptions: 1) never, 2) rarely, 3) sometimes, 4) often, and 5) very often.

Figure 3 and Table 2 show the distributions of scores in the three subtests and the total score. In the Scale subtest, the Off-beat subtest, and the total score the scores follow the normal distribution. In the Out-of-tune subtest the distribution of scores is slightly negatively skewed as a result of a ceiling effect. This effect is similar in MZ and DZ twins.

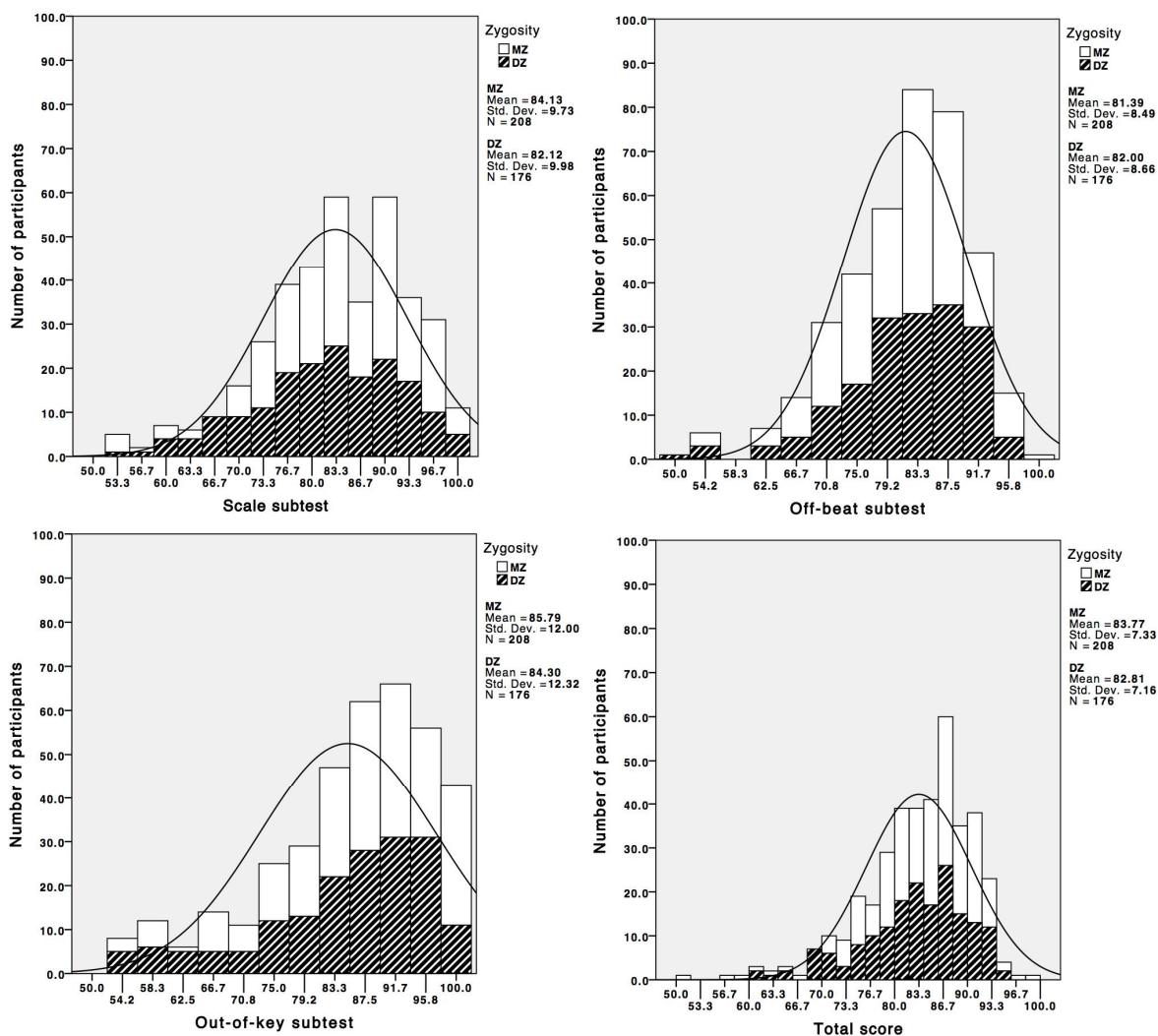


Figure 3. Distributions of the subtests and the total score. The curve in each histogram represents the normal distribution. The scores are presented in percentages.

Table 2. Distribution of scores in the subtests and the total score (N = 384).

	Minimum	Maximum	PS	Mean	Std. Dev.
Scale subtest	16 (53.3 %)	30 (100 %)	2 (2.9 %)	24.9 (83.2 %)	3.0 (9.9 %)
Off-beat subtest	12 (50.0 %)	24 (100 %)	1 (0.3 %)	19.6 (81.7 %)	2.1 (8.6 %)
Out-of-tune subtest	11 (45.8 %)	24 (100 %)	43 (11.2 %)	20.4 (85.1 %)	2.9 (12.2 %)
Total score	40 (51.1 %)	77 (98.6 %)	0 (0 %)	64.7 (83.3 %)	5.7 (7.3 %)

Minimum score; Maximum score; PS, Participants with perfect score; Mean; Standard Deviation.

Scatter plots indicating the relationships between the subtests are presented in Figure 4. There is a positive correlation ($r = .49$, $p < .001$) between the Scale and Out-of-key subtests, a modest correlation ($r = .14$, $p = .007$) between the Off-beat and Out-of-key subtests, and no statistically significant correlation between the Scale and Off-beat subtests.

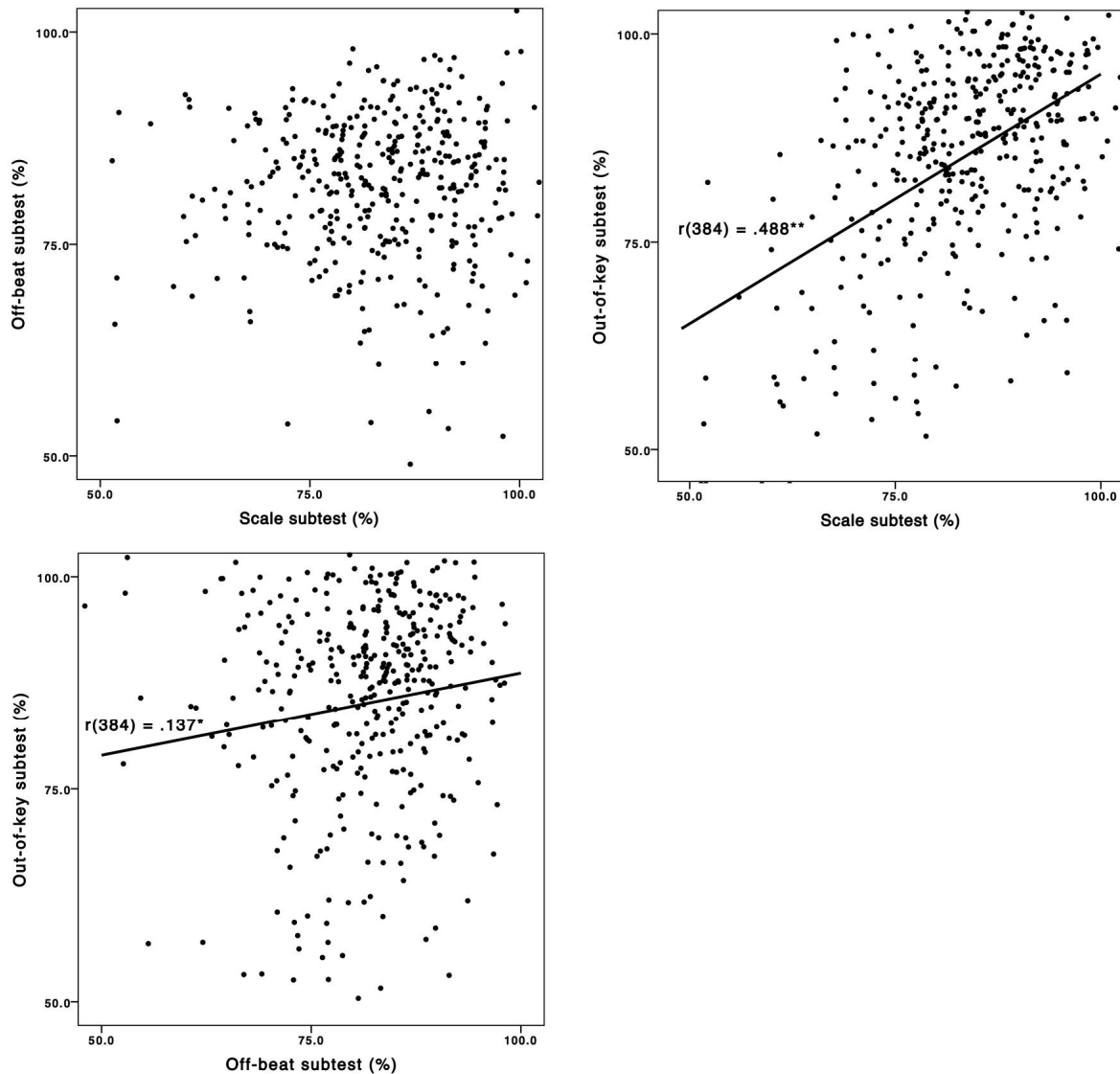


Figure 4. Scatter plots indicating the relationships between the subtests. The lines indicate the correlation coefficients. * = $p < .05$; ** = $p < .001$. There is no statistically significant correlation between Scale and Off-beat subtests. A small amount of artificial error has been added to the scatter plots in order to prevent the collision of data points.

Twin correlations are presented in Table 3. In the Scale subtest, the correlations are high for MZ pairs, moderate for DZ pairs, and statistically significant for both groups. In the Off-beat subtest, the correlation is moderate for MZ pairs and statistically significant, but for DZ twin pairs the correlation is non-significant. In the Out-of-key subtest and the total score, the correlations are high and statistically significant for both zygosity groups.

Table 3. Twin correlations in the subtests and the total score.

		MZ (n=69)	DZ (n=44)
Scale subtest	r (95 % CI)	.58 (.40–.72)	.38 (.09–.61)
	<i>p</i>	< .001	.010
Off-beat subtest	r (95 % CI)	.31 (.08–.51)	-.20 (-.47–.10)
	<i>p</i>	.010	.202
Out-of-key subtest	r (95 % CI)	.63 (.46–.75)	.67 (.47–.81)
	<i>p</i>	< .001	< .001
Total score	r (95 % CI)	.66 (.50–.78)	.65 (.44–.79)
	<i>p</i>	< .001	< .001

Figures in parentheses indicate the 95 % confidence interval. Significant correlations ($p < .05$) are written in bold.

3.2. Model fitting results

In all subtests and total score, the fit of most of the models testing the assumptions of genetic modelling was equal to the corresponding saturated model, and in these cases the homogeneity of variances and equality of means was assumed. The only exception was in the Off-beat subtest, where the model, in which the variances between the groups first twins and second twins were constrained to be equal, had a poorer fit to the data than the corresponding saturated model. Since the order of the twins within each twin pair was random, this exception was considered a coincidence. However, it means that the results regarding the Off-beat subtest could be slightly distorted, since the homogeneity of variances assumption is violated. The test statistics are presented in Appendix I.

When estimating the models with latent variance components, the full model was chosen based on twin correlations presented in Table 3 (See Chapter 1.5 for the rationale). An ADE model was used as the full model for the Off-beat subtest while an ACE model was

used for the other subtests and the total score. The fit of each full model was equal to the corresponding saturated model (See Appendix I for test statistics).

The comparison statistics between the full models and corresponding simpler models are shown in Table 4. The estimates for variance explained by the different latent components are shown in Table 5. Estimates are shown only for full models and those whose fit to the data was equal to the corresponding full model.

In the Scale subtest, the AE model was the most parsimonious model that fit the data well. In both the ACE and AE model the estimates of additive genetic effects were strong (.50 and .58 respectively). In the ACE model, the estimate of common environmental effects was modest (.07).

In the Off-beat subtest, both AE and E models' fit to the data was not significantly worse than that of ADE model, thus the E model was the most parsimonious model that fit the data well. In the ADE model, the estimate of dominant genetic effects was modest (.24), while the estimate of additive genetic effects was zero. In the AE model, the estimate of additive genetic effects was modest (.18).

In the Out-of-key subtest, the CE model was the most parsimonious model that fit the data well. In the ACE model, the estimate of additive genetic effects was negligible (.03). In both the ACE and CE model, the estimate of common environmental effects was strong (.59 and .61, respectively).

In the total score, the results were similar to the Out-of-key subtest. The CE model was the most parsimonious model that fit the data well. In the ACE model, the estimate of additive genetic effects was zero. In both the ACE and CE model, the estimate of common environmental effects was strong (.63 and .64, respectively).

Table 4. Model fitting results for each subtest and the total score.

	Model	-2LL	df	Δ -2LL	Δ df	p	p*	BIC
Scale subtest	ACE	2813.01	378					347.70
	AE	2813.09	379	0.08	1	.781	.391	344.94
	CE	2816.69	379	3.27	1	.080	.040	346.74
	E	2846.36	380	33.35	2	<.001	<.001	358.78
Off-beat subtest	ADE	2726.71	378					304.55
	AE	2728.25	379	1.55	1	.214	.107	302.53
	E	2730.88	380	4.17	2	.124	.056	301.04
Out-of-key subtest	ACE	2946.35	378					414.37
	AE	2953.82	379	7.47	1	.006	.003	415.31
	CE	2946.35	379	0	1	1.00		411.58
	E	3004.57	380	58.22	2	<.001	<.001	437.88
Total score	ACE	2548.41	378					215.40
	AE	2556.46	379	8.05	1	.005	.003	216.63
	CE	2548.41	379	0	1	1.00		212.61
	E	2611.17	380	62.76	2	<.001	<.001	241.18

Models with $p > .05$ in the χ^2 test are written in bold text. -2LL, minus 2 log likelihood; df, degrees of freedom; Δ -2LL, the difference in the -2LL values against the full model; Δ df, the difference in degrees of freedom against the full model; p, "naive", non-corrected p-value; p*, corrected p-value, usually $p/2$, see Dominicus et al. (2006); BIC, Bayesian Information Criterion.

Table 5. The estimates for variance explained by different latent components in different models.

		A	C/D	E
Scale subtest	ACE	.50 (.00-.69)	.07 (.00-.50)	.43 (.31-.60)
	AE	.58 (.42-.70)	-	.42 (.30-.58)
Off-beat subtest	ADE	.00 (.00-.34)	.24 (.00-.44)	.76 (.56-.99)
	AE	.18 (.00-.39)	-	.82 (.61-1.00)
	E	-	-	1.00 (1.00-1.00)
Out-of-key subtest	ACE	.03 (.00-.46)	.59 (.19-.70)	.38 (.27-.51)
	CE	-	.61 (.49-.70)	.39 (.30-.51)
Total score	ACE	.01 (.00-.44)	.63 (.23-.73)	.36 (.25-.47)
	CE	-	.64 (.53-.73)	.36 (.27-.47)

Estimates are shown only for full models and those models whose fit to the data was not significantly worse than that of the corresponding full model. A = additive genetic effects; D = dominant genetic effects; C = common environmental effects; E = unique environmental effects

4. Discussion

The goal of the current study was to obtain new information about genetic and environmental effects on musical pitch and rhythm perception ability in the general population. The main finding of the study was the highly dissimilar pattern of results in the two subtests related to pitch perception. Common environmental effects were strong (.59 or .61 depending on the model) in the Out-of-key subtest, where the task was to detect pitch incongruences in single melodies. Genetic effects were only negligible (.03) in this subtest, which is surprising and somewhat contrary to previous findings (Drayna et al., 2001; Mosing et al., 2014; Pulli et al., 2008; Ukkola et al., 2009; Ullén et al., 2014). In contrast, genetic effects were strong (.50 or .58) in the Scale subtest, where the task was to compare two consecutive melodies for any pitch differences. Here, common environmental effects were marginal (.07). The differences in the cognitive demands of these two subtests are discussed more throughout below. In the Off-beat subtest, the task was to detect rhythmic incongruences in single melodies. Unique environmental effects were strong (.76–1.00) while genetic effects were small (.18 or .24) and common environmental effects were absent. This result is similar to that acquired by Ullén et al. (2014), although their data suggested stronger genetic effects. In the total score, common environmental effects were strong (.63 or .64) while genetic effects were minimal (.01). It is possible that the different proportions of genetic and environmental effects in the different subtests reflect the differences in the developmental and perceptual processes related to pitch and rhythm.

A strong correlation was found between the pitch-related subtests while the rhythm-related subtest showed only small or no correlation to the pitch-related subtests. Also Hausen et al. (2013) reported a similar pattern of correlations. These results support the distinction between pitch and rhythm perception abilities, which has been previously suggested by brain lesion studies (Ayotte et al., 2000; Di Pietro et al., 2004; Liegeois-Chauvel et al., 1998; Peretz, 1990; Peretz & Kolinsky, 1993; Piccirilli et al., 2000; Vignolo, 2003).

4.1. Genetic and environmental effects on pitch perception

Two of the three subtests, namely the Scale and Out-of-key subtests, were pitch-related. Interestingly, their results were very different. As was expected, genetic effects were

strong in the Scale subtest. This result is in line with previous findings (Drayna et al., 2001; Pulli et al., 2008; Ukkola et al., 2009; Ullén et al., 2014). The data also supported marginal common environmental effects. It is surprising that in the Out-of-key subtest common environmental effects were very strong while genetic effects were marginal. Common environmental effects of this magnitude are rare in twin studies (Boomsma et al., 2002; McGue & Bouchard Jr, 1998; Vinkhuyzen et al., 2009). The age of the participants makes this result even more astonishing: since they are already in their mid-thirties, most of them do not live together anymore and have more limited contact with each other than when they were children. The result is unlikely to be due to selection bias or methodological artefacts, because the other subtests showed distinct differences between the MZ and DZ pairs. Furthermore, the distribution of scores was similar for twins and healthy volunteers (Hausen et al., 2013) and did not differ by zygosity or between twins from pairs in which both participated and twins from pairs in which only one twin participated. Thus, the reason for these distinctive results most likely lies with differences in the cognitive demands of the tasks.

Some of the essential differences in the cognitive demands of the Scale and Out-of-key subtests are related to memory. In the Scale subtest, two consecutive melodies are compared, which requires working memory⁹. In the Out-of-key subtest, only one melody is presented and the participant has to rely on his/her own long-term memory trace (i.e. model) of musical key in order to recognize, whether a mistuned tone was introduced to a melody or not. Since these models of musical keys are not inborn but learned (Hannon & Trainor, 2007; Trainor, 2005; Trainor & Corrigall, 2010; Trainor et al., 2012; Trehub, 2003), the only way to acquire them is to be exposed to music, i.e. by environmental events. This might be the reason why common environmental effects explain most of the individual differences in the Out-of-key subtest, i.e. in the ability to detect tones that are out of key. While strong long-term memory traces of musical keys might enhance the performance in the Scale subtest too, it seems that other factors, which are more strictly controlled by genes, explain most of the individual differences. These factors could include e.g. pitch-related working memory capacity (see Marin & Perry, 1999) but also attention-related processes, such as vigilance or focused attention, could be important.

⁹ The association between working memory and music perception deficits have been found in patients who had acquired amusia following a stroke in the middle cerebral artery (Särkämö et al., 2008, Särkämö et al., 2010).

If the individual differences in the learned models of musical keys are the central factor causing common environmental effects in the Out-of-key subtest, then what are the environmental factors that affect the development of these models? Since the concept of musical key seems to be learned during early childhood (Corrigall & Trainor, 2010; Trainor & Trehub, 1992; Trainor & Trehub, 1994), it can be hypothesized that the amount and the quality of musical exposure during early childhood could be a major factor causing these differences. However, there seems to be no evidence about critical learning periods for musical scale structure, even though such evidence exists for more basic pitch acquisition (Trainor, 2005). Hausen et al. (2013) found that the years of received music education predicted the performance in the OTA. Especially the total score and the Off-beat subtest score were affected by the amount of *formal* music education: In both, each additional year of music education increased the score by approximately 0.3 points. In another study, Hausen, Saarikallio, Särkämö, and Tervaniemi (in preparation) found out that self-reported amount of daily musical activities (e.g. active listening to music, going to concerts, playing an instrument) did not correlate with the Out-of-key subtest. Thus, especially formal music education is could to be one of the major environmental factors causing individual differences in the Out-of-key subtest.

The results raise interesting questions when compared to the previous study by Drayna et al. (2001). They acquired notable genetic effects (.71 to .80) and no support for common environmental effects with a task that required participants to listen to popular melodies and to judge, whether they were played correctly or not. This task is more similar to the Out-of-key subtest than to the Scale subtest since "correctness" of a melody had to be judged using prior knowledge and not by comparing it with the one played just before it. Why this result was then not replicated in the current study with the Out-of-key subtest? The reason might be related to the difference in the nature of the stimuli used and participants' earlier exposure to them. Since Drayna et al. used highly popular and common melodies, it is most probable that even those participants' who were not active music listeners or practitioners had heard them many times during their lives. This goes to say that probably all listeners had acquired some form of a memory trace for most of the melodies. On the other hand, the stimuli used in the current study were unfamiliar melodies. This means that in the Out-of-key subtest, one actually has to compare each melody to a memory trace of musical key and not to a memory trace of a particular

melody, which might be more cognitively demanding. Interestingly, lesion studies and imaging studies of healthy participants support differences in the processing of highly familiar melodies to unfamiliar melodies. As mentioned in the introduction, the right temporal structures seem to be less critical in recognition of highly familiar tunes than in tasks involving new tunes (Peretz & Zatorre, 2005). Thus, tasks in the DTT and in the Out-of-key subtest are likely to be quite different after all, which would be in line with the studies mentioned in the Chapter 1.3.

The twin study by Ullén et al. (2014) provides another interesting comparison to the results of the current study. The Pitch subtest used in their study included a comparison task of two sinusoidal tone pairs, which are much shorter and simpler stimuli than the melodies used here. The estimates of genetic effects were .30 and .12, respectively for females and males, and the estimates of common environmental effects were .19 and .38, respectively for females and males. The Melody subtest used by Ullén et al. included a comparison task of two isochronous sequences of four to nine tones. The data supported strong genetic effects (.59) and no common environmental effects. The Melody subtest is more close to the Scale subtest used here, and the data support strong genetic effects for both. Thus, also the results of Ullén et al. seem to support the hypothesis that more working memory-dependent tasks are more heritable. Furthermore, they interpret that the even higher estimate of genetic effects acquired by Drayna et al. (2001) using the DTT is related to the fact that the DTT has longer stimuli than their Memory subtest, and thus would place a higher load on working memory and intelligence.

4.2. Genetic and environmental effects on rhythm perception

According to the results of this study, common environmental factors do not seem explain the individual differences in rhythm perception ability as measured with the Off-beat subtest. Also genetic effects seem to be fairly modest while unique environmental effects were estimated to be much larger in the Off-beat subtest than in the pitch-related subtests. Strong effects of unique environment can be interpreted in several ways.

One possible interpretation is that these unique environmental effects are caused by non-shared events. For example, twins could differ in the amount and the quality of their

musical activities, such as playing an instrument and listening to music, which could affect their rhythm perception ability. Given that the twins studied here were already in their mid-thirties, this is certainly possible. Hausen et al. (in preparation) found that the performance in the Off-beat subtest is positively associated with music-related activities in everyday life ($r = .39$), but not with formal music education. In their study, this association was strong especially between the Off-beat performance and the self-evaluated importance of music in daily life ($r = .49$), and the Off-beat performance and the amount of dancing or music-related sports activities ($r = .39$). The differences in these music-related activities could make twins different regarding rhythm perception. This interpretation naturally remains speculative and cannot be proven with the current data, which does not include information about twins' music-related activities in everyday life.

Another interpretation of the results is related to the validity and reliability of the subtest. It is possible that the task is more difficult to comprehend than other the tasks, and some of the participants simply failed to do so. This could create artificial variance in the subtest, which would not be explained by genetic or common environmental factors. The test-retest reliability of the separate subtests of the OTA has not been examined, and therefore this interpretation is difficult to test. Also, since the participants did the test at home and not in a laboratory, the possible problems understanding the task could not be discussed with the participants. It is worthwhile to note that the distribution of the scores in the Off-beat subtest was normal, which makes this interpretation unlikely. If some of the participants did not understand the task, one would expect to find a number of scores around random chance level, which was not the case here.

It should be kept in mind that homogeneity of variances could not be assumed when modelling this subtest regarding the distributions between the groups of first twins from each pair and of second twins from each pair. Since the order of the twins within each pair was randomized (i.e. based on their order in the data), it is likely that this violation is a result of a coincidence. This violation of assumptions can skew the estimates, but given the larger uncertainty related to them, it is unlikely to alter the way they are interpreted.

Ullén et al. (2014) reported fairly high (.50) estimate of heritability of rhythm perception with a task that involved comparison of two short rhythm sequences. If the estimate obtained here for genetic dominant effects (.24) for the Off-beat subtest is accurate, one

could hypothesize that the reason for higher heritability of the task used by Ullén et al. could be related to the fact that it is likely to be more demanding regarding the working memory. This explanation is parallel with the one suggested earlier for the results of the pitch-related subtests. However, one should be cautious about making too bold deductions based on the result obtained with the Off-beat subtest, because of the fact that the effects of unique environment, which seem to explain a great majority of the individual differences, could be a result of number of different factors.

4.3. Limitations of the study

The current study has some limitations related to the number of participants, the background information of the participants, control of the test situation, and to the representativeness of the Finnish population. These are all discussed separately.

The analyses were conducted by using data from 384 participants, 113 full pairs and 158 unpaired twins. Even though strong effects can be detected with this sample, smaller effects cannot be reliably confirmed. For example, in the Off-beat subtest the E model, which expects the genetic effects to be zero, did not have a poorer fit to the data than the ADE model, meaning that genetic effects could be absent in the subtest. A larger set of data would result in more accurate results (i.e. in narrower confidence intervals of the parameter estimates) in this and other subtests, as well as in the total score.

Another limitation is related to the information about the participants' musical background. In earlier FinnTwin16 studies extensive information from medical-social questionnaires have been collected during 5 waves at ages of 16, 17, 18.5, 24 and 34 years. For the current analysis, only minimal information of the participants was acquired, i.e. their age, sex, place of residence, education, information about their long-term illnesses and four self-evaluation questions about music. Since common environment explained over half of the population variation in the Out-of-key subtest and total score, it would be interesting to know what are the environment factors causing this phenomenon. For example, it would be beneficial to have more information about their formal music education, hobbies and music listening habits, but these were not the focus of the FinnTwin16 cohort study (Kaprio et al., 2002; Kaprio, 2013).

Third limitation of this study is related to the limited controllability of the test situation. It was possible to rule out participants who had not completed the whole test, had done so without using time required to listen to all of the stimuli or had performed markedly under chance level. However, since the participants did the test online, it was not possible to control whether the test situation was uninterrupted or not, what kind of audio equipment they used, did they properly understand the different tasks to name few examples. It is possible that these uncontrolled factors could have caused additional variance to the data. However, there is no reason to assume that these factors would differ for the groups of MZ and DZ twins. Thus, it is unlikely that any estimates of genetic or common environmental effects would be increased by such factors.

Fourth limitation is related to the representativeness of the study. The participants represent all geographical regions of Finland and all educational levels. Sixty-one per cent of the participants had a Bachelor's or higher degree, which means that the higher educational levels were overrepresented. However, the genetic modelling was also performed with educational level as a third covariate, and these results were almost identical to those presented here. Therefore, it is concluded that this limitation has only a minor or no effect on the estimates of genetic and environmental effects presented here.

As a final note, one should also keep in mind that the sample presented here consisted of twins aged 32–38 years, i.e. the age range was quite small. However, this should be considered as a deliberate choice rather than a limitation. Even though a wider age range would make it possible to study whether proportional amount of genetic and environmental effects vary as a function of age, this analysis would have required a much larger sample that was studied here.

4.4. Summary

The current study makes a significant contribution to the ongoing scientific discussion about the proportional genetic and environmental effects on music perception by being the first study to present evidence for strong common environmental effects in a pitch-related music perception task (the Out-of-key subtest). There were also strong common

environmental effects on the total score of the OTA music perception test. These results have an interesting novelty value as previous studies (Drayna et al., 2001; Ullén et al., 2014; Mosing et al., 2014) have provided support for genetic factors as the main contributors to the individual differences. However, there are differences in the cognitive demands of the Out-of-key subtest and the tests applied in the previous studies. Perhaps the most important of these differences is the need to rely on the long-term memory traces of musical scales or keys in the Out-of-key subtest. It is proposed that these differences explain the different results between the current study and the previous studies. Based on previous research (Hausen et al., 2013; Hausen et al., in preparation), it is presumed that formal music education is one of the key environmental factors causing the results in the Out-of-key subtest and total score.

The results of the current study regarding rhythm perception could be interpreted so that unique environmental factors cause most of the individual differences in this ability while the role of genetic factors would be less pronounced. However, given the potentially large measurement error in the Off-beat subtest, one should be cautious about making this conclusion. Furthermore, the difference in the twin correlations between the groups of MZ and DZ twins suggest the existence of dominant genetic effects, even though these were estimated to be modest. More research on rhythm perception in melodically complex material is required in order to clarify these results. In any case, music activity in everyday life seems to have some influence in this ability, while formal music education seems to have little effect (Hausen et al., in preparation).

Ultimately, quantitative genetic studies answer the question about the relative contributions of nature (genes) and nurture (environment) to a given trait. In their recent study, Mosing et al. (2014) took the position that nature (i.e. genetic effects) would prevail over nurture (i.e. amount of practise) in explaining the individual differences in music perception ability. The current study shows that with sufficiently complex melodic material environmental factors, perhaps including formal music education and music in everyday life, can in some music perception tasks explain more of the individual differences than genetic factors do. The sample of this study should be regarded as representing the Finnish population of the corresponding cohort, i.e. individuals aged 32–38 years. It is presumed that also in other countries embracing the tradition of Western music the results would be similar. However, this remains to be proven by future research.

References

- Abrahamson, A. C., Baker, L. A., & Caspi, A. (2002). Rebellious teens? Genetic and environmental influences on the social attitudes of adolescents. *Journal of Personality and Social Psychology*, *83*(6), 1392-1408.
- Ayotte, J., Peretz, I., Rousseau, I., Bard, C., & Bojanowski, M. (2000). Patterns of music agnosia associated with middle cerebral artery infarcts. *Brain: A Journal of Neurology*, *123*(9), 1926-1938.
- Baharloo, S., Service, S. K., Risch, N., Gitschier, J., & Freimer, N. B. (2000). Familial aggregation of absolute pitch. *The American Journal of Human Genetics*, *67*(3), 755-758.
- Brattico, E., Tervaniemi, M., Näätänen, R., & Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Research*, *1117*(1), 162-174.
- 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*(1), 100-130.
- Bischoff Renninger, L., Wilson, M. P., & Donchin, E. (2006). The processing of pitch and scale: An ERP study of musicians trained outside of the western musical system. *Empirical Musicology Review*, *1*(4), 185-197.
- Boomsma, D. I., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews Genetics*, *3*(11), 872-882.
- Boomsma, D. I., de Geus, E. J. C., van Baal, C. M., & Koopmans, J. R. (1999). A religious upbringing reduces the influence of genetic factors on disinhibition: Evidence for interaction between genotype and environment on personality. *Twin Research*, *2*(2), 115-125.
- Bouchard, T. J., & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of Neurobiology*, *54*(1), 4-45.
- Chang, H., & Trehub, S. E. (1977). Auditory processing of relational information by young infants. *Journal of Experimental Child Psychology*, *24*(2), 324-331.
- Corrigall, K. A., & Trainor, L. J. (2010). Musical enculturation in preschool children: Acquisition of key and harmonic knowledge. *Music Perception*, *28*(2), 195-200.
- D'Amato, M. R. (1988). A search for tonal pattern perception in cebus monkeys: Why monkeys can't hum a tune. *Music Perception*, *1*, 453-480.
- Di Pietro, M., Laganaro, M., Leemann, B., & Schnider, A. (2004). Receptive amusia: Temporal auditory processing deficit in a professional musician following a left temporo-parietal lesion. *Neuropsychologia*, *42*(7), 868-877.
- Dominicus, A., Skrondal, A., Gjessing, H. K., Pedersen, N. L., & Palmgren, J. (2006). Likelihood ratio tests in behavioral genetics: Problems and solutions. *Behavior Genetics*, *36*(2), 331-340.
- Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., & Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science*, *291*(5510), 1969-1972.

- Fisher, R. A. (1919). XV.—The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52(2), 399-433.
- Fitch, W. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, 100(1), 173-215.
- Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed., pp. 149-180). London: Academic Press.
- Fries, W., & Swihart, A. A. (1990). Disturbance of rhythm sense following right hemisphere damage. *Neuropsychologia*, 28(12), 1317-1323.
- Fry, D. (1948). An experimental study of tone deafness. *Speech*, 1, 4-11.
- Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: Premotor and striatal interactions in musicians and nonmusicians during beat perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(23), 7540-7548.
- Grahn, J. A., & Rowe, J. B. (2013). Finding and feeling the musical beat: Striatal dissociations between detection and prediction of regularity. *Cerebral Cortex*, 23(4), 913-921.
- Griffiths, T. D., Büchel, C., Frackowiak, R. S., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, 1(5), 422-427.
- Hagmann, C. E., & Cook, R. G. (2010). Testing meter, rhythm, and tempo discriminations in pigeons. *Behavioural Processes*, 85(2), 99-110.
- Hall, D. A., Johnsrude, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A., & Summerfield, A. Q. (2002). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 12(2), 140-149.
- Halsband, U., Ito, N., Tanji, J., & Freund, H. J. (1993). The role of premotor cortex and the supplementary motor area in the temporal control of movement in man. *Brain: A Journal of Neurology*, 116(1), 243-266.
- Hannon, E. E., & Johnson, S. P. (2005). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cognitive Psychology*, 50(4), 354-377.
- Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, 11(11), 466-472.
- Harrington, D. L., Haaland, K. Y., & Hermanowitz, N. (1998). Temporal processing in the basal ganglia. *Neuropsychology*, 12(1), 3-12.
- Hart, H. C., Palmer, A. R., & Hall, D. A. (2003). Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cerebral Cortex*, 13(7), 773-781.
- Hasegawa, A., Okanoya, K., Hasegawa, T., & Seki, Y. (2011). Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Scientific Reports*, 1(120), 1-12.
- Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., & Särkämö, T. (2013). Music and speech prosody: A common rhythm. *Frontiers in Psychology*, 4(566), 1-16.

- Hausen, M., Saarikallio, S., Särkämö, T. & Tervaniemi, M. (in preparation). Listen to the Beat: Perception of Musical Meter is Associated with Everyday Use of Music.
- Hauser, M. D., & McDermott, J. (2003). The evolution of the music faculty: A comparative perspective. *Nature Neuroscience*, 6(7), 663-668.
- Honing, H. (2012). Structure and interpretation of rhythm in music. In D. Deutsch (Ed.), *The Psychology of Music* (3rd ed., pp. 369-404). London: Academic Press.
- Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, 46(2), 632-639.
- Ibbotson, N., & Morton, J. (1981). Rhythm and dominance. *Cognition*, 9(2), 125-138.
- IBM Corp. (2013). *IBM SPSS Statistics for Macintosh* (22.0th ed.). Armonk, NY: IBM Corp.
- Ivry, R. B., & Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience*, 1(2), 136-152.
- Ivry, R. B., Keele, S., & Diener, H. (1988). Dissociation of the lateral and medial cerebellum in movement timing and movement execution. *Experimental Brain Research*, 73(1), 167-180.
- Johnson, W., Turkheimer, E., Gottesman, I. I., & Bouchard, T. J., Jr. (2010). Beyond heritability: Twin studies in behavioral research. *Current Directions in Psychological Science*, 18(4), 217-220.
- Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain: A Journal of Neurology*, 123(1), 155-163.
- Jones, M. R., Boltz, M., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. *Perception & Psychophysics*, 32(3), 211-218.
- Kalmus, H., & Fry, D. (1980). On tune deafness (dysmelodia): Frequency, development, genetics and musical background. *Annals of Human Genetics*, 43(4), 369-382.
- Kaprio, J. (2013). The Finnish twin cohort study: An update. *Twin Research and Human Genetics*, 16(1), 157-162.
- Kaprio, J., Pulkkinen, L., & Rose, R. J. (2002). Genetic and environmental factors in health-related behaviors: Studies on Finnish twins and twin families. *Twin Research*, 5(5), 366-371.
- Karma, K. (1984). Musical aptitude as the ability to structure acoustic material. *International Journal of Music Education*, 3, 27-30.
- Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience*, 17(10), 1565-1577.
- Lee, Y., Janata, P., Frost, C., Hanke, M., & Granger, R. (2011). Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *NeuroImage*, 57(1), 293-300.
- Liegeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain: A Journal of Neurology*, 121(10), 1853-1867.

- Loehlin, J. C., McCrae, R. R., Costa Jr, P. T., & John, O. P. (1998). Heritabilities of common and measure-specific components of the big five personality factors. *Journal of Research in Personality*, 32(4), 431-453.
- Marin, O. S. M., & Perry, D. W. (1999). Neurological aspects of music perception and performance. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed., pp. 653-724). London: Academic Press.
- McDermott, J., & Hauser, M. D. (2007). Nonhuman primates prefer slow tempos but dislike music overall. *Cognition*, 104(3), 654-668.
- McGue, M., & Bouchard Jr, T. J. (1998). Genetic and environmental influences on human behavioral differences. *Annual Review of Neuroscience*, 21(1), 1-24.
- Milner, B. (1962). Laterality effects in audition. *Interhemispheric Relations and Cerebral Dominance*, 1, 177-195.
- Mithen, S. J. (2005). *The singing neanderthals: The origins of music, language, mind, and body*. Cambridge: Harvard University Press.
- Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullen, F. (2014). Practice does not make perfect: No causal effect of music practice on music ability. *Psychological Science*, doi:0956797614541990 [pii]
- Neale, M. (2004). *Mx software*. Richmond, VA: Virginia Commonwealth University.
- Neale, M., & Cardon, L. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publishers.
- Patel, A. D. (2010). *Music, language, and the brain*. New York: Oxford University Press.
- Patel, A. D., & Demorest, S. M. (2012). Comparative music cognition: Cross-species and cross-cultural studies. In D. Deutsch (Ed.), *The Psychology of Music* (3rd ed., pp. 647-681). London: Academic Press.
- Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, 19(10), 827-830.
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4), 767-776.
- Penhune, V. B., Zatorre, R., & Feindel, W. (1999). The role of auditory cortex in retention of rhythmic patterns as studied in patients with temporal lobe removals including heschls gyrus. *Neuropsychologia*, 37(3), 315-331.
- Penhune, V. B., Zatorre, R. J., & Evans, A. C. (1998). Cerebellar contributions to motor timing: A PET study of auditory and visual rhythm reproduction. *Journal of Cognitive Neuroscience*, 10(6), 752-765.
- Penhune, V. B., & Doyon, J. (2002). Dynamic cortical and subcortical networks in learning and delayed recall of timed motor sequences. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 22(4), 1397-1406.

- Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, *113*(4), 1185-1205.
- Peretz, I. (2008). Musical disorders from behavior to genes. *Current Directions in Psychological Science*, *17*(5), 329-333.
- Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. *Annals of the New York Academy of Sciences*, *999*(1), 58-75.
- Peretz, I., Cummings, S., & Dubé, M. (2007). The genetics of congenital amusia (tone deafness): A family-aggregation study. *The American Journal of Human Genetics*, *81*(3), 582-588.
- Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., . . . Bouchard, B. (2008). On-line identification of congenital amusia. *Music Perception: An Interdisciplinary Journal*, *25*(4), 331-343.
- Peretz, I., & Kolinsky, R. (1993). Boundaries of separability between melody and rhythm in music discrimination: A neuropsychological perspective. *The Quarterly Journal of Experimental Psychology*, *46*(2), 301-325.
- Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology*, *56*, 89-114.
- Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., & Peretz, I. (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia*, *49*(5), 961-969.
- Piccirilli, M., Sciarra, T., & Luzzi, S. (2000). Modularity of music: Evidence from a case of pure amusia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *69*(4), 541-545.
- Pinker, S. (1999). How the mind works. *Annals of the New York Academy of Sciences*, *882*(1), 119-127.
- Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition*, *98*(1), 1-11.
- Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2008). DNA: The basis of heredity. In R. Plomin, J. C. DeFries, V. S. Knopik & J. M. Neiderhiser (Eds.), *Behavioral Genetics* (5th ed., pp. 39-49). New York: Worth Publishers.
- Pulli, K., Karma, K., Norio, R., Sistonen, P., Goring, H. H., & Jarvela, I. (2008). Genome-wide linkage scan for loci of musical aptitude in finnish families: Evidence for a major locus at 4q22. *Journal of Medical Genetics*, *45*(7), 451-456.
- Purcell, S. (2008). Statistical methods in behavioral genetics. In R. Plomin, J. C. DeFries, V. S. Knopik & J. M. Neiderhiser (Eds.), *Behavioral Genetics* (5th ed., pp. 357-411). New York: Worth Publishers.
- Rao, S. M., Mayer, A. R., & Harrington, D. L. (2001). The evolution of brain activation during temporal processing. *Nature Neuroscience*, *4*(3), 317-323.

- Rao, S. M., Harrington, D. L., Haaland, K. Y., Bobholz, J. A., Cox, R. W., & Binder, J. R. (1997). Distributed neural systems underlying the timing of movements. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(14), 5528-5535.
- Sarna, S., Kaprio, J., Sistonen, P., & Koskenvuo, M. (1978). Diagnosis of twin zygosity by mailed questionnaire. *Human Heredity*, *28*(4), 241-254.
- Schubotz, R. I., Friederici, A. D., & Yves von Cramon, D. (2000). Time perception and motor timing: A common cortical and subcortical basis revealed by fMRI. *NeuroImage*, *11*(1), 1-12.
- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, *32*(5), 771-783.
- Schwekendiek, D., & Pak, S. (2009). Recent growth of children in the two Koreas: A meta-analysis. *Economics & Human Biology*, *7*(1), 109-112.
- Seashore, C. E., Lewis, D., & Saetveit, J. (1960). *Measures of musical talents (test)*. New York: Psychological Corporation.
- Shofner, W. P. (2005). Comparative aspects of pitch perception. In C. J. Plack, A. J. Oxenham, R. R. Fay & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 56-98). New York: Springer.
- Silventoinen, K., & Kaprio, J. (2008). Kaksos- ja perhetutkimukset geneettisten ja ympäristötekijöiden vaikutuksen arvioimisessa. *Sosiaalilääketieteellinen Aikakauslehti*, *45*(3).
- StataCorp. (2009). *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.
- Thivard, L., Belin, P., Zilbovicius, M., Poline, J., & Samson, Y. (2000). A cortical region sensitive to auditory spectral motion. *Neuroreport*, *11*(13), 2969-2972.
- Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, *107*(4), 885.
- Trainor, L. J. (2005). Are there critical periods for musical development? *Developmental Psychobiology*, *46*(3), 262-278.
- Trainor, L. J., & Hannon, E. E. (2012). Musical development. In D. Deutsch (Ed.), *The Psychology of Music* (3rd ed., pp. 423-497). London: Academic Press.
- Trainor, L. J., & Corrigall, K. A. (2010). Music acquisition and effects of musical experience. *Music Perception*, *36*, 89-127.
- Trainor, L. J., Marie, C., Gerry, D., Whiskin, E., & Unrau, A. (2012). Becoming musically enculturated: Effects of music classes for infants on brain and behavior. *Annals of the New York Academy of Sciences*, *1252*(1), 129-138.
- Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(2), 394.

- Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in western tonal music: Developmental perspectives. *Perception & Psychophysics*, *56*(2), 125-132.
- Trehub, S. E. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences*, *930*(1), 1-16.
- Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience*, *6*(7), 669-673.
- Trehub, S. E., Schellenberg, E. G., & Kamenetsky, S. B. (1999). Infants' and adults' perception of scale structure. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 965-975.
- Trehub, S. E., & Thorpe, L. A. (1989). Infants' perception of rhythm: Categorization of auditory sequences by temporal structure. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, *43*(2), 217-229.
- Trehub, S. E., & Trainor, L. (1998). Singing to infants: Lullabies and play songs. *Advances in Infancy Research*, *12*, 43-78.
- Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, *9*(5), 160-164.
- Ukkola, L. T., Onkamo, P., Raijas, P., Karma, K., & Järvelä, I. (2009). Musical aptitude is associated with AVPR1A-haplotypes. *PLoS One*, *4*(5), e5534.
- Ullén, F., Mosing, M. A., Holm, L., Eriksson, H., & Madison, G. (2014). Psychometric properties and heritability of a new online test for musicality, the Swedish Musical Discrimination Test. *Personality and Individual Differences*, *63*, 87-93.
- Vandenberg, S. G. (1962). The hereditary abilities study: Hereditary components in a psychological test battery. *American Journal of Human Genetics*, *14*(2), 220-237.
- Vignolo, L. A. (2003). Music agnosia and auditory agnosia. *Annals of the New York Academy of Sciences*, *999*(1), 50-57.
- Vinkhuyzen, A. A. E., Van der Sluis, S., Posthuma, D., & Boomsma, D. I. (2009). The heritability of aptitude and exceptional talent across different domains in adolescents and young adults. *Behavior Genetics*, *39*(4), 380-392.
- Visscher, P. M., Medland, S. E., Ferreira, M. A., Morley, K. I., Zhu, G., Cornes, B. K., . . . Martin, N. G. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genetics*, *2*(3), e41.
- Wilson, S. J., Pressing, J. L., & Wales, R. J. (2002). Modelling rhythmic function in a musician post-stroke. *Neuropsychologia*, *40*(8), 1494-1505.
- Wing, H. D. (1970). *Standardised tests of musical intelligence*. London: Nelson Publishing Company.
- Winkler, I., Haden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(7), 2468-2471.

- Zarco, W., Merchant, H., Prado, L., & Mendez, J. C. (2009). Subsecond timing in primates: Comparison of interval production between human subjects and rhesus monkeys. *Journal of Neurophysiology*, *102*(6), 3191-3202.
- Zatorre, R. J. (1985). Discrimination and recognition of tonal melodies after unilateral cerebral excisions. *Neuropsychologia*, *23*(1), 31-41.
- Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *The Journal of the Acoustical Society of America*, *84*(2), 566-572.
- Zatorre, R. J. (2003). Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience*, *6*(7), 692-695.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*(1), 37-46.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8*(7), 547-558.
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*(10), 946-953.
- Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *14*(4), 1908-1919.
- Zatorre, R. J., & Samson, S. (1991). Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain: A Journal of Neurology*, *114*(6), 2403-2417.
- Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(13), 5768-5773.

Appendix I. Additional model-comparison test statistics

Model	-2LL	df	Δ -2LL	Δ df	p
Scale subtest					
saturated	2807.41	372			
fixed variances between groups of MZ twins and DZ twins	2809.00	375	1.60	3	.660
fixed variances between groups of 1st twins and 2nd twins	2808.83	374	1.42	2	.491
fixed means between groups of MZ twins and DZ twins	2811.29	375	3.88	3	.274
fixed means between groups of 1st twins and 2nd twins	2808.04	374	0.63	2	.728
ACE	2813.01	378	5.60	6	.469
Off-beat subtest					
saturated	2714.61	372			
fixed variances between groups of MZ twins and DZ twins	2721.23	375	6.62	3	.085
fixed variances between groups of 1st twins and 2nd twins	2721.13	374	6.52	2	.038
fixed means between groups of MZ twins and DZ twins	2716.01	375	1.40	3	.707
fixed means between groups of 1st twins and 2nd twins	2715.83	374	1.22	2	.543
ACE	2726.71	378	12.10	6	.060
Out-of-key subtest					
saturated	2941.34	372			
fixed variances between groups of MZ twins and DZ twins	2941.85	375	0.51	3	.916
fixed variances between groups of 1st twins and 2nd twins	2941.78	374	0.45	2	.800
fixed means between groups of MZ twins and DZ twins	2945.89	375	4.56	3	.207
fixed means between groups of 1st twins and 2nd twins	2945.24	374	3.91	2	.142
ACE	2946.35	378	5.01	6	.542
Total score					
saturated	2546.69	372			
fixed variances between groups of MZ twins and DZ twins	2547.03	375	0.34	3	.952
fixed variances between groups of 1st twins and 2nd twins	2546.93	374	0.23	2	.891
fixed means between groups of MZ twins and DZ twins	2548.10	375	1.40	3	.705
fixed means between groups of 1st twins and 2nd twins	2546.75	374	0.05	2	.975
ACE	2548.41	378	1.71	6	.944
Models with $p < .05$ in the χ^2 test are written in bold text. -2LL, minus 2 log likelihood; df, degrees of freedom; Δ -2LL, difference in -2LL against the saturated model. It follows a χ^2 distribution of Δ df degrees of freedom; Δ df, difference in degrees of freedom against the saturated model; p, p-value.					