JYVÄSKYLÄ STUDIES IN HUMANITIES
173

Vinoo Alluri

Acoustic, Neural, and Perceptual Correlates of Polyphonic Timbre

JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN HUMANITIES 173

Vinoo Alluri

Acoustic, Neural, and Perceptual Correlates of Polyphonic Timbre

Esitetään Jyväskylän yliopiston humanistisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa S212
helmikuun 28. päivänä 2012 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Humanities of the University of Jyväskylä,
in building Seminarium, Auditorium S212, on February 28, 2012 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2012

Acoustic, Neural, and Perceptual Correlates of Polyphonic Timbre

JYVÄSKYLÄ STUDIES IN HUMANITIES 173

Vinoo Alluri

Acoustic, Neural, and Perceptual Correlates of Polyphonic Timbre

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2012

Editors
Petri Toiviainen
Department of Music, University of Jyväskylä
Pekka Olsbo, Ville Korkiakangas
Publishing Unit, University Library of Jyväskylä

Cover picture by Varun Kapoor

URN:ISBN:978-951-39-4654-8
ISBN 978-951-39-4654-8 (PDF)
ISSN 1459-4331

ISBN 978-951-39-4653-1 (nid.)
ISSN 1459-4323

Jyväskylä University Printing House, Jyväskylä 2012

To Neha

"There is a creeper - I think, it is called the morning glory - which has that extraordinary pale blue color that only flowers have, or a deep purple with a touch of mauve, or a peculiar white. Only living flowers have those colors. They come, they bloom in the morning - the trumpet-shaped flowers - and then within a few hours they die. You must have seen those flowers. In their death they are almost as beautiful as when they are alive. They bloom for a few hours and cease to be, and in their death they do not lose the quality of a flower."

-- Jiddu Krishnamurti, 1965

ABSTRACT

Alluri, Vinoo
Acoustic, Neural, and Perceptual Correlates of Polyphonic Timbre
Jyväskylä: University of Jyväskylä, 2012, 76 p.
(Jyväskylä Studies in Humanities
ISSN 1459-4331; 173)
ISBN 978-951-39-4653-1 (nid.)
ISBN 978-951-39-4654-8 (PDF)

The central focus of the thesis is to investigate timbre perception in a polyphonic context using an interdisciplinary approach that aims at combining the behavioural, acoustic and neural domains. The current work first deals with the gap concerning the connection between monophonic timbre and emotions followed by a thorough investigation of polyphonic timbre perception. The investigation concerning the link between monophonic timbre and emotion-detection revealed that two main dimensions, representing perceived Valence and Energy Arousal sufficiently captured the monophonic timbre affect space. The consistent patterns displayed by the acoustic features suggested that there exist universal perceptual mechanisms in play during affect assessments.

With respect to determining the dimensionality of the perceptual polyphonic timbre space, two main dimensions were established, one representing perceived activity and the other representing perceived brightness. Subsequently, spectrotemporal modulations, represented by the new set of acoustic features introduced, the Sub-Band Fluxes, were found to significantly correlate with perceptual ratings. The cross-cultural investigation performed for the sake of generalization of the perceptual dimensions revealed that the dimensionality of the polyphonic timbre space increased from two, representing Activity and Brightness again, to three based on prior exposure via enculturation and musical training. The third dimension, however, did not show any clear pattern although it was found to represent perceived fullness in the case of musicians. The acoustic features displayed similar correlation patterns with the perceptual dimensions across participant groups suggesting common perceptual mechanisms at play.

The neural underpinnings of timbre processing were uncovered through a functional magnetic resonance imaging (fMRI) study conducted in a naturalistic setting by using realistic stimuli. Results revealed, in addition to sensory areas, the involvement of cognitive areas of the cerebellum, and default-mode network cerebrocortical areas in timbre processing. Overall, the results lend support to the notion that the perceptual mechanisms involved in the processing of polyphonic timbre dimensions are fairly universal albeit with minor differences and that they highly respond to spectrotemporal modulations or change. Finally, the methodological contributions of this dissertation have potential implications in the field of cognitive neuroscience of music.

Keywords: monophonic timbre, polyphonic timbre, acoustic features, perceptual dimensions, fMRI, non-parametric statistical significance estimation

Author's address Vinoo Alluri
Department of Music
University of Jyväskylä
Finland

Supervisors Professor Petri Toiviainen
Department of Music
University of Jyväskylä
Finland

Docent Elvira Brattico
Cognitive Brain Research Unit
University of Helsinki
Finland

Reviewers Professor Jean-Julien Aucouturier
Department of Computer Science
Temple University
Japan

Professor Peter Vuust
Center for Functionally Integrative Neuroscience
Aarhus University
Danish Royal Academy of Music
Denmark

Opponent Professor Stephen E. McAdams
Department of Music Theory
Schulich School of Music
McGill University
Canada

ACKNOWLEDGEMENTS

This research was conducted from January 2008 to September 2011 at the Department of Music at the University of Jyväskylä. The research was supported by the Centre of Excellence in Interdisciplinary Music Research funded by the Academy of Finland (project number 7118616). I am fortunate to have been jointly supervised by Professor Petri Toiviainen and Docent Elvira Brattico. Prof. Toiviainen's mentoring has been invaluable. My deepest respects for his clarity in thought and humility. I cherish all the discussions I had with him over coffee that resulted in the birth of new perspectives and ideas. I thank him for being so patient with me, trusting me, and giving the freedom to explore and experiment, and make mistakes! I highly appreciate the guidance provided by Docent Brattico, especially in the field of neuroscience, which I was alien to at the beginning of my PhD. The enthusiasm that she exudes whenever a new idea is discussed makes one believe that anything is possible, and has often been a source of encouragement. I learnt a lot from both of them and my heartfelt gratitude to them for their scientific guidance and support. Many thanks to collaborators Professor Iiro Jääskeläinen, Enrico Glerean and Professor Mikko Sams for their critical insight.

I consider it an honour that my thesis was reviewed by eminent professors, Professor Peter Vuust and Professor Jean-Julien Aucouturier, and extend my gratitude for their time and interest. Their appreciation for my work has made me value the four years of hard work even more! I wish to convey my sincere gratitude to Professor Stephen McAdams for agreeing to be my opponent. It is truly an honour.

A big thank you to my roommates at work, Jonna, Marc and Rafael, for tolerating my occasional verbal outbursts and messy desk for all these years. My gratitude goes to Tuomas Eerola for facilitating my initial integration into the CoE team. Working with him has played a vital role in shaping my thesis. Thank you Geoff for all the proofreading, Olivier Lartillot for his suggestions and help in computational analysis, and Markku Pöyhönen for all the support.

Most importantly, I extend my heartfelt gratitude to my friends for their love and support who made me feel at home. Thanks Maria, Kimmo, Jonna and Kaisa for giving me a Finnish family, and Ananda for her gastronomical support that often convinced me that I am at the right place at the right time! Finally, none of this would have been possible if not for all the love and support from my family.

Jyväskylä, 26 January, 2012

CONTENTS

LIST OF PUBLICATIONS

List of publications which are included in this thesis:

I Eerola, T., Ferrer, R., & Alluri, V. (In press). Timbre and affect dimensions: Evidence form affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception*.

II Alluri, V. & Toiviainen, P. 2010. Exploring Perceptual and Acoustic Correlates of Polyphonic Timbre. *Music Perception*, *27(3),* 223–241.

III Alluri, V. & Toiviainen, P. 2012. Effect of enculturation on the semantic and acoustic correlates of polyphonic timbre. *Music Perception*, *29(3),* 297-310.

IV Alluri, V., Toiviainen, P., Jääskeläinen, I., Glerean, E., Sams, M. & Brattico, E. 2012. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, *59*, 3677-3689. doi:10.1016/j.neuroimage.2011.11.019.

AUTHOR'S CONTRIBUTIONS

In Publication I, the author was responsible for the design of stimuli for experiment 3, and co-contributed to data collection, and part of the writing. In Publication II, the author was responsible for designing the study, collecting empirical data, analysis of behavioural and acoustic data and writing the article. In Publication III, the author was responsible for designing the study, collecting empirical data, analysis of behavioural and acoustic data and writing the article. In Publication IV, the author was responsible for conducting the behavioural experiment, analysing the behavioural and acoustic data. Additionally, all the processing and analyses performed on the neural data following the preprocessing stage of the fMRI data was designed and implemented by the author. The author played the main role in writing.

1 INTRODUCTION

Music is known to be an important facet of all human cultures (Merriam, 1964). From an evolutionary context, two opposing views of music have been proposed. Pinker (1997) contends that music fails to have evolutionary significance and could be considered merely as ”auditory cheesecake” and has no survival value. Huron (2001) outlines several perspectives concerning the role of music making and listening and contends that despite any convincing evidence of the evolutionary origins of music, its ubiquity since ancient times renders it a vital element of human existence and would be unjustified to rule out a possible evolutionary origin. In addition, several studies have proposed that music making and listening are fundamentally inherent processes in humans (Miller, 2000). For instance, Zentner and Eerola (2010) found that moving to music is an inherent trait displayed by humans beginning as early as infancy. Trehub (2000) outlines several studies on infants' perception of musical elements that purport the notion of music processing as an innate automatic process. More importantly, recent evidence from the neural perspective presents the vital role of music in evoking pleasurable feelings as those of euphoria and craving which occur as a result of the activation of the reward circuitry of the brain (Salimpoor et al., 2011). Although it can be argued that pleasure may not the most fundamental need of human life, needless to say it remains an important desirable human experience. Furthermore, music has been considered as a means for investigating other cognitive functions of the human brain and hence human behaviour (Brown, Merker & Wallin, 2001, Origins of music).

Music is composite of several perceivable elements of varying levels of abstraction, such as loudness, timbre, pitch, and rhythm. These elements can be classified based on their abstraction levels as being low-level or high-level. Low-level elements are those that are perceived in a bottom-up fashion without a need for domain-specific knowledge. For instance, loudness, pitch and timbre processing automatically recruit sensory mechanisms, and are performed rapidly in very short-time spans. On the other hand, rhythm encapsulates context-dependent aspects of music and recruits perceptual processes that are top-down in nature, and require a longer time-span. A fundamental question often posed is concerning the universality in music structure and its perception, in other words to what extent it is innate and hard-wired or is shaped by the environment, and is regarded as an instance thereof of the nature versus nurture problem. Despite the diversity in music across cultures, do the perceptual processes share fundamental biological sensory processing mechanisms? Moreover, the perceptive and cognitive mechanisms of a person can be affected by prior exposure in the form of enculturation or musical expertise. Evidence from infant studies indicates the presence of universals in the processing of some music principles (see Trehub, 2000 for an overview), such as categorization of successive events based on similarity in pitch, timbre or loudness, interval categorization based on consonance or dissonance, melody contour equivalence regardless of pitch transposition, and rhythm equivalence despite tempo changes. Harwood (1976) delineates a few elements of auditory perception that are found to be similar across cultures such as categorical pitch perception, judgments of octave equivalence, auditory stream segregation, and melodic contour processing. Carterette & Kendall (1999) outline some structural elements of music that are common across cultures such as the use of stable reference pitch and reference pulse, and the division of octave into scale steps, for instance. The authors contend that the process of music cognition in fact shares universals and advocate the notion that the differences in musical elements across cultures are but an "elaboration of a few universals".

Results of some perceptual studies support the view that there exist cross-cultural regularities in the perception of certain structural elements of music such as beat (Toiviainen & Eerola, 2003), and intervals (Smith & Williams, 1999), as well as in the elicitation of melodic expectations (Krumhansl, Toivanen, Eerola, Toiviainen, Jarvinen & Louhivuori, 2000; Eerola, Louhivuori & Lebaka, 2009; Castellano, Bharucha, & Krumhansl, 1984), and affect identification (Balkwill & Thompson, 1999; Balkwill, Thompson & Matsunaga, 2004; Gregory & Varney, 1996; Meyer, Palmer & Mazo, 1998; Adachi, Trehub & Abe, 2004; Zacharopoulou & Kyriakidou, 2009). Nevertheless, in addition to commonalities, some of these studies highlight differences in music perception and cognition, and associate them to one's culture-specific knowledge, for instance concerning melodic expectancies (Krumhansl et al., 2000; Curtis & Bharucha, 2009).

Furthermore, Demorest, Morrison, Beken & Jungbluth (2008) contend that the universality in the perception of musical properties suggested by several studies may be attributed to their major reliance on music following rules of the Western diatonic system. Evidence from cross-cultural studies in the field of neuroscience that have investigated the issue of universality in music processing suggest the existence of similarities in neural processing despite varying cultural backgrounds, albeit with minor differences (Nan, Knosche & Friederici, 2006; Morrison, Demorest, Aylward, Cramer and Maravilla, 2003). Recently, Nan, Knosche, Zysset & Friederici (2008) provided evidence for the first time illustrating effects of musical style acculturation on some motor and auditory areas of the brain. In summary, evidence from both behavioral and neural studies suggests macro level similarities with differences being observed on a more micro level.

These differences could be attributed to one of the mechanisms involved in perceptual learning, that is, differentiation (Goldstone, 1998). Goldstone (1998) discusses *perceptual learning* as a process how an organism's perceptual system undergoes long-lasting changes during the process of perceiving and interacting with its surrounding. The mechanism that appears to be analogous to the process of perceptual dimension identification is *differentiation.* Goldstone (1998) discusses *differentiation* as a mechanism enabling perceptual learning wherein prior exposure or training allows for the dissociation of perceptual elements from a single stimulus. A subcategory of this kind of perceptual learning is *differentiation of dimensions*, which refers to separation of perceptual dimensions and increase in the number thereof as a result of prior exposure or training. Differentiation of dimensions as a form of perceptual learning has not been widely studied in musical contexts to date.

Musical elements possess representations in three domains, namely the acoustic, neural and behavioural. The behavioural domain deals with mental representations of musical elements. The acoustical domain deals with physical representations of musical elements. Acoustic features that are based on psychoacoustic principles serve as the physical representations. Perceptual modelling aims at bridging the gap between the behavioural and computational domains by enabling associations between semantics and physical parameters. On the other hand, in the neural domain, studies aim at localizing brain areas pertaining to the processing of musical elements. Specifically, the neural studies have investigated the processing of musical elements employing highly controlled auditory paradigms in which these elements have been presented in isolation and manipulated artificially. Thus, as a result, such settings hinder us from obtaining a true understanding of the neural underpinnings of the perceptual processes involved in music feature processing.

In order to acquire a comprehensive picture of perceptual processes, as is the need in the field of music cognition, there exists a dire need to find the links between these three domains. However, the dearth in the number of studies that combine knowledge from all three domains poses a challenge in establishing the study designs in addition to methodological concerns. Nevertheless, in order to uncover in a holistic fashion the mechanisms involved in the processing of musical elements, investigations by integrating all three domains is elemental.

While many studies have focused on understanding the universality of the perceptual and cognitive processes of higher-level features such as harmony, melody, and rhythm, the perceptual aspects of timbre, have received much less attention. The aim of the thesis is to investigate timbre perception by employing an interdisciplinary approach combining the behavioural, acoustic and neural domains. The following chapter outlines the notion and importance of timbre, and summarizes the current state of affairs in the field of timbre perception.

Chapter three provides an overview of the research aims. Chapter four describes the methods used and developed employed in the studies that have been carried out. Chapter five provides brief summaries of all the studies. Chapter six concludes the dissertation by discussing the results, implications, limitations and future plans of the thesis.

2 TIMBRE

Timbre has been one of the most ill-defined musical concepts. Several definitions of timbre have been proposed over the years but there fails to exist one single, widely accepted definition. The classical definition of timbre put forth by the American Standards Association (ASA, 1960) is “that attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar”. Timbre is also the attribute of sound that has been termed as ‘a vehicle for source identity’ (McAdams and Giordano, 2009; McAdams, 1993; Handel 1995). It has also been considered as a “major structuring force in music and one of the most important and ecologically relevant features of auditory events” (Menon, Levitin, Smith, Lembke, Krasnow, Glazer et al., 2002, p. 1742). Timbre is the property that allows listeners to categorize and stream sound information and thereby form a mental representation of one’s surroundings. Among all the basic elements of music, timbre can considered to play a vital role in human survival. For instance, imagine you are crossing a busy street full of cars and the sound of a speeding car alerts you to make sure you are not in its way. The process of escaping danger elicits affect related to the feeling of reward, which can be thought of as a fundamental incentive to human actions. The affect-elicitation capabilities of timbre can be further discussed in a musical context. The same song sung by different artists despite the exact harmonic and rhythmic arrangements can engender different affects in the listeners. Apart from its survival value and affect-elicitation capabilities, the physical properties of timbre have also been shown to be important in enabling the communication and identification of affect in speech studies (Juslin & Laukka, 2003; Laukka, Juslin & Bresin, 2005; Scherer & Oshinsky 1977) and nonverbal affect vocalizations (Belin, Fillion-Bilodeau, & Gosselin, 2008; Bradley, 2000; Redendo, Fraga, Padron, & Pineiro, 2008).

Historically timbre has generally been defined and investigated as a phenomenon in the context of isolated instrument sounds. However, in real life, we generally encounter combinations of sounds resulting in complex emerging timbral soundscapes, for instance that of a bustling market place, or a busy highway. Furthermore, in the context of music, several instrument timbres blend resulting in an overall emergent timbre, such as that of a jazz ensemble, a rock concert, or a symphony. This approach to timbre perception as a conglomerate of individual timbres has been largely neglected.

Ensemble timbres or timbre blends can generally be divided into three categories, namely *timbral heterogeneity, timbral augmentation* and *emergent timbres* (Sandell, 1995). Timbre heterogeneity occurs when concurrently sounding timbres are perceived as separate entities or distinct sources, for example a piano and horn. Timbral augmentation occurs when one instrument merely acts as an embellishment to another dominant timbre, such as a flute and violin, which may sound more like a violin with an extra airy sound. Timbral emergence transpires when the constituent timbres fuse to form a novel overall timbre.

These three divisions of timbre can be further discussed in light of auditory grouping, that is, *perceptual segregation* and *perceptual fusion,* put forth by Bregman (1990) and Darwin and Carlyon (1995). Auditory grouping is the process wherein the brain perceives sounds and decides to group them either as belonging to individual sources or into one overall 'chimeric' or 'fictitious' (Bregman, 1990) or 'virtual' (McAdams, 1984) sound source. For instance, two violins playing two different pieces simultaneously may be perceived by listeners as two separate perceptual auditory streams or sound sources despite them belonging to the same timbre class, thereby in a way maintaining timbral heterogeneity. Another instance of perceptual segregation is the cocktail party effect, which is the ability to focus on one talker in a party amidst several other conversations and background noises. On similar lines, the listener possesses the ability to employ different listening strategies to either focus on one single sound source or to group them to perceive them as a whole. On the other hand, multiple instruments can perceptually fuse in listeners' minds to form a coherent concept, for instance, the overall sound of a string orchestra. Such Gestalt-like grouping leads to what is described by Bregman as a chimeric sound source. Another example is the emerging sound of a busy market place, wherein a combination of several individual timbres results in a soundscape or timbral environment (Ferrer, 2009), which forms a basis for identifying and categorizing sensory auditory information on a macroscopic or Gestalt level. The above instances are apt examples of timbral augmentation and emergent timbres. Moreover, timbral augmentation can be considered as a special reduced case of timbral emergence wherein the dominant timbre plays the main role in the emerging timbre.

For the sake of differentiating these two main categories of individual timbres and emerging timbres, they will be termed for the reminder of this work as monophonic timbre and polyphonic timbre. The terms 'monophonic' and 'polyphonic' should not be confused with the music theoretical term of polyphony versus homophony or monophony, rather only as a means to distinguish the two notions of timbre blends.

2.1 Monophonic Timbre

2.1.1 Perceptual and Acoustic Correlates of Monophonic Timbre

Research on monophonic timbre has a history, starting from the 70s, of perceptual similarity experiments of single instrument sounds, which concentrated on creating meaningful timbre spaces (Grey, 1977; Grey & Gordon, 1978; Iverson & Krumhansl, 1993; Lakatos, 2000; McAdams, Winsberg, de Soete, & Krimphoff, 1995). A timbre space can be thought of as a geometrical construct that aims at capturing a mental representation of the stimuli through the ratings of perceived similarity of the sounds. This was achieved via multidimensional scaling of similarity ratings of the stimuli collected from a number of listeners. In addition, the aforementioned studies aimed at finding acoustic features that best correlate with the perceptual dimensions.

Several studies performed on monophonic timbre have reported that the spectral centroid, which can be conceptualized as the center of gravity of the spectrum, explains one of the timbre space dimensions (Grey & Gordon, 1978; Iverson & Krumhansl, 1993; McAdams et al., 1995) and is often referred to as a measure of perceived 'brightness' (Beauchamp, 1982; De Poli & Pradoni, 1997). However, interpretations of the other timbral space dimensions have lacked consensus, probably because the nature of the stimuli used in each experiment determines the perceptual and acoustic features that explain the resulting dimensions and thus cannot be generalized across other sets of stimuli. Nonetheless, in a meta-analysis of timbre spaces presented by McAdams (1999), the author argues that despite the diversity in the stimuli, the actual perceived similarity remains fairly constant irrespective of the context it is presented in. Various spectral and temporal features of the stimuli have been found to explain the dimensions of perceptual timbre spaces, such as the log-attack time, spectral flux, attack synchrony, and spectral irregularity, to name a few. In addition to the aforementioned features, the Mel-Frequency Cepstral Coefficients (MFCC), which are described in later sections, appear to be quite prominent in characterizing timbre (De Poli & Pradoni, 1997; Terasawa, Slaney, & Berger, 2005).

A few studies have aimed at creating 'physical' timbre spaces through self-organizing maps (SOM) from features of the audio in order to find correlations with perceptual timbre spaces (Cosi, De Poli, & Lauzzana, 1994; De Poli, Prandoni, & Tonella, 1993; Loureiro, de Paula, & Yehia, 2004; Toiviainen, Kaipainen, & Louhivuori, 1995). The clustering of sounds in these studies has been found to be comparable to human similarity judgments.

2.1.2 Monophonic Timbre Semantics

Previous studies concerning timbre semantics mainly focused on finding semantic labels that best characterize monophonic timbre. Commonly cited is the work of von Bismarck (1974) regarding adjectives that describe timbre

(Darke, 2005; Disley & Howard, 2004; Disley, Howard, & Hunt, 2006; Movarec & Stepanek, 2003; Nykanen & Johansson, 2003; Pratt & Doak, 1976). Bismarck suggested a subset of four scales (dull-sharp, compact-scattered, full-empty, and colorless-colorful) to describe the timbre of single instrument sounds. Some studies have aimed at finding consistently used adjectives to describe timbre (Darke, 2005; Disley et al., 2006; Lukasik, 2005; Moravec & Stepanek, 2003; Nykanen & Johansson, 2003; Sarkar, Vercoe, & Yang, 2007). Commonly, such studies have used one of the two rating methods: Verbal Attribute Magnitude Estimation (VAME) (Kendall & Carterette, 1993a) and Semantic Differentials (Osgood, Suci, & Tannenbaum, 1957). The VAME uses scales that quantify the applicability of each adjective or descriptor (e.g., bright <-> not bright) rather than using opposites or bipolar scales (e.g., bright <-> dull). The participants in these experiments are typically required to rate the applicability of these adjectives to monophonic instrument sounds on such scales. Some studies that used the VAME approach report significant negative correlation between descriptors such as bright and dull (Disley & Howard, 2004; Disley et al., 2006) as well as bright and dark (Lukasik, 2005). Nevertheless, the use of semantic differentials, or bipolar scales, has been called into question, since the descriptors on each end of the scale are not necessarily opposites of each other (Darke, 2005; Kendall & Carterette, 1993a). However, this technique does enable controlled judgment of semantic concepts, and is advocated since it proves to be quite adaptable in these kinds of perceptual tasks (Pratt & Doak, 1976).

In several studies, the most common descriptor of monophonic timbre has been reported to be brightness. Helmholtz (1885/1954) uses the term 'bright' or 'brilliant' to classify the musical quality of a tone. In more recent studies, either the word 'bright' (Darke, 2005; Disley & Howard, 2004; Disley et al., 2006; Gounaropoulus & Johnson, 2006) or bipolar scales such as 'dark-bright' (Lukasik, 2005; Sethares, 1998) or variants of it such as 'brilliant-dull' (Pratt & Doak, 1976) have been used to describe monophonic timbre. 'Fullness' has also been used to describe a perceptual aspect of timbre (von Bismarck, 1974; Darke 2005). The term 'thin,' which can be regarded as an antonym for full, appears in some studies (Darke, 2005; Disley & Howard, 2004; Disley et al., 2006). Additionally, Helmholtz (1885/1954) used the term 'full' as being descriptive of the musical quality of a tone, or in a simpler word, timbre. In Fitzgerald & Lindsay's (2004) study on semantic labels and acoustic correlates of oboe, 'Power' and 'Vibrancy' were reported as two of their three perceptual dimensions. 'Power' was found to have high loadings of the semantic label 'Strong.' Kendall and Carterette (1993b) reported the same finding of the word 'Strong' being associated with the 'Power' dimension. Despite the fact that these studies have used very limited sets of instrument sounds, sometimes confined to a single instrument family, some adjectives have been found to be used repeatedly to describe timbre. These include, among others, harsh, bright, full, and warm. Following these results, Gounaropoulos & Johnson (2006) created an interface that allows users to synthesize new timbres through the manipulation of some existing acoustic features that correlate to perceptual qualityies of the sound. Though the semantic labels discussed above have been used for monophonic timbre, it is interesting to see whether similar patterns may be observed in describing polyphonic timbre.

2.1.3 Neural Correlates of monophonic timbre

Electrophysiological studies that employed the mismatch negativity (MMN) paradigm to investigate the neural correlates of perception of monophonic timbre have hinted at a high correlation between the amplitude of brain responses and the magnitude of changes in certain timbral features. These include brightness (Toiviainen, Tervaniemi, Luohivuori, Saher, Huotilainen et al., 1998), spectral center of gravity, and even harmonic attenuation (Caclin, McAdams, Smith, & Giard, 2008). In addition, neuroimaging studies have aided in localizing regions of the brain that deal with monophonic timbre processing. Platel, Price, Baron, Wise, Lambert et al. (1997) report the involvement of the right superior and middle frontal gyrus during selective attention to timbre while pitch and rhythmic content varied simultaneously. Menon et al. (2002) found that posterior regions of the Heschl's gyrus and superior temporal sulcus were mainly involved in the processing of timbre. However, majority of these studies have used artificial stimuli in controlled auditory paradigms in which the single instrument sounds have been presented in isolation and manipulated artificially, thereby compromising ecological validity.

The central focus of monophonic timbre research has been to uncover its perceptual and acoustical correlates. As discussed in earlier sections about the importance of music in everyday life, especially in engendering pleasurable feelings, which can be considered very fundamental to humans, the role of the intrinsic musical elements, in this case timbre, in causing this has not been investigated. No studies exist that have investigated the link between monophonic timbre and emotions.

2.2 Polyphonic Timbre

Polyphonic timbre refers to the overall timbral mixture in a music signal, or in simple words, the 'global sound' of a piece of music (Aucouturier, 2006). As described earlier in this thesis, polyphonic timbre can be regarded to be synonymous with emerging timbre. An interesting analogy is drawn by Bregman (1990) to explain the phenomenon of emergence of a higher-order form from lower-order constituent elements. The author uses a knife as an example, wherein the molecules that constitute a knife are not sharp, however the 'sharpness' of a knife can be considered as an emergent property of a knife. In this regard, polyphonic timbre can be explained as the *emergent property* of monophonic timbre.

2.2.1 Perceptual and Neural Significance of Polyphonic Timbre

Polyphonic timbre has been found to be a significant perceptual component of music, especially in studies that involve tasks such as genre identification, categorization, or emotional affect attribution. The study carried out by Gjerdingen & Perrott (2008) examined the time required for people to identify or classify into genres very short music excerpts. They reported that musical extracts as short as 250 ms were sufficient for genre identification. The authors highlight the importance of the overall timbre in the perceptual process of identification and categorization. The authors describe overall timbre as an agglomerate "of spectral and rapid time-domain variability in an acoustic signal" which is put together by the listener in a Gestalt-like manner that thereby enables listeners to identify, classify, and categorize the heard piece of music. In other words, overall timbre possesses cues that enable identification and classification. For instance, the presence of high amounts of acoustic energy in the lower end of the spectrum points more towards rap and hip-hop rather than classical music (Gjerdingen & Perrott, 2008).

In a perceptual song identification study performed by Schellenberg, Iverson &McKinnon (1999), the authors reported successful identification rates of songs from excerpts as short as 200ms and 100ms. Similar results were obtained by Krumhansl (2010) in her study on genre and song identification of 400ms and 300ms excerpts. Peretz, Gagnon, & Bouchard (1998) add to this by reporting that participants were successful in identifying affective connotations of 500ms long sad or happy musical excerpts taken from the Western classical repertoire. In addition, Bigand, Vieillard, Madurell, Marozeau, & Dacquet (2005) report that a duration of 1000ms sufficed in making successful affect judgements. Based on these studies, it can be inferred that short durations suffice to perform recognition and categorization tasks. Such short durations render it difficult to obtain a clear sense of musical elements such as rhythmic, melodic and harmonic progressions. The only information that can be gathered relates to the overall sound, or in other words, polyphonic timbre. Hence, it can be deduced that polyphonic timbre indeed is an essential element of music and thus calls for uncovering the underlying mechanisms involved in its perception.

2.2.2 Polyphonic Timbre in Music Computing

Features representing polyphonic timbre have been demonstrated to be vital in the design of computational systems that perform genre-, style-, and mood-based categorization of music (e.g., Barbedo & Lopes, 2007; Jiang, Lu, Zhang, Tao, & Cai, 2002; Liu, Lu, & Zhang, 2003; Tzanetakis & Cook, 2002). Computational features representing timbral and rhythmic features have been the most popular choices when designing such computational systems since they were found to maximize the performance of the system to accomplish the respective task. Commonly used timbral features comprise spectral flux, spectral roll-off, and spectral centroid (Barbedo & Lopes, 2007; Tzanetakis & Cook, 2002). In addition, other repeatedly used features known to represent the overall timbre of music signals are the Mel-Frequency Cepstral Coefficients (MFCCs) (Aucouturier & Pachet, 2003; Logan & Salomon, 2001; Lu, Liu, & Zhang, 2006; McKinney & Breebaart, 2003; Pampalk, Flexer, & Widmer, 2005; Pye, 2000; Zhu, Xue, & Lu, 2004). The MFCCs were first used in the domain of speech recognition (Davis & Mermelstein, 1980) and are derived by performing discrete cosine transform on spectra shaped by the Mel- Scale that approximates the human auditory response. They have often turned out to be the best predictors of class membership, which might suggest that they encapsulate some important properties of the overall timbre. Studies performed using the query-by-semantic description paradigm have focused on creating a meaningful vocabulary to describe aspects of music such as instrumentation, emotional content, genre, song concepts, and usage terms. For instance, recent work by Turnbull and colleagues (2006, 2007, 2008) has dealt with semantic annotation and retrieval of music. They used the delta-MFCCs (ΔMFCCs), which capture the temporal fluctuation of the MFCCs, to represent features of the music tracks. Whitman and Ellis (2004) created an automatic record review system that annotated music excerpts based on web-reviews of artists. They made use of the features derived from the MFCCs to characterize short music excerpts. By way of machine learning, the system then learned to create reviews based on acoustic properties and their associated semantic descriptions. For generalizability, the authors emphasized the need to use features that capture the overall perceptual quality of the music excerpts. Although not explicitly mentioned in their study, one might speculate that it is the overall timbre of the music that causes listeners to identify and describe its emotive, situational, or structural aspects.

Feature selection in the studies above was using heuristic approaches rather than systematic approaches. This may bring into question the true usefulness of timbral features in perceptual classifiers. However, it has also been shown that in the design of systems that model perceptual data wherein systematic feature selection, such as backward elimination or forward selection, is employed have also reported the emergence of timbral features as good predictors. For instance, recent studies dealing with the prediction of affect in music have reported that timbral features have emerged as good predictors of emotional and mood ratings of music (Eerola, Lartillot & Toiviainen, 2009; Laurier, Sordo, Bozzon, Brambilla & Fraternali, 2009; Schmidt & Kim, 2010).

In sum polyphonic timbre serves as an indispensable element of music that serves as a basis for several cognitive processes such as categorization based on genre, and affect-identification. While many studies have focused on understanding the perceptual and cognitive processes of higher-level features such as harmony, melody, and rhythm, the perceptual aspects of polyphonic timbre have received much less attention. Furthermore, on a more general note, an interesting development is that of contemporary music, which appears to be deviating from the well-known theories of Western melodic, harmonic, and rhythmic progressions. This music seems to move towards creating new sounds and textures by focusing on the blending of varied timbres. These results simply emphasize the importance of delving into the realm of polyphonic timbre perception.

2.3 Gaps and Limitations in Timbre Perception

This section outlines some of the important gaps and limitations that exist in the fields of monophonic timbre and polyphonic timbre perception.

- Despite the large corpus of work investigating the perceptual and acoustical correlates of monophonic timbre, there exists a lack of studies examining the connection between monophonic timbre and emotions. The importance of timbre in emotion-detection has been established in a polyphonic context but has not been thoroughly investigated in a monophonic context.
- In the context of the two divisions of timbre considered in this thesis, that is, of monophony and polyphony, no studies exist investigating the perceptual and corresponding acoustic correlates of polyphonic timbre. The first step is to identify the perceptual dimensions of the polyphonic timbre space. Following this, the acoustic correlates of the respective dimensions need to be determined.
- To allow for generalizations of the perceptual dimensions of polyphonic timbre space, there is a need to broaden the scope of research to accommodate a more global palette of both music and listeners. A cross-cultural setting provides an opportunity to examine the effect enculturation on polyphonic timbre perception. Moreover, dividing the population based on musical-expertise also aids in examining the effects of musicality. Additionally, this investigation would serve as a means to investigate the phenomenon of perceptual learning and the effect of enculturation and musical expertise on it.
- Previous neuroimaging studies have attempted at localizing monophonic timbre processing in the brain by means of controlled paradigms that are far from real world situations wherein we listen commonly to continuous music, which is majorly polyphonic. This calls for investigating polyphonic timbre processing in naturalistic settings by using more realistic stimuli.
- Finally, in order to obtain a holistic understanding of timbre perception, it is necessary to find the links between its behavioral, computational and neural representations.

2.4 Challenges

Investigating any phenomenon, especially as complex as timbre, will be accompanied by its own set of challenges due to its inherent abstractness and multidimensional nature. A few of these challenges are discussed in this section.

2.4.1 The 'Semantic Gap'

One of the main challenges that entails understanding the perceptual and cognitive aspects of musical elements via perceptual studies is the 'Semantic Gap', which is defined as the discrepancy between what can be recognized in music signals by the existing state-of-the-art methods and what human listeners associate with music (Celma, Herrera & Serra, 2006). Restricting the language of semantic description to English can introduce 'semantic noise' due to the variance in subjective understanding of the connotative meanings of words from person to person. Hence, design of behavioural studies and choice of descriptors needs to be done keeping this limitation in mind.

2.4.2 Generalizability

A key requirement while designing perceptual experiments is to strike a balance between systematicness and representativeness. Systematic or methodological experimental designs allow for more controlled investigations, however compromising generalizability. There are two needs to be met to allow for generalizations, that is, in relation to the population and the settings.

Extending the population set to include people from different backgrounds, be it culture- or musical expertise-based, would help obtain a deeper insight into the process of timbre perception. An ideal cross-cultural setting would involve homogeneous groups wherein each participant group would be extremely familiar with music from one culture and completely ignorant with that of the other culture. However, globalization and the exponentially increasing accessibility to music from various parts of the world make it increasingly challenging to meet such a requirement. In comparison musical expertise poses to be less of a challenge to address.

Generalizations concerning the settings can be further discussed in light of two aspects, that is, stimuli and context. Using natural stimuli instead of synthesized ones, and extending them to include music from other cultures in addition to the hitherto Western Tonal theory-based music, would lend more weight to the representativeness of the observed results. Moreover, in the neural domain, studies have relied on controlled auditory paradigms in which musical elements under investigation, either natural or synthesized, have been presented in isolation and manipulated artificially, providing then a limited perspective on timbre processing. Often the presentation of the musical stimuli lasts only few seconds followed by a rest or baseline condition, and hence is from a realistic reproduction of natural music listening situations. Furthermore, in such conditions the participants sometimes are required to perform an unrelated task, which further adds to the non-representativeness of the context. Therefore, there is a pressing need to alter the existing research paradigm to perform investigations with more naturalistic settings. For the sake of improving the representativeness of the results, the above-mentioned factors need to be deliberated on.

2.4.3 Choice of Acoustic features

When attempting to link semantics of polyphonic timbre to acoustic content, selection of acoustic features is a crucial step. Feature selection varies depending on the task at hand and the modelling approach used and as a result may compromise the generalizability of the selected acoustic features in their predictive capacities for other corpora of audio. In addition, the choice of parameters during computational feature extraction, such as analysis window sizes, default thresholds and choice of algorithms may cause variability in the predictive power of the respective acoustic feature.

2.4.4 Neural domain limitations

Investigating polyphonic timbre perception in the neural domain would allow us to gain a more holistic comprehension of this phenomenon. However, brain research suffers from several problems concerning the noisiness of the data, which is caused due to various factors. The inter-subject variability of the anatomy and physiology of the brain, within-subject variability of the hemodynamic response function, Nyquist ghosting, noise due to physiological responses such as respiration and minute body movements, to name a few, serve as sources of noise. Furthermore, neuroimaging methods such as fMRI possess poor temporal resolution. However, these factors need not be seen as obstacles, rather as parameters that need to be compensated to the maximum possible extent while designing and analyzing studies.

3 AIMS OF THE THESIS

The main aim of this thesis is to investigate the perception of the most elusive of musical elements, timbre. Specifically, it aims at filling existing gaps in the field of monophonic timbre and extends timbre perception investigations to polyphony in an interdisciplinary setting. The studies are divided based on the following sub-topics.

1. Link between monophonic timbre and emotions
2. Polyphonic timbre semantics
3. Perceptual dimensions of polyphonic timbre space
4. Link between perceptual dimensions and acoustic features
5. Generalizability in light of representativeness of population and stimuli
6. Perceptual learning in a musical context
7. Generalizability in light of representativeness of context and stimuli
8. Neural correlates of acoustic features of polyphonic timbre
9. Perceptual validation of acoustic features of polyphonic timbre
10. Associations between perceptually validated features and corresponding neural correlates

Figure 1 illustrates the distribution of the above topics based on the domains their investigation encompasses.

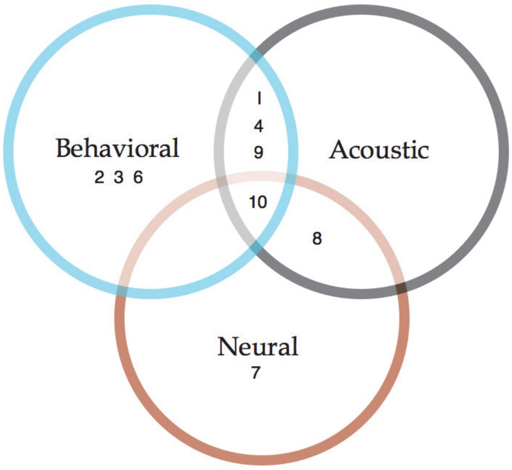

FIGURE 1 Distribution of sub-topics of investigation among the behavioural, acoustic and neural domains

An interdisciplinary approach, as illustrated above, acts as an ideal setting to investigate perceptual and cognitive phenomena. However, the paucity of studies that elect this approach renders it challenging to deal with the methodological concerns that arise during the span of the research investigations. Specifically, no studies exist that investigate musical feature processing using natural stimuli that combine all three domains on investigation. This calls for a need to develop novel methodologies that allow for such investigations and as a result raises several methodological concerns. Some of the methodological aims are mentioned below:

1. Processing acoustic data to make it compatible with fMRI data
2. Estimation of statistical significance of the correlations between acoustic feature time-series and fMRI time-series
3. Estimation of cluster-sizes to correct for multiple comparison

The following chapter explains in detail the methods employed and novel methods adapted for the four studies.

4 METHODS

The Methods section describes the techniques that were used for this interdisciplinary investigation of polyphonic timbre perception. This section deals broadly with four main themes: intrinsic dimensionality estimation, computational acoustic feature extraction, fMRI data acquisition and pre-processing, and statistical analyses techniques. There are three key contributions of the work presented here. First, techniques for estimating the intrinsic dimensionality of perceptual spaces are proposed. Second, a new set of acoustic features, namely the Sub-Band fluxes, is introduced whose perceptual relevance has been validated as a result of Studies II-IV. Third, novel statistical analysis methods are developed that enable investigations of musical feature processing using more naturalistic stimuli unlike previous fMRI studies that have used controlled, reduced and artificially manipulated auditory stimuli. They are explained in detail in what is to follow.

Section 4.1 describes the methods used in Study III. Section 4.2 describes the common method used for computational feature extraction for all the Studies (I-IV). Section 4.3 onwards explains the methods employed for Study IV only.

4.1 Intrinsic Dimensionality Estimation

In order to explore the phenomenon of perceptual learning and the effect of enculturation and musical expertise on the dimensionality of the polyphonic timbre perceptual space, the intrinsic dimensionality of the perceptual ratings was calculated. The intrinsic dimensionality (ID) of a dataset can be described as an estimate of the number of “independent” variables required to represent it. Studies II and III required the investigation of the underlying perceptual structure of the perceptual ratings in order to determine the perceptual dimensions of the polyphonic timbre space. Specifically, the cross-cultural context in Study III required the comparison of the perceptual dimensions across datasets. The most commonly used traditional estimation method is based on the Eigenvalues satisfying the Kaiser criterion (1960). However, there exist other intrinsic dimensionality estimation techniques offered by the fields of chaos theory, complex systems theory, and machine learning, which have seldom been used for perceptual data. To this end, three methods were chosen to calculate the intrinsic dimensionality of the datasets, namely Maximum Likelihood estimation (MLE) (Levina & Bickel, 2004), Correlation Dimension estimation (CD) (Camastra & Vinciarelli, 2002), and Nearest Neighbor (NN) (Pettis, Bailey Jain & Dubes, 1979). MLE approach was found to provide more reliable estimates than the CD and NN approaches (Levina & Bickel, 2004). However, for the sake of comparison, all three procedures in addition to the Eigen Value Estimation were employed in order to estimate the intrinsic dimensionality.

4.2 Acoustic Features and Computational Extraction

Parameterization of audio is an important step in computational modelling. A plethora of features exist in literature defining spectral, temporal, and spectrotemporal aspects of audio. When attempting to link semantics of polyphonic timbre to acoustic content, selection of acoustic features is a crucial step. Feature selection varies depending on the task at hand and the modelling approach used. Most often features are selected based on existing theoretical knowledge or other criteria such as interpretability. Commonly, automatic feature selection is subsequently performed by employing methods such as step-wise regression or sequential feature selection. These take in all the features as input to produce a desired output, and then through iterative processing select the features that best help predict or provide better results. These approaches are commonly found in a few studies that focus on mood, emotion or expression prediction using audio features (Eerola, Alluri, & Ferrer, 2008; Leman, Vermeulen, De Voogdt, Moelants, & Lesaffre, 2005; Mion & De Poli, 2008; Yang, Lin, Su, & Chen, 2008). In addition, there exist methods for automatic generation of acoustic features (Li & Stern, 2004; Schuller, Wallhoff, Arsic, & Rigoll, 2006).

Timbre-related computational features commonly used in previous work (Grey, 1977; Grey & Gordon, 1978; Iverson & Krumhansl 1993; Lakatos, 2000; McAdams et al., 1995; Aucouturier, 2006; Aucouturier & Pachet, 2003; McAdams, Winsberg, de Soete, & Krimphoff, 1995; Tzanetakis & Cook, 2002) were used as a starting point. An important distinction that determines which features to opt for is the monophonic or polyphonic nature of the stimuli. For instance, in previous timbre studies, various spectral and temporal features such as the spectral centroid, log-attack time, spectral flux, attack synchrony, and spectral irregularity, to name a few, have been found to explain the dimensions of perceptual monophonic timbre spaces. In addition to the aforementioned features, the Mel-Frequency Cepstral Coefficients (MFCC),

which are described in later sections, appear to be quite prominent in characterizing monophonic timbre (De Poli & Pradoni, 1997; Terasawa, Slaney, & Berger, 2005). However, estimating some features such as the log-attack time may not be sufficiently reliable due to the superposition and interleaving of various timbres as is in the case of polyphonic timbre. Similarly, calculation of features such as inharmonicity and irregularity is performed under the assumption of the presence of a single tone. Moreover, it is important that any feature strikes a balance between computability and interpretability. As a result, some features representing higher order moments such as spectral kurtosis and skewness were excluded for polyphonic mixtures. All the features that were chosen are summarized below.
The (*) signifies that the features are calculated under the assumption of monophony or the presence of a single tone. Hence they were used only in Study I.

<u>Temporal</u>

Zero Crossing Rate (ZCR): Number of time-domain zero-crossings of the signal per time unit.

Root Mean Square Energy (RMS): Measure of instantaneous energy contained in the signal, obtained by taking the square root of sum of the squares of the amplitude.

**Attack Slope (*AS): Number of time-domain zero-crossings of the signal per time unit.

**Envelope Centroid (*EC): Centroid of the temporal envelope

**Envelope Fluctuation (*EF): Standard deviation of the temporal envelope.

<u>Spectral</u>

*Spectral Centroid (*SC): Geometric centre on the frequency scale of the amplitude spectrum.

*High Energy – Low Energy Ratio (*HLR): Ratio of energy content below and above 1500 Hz.

*Spectral Entropy (*ENT): The relative Shannon entropy (1948) calculated using the equation below

$$H_t = -\frac{\sum_{n=1}^{N} A_t[n] \log A_t[n]}{\log N} \tag{1}$$

where A_t is the amplitude spectrum of audio frame at time t and N the number of frequency bins in the amplitude spectrum. The relative Shannon entropy indicates whether the spectrum contains predominant peaks or not. For example, a single sine tone has minimal entropy and white noise maximal.

Spectral Roll-off (SRO): Frequency below which 85% of the total energy exists.

Spectral Spread (SS): Standard deviation of the spectrum.

Spectral Flatness (SF): Wiener entropy of the spectrum, defined as the ratio of its geometric mean to its arithmetic mean.

**Spectral Regularity* (SR): Measure of variation of the successive peaks in the spectrum, also known as Spectral Smoothness (McAdams et al., 1999).

**Inharmonicity* (Inharm): Measure of the deviation of the partials from the harmonic frequencies (Jensen, 1999).

Spectrotemporal

Roughness (R): Estimate of Sensory dissonance (Sethares, 1998).

Spectral Flux (SFlux): Measure of temporal change in the spectrum, obtained by calculating the Euclidian distance between subsequent window-based amplitude spectra.

Sub-Band Flux (SubFlux): A new set of features are introduced, namely the Sub-Band Flux. The Sub-Band Flux represents the fluctuation of frequency content in ten octave-scaled bands of the spectrum. Figure 2 displays schematically the calculation of the Sub-Band Fluxes.

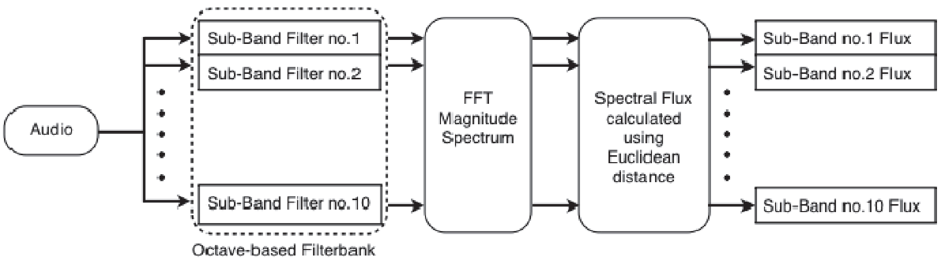

FIGURE 2 Sub-band Fluxes extraction procedure

The division into Sub-Bands was obtained using a 10-channel filterbank of octave-scaled second-order elliptical filters. The frequency response of the filter-bank can be seen in Figure 3.

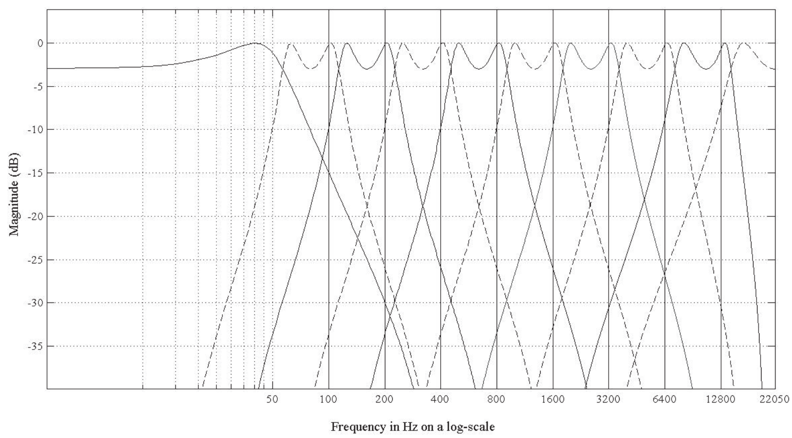

FIGURE 3 Frequency response of the 10-channel filterbank of octave-scaled second-order elliptical filters used for obtaining the Sub-Band fluxes.

For each of the ten channels the spectral flux was calculated as the Euclidean distance between successive amplitude spectra obtained using Short-Time Fourier Transform. Table 1 displays the boundaries of each of the Sub-Bands.

TABLE 1 Frequency ranges of the Sub-Bands

	Frequency Range
Sub-Band No. 1	0 – 50 Hz
Sub-Band No. 2	50 – 100 Hz
Sub-Band No. 3	100 – 200 Hz
Sub-Band No. 4	200 – 400 Hz
Sub-Band No. 5	400 – 800 Hz
Sub-Band No. 6	800 – 1600 Hz
Sub-Band No. 7	1600 – 3200 Hz
Sub-Band No. 8	3200 – 6400 Hz
Sub-Band No. 9	6400 – 12800 Hz
Sub-Band No. 10	12800 – 22050 Hz

Cepstral

Mel-Frequency Cepstral Coefficients (MFCCs, 13 features in total):

MFCCs offer a description of the spectral shape of the sound. MFCCs are a result of the discrete Cosine Transform (DCT) of the log amplitude of the Mel-frequency spectrum of an audio signal. Figure 4 illustrates the process of obtaining the MFCCs.

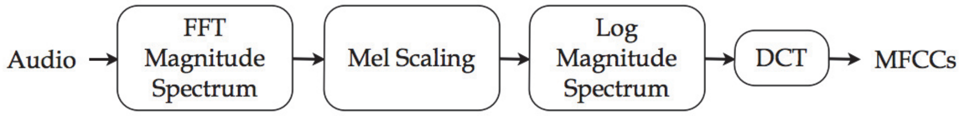

FIGURE 4 Extraction process of MFCCs

The magnitude spectrum of the audio is scaled according to the Mel scale, which approximates the human auditory system's response more closely than the linearly spaced frequency bands. This weighted spectrum is then subjected to a log operation followed by a discrete cosine transform (DCT). Most of the signal information tends to be concentrated in the first few components of the DCT. Following the convention the first 13 components were used.

Features were extracted using a window length of 25 ms and an overlap of 50% between successive windows. Window lengths of this magnitude often are used for feature extraction in Music Information Retrieval (Barbedo & Lopes, 2007; Lu, Liu, Zhang, 2006; Pampalk et al., 2005; Tzanetakis & Cook, 2002).
For Studies I, II and III the respective feature sets comprised the mean of each acoustic feature across all frames due to the nature of the studies. On the other hand, due to the paradigm opted in Study IV, the time-series of each feature was retained. All the features were extracted using the MIRToolbox (Lartillot and Toiviainen, 2007) in the MATLAB environment with functions written by the present author. In the current work different sets of features were chosen for each of the four studies based on the nature of the stimuli and the paradigm of the study. Table 2 summarizes these feature sets.

TABLE 2 List of acoustic features and their usage in the four studies.

Domain		STUDY I	STUDY II	STUDY III	STUDY IV
T	ZCR	X	X	X	X
T	RMS				X
T	AS	X			
T	EC	X			
T	EF	X			
S	SC	X	X	X	X
S	HLR	X	X	X	X
S	ENT	X	X	X	X
S	SRO	X	X	X	X
S	Spread	X			X
S	Skewness	X			
S	Kurtosis	X			
S	Flatness	X			X
S	Irregularity	X			
S	Inharm	X			
ST	Rough	X	X	X	X
ST	SFlux	X	X	X	X
ST	SubFlux	X	X	X	X
C	MFCC		X		

T - Temporal, S - Spectral, ST - Spectrotemporal, C - Cepstral

4.3 Functional Magnetic Resonance Imaging

One of the most popular emerging imaging methods in the field of cognitive neuroscience is functional Magnetic Resonance Imaging (fMRI). fMRI measures the blood oxygenation level dependent or BOLD contrast, which can be considered as an indicator of neuronal activity. Deoxygenation of the blood occurs as a result of consumption of oxygen by activated neurons. This deoxygenation is compensated by a flow of freshly oxygenated blood to the respective deoxygenated brain regions. This process takes place in the span of a few seconds. Due to this delay in the measured response, the BOLD signal is considered as an indirect measure of neuronal activity. The delay is characterized by what is known as the hemodynamic response function (HRF). The HRF is known to vary within and between brains. However, one of the commonly used models of the HRF is the double gamma HRF, illustrated in Figure 5.

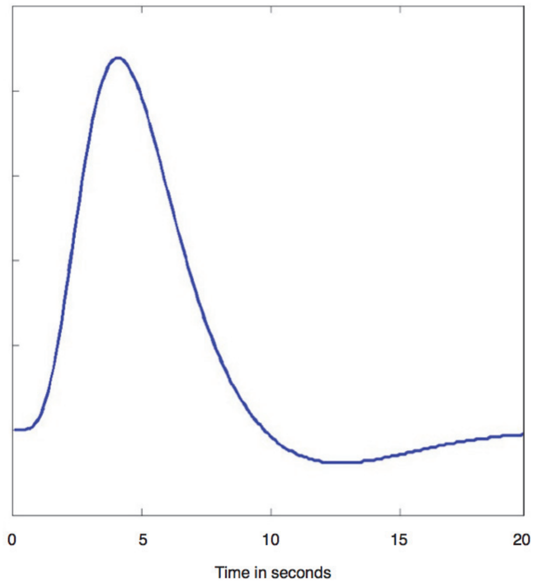

FIGURE 5 Double gamma hemodynamic response function.

The double gamma HRF is characterized by a peak at 5 seconds with an undershoot occurring at around 15 seconds.

Commonly, fMRI studies in the field of cognitive neurosciences of music utilize controlled paradigms wherein the feature under investigation is usually presented in isolation and manipulated artificially. This approach fails to emulate real-world situations wherein we are constantly bombarded with continuous streams of sensory information and hence reveals only an incomplete picture of the neural mechanisms involved in the processing any stimulus. Therefore, studying music listening as a continuous process using naturalistic stimuli could provide more accurate accounts of the processing of musical features in the brain. The following sections explain in detail the process of obtaining and methods developed for analyzing fMRI data in a novel paradigm using naturalistic stimuli.

4.4 fMRI data acquisition and pre-processing

The fMRI measurements were conducted with the 3-Tesla scanner (3.0 T Signa VH/I General Electric) at the Advanced Magnetic Imaging (AMI) Centre of the Aalto University and were approved by the local ethical committee. To prevent postural adjustments and to attenuate the noise and vibration of the scanner, foam cushions were placed around the arms of the participants. Music was presented through audio headphones with about 30 dB of gradient noise attenuation. Further attenuation was achieved with cotton inserted in the headset. Thirty-three oblique slices covering the whole brain (field of view 200 x 200 mm; 64 x 64 matrix; slice thickness 4 mm; gap 0 mm) were acquired using a single-shot gradient echo-planar imaging (EPI) sequence (TR = 2 sec; echo time, 32 ms; flip angle, 75°) sensitive to blood oxygenation level-dependent (BOLD) contrast. During the fMRI measurement, participants listened to the stimulus presented at an average sound level of 80 dB. The participants were instructed to stay still and to relax while listening to the musical stimulus and to maintain their gaze on the screen. Subsequent to a short break after fMRI recording, anatomical T1 weighted MR images (field of view 260 x 260 mm; 256 x 256 matrix; thickness 1 mm; spacing 0 mm) were acquired.

In the pre-processing stage, whole-brain image analysis was carried out using Statistical Parametric Mapping 5 (SPM5 - http://www.fil.ion.ucl.ac.uk/spm). Images for each subject were realigned, spatially normalized into the Montreal Neurological Institute template (12 parameter affine model, gray matter segmentation; realignment: translation components < 2mm, rotation components < 2°), and spatially smoothed (Gaussian filter with FWHM of 6 mm). fMRI responses were detrended using a high-pass filter with a cut-off frequency of .008 Hz, which conforms to the standards used to reduce the effects the scanner drift typically occurring at a timescale of 128 seconds (Smith & Williams, 1999). Following this, Gaussian smoothing was performed as it provides a good compromise between efficiency and bias (Friston, Josephs, Zarahn, Holmes, Rouquette, et al., 2000). The smoothing kernel had a width of 5 sec, which was found to maximize the correlation between the frequency content of the HRF and the smoothing kernel. The effect of the participants' movements was removed by modelling the 6 movement parameters as regressors of no interest.

4.5 Acoustic feature post-processing

Prior to performing correlation between the acoustic features and the voxel time-series, the acoustic features were post-processed for the following reasons. First, in order to account for the lag and smoothing caused due to the hemodynamic response in the fMRI data, the acoustic feature time-series were convolved with the aforementioned double-gamma HRF. Following this, the convolved acoustic feature time-series were filtered employing the detrending filter used in the post-processing stage of the fMRI data. This operation was performed to eliminate those low-frequency components whose eventual brain correlates were eliminated during the preprocessing stage of the fMRI time-series. For subsequent analysis, the features of only the part of the stimulus that contained music were used, hence leading to the exclusion of the last 24 seconds due to the presence of applause in the audio recording. In addition, the first 26 seconds corresponding to the length of the HRF were excluded in order to avoid any artefacts caused due to the convolution operation. Following this, all the acoustic features were downsampled to match the sampling rate of the fMRI data.

4.6 Statistical Analysis

The preprocessing stage of the fMRI serves to configure the data by increasing the Signal-to-Noise ratio (SNR) for further statistical analyses. Statistical analyses commonly comprise correlation analysis. A crucial step ensuing correlation analysis is to estimate the significance of the observed correlations and thereby control the false positive rate due to multiple comparisons. Several theoretical approaches have been suggested in order to estimate significance levels of activations in fMRI data. The most common approach is the use of Bonferroni correction and Gaussian random field theory (RFT) in order to estimate statistical significances. The Bonferroni correction determines the corrected significance values at a voxel-level whereas RFT does the same at a group-level or over a family of voxels.

Bonferroni correction is performed to estimate the corrected significance value and it is calculated under the assumption that each voxel's activation is uncorrelated with another. This correction proves to be considered to be too stringent due to two main reasons. Firstly, the functionality of the brain is such that neighbouring voxels belonging to certain regions in the brain tend to perform or processes information in a similar fashion. Second, due to the spatial correction and smoothing operations performed on the fMRI data in the preprocessing stage already introduce a certain amount of spatial correlation.

RFT is used to calculate the Euler characteristic (EC) for a statistical map of *N*-dimensions (3-D in this case) which then is used to determine the Family-Wise Error rate (FWE), that is, the likelihood that a group or family of voxels would have been found to correlate significantly by chance. In order to employ RFT to 3-D images as summarized by Brett, Penny & Kiebel (2003) and Worsley (2003), several underlying assumptions about the data are made. The EC is calculated based on the assumption that there exists a baseline or rest condition and a task condition (Worsley, Evans, Marrett, & Neelin, 1992). Due to the experimental design employed in Study IV which emulated real life situation of continuously listening to real music without any other rest or baseline condition, the choice of RFT for estimating statistical significances proves to be inapplicable. Furthermore, although it is less conservative than Bonferroni correction, RFT is not assumed to be the best choice with low numbers of subjects and scans (Holmes, Blair, Watson, & Ford, 1996). This calls for applying non-parametric methods for dealing with the problem of statistical significances.

4.6.1 The Statistical Significance problem

The aim of this section is to highlight some issues that need to be taken into account while determining statistical significances and propose solutions to overcome them. A common problem plaguing any time-series is the presence of serial correlation. Serial correlation causes the number of effective degrees of freedom to decrease from the total number of sample points. It is vital that the significance of the correlations between the fMRI time-series and the respective continuous data (e.g., continuous behavioural ratings or stimulus-dependent continuous features) be calculated based on the correct effective degrees of freedom. The effective degrees of freedom vary based on the nature of the data and hence need to be determined empirically as different data have different amounts of temporal correlation.

The substantial number of voxels contained in an fMRI volume (~200,000) ascertains the existence of false positives. Different degrees of spatial smoothness exist in different sets of fMRI data, typically affected for instance by the brain template matching and Gaussian spatial smoothing operations performed during the pre-processing stage and has not been generally accounted for systematically in previous studies. This is reflected in the fact that various studies report the exact same threshold for cluster-correction (For instance, 17 voxels reported by Fox, Snyder, Vincent, Corbetta, Van Essen, et al., 2005; Fox, Zhang, Snyder, & Raichle, 2009 based on McAvoy, Ollinger, & Buckner, 2001). Moreover, non-parametric statistical significance estimation methods have been considered to outperform parametrical approaches in fMRI settings wherein the data possess low degrees of freedom (Nichols & Holmes, 2001).

Neglecting the nature of the data at hand and assuming standard thresholds for multiple comparisons compromises the validity of the results and causes an increase in the number of false positives if the assumed threshold is lower and a decrease in the statistical power (with consequent risk of Type II errors or false negatives) for a higher threshold. This calls for a more non-parametric approach of obtaining an accurate estimate of the significance of obtaining a certain cluster-size of neighbouring voxels that correlate significantly than utilize theoretically suggested standards. The above-mentioned problems are some of the many that have been neglected by several researchers during the statistical analysis stage of the fMRI data. The following section describes the methods that have been employed in the Study IV to tackle the statistical analysis segment.

4.6.2 Employed Methodology

As there exist no well-established statistical methods for investigating continuous data obtained in a naturalistic paradigm as in Study IV, non-parametric methods were adopted to solve the problem of statistical significances in correlation analysis. The flowchart (Figure 6) illustrates the various steps followed in the statistical analysis stage of Study IV. The entire correlation analysis, including implementation of the non-parametrical methods, was performed in the MATLAB environment with functions written by the present author.

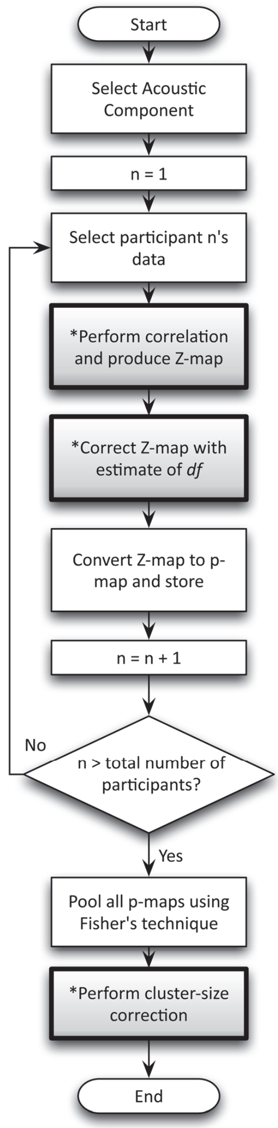

FIGURE 6 Schematic illustration of the statistical analysis employed in Study IV

The stages in the flowchart highlighted indicate the segments that required non-parametric approaches while estimating the respective parameters. The following section explains the detail of the empirical approaches used to fulfil these tasks.

4.7 Empirical Approaches to Estimation of Statistical Significance of fMRI Data

In order to estimate the significance values, stochastic simulation or Monte-Carlo method was employed instead of theoretical approaches. Stochastic simulation or the Monte-Carlo approach is known to be an attractive option in the field of Inferential Statistics especially whilst dealing with large number of comparisons (See Gamerman for an overview, 1997). One specific method that falls under this approach is *Significance Estimation Utilizing Permutation Tests.* Significance estimation utilizing permutation tests operates on the principle of repeatedly calculating a statistic by randomly permuting the data at hand in order to obtain a probability of obtaining the observed statistic under investigation. This approach was employed to empirically estimate the significance of observed correlations, and the effective degrees of freedom which were then used to correct the significance values in the first level of analysis. Subsequently, it was used to determine the significance of obtaining different cluster sizes in order to minimize Type I errors (false positives) due to multiple comparisons in the second-level. The processes are explained in detail in what is to follow.

4.7.1 Significance Estimation of Observed Correlations

The significance of the observed correlations was estimated in two stages. The first step comprised voxel-wise time-series correlations with the acoustic features at an individual level. The Flowchart shown in Figure 7 depicts the Monte-Carlo simulation that was performed in order to estimate the significance of the observed correlation value at any voxel.

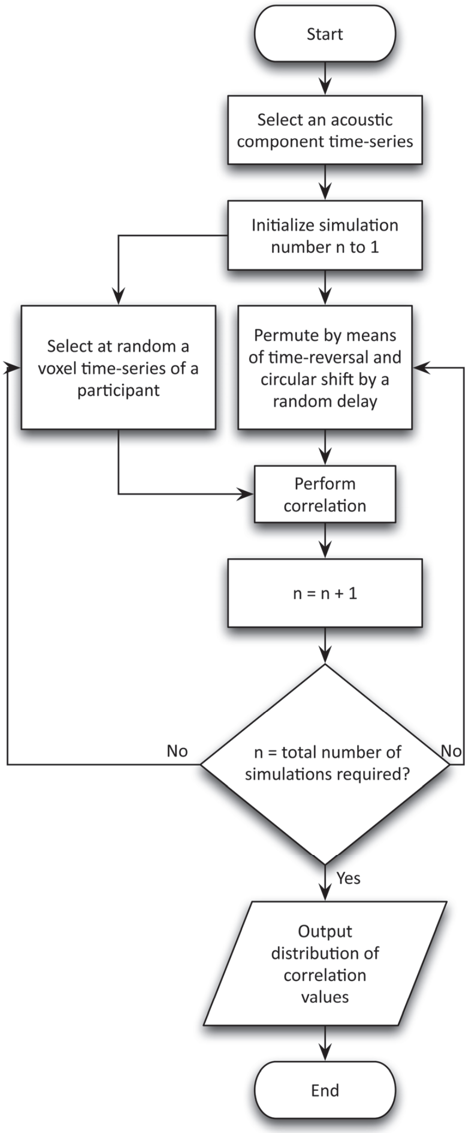

FIGURE 7 Schematic illustration of the significance estimation method employed in Study IV

The above process was executed for each acoustic component. As a result, an estimate was obtained of the distribution of correlation values between the fMRI time-series and any random signal with similar spectral properties as the acoustic component in question. Following this, for each acoustic component, the respective correlation distribution curve is utilized to assign the significance value for the observed correlation values between each voxel time-series and the respective acoustic component. This procedure results in significance maps

(p-maps) per participant for each of the acoustic components. The p-maps were then converted to respective Z-score maps using Fisher's Z transformation (Equation 2).

$$z = \frac{1}{2}\ln\frac{1+r}{1-r} = \tanh^{-1}(r) \quad (2)$$

Following this, taking into account the presence of serial correlation caused due to the nature of the stimulus and the additional smoothness introduced due to the HRF convolution operation, the Z-statistic was normalized by the factor $1/\sqrt{(df-3)}$, where df represents the estimated number of effective degrees of freedom.

In order to estimate the effective degrees of freedom for every acoustic component another Monte-Carlo simulation was performed of the method adopted by Pyper and Peterman (1998). In this approach, the effective degrees of freedom (df) are estimated using equation 3,

$$\frac{1}{df} \approx \frac{1}{N} + \frac{2}{N}\sum_j \frac{N-j}{N}\rho_{XX}(j)\rho_{YY}(j) \quad (3)$$

where $\rho_{XX}(j)$ and $\rho_{YY}(j)$ are the normalized autocorrelations of the signals of N observations at lag j. The maximal lag j chosen was N/5 because it is known to yield relatively accurate results in terms of error rates (Pyper and Peterman, 1998). The procedure is illustrated in the flowchart shown in Figure 8.

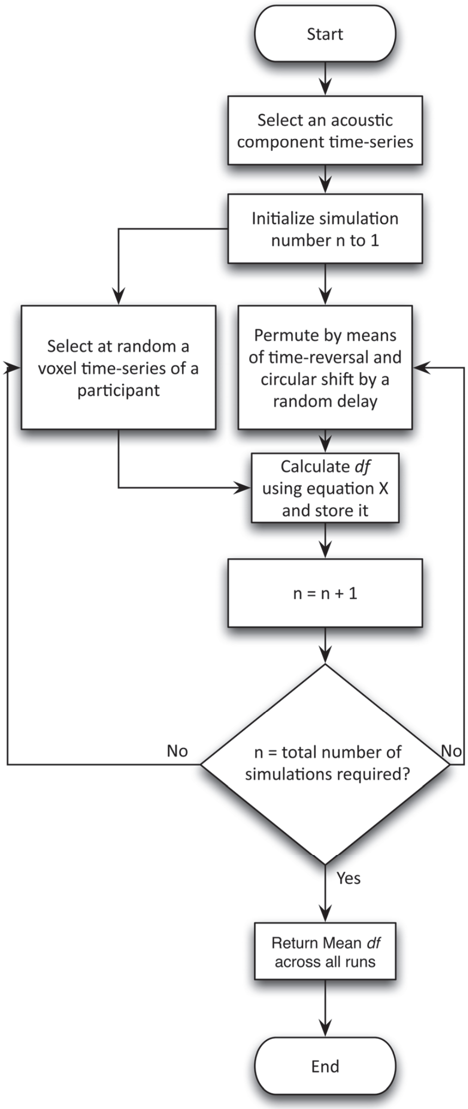

FIGURE 8 Schematic illustration of the method to estimate the effective degrees of freedom

The above procedure was executed for each acoustic component separately.

Following this, second-level analysis involved pooling individual results to obtain group maps for each acoustic component. In order to obtain the group maps we employed the combining tests procedure described by Lazar (2008). In this approach, individual corrected Z-score maps were first converted to p-maps and were then pooled using the Fisher's p-value technique (Fisher, 1950) to create the group maps. To this end, the individual p-values of each voxel were pooled using the equation below to obtain a T-statistic (Equation 4) that is modeled as a Chi-square distribution with 2k degrees of freedom where k represents the number of participants.

$$T = -2\sum_{i=1}^{k} \log p_i \tag{4}$$

The group maps hence obtained for each component were thresholded at a significance level of p < .001.

4.7.2 Multiple Comparison Correction by cluster-size Thresholding

Following this, in order to minimize Type I errors, the group maps were corrected for multiple comparisons using cluster size thresholding which was determined empirically by a Monte-Carlo simulation of the approach put forth by Ledberg, Akerman, & Roland (1998). This method aims at finding an estimation of the distribution of cluster sizes from which one can estimate the cluster-size threshold to be used to correct for multiple comparisons at any particular significance level. In this approach, first, pseudo noise-Statistical Images (pn-SIs), which contain the same spatial spectral properties as the signal-Statistical Image (signal-SI) but contain no stimulus-dependent activations, are used to obtain an estimate of the auto-correlation function (ACF) kernel *K*. The pseudo noise-SI is created as the Z-score statistic image obtained by randomly choosing a subject's data, correlating it with the reversed and circularly time-shifted acoustic component time series. Then, the ACF kernel *K* is determined according to Equation 5.

$$K = IFT\left|FFT(P)\right| \tag{5}$$

where *P* represents the pseudo noise-Statistical Image. To account for the variance in the estimate of the ACF, *K* is determined as an average of all the ACF kernels obtained via a Monte Carlo simulation illustrated in the flowchart to follow (Figure 9).

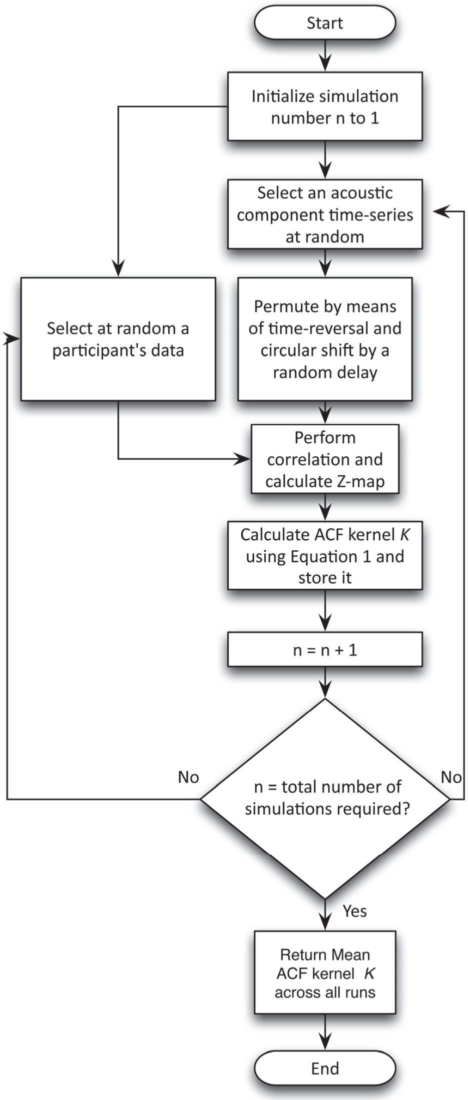

FIGURE 9 Schematic illustration of the method used to estimate the ACF kernel

After obtaining the mean ACF kernel *K*, the next step is to create Noise-Statistical Image and estimate the probabilities of obtaining different cluster-sizes. The process of obtaining the cluster-size distribution is illustrated in the following flowchart (Figure 10).

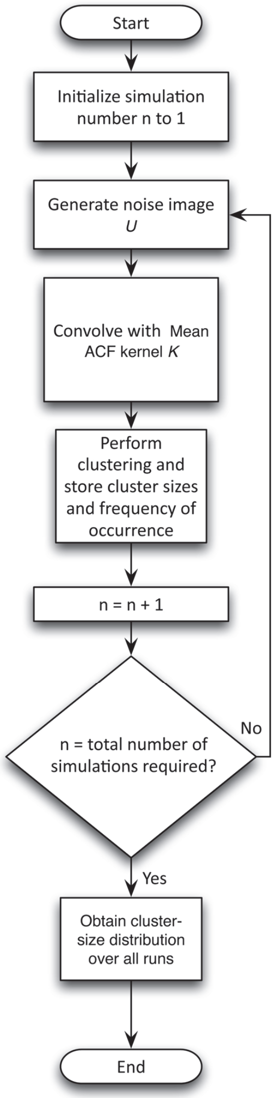

FIGURE 10 Schematic illustration of the method used to estimate the cluster size threshold

In order to create a Noise-Statistical Image (noise-SI), a randomly generated image *U* is convolved with the ACF kernel, *K*, using equation 6.

$$SI = U * K \tag{6}$$

To estimate the distribution of the cluster sizes, 1000 Noise-SIs were generated using the equation above. From all the Noise-SIs, the distribution of cluster sizes is obtained.

5 STUDY SUMMARIES

This section gives an overview of the design of each of the four studies and summarizes the results. The first study dealt with investigating the missing link between monophonic timbre and emotions. Following this, the perceptual and acoustic correlates of polyphonic timbre were investigated in the second study whose results are tested in a cross-cultural setting in the third study. Finally, the neural underpinnings of polyphonic timbre were investigated by employing an interdisciplinary approach, which is a combination of behavioural psychology, computational feature extraction, statistical modelling and neuroimaging.

STUDY I

Introduction: Study I comprised three experiments wherein the role of monophonic timbre in emotional expression to music was investigated. Specific research questions included:

- What is the underlying dimensionality of the monophonic timbre affect space?
- How consistent are affect ratings of monophonic timbre?
- What are the acoustic features that contribute to the affect ratings?
- How well can the acoustic features of isolated instrument sounds predict the affect ratings?

Methods: The study comprised three experiments. In Experiment 1, isolated natural instrument sounds were evaluated in the three-dimensional affect space comprising perceived valence, energy arousal and tension arousal respectively. In Experiment 2, the underlying dimensionality of the monophonic timbre affect space was investigated employing a similarity-rating task followed by multidimensional scaling (MDS). Experiment 3 comprised rating perceived affect of natural sounds and their artificially filtered versions using the same paradigm employed in Experiment 1.

Results: High agreement was observed in participants' ratings for the three dimensions of valence, energy arousal and tension arousal in Experiments 1 (Cronbach's α = .96, .92, and .94 respectively). Experiment 2 revealed two sufficient and interpretable dimensions (stress = .27), relating to Valence and Energy Arousal, which represented sufficiently the monophonic timbre affect space. Experiment 3 revealed similar consensus among participants' ratings for the three affect dimensions. Acoustic features representing the ratio of high frequency energy and low frequency energy, attack slope and spectral regularity were able to predict the perceptual ratings of Valence and Energy Arousal relatively well. Table 3 displays the regression coefficients for these models. Cross-validation of the models resulted in more robust models for Energy Arousal than Valence models (*r* = .77 and *r* = .45 respectively).

TABLE 3 Summary of Regression Analyses in Experiments 1 and 3

	Valence		Energy	
	Expt. 1	Expt. 3	Expt. 1	Expt. 3
R^{2adj}	.58	.54	.59	.74
	β	β	β	β
Attack Slope	.04	-.17	.28	.32
Envelope Centr.	.33	.71	-.50	-.42
HF - LF energy ratio	-1.09	-.52	.91	.71
Spectral Skewness	-.28	-.30	-.01	-.30
Spectral Regularity	-.01	.03	.01	-.13
Spectral Flux	-.15	.81	-.04	-.09
Sub-Band No.6 Flux	.03	-.75	.11	.33

Conclusion: The high consensus obtained in participants' ratings suggests that monophonic timbre indeed contains specific cues to musical expression. The MDS solutions revealed that two main dimensions representing perceived Valence and Energy Arousal sufficiently captured the monophonic timbre affect space. A subset of acoustic features were found to explain the affect ratings relatively well. The features display consistent correlation patterns with Valence and Energy Arousal appear across stimulus sets suggesting the presence of universal perceptual mechanisms in play during affect assessments.

STUDY II

Introduction: Study II comprised two experiments. The aim of Experiment 1 was to devise a framework of subjective rating scales for quantifying the perceptual qualities of polyphonic timbre. Experiment 2 comprised a listening experiment wherein short excerpts of Indian popular music were rated on the perceptual scales obtained from Experiment 1, followed by identification of perceptual dimensions of polyphonic timbre space and the corresponding acoustic correlates.

Specific research questions included:

- Which adjectives are used to describe polyphonic timbre?
- How consistently are these terms used?
- How do these descriptors compare to those of monophonic timbre?
- What are the perceptual dimensions of polyphonic timbre space?
- What are the salient acoustic correlates of the perceptual dimensions and how well can they be used to predict the perceptual ratings?
- How perceptually relevant are the already existing computational polyphonic timbre features?

Methods: The goal of Experiment 1 was to devise a framework for acquiring quantitative assessments of polyphonic timbre. This was achieved by means of a survey, which helped in collecting perceptual scales that could be used to describe polyphonic timbre. The participants comprised students or staff members at the Department of Music. Subsequently, in order to reduce the number of perceptual scales, a preliminary listening test was conducted. The perceptual scales were determined via squared multiple correlation (SMC) analysis. Experiment 2 comprised a listening test using the perceptual scales obtained from Experiment 1. The participants comprised majorly musicians. The behavioural data was subjected to factor analysis to determine the main perceptual dimensions. Following this, computational extraction of acoustic features and correlation analysis helped determine the salient acoustic correlates of the obtained perceptual dimensions. A new set of features, namely the Sub-Band Fluxes were introduced. Regression analysis was used to assess the predictability of the perceptual dimensions using acoustic features.

Results: As a result of Experiment 1 a subset of eight perceptual scales was identified that were applicable for describing polyphonic timbre. Analysis of the perceptual ratings obtained from Experiment 2 revealed high agreement between the participants' perceptual ratings. Subsequently, analysis of the perceptual ratings resulted in the identification of three perceptual dimensions: Activity, Brightness, and Fullness. Table 1 displays the eight scales and their respective squared multiple correlation values (representative of the uniqueness of the scale) obtained from Experiment 1, Cronbach's alphas obtained from Experiment 2, and the perceptual dimensions that encompass them.

TABLE 4 Squared Multiple Correlation Values and Corresponding Cronbach's Alphas for each of the Perceptual Scales Comprising the Perceptual Dimensions

		SMC	Cronbach's Alpha
Activity	Soft – Hard	.88	.96
	Cold – Warm	.75	.90
	Strong – Weak	.71	.91
	High Energy - Low Energy	.69	.96
Brightness	Dark – Bright	.78	.96
	Colorful-Colorless	.75	.91
	Acoustic – Synthetic	.72	.96
Fullness	Full – Empty	.51	.84

Relatively high correlations were observed between certain acoustic features and these perceptual dimensions in particular for the new set of acoustic features that quantify spectrotemporal modulations, that is, the sub-band fluxes. Table 5 displays the five acoustic features having the highest correlation values with each of the three perceptual dimensions.

TABLE 5. Acoustic Features Exhibiting the Highest Correlation with the Perceptual Dimensions

Activity	**Brightness**	**Fullness**
Sub-Band No.7 Flux 0.76***	Zero-crossing Rate 0.44***	Sub-Band No.2 Flux 0.60***
Entropy 0.73***	Sub-Band No.1 Flux -0.40***	Sub-Band No.1 Flux 0.47***
HE-LE ratio 0.71***	Sub-Band No.6 Flux 0.38***	Sub-Band No.3 Flux 0.41***
Zero-crossing Rate 0.67***	Sub-Band No.7 Flux 0.36***	HE-LE ratio -0.32**
Sub-Band No.8 Flux 0.65***	HE-LE ratio 0.34***	MFCC13 -0.32**

* p<0.05, ** p<0.01, *** p<0.001

As can be seen in table 5, spectrotemporal modulations, represented by the Sub-Band Fluxes, correlate significantly with the perceptual dimensions. The MFCCs, on the other hand, do not seem to contribute significantly to any of the perceptual dimensions. The perceptual dimension of Activity could be predicted to a high degree of accuracy (R^2 = .70) using linear regression on acoustic features extracted from the sound. On the other hand, the dimensions of Brightness and Fullness dimensions exhibited lower accuracies of predictability (R^2 = .31 and R^2 = .51 respectively). Results of cross-validation of the regression models for the dimensions suggested that the regression model for Activity is more robust in comparison to the Brightness and Fullness models. A potentially interesting observation is that of the increase in the variance explained by the Fullness model as the window size for feature extraction increases to 700 ms. This trend seem to suggest that different perceptual dimensions operate on different timescales and deserves further investigation.

Conclusion: The aim of this study was to explore the perceptual components of polyphonic timbre and look for eventual regularities. Similarities were found between semantic associations of monophonic and polyphonic timbre. The consistency in participants' ratings suggested the presence of regularities and patterns in the way people perceive polyphonic timbre. Three main perceptual dimensions were found, namely Activity, Brightness and Fullness. The new set of acoustic features introduced representing spectrotemporal modulations, that is, the sub-band fluxes were found to be most relevant, while the well known polyphonic timbre descriptors, the Mel-Frequency Cepstral Coefficients failed to correlate significantly with any of the perceptual dimensions. Furthermore, the perceptual dimensions can be predicted relatively well by the regression models, specifically for the Activity dimension. In addition, for the dimensions of Brightness and Fullness, the regression models could explain a higher proportion of variance if the window length during feature extraction was increased suggesting that different perceptual dimensions operate on different timescales. One main limitation of the study was the use of only one style of music (Indian popular music) and only one type of listeners (Western), which may not allow for generalization of the results. This limitation was overcome by extending the experimental setting in Study III to include a wider range of listeners and musical material.

STUDY III

Introduction: Study III served as a cross-cultural extension to study II in order to lend more weight to the representativeness of the findings. Polyphonic timbre perception was investigated in a cross-cultural context to identify eventual universal patterns and culture-dependent differences. Additionally, the phenomenon of perceptual learning in the context of polyphonic timbre was investigated. Specific research questions included:

- How does the dimensionality of the polyphonic timbre space vary cross-culturally?

- How does the underlying structure or configuration of the perceptual dimensions vary cross-culturally?
- How variant are the acoustic correlates of these perceptual dimensions for the same stimulus set?
- How variant are the acoustic correlates of these perceptual dimensions for the same listener type?

In addition to these questions, the effect of musicality on the dimensionality and configuration of the polyphonic timbre space was also investigated. Additional methods to calculate the intrinsic dimensionality are also suggested. These details were excluded from the published article but will be discussed here nevertheless.

Methods: A balanced design was employed wherein two sets of stimuli comprising short Indian and Western popular music excerpts were rated by non-musicians of Indian and Western origins on the eight perceptual scales derived from Study II. The design is illustrated in Figure 11.

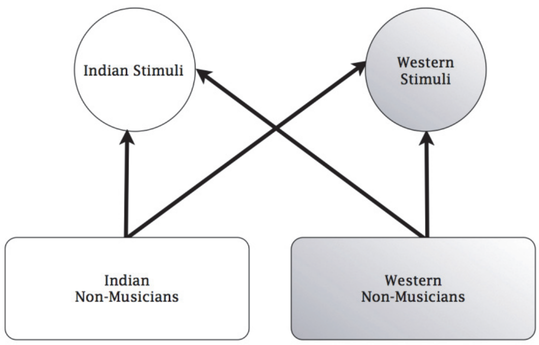

FIGURE 11 Cross-cultural Experimental Design

Factor analysis was subsequently performed on the perceptual ratings to identify the underlying perceptual dimensions for all the datasets. Following this, features extracted from the music stimuli were correlated with the thus obtained perceptual dimensions. Regression analyses were then performed to investigate how well the perceptual dimensions could be predicted from the acoustic features and how well they generalized across the datasets. Finally, cross-validation of the models across groups was performed in order to investigate the universality of the perceptual dimensions.

The methods omitted from the published paper comprised ID estimation techniques in order to unravel the dimensionality of the polyphonic timbre spaces. The intrinsic dimensionality of the four datasets (II – Indian music, Indian listeners; IW - Indian music, Western listeners; WI - Western music, Indian listeners; WW - Western music, Western listeners) was estimated using, in addition to the Eigen Value criterion, methods commonly used in fields such as chaos theory, complex systems theory, and machine learning, namely Maximum Likelihood estimation (MLE) (Levina & Bickel, 2004), Correlation Dimension estimation (CD) (Camastra & Vinciarelli, 2002), and Nearest Neighbor (NN) (Pettis et al. 1979). Furthermore, in order to investigate the role of familiarity and musical training in perceptual judgements of polyphonic timbre, the data obtained by using Indian stimuli from the current study was compared to that of Study II.

Results: Participants displayed a high degree of consistency in their ratings both within and between groups reflected by the high Cronbach's alphas (Group II .96 ≥ α ≥ .74 for all scales, except for 'Warm—Cold' with α = .67; IW .97 ≥ α ≥ .87 for all scales; WI .96 ≥ α ≥ .71 for all scales & WW.97 ≥ α ≥ .90 for all scales). Participants' high level of familiarity with the musical style due to enculturation was associated with increased intrinsic dimensionality, from two dimensions to three based on the Eigen Value criterion, of their respective timbre spaces. The two-factor solutions displayed highly similar loading patterns across participant groups. The three-factor solutions revealed differences in the underlying structure of the perceptual dimensions in the groups familiar with the musical style. Correlation analyses between acoustic features and perceptual dimensions revealed acoustic profiles that were similar across ethnic backgrounds and stimulus types. Figure 12 displays the acoustic profiles for all groups.

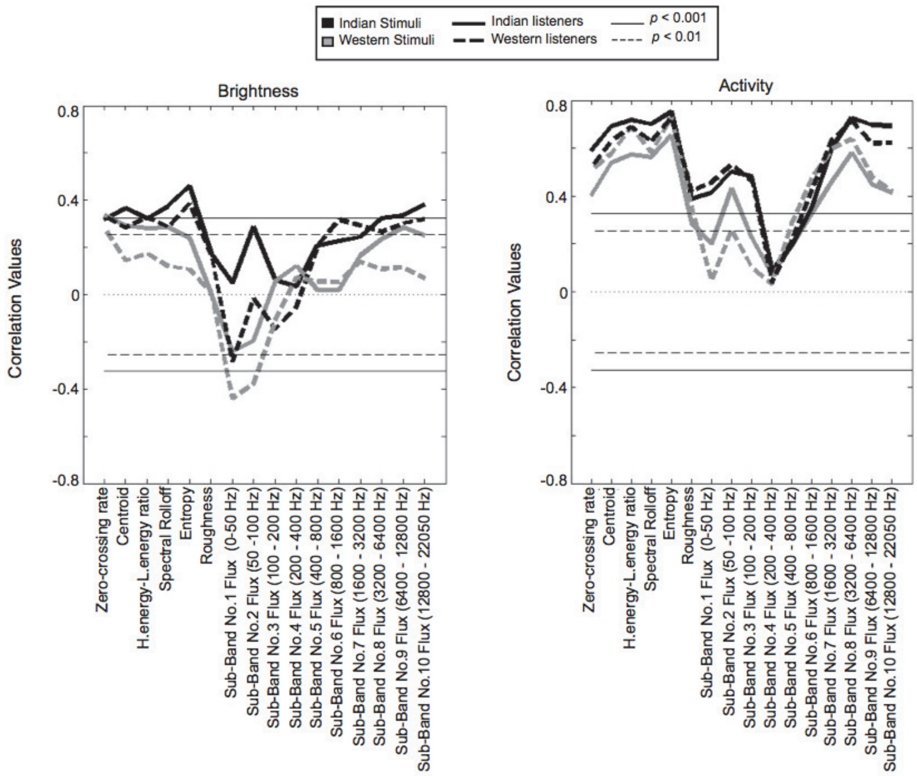

FIGURE 12 Acoustic profiles for each perceptual dimension and group (correlations with absolute values higher than those indicated with the dashed and solid lines are significant at the levels $p < .01$ and $p < .001$, respectively).

Overall, the acoustic features correlated more with Activity than with Brightness. Cross-validations within each stimulus set yielded highly significant correlations. However, across stimulus sets the Activity dimension proved to be more robust.

Results from the additional ID estimation techniques are summarized below. All the dimension estimation methods give rise to the same pattern in which the Indian participants' ratings have higher intrinsic dimensionality than those of the Western participants. For the Western stimuli, the Eigen Value criterion shows a difference between the participant groups wherein higher intrinsic dimensionality is found in the Western participants' ratings. Figure 13 displays the results of ID investigation using all the abovementioned methods.

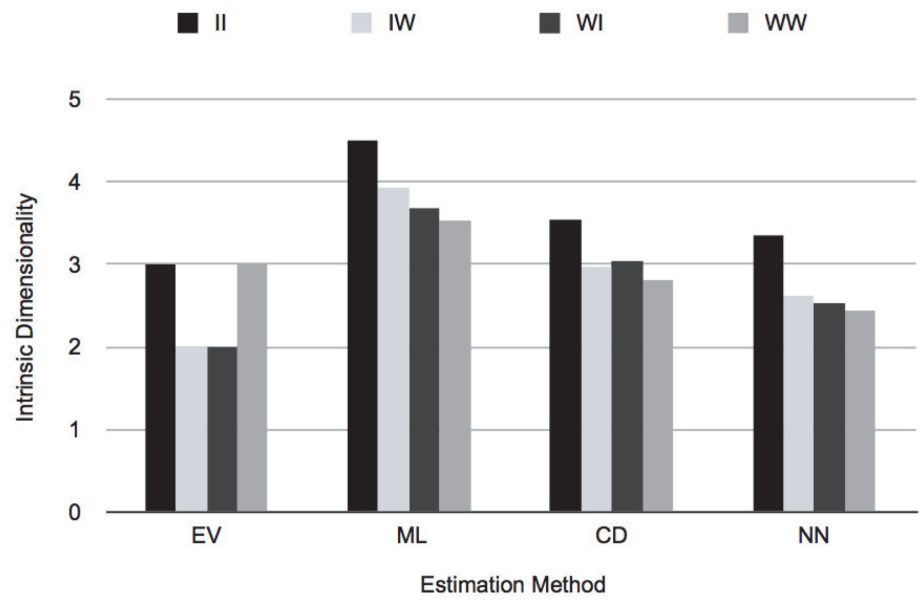

FIGURE 13 Intrinsic Dimensionality estimation results for groups II, IW, WI and WW.

Figure 14 displays the ID estimation results for all the groups that rated only the Indian stimuli (II, IW, and IWM - Indian stimuli, Western musicians: obtained from Study II).

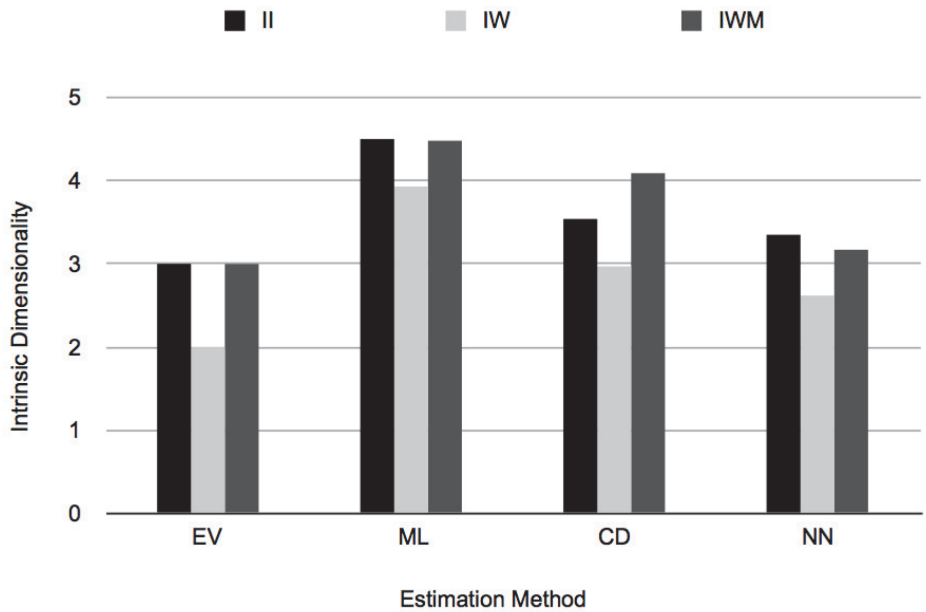

FIGURE 14 Intrinsic Dimensionality estimation comparison for groups II, IW and IWM.

All estimation methods revealed the same underlying patterns with the intrinsic dimensionality of IWM on a level comparable to that of II and notably higher than that of IW. Due to the lack of data from Indian musicians, the comparison was not balanced. Hence the effect of musical training was omitted in the published paper.

Conclusion: The results suggest that there exist commonalities in semantic associations to polyphonic timbre despite cultural differences. An interesting conclusion is that there is increase in the dimensionality of the perceptual timbre space as a result of both prior exposure through enculturation and musical expertise, although there are differences in the underlying structure of the perceptual dimensions. This is in accordance with the notion of perceptual learning via differentiation of dimensions as put forward by Goldstone (1998). However, two main dimensions were repeatedly found for all the datasets, namely, Activity and Brightness whose configurations matched closely with the results of Study II that used musically trained Western participants. This finding suggests that these two dimensions are well representative of two of the perceptual dimensions of the polyphonic timbre space. The significant correlations observed between the acoustic profiles suggest that the perception of polyphonic timbre with regard to its acoustic correlates involves common underlying mechanisms. Cross-validations of the Activity and Brightness models for the Indian and Western participant groups within each stimulus set yielded highly significant correlations (See Tables 7 and 8 in the original paper). However, Activity model proved to be more robust than Brightness suggesting that Activity possess a higher degree of universality in its perception.

Brightness again failed to correlate highly with the present set of acoustic features thereby rendering it a more complex dimension whose model could be improved by using more relevant features, possibly even a non-linear combination of features. Overall, the results suggest the presence of common perceptual mechanisms involved in polyphonic timbre processing for the two dimensions representing Activity and Brightness, however with minor differences being observed due to prior exposure via enculturation and musical training.

STUDY IV

Introduction: Study IV comprised an investigation to uncover the neural underpinnings of timbre processing. Importantly, in order to obtain a more accurate picture of how the brain functions under more realistic conditions, a novel paradigm was introduced that combines methods that facilitate investigating neural processing in more naturalistic settings wherein participants listen to continuous music. Specific questions included:

- What brain areas are recruited during timbre processing?
- How do these brain areas compare to those recruited while processing monophonic timbre and in more controlled conditions?

In addition to timbre, rhythm and tonality processing were also studied but the corresponding results are not discussed because it is beyond the scope of the thesis. The novel paradigm that was introduced is a combination of functional magnetic resonance imaging, computational acoustic feature extraction, behavioural psychology, and non-parametric approaches to estimating statistical significance.

Methods: Brain responses of participants were measured using fMRI while they were listening to an eight minute long piece of music. By employing computational acoustic feature extraction procedures, the temporal evolution of timbral, tonal and rhythmical features of the piece were quantified, whose perceptual relevance was validated through a behavioural experiment. Subsequently the musical features were correlated with the brain responses in order to determine brain areas involved in the concurrent processing of a range of musical features. Novel methods for estimating the significance of the observed correlations, and cluster-size correction for multiple comparisons, were proposed.

Results: Perceptual validation of the acoustic components resulted in four interpretable timbre dimensions representing Activity, Fullness, Brightness and Timbral Complexity. Subsequent correlation analysis revealed brain areas additional to those that were known previously to be recruited while processing controlled stimuli (See Figure 15). Timbre-related acoustic components correlated positively with activations in large areas of the temporal lobe (Superior Temporal Gyrus, Heschl's Gyrus, and Middle Temporal Gyrus), which were previously reported to process monophonic timbre (Halpern, Zatorre, Bouffard, Johnson, 2004; Caclin, Brattico, Tervaniemi, Näätänen, Morlet, et al., 2006). Second, interhemispheric specialization in the auditory cortices was observed, specifically the caudolateral and anteriolateral parts of the superior temporal gyrus (STG), in addition to the right temporal lobe displaying overall larger areas with significant correlations.

Cortical areas belonging to the default-mode-network-related (DMN) cortical areas were found to be involved in the processing of timbral dimensions. Particularly, negative correlations between Activity and Fullness and brain activity were observed in DMN-related cerebrocortical regions in the vicinity of the left superior frontal gyrus (BA 9), the left precuneus and surrounding parietal areas, and the ventral medial prefrontal cortex. The left posterior cingulate, one of the central structures of the DMN, was observed to be deactivated during moments in the stimulus with high Brightness. Interestingly, activations in the right putamen and the bilateral precentral gyrus as well as the right medial frontal gyrus, which appear to be related to movement (Grahn and Rowe, 2009), were found to correlate positively with Brightness and Activity, respectively. In addition to the DMN, negative correlations for timbral features of Activity and Fullness were observed in the somatosensory areas. Furthermore, areas of the cerebellum, including lobule VI, Crus I and II, were found for the first time to be involved in processing timbre-related acoustic components.

Conclusion: The current study introduced a new paradigm to investigate and predict the brain areas related to the processing of timbral, tonal, and rhythmical features of a stimulus under naturalistic conditions. A notable result of this novel naturalistic approach employed is that, in addition to corroborating findings from previous controlled settings, it revealed additional brain areas involved in timbral feature processing. In addition to the auditory cortices with a right-hemispheric specialization, the results for the first time demonstrate that timbral features activated DMN areas of the cerebrum and cognitive areas of the cerebellum under naturalistic settings. The DMN is a neural circuit constantly monitoring the sensory environment and displaying high activity during lack of focused attention on external events (Fox et al., 2009; Mcavoy, Larson-Prior, Nolan, Vaishnavi, Raichle, et al., 2008). As low values in Activity and Fullness were mostly associated with sections in the stimulus with sparse texture played by the piano, thereby resulting in lower levels of auditory-cognitive load, the activation of the DMN during these moments is in line with previous results (Levitin & Menon, 2003; Pallesen, Brattico, Bailey, Korvenoja, Gjedde, 2009; Uddin, Kelly, Biswal, Xavier Castellanos, & Milham, 2009).

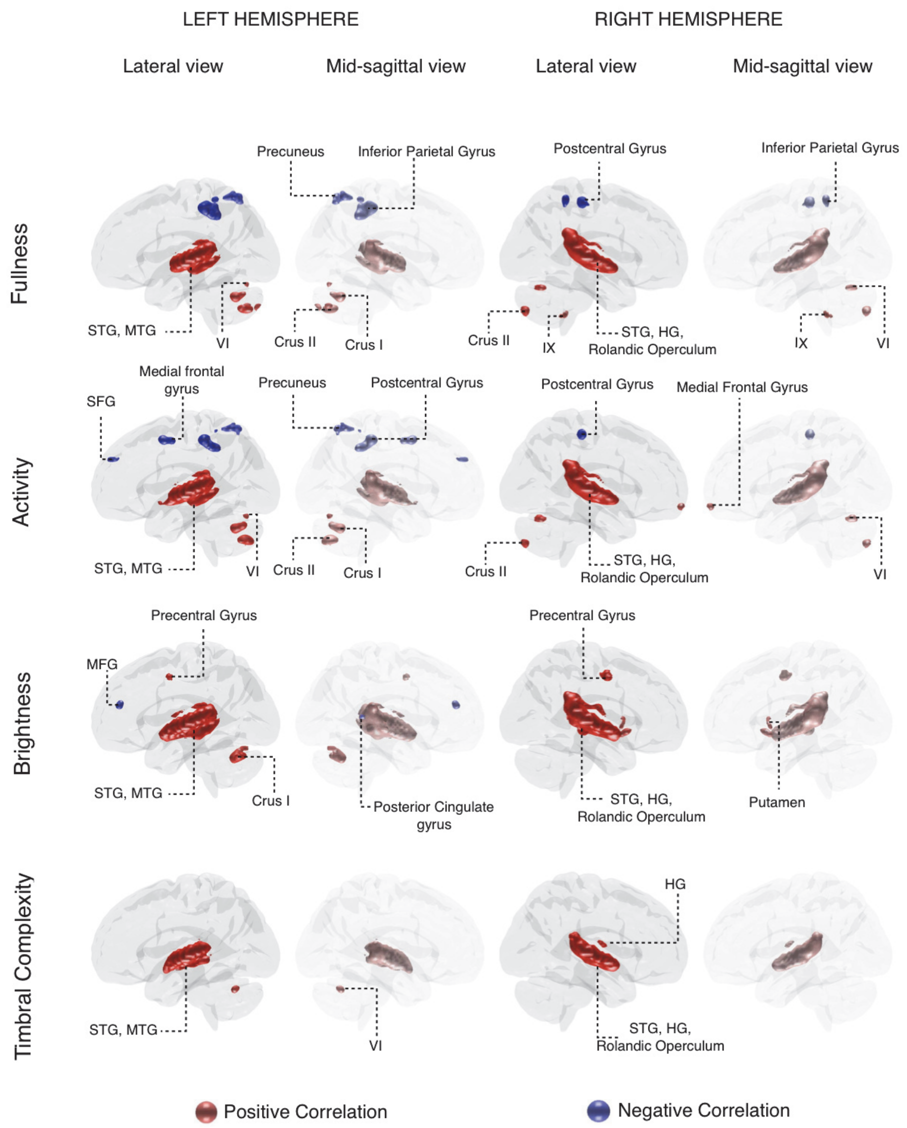

FIGURE 15 Lateral and mid-sagittal views of the left and right hemispheres of the brain showing regions correlating significantly with timbral components. The significance threshold for the correlations was set at $p = .001$. Cluster correction was performed at a significance level of $p = .001$ ($Z = 3.29$) corresponding to a cluster size of 22 voxels. The areas indicated in red and blue correspond to the brain areas that correlated positively and negatively to each of the timbral components, respectively.

The correlations between Activity and Brightness with activations in motor-related brain areas appear to concur partly with the findings of Uddin et al.

(2009) who report the existence of functional connectivity of the motor-control areas to the anticorrelated DMN associated with the posterior cingulate cortex. Further research, however, is called for to clarify the link between timbral features and motor-related brain activity. A potential explanation for this finding is the presence of covariance between the acoustic components representing timbre and rhythm, the latter of which was found to activate motor cortical and subcortical areas of the brain (such as the premotor cortex and basal ganglia). Interestingly, negative correlations for Activity and Fullness were observed in the somatosensory areas, considered to be part of the mirror neuron system (Keysers, Kaas, & Gazzola, 2010; Koelsch, Fritz, von Cramon, Muller, & Friederici, 2006), and were previously known to be recruited while hearing sounds resulting from other people's actions. In the present study, such moments were characterized by the presence of sustained notes. These results suggest lend support to the action-perception coupling mechanism in musicians while listening to music.

Furthermore, the cognitive regions of the cerebellum that were found to correlate with Activity and Fullness have been previously associated with processing high cognitive load in an auditory task with chords (Salmi, Pallesen, Neuvonen, Brattico, Korvenoja et al., 2010). The results for the first time demonstrate the role of the cerebellum in cognitive processing of timbral dimensions under naturalistic music listening conditions. Conducting further experiments with larger sets of naturalistic stimuli representing, for instance, different genres, would allow for generalizations concerning the brain networks involved in processing timbral features in real time.

6 DISCUSSION

The aims of this thesis were to investigate timbre perception in an interdisciplinary setting. The dissertation comprises four studies. The first study investigated the missing link between monophonic timbre and affect. The subsequent three studies explored the perceptual, acoustical and neural correlates of polyphonic timbre.

6.1 Summary of Contributions

The key contributions are listed below.

- This work uncovers the dimensions of the monophonic timbre affect space for the first time.

- It is the first work that deals with polyphonic timbre perception. It reveals the perceptual dimensions of the perceptual polyphonic timbre space.

- It further investigates the effect of enculturation and musical expertise on the dimensionality and configuration of the polyphonic timbre space.

- It uncovers the acoustic correlates of the perceptual dimensions and further examines how they differ based on stimulus-type and listener-type.

- Finally, this dissertation comprises a seminal study that introduces a new paradigm that combines neuroimaging, acoustic feature extraction, behavioural methods, and a novel neuroinformatic procedure for the investigation of music feature processing in the brain in naturalistic settings. As a result, neural correlates of polyphonic timbre are revealed.

In addition to the above contributions, the interdisciplinary nature of the dissertation required the development of new methodologies to perform part of the investigations. The methodological contributions are listed below.

- In order to investigate the intrinsic dimensionality of the perceptual timbre space, methods from alternative fields were adopted. Never before have these alternative methods been used for estimating the dimensionality of a perceptual space, especially in investigating the phenomenon of perceptual learning of musical elements.

- Novel acoustic features that captured spectrotemporal modulations were developed. They proved to be among the topmost correlating acoustic features with the perceptual dimensions.

- The new interdisciplinary paradigm introduced required the development of methods to perform the analysis. This resulted in devising a novel neuroinformatic procedure that allows analyzing fMRI data under naturalistic condition. This comprised specifically of the following:

    - processing acoustic feature time-series in order to make them comparable with the fMRI time-series
    - adapting non-parametric approaches for the estimation of statistical significance of the observed correlations between the fMRI and acoustic feature time-series
    - adapting non-parametric approaches for the estimation of cluster-size correction addressing the problem of multiple comparisons.

The overall results and implications of this dissertation will be discussed in this section.

Polyphonic timbre shares semantic associations similar to those of monophonic timbre. Semantic associations can be thought of as our ability to perceive and subsequently describe sensory information in words. This finding suggests that both phenomena share common underlying perceptual mechanisms. The dimensionality of the polyphonic timbre space was established to be two, and was found to increase to three based on prior exposure and training. This finding lends support to the notion of increase in the dimensions of the perceptual space as a result of prior exposure or training (Goldstone, 1998) albeit with minor differences in the configuration of the dimensions.

Two main dimensions of polyphonic timbre were established, one representing the perceived activity and the other representing the perceived brightness (Study II and Study III). The third dimension was found to represent perceived fullness in the case of musicians (Study II and Study IV). The Fullness dimension however displayed clear patterns for Finnish musicians. However, in order to generalize this dimension, it would be necessary to investigate if this dimension emerges for Indian musicians as well.

An important result is the contribution of the novel dynamic spectrotemporal features introduced, that is the Sub-Band Fluxes, in explaining the perceptual dimensions, in particular of Activity and Fullness. Activity was found to correlate highly with the flux in the frequency range of around 1600 ~ 6400 Hz, represented by Sub-Bands 7 and 8. Fullness was found to correlate with the spectral flux in the lower frequency regions (50 Hz ~ 200 Hz) of the spectrum. Brightness on the other hand was found to generally correlate highest with static[1] features represented by Zero-Crossing Rate and High Energy-Low Energy Ratio. In addition, some significant correlations were also observed with a few of the Sub-Band fluxes. However the overall lower correlation values may suggest that the Brightness dimension is more complex than Activity or Fullness. For instance, the presence of a few notes in the high registers, may render a stimulus perceptually bright although the acoustic features that have been previously associated with perceptual brightness in the monophonic timbre domain, such as spectral centroid, might fail to capture this in a polyphonic mixture. This calls for an extension of the existing feature set to capture the Brightness dimension.

The perceptual dimensions have been derived from semantic data in Studies II and III. Additionally, these dimensions were also obtained by perceptual validation of the acoustic components derived from the stimulus in Study IV. Moreover, the stimuli varied in duration in these studies (1.5 seconds in studies II and III; 6 seconds in study IV). Two inferences can be made at this point. First inference concerns the perceptual relevance of polyphonic timbre as described in section 2.2.1. Till date the studies that investigated the perceptual relevance of polyphonic timbre have commonly used very short durations of musical excerpts (of the order of a few hundred milliseconds). They appear to have taken for granted that polyphonic timbre is what remains when a clear sense of rhythmic, melodic and harmonic progressions due to the short durations. The emergence of similar perceptual dimensions for short (Studies II and III) and longer musical excerpts (Study IV) drive the point home that polyphonic timbre as a property neither is that which exists in the absence of clear rhythmic, melodic and harmonic patterns nor is it an agglomerate of them. Rather the results denote that polyphonic timbre is a perceivable musical element that can be described by its constituent dimensions whose evolution in time can be quantified. Furthermore the results appear to extend the paradox put forth by Bregman (1990, pp.488) that "timbre influences scene analysis, but scene analysis creates timbre" from the monophonic to polyphonic domain. And importantly, polyphonic timbre indeed can be considered an agglomerate of its individual perceptual dimensions of Activity and Brightness for non-musicians irrespective of their cultural origins with an additional dimension of Fullness for musicians.

---

1 The static nature refers to the temporal independency of the features while extraction, which holds good for all the spectral features. Dynamic features on the other hand comprise temporal and spectrotemporal features.

The second inference concerns the perceptual mechanisms involved in processing the perceptual polyphonic timbre dimensions. The common elements binding the behavioural and neural domain are the acoustic features, particularly in this case, the Sub-Band Fluxes. Particularly, the PCA performed on the stimulus in study IV with the perceptual validation of the obtained PCs thereof again suggest that Sub-Band flux related features indeed capture something perceptual. The results indicate that the underlying perceptual mechanisms rely on spectrotemporal modulations. In addition, the Sub-Band Fluxes were also found to be important in explaining the monophonic timbre affect dimensions (Study I). These findings are highly interesting in light of the theory of perception discussed by Hawkins (2004). The author describes the phenomenon of perception as one comprising the integration of spatiotemporal patterns obtained from our senses. In the auditory modality, the spatial element is analogous to the frequency-to-place transformation on the basilar membrane. The brain perceives music by integrating spatiotemporal, or in light of the present thesis, spectrotemporal information, thereby facilitating the formation of a mental representation of its constituent elements. On similar lines, Carterette and Kendall (1999) describe perceptual process as that operating on the principle of contrast or change. To further elaborate, the authors contend that the sensory mechanisms look for changes in order to make sense of the information. Kluender, Coady, & Kiefte (2003) add to this by stating that the perceptual systems of all sensory modalities respond mainly to change. The author also cites work pertaining to the modalities of vision, smell, taste, and touch, which advocate the notion that perceptual systems regulate their reception in a way that helps maximize their sensitivity to change. Spectrotemporal modulations, represented by Sub-Band Fluxes seem to capture this very aspect of perceivable change in polyphonic timbre.

Finally, the methodological contributions of this dissertation have potential implications in the field of music and neuroscience. It could pave way for investigating several music-related phenomena using real music in a manner that can be considered relatively faithful representation of real-life music listening situations.

6.2 Limitations and Future Studies

The current thesis possesses certain limitations that need to be acknowledged. This section outlines these limitations and provides suggestions to overcome them via future studies.

6.2.1 Behavioural Domain

The framework of subjective rating scales that was developed for quantifying the perceptual qualities of polyphonic timbre employed participants mostly of Finnish origin. This can introduce “semantic noise” due to the variance in subjective understanding of the connotative meanings of words from person to person. A natural extension would be to devise the framework of perceptual scales using participants of Indian origin. Furthermore, all the listening experiments were conducted in English while most of the participants were non-native English speakers. However, this limitation may not pose a problem as the results nevertheless displayed similar patterns in semantic associations across all participant groups.

Another potential limitation of the study is that the effect of musical expertise on the dimensionality of the perceptual polyphonic timbre space was not investigated in a balanced fashion due to the lack of perceptual data from Indian musicians. A natural extension to this work would be to include Indian musicians as participants in order to compare the relative importance of enculturation and musical expertise on polyphonic timbre perception.

6.2.2 Acoustic Domain

Despite the careful selection and introduction of novel acoustic features in this thesis, the features correlating with Brightness seem insufficient. This calls for developing new acoustic features or non-linear combinations of existing ones in order to capture perceived brightness. Furthermore, the results of correlation analyses between acoustic features and perceptual dimensions seem to depend on the length of the window used for feature extraction, suggesting that the perceptual dimensions may operate on mutually different time scales. This is a potentially interesting finding that deserves further investigation.

6.2.3 Neural Domain

The generalizability of the neural correlates of perceptual dimensions of polyphonic timbre may be compromised due to three reasons. Firstly the participant pool was limited to Finns. It would be interesting to investigate the generalizability of the findings across a wider set of people with different cultural backgrounds. Second, the participant pool was limited to musicians. It has been shown previously that neural responses to musical elements may vary, albeit in a minor fashion, based on familiarity due to enculturation or musical expertise (Nan et al., 2006; Morrison et al., 2003; Nan et al., 2008; Pantev et al., 2001, Wong et al., 2007). Finally, the use of one piece of music as stimulus may not allow us to generalize over music belonging to different genres. However, the results obtained using the data-driven paradigm could serve as a basis for future hypothesis-driven studies investigating polyphonic timbre perception in an extended context accommodating wider sets of stimuli representing different genres and wider participant pool varying based on their cultural background or musical expertise.

Another potential limitation concerns the assumption of linearity between external stimulus properties and brain responses. However, GLM has been the prevalent default choice in fMRI studies. To investigate the linearity of the relationship a complete comparison of the Pearson vs. Spearman correlations of the data in Study IV was performed. The two kinds of correlations yield very similar outcomes (mean error between Pearson and Spearman correlations, averaged over participants and acoustic components is .0340 with a range of .0224 - .0643). A high mean error would question the linear nature of the relationship. However, since that was not the case, the assumption of linearity appears to hold good.

Furthermore, the brain is active all the time and this makes the activations at any voxel a composite of stimulus-based and non-stimulus-based activations, such as mind wandering and attention shifts. Before performing correlation between the brain responses and the stimulus, it could be beneficial to separate the stimulus-based activations from the non-stimulus based activations. A possible solution could be to tease apart the individual components better is by using blind source separation algorithms, such as Independent Component Analysis (ICA). ICA provides means to separate networks in the brain that process information in a manner statistically independent from other regions in the brain. Furthermore, ICA can be a powerful tool in separating 'noisy' activations (such as physiological noise coming from breathing or cerebral arteries) from task-related activations.

A further limitation is that fMRI serves as an indirect measurement of neuronal activity due to its limited temporal resolution (of the order of a couple of seconds) unlike other brain activity measurement methods such as electroencephalography (EEG) and magnetoencephalography (MEG) (of the order of milliseconds). An option would be to utilize fMRI and EEG simultaneously to allow investigating brain functioning with high spatial and temporal resolution. Measuring EEG in an fMRI scanner comes with its own complications (See Gutberlet, 2010); nevertheless this combination has recently gained attention and provides an attractive option to uncover in greater detail the working of the brain.

CONCLUSION

The main focus of the thesis was to investigate the perception of polyphonic timbre in an interdisciplinary setting. Consistent patterns observed in relation to the perceptual dimensions and the respective acoustic correlates obtained from all the studies advocate to a relatively large extent the universality of the results. The results add to the existing notion that change indeed is what perceptual systems respond to. Furthermore, prior exposure via enculturation or musical expertise plays an important role in altering perceptual mechanisms to respond to different changes. Overall, perceptual mechanisms appear to be highly sensitive to change, which in current work are characterized by spectrotemporal modulations. The neural investigation results revealed that timbral components' processing in a naturalistic condition was associated with brain areas previously linked with processing cognitive and attentional load. Additionally, the recruitment of the somatomotor areas of the brain also indicates that listening to natural timbres triggers action-perception mechanisms in musicians' brains. To conclude, a multifaceted investigation approach is recommended in order to obtain a holistic comprehension of the perceptual processes involved in musical element processing.

YHTEENVETO

Polyfonisen sointivärin hahmottamisen akustiset ja hermostolliset vastineet

Väitöskirja tutkii polyfonisen sointivärin havaitsemista käyttäen monitieteistä lähestymistapaa, joka pyrkii yhdistämään behavioraalisen, akustisen ja neuraalisen näkökulman. Työssä käsitellään ensin monofonisen sointivärin ja tunteiden välistä yhteyttä, minkä jälkeen tarkastellaan polyfonisen sointivärin havaitsemista. Monofoniseen sointiväriin ja tunteisiin keskittynyt tutkimus paljasti, että kaksi päädimensiota, miellyttävyys ja aktiivisuus, selittävät hyvin monofonisen sointivärin tunneavaruuden. Akustisten piirteiden rooli havaintojen selittämisessä oli samankaltainen eri ärsykejoukoilla. Tämä viittaa siihen, että sointivärin välittämien tunteiden havaitsemisessa on universaaleja mekanismeja.

Polyfonisessa sointiväriavaruudessa havaittiin kaksi pääulottuvuutta, aktiivisuus ja kirkkaus. Spektrotemporaalisten muutosten mittaamiseen kehitettiin uusi piirrejoukko, alikanavavaihtelu (*subband flux*). Tämän piirrejoukon havaittiin korreloivan merkitsevästi kuulijoiden arvioiden kanssa. Havaintoulottuvuuksien yleisyyden testaamiseksi suoritettiin kulttuurienvälinen tutkimus. Tässä havaittiin, että kuulijoiden kulttuuritausta ja musiikillinen koulutus vaikuttivat polyfonisen äänenväriavaruuden havaintoulottuvuuksien määrään (2-3). Kolmas ulottuvuus ei kuitenkaan ollut helposti tulkittavissa, joskin muusikoilla sen havaittiin edustavan äänen täyteläisyyttä. Akustisilla piirteillä oli samankaltaiset korrelaatiorakenteet kaikkien ryhmien osalta, mikä viittaa universaalien havaintomekanismien olemassaoloon.

Sointivärin neuraalista prosessointia tutkittiin toiminnallisella magneettikuvauksella (fMRI) yhdistettynä naturalistiseen asetelmaan, jossa käytettiin luonnollisia musiikkiärsykkeitä. Tulokset paljastivat ensimmäistä kertaa, että sointiväriä prosessoidaan kuuloaivokuoren lisäksi myös pikkuaivojen kognitiivisissa osissa, sekä oletusmoodiverkon (*default mode network*) kuorialueilla. Yleisesti ottaen tutkimustulokset viittaavat siihen, että polyfonisen sointivärin ulottuvuuksien prosessointiin liittyvät havaintomekanismit ovat melko universaaleja, vaikkakin pieniä eroja voidaan havaita eri ryhmien välillä. Lisäksi tulokset osoittavat, että nämä mekanismit ovat voimakkaasti yhteydessä spektrotemporaalisiin modulaatioihin, eli ärsykkeessä oleviin ajallisiin muutoksiin. Väitöskirjassa on myös kehitetty uusia menetelmiä, joita voidaan soveltaa musiikin neurotieteessä.

REFERENCES

Adachi, M., Trehub, S. E., & Abe, J. 2004. Perceiving emotion in children's songs across age and culture. *Japanese Psychological Research*, *46 (4)*, 322-336.

Aucouturier, J.J. 2006. *Dix Expériences sur la Modélisation du Timbre Polyphonique*. (Doctoral Dissertation, Universite Paris 6, France).

Aucouturier, J.J. & Pachet, F. 2003. Representing Musical Genre: A State of the Art. *Journal of New Music Research*, *32(1),* 83-93.

Balkwill, L. L., & Thompson, W. F. 1999. A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, *17*, 43-64.

Balkwill, L.L., Thompson, W. F., & Matsunaga, R. 2004. Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research*, *46*, 337-349.

Barbedo, J.G.A., Lopes, A., 2007. Automatic Genre Classification of Musical Signals. *EURASIP Journal on Advances in Signal Processing*, 1-12.

Beauchamp, J. W. 1982. Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones. *The Journal of Audio Engineering Society*, *30(6),* 396-406.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. 2008. The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40(2)*, 531-539.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. 2005. Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion*, *19*, 1113–1139.

Bradley, M. M. 2000. Affective reactions to acoustic stimuli. *Psychophysiology*, *37(2)*, 204-215.

Bregman, A.S. 1990. *Auditory Scene Analysis: The perceptual organization of sound*. Cambridge (MA): MIT Press.

Brett, M., Penny, W., Kiebel, S., 2003. An Introduction to random field theory, In: Frackowiak, R.S.J., Friston, K.J., Frith, C., Dolan, R., Friston, K.J., Price, C.J., Zeki, S., Ashburner, J., Penny, W.D. (Eds.), *Human Brain Function*, 2nd edition. Academic Press.

Brown, S., Merker, B., & Wallin, N. 2000. Evolution of human music through sexual selection. *In The origins of music*, N. Wallin, B. Merker, and S. Brown, Eds. Cambridge, MA: MIT Press. 3-24.

Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M. H., McAdams, S. 2006. Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience, 18,* 1959-1972.

Caclin, A., McAdams, S., Smith, B. K., & Giard, M. H. 2008. Interactive processing of timbre dimensions: An exploration with event-related potentials. *Journal of Cognitive Neuroscience*, *20(1)*, 49-64.

Camastra, F., & Vinciarelli, A. 2002. Estimating the intrinsic dimension of data with a fractal-based approach. *Institute of Electrical and Electronics*

Engineers. Transactions on Pattern Analysis and Machine Intelligence, *24(10)*, 1404–1407.

Carterette, E.C. & Kendall, R. A. 1999. Comparative Music Perception and Cognition. In D. Deutsch, *The Psychology of Music* (pp. 725-791). San Diego: Academic.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. l. 1984. Tonal hierarchies in the music of north India. *Journal of Experimental Psychology*, *113*, 394-412.

Celma, Ò., Herrera P., & Serra X. 2006. Bridging the Music Semantic Gap. Workshop on Mastering the Gap: From Information Extraction to Semantic Representation. 187.

Cosi, P., De Poli, G., & Lauzzana, G. 1994. Auditory modelling and self-organizing neural networks for timbre classification. *Journal of New Music Research*, *23(1)*, 71-98.

Curtis, M. E. & Bharucha, J. J. 2009. Memory and musical expectation for tones in cultural context. *Music Perception*, *26*, 365-375.

Darke. G. 2005. Assessment of timbre using verbal attributes. In C. Traube & S. Lacasse (Eds.), *Proceedings of the Second Conference on interdisciplinary Musicology* (pp. 1-12). Montreal, Quebec, Canada.

Darwin, C. J. and Carlyon, R. P. 1995. Auditory grouping, in *The handbook of perception and cognition,* Volume 6, Hearing edited by B. C. J. Moore (Academic, London), pp. 387-424.

Davis, S.B. & Mermelstein, P. 1980. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *28(4)*, 357–366.

De Poli, G. & Prandoni, P. 1997. Sonological Models for Timbre Characterization. *Journal of New Music Research*, *26(2)*, 170-197.

De Poli, G., Prandoni, P. & Tonella, P. 1993. Timbre clustering by self-organizing neural networks. In G. Haus & I. Pighi (Eds.) *Proceedings of X Colloquium on Musical Informatics* (pp. 102–108).

Demorest, S. M., Morrison, S. J., Beken, M. N., & Jungbluth, D. 2008. Lost in translation: An enculturation effect in music memory performance. *Music Perception*, *25*, 213–223.

Disley, A. C., & Howard, D. M. 2004. Spectral correlates of timbral semantics relating to the pipe organ. *Speech, Music and Hearing*, *46*, 25-39.

Disley, A. C., Howard, D. M., & Hunt, A. D. 2006. Timbral description of musical instruments. *International Conference on Music Perception and Cognition*, 61-68.

Eerola, T., Alluri,V., and Ferrer, R. 2008. Emotional connotations of isolated instruments sounds. In Suk Won Yi (Ed.) *Proceedings of the 10th International Conference on Music Perception and Cognition*, 483–489, Sapporo, Japan. University of Hokkaido.

Eerola, T., Lartillot, O., & Toiviainen, P. 2009. Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. In *Proceedings of 10th International Conference on Music Information Retrieval*.

Eerola, T., Louhivuori, J., & Lebaka, E. 2009. Expectancy in north sami yoiks revisited: the role of data-driven and schema-driven knowledge in the formation of melodic expectations. *Musica Scientiae*, *13(2)*, 39-70.

Ferrer, R. 2009. Embodied Cognition Applied to Timbre and Musical Appreciation: Theoretical Foundation. *British Postgraduate Musicology*, Vol 10.

Fisher, R. A. 1950. *Statistical Methods for Research Workers*. 11th edn. Oliver and Boy, London.

Fitzgerald, R. A., & Lindsay, A. T. 2004. Tying semantic labels to computational descriptors of similar timbres. In *Proceedings of Sound and Music Computing*. Paris: IRCAM.

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., and Raichle, M.E. 2005. The human brain is intrinsically organized into dynamic, anti-correlated functional networks. *Proceedings of the National Academy of Sciences. USA 102*, 9673–9678.

Fox, M. D., Zhang, D., Snyder, A. Z., Raichle, M. E. 2009. The Global Signal and Observed Anticorrelated Resting State Brain Networks. *Journal of Neurophysiology, 101(6)*: 3270-3283.

Friston, K. J., Josephs, O., Zarahn, E., Holmes, A. P., Rouquette, S., Poline, J-B. 2000. To smooth or not to smooth? *Neuroimage, 12*, 196–208.

Gamerman, D. 1997. *Markov Chain Monte Carlo: Stochastic simulation of Bayesian Inference.* Chapman and Hall, London.

Gjerdingen, R, O. & Perrott, D. 2008. Scanning the Dial: The Rapid Recognition of Music Genres. *Journal of New Music Research*. *37 (2)*, 93-100.

Goldstone, R.L. 1998. Perceptual learning. *Annual Review of Psychology, 49,* 585–612.

Gounaropoulos, A and Johnson, C. G. 2006. Synthesising Timbres and Timbre-Changes from Adjectives/Adverbs. *Lecture Notes in Computer Science*, *3907*, 664–675.

Grahn, J. A., Rowe, J. B. 2009. Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *Journal of Neuroscience, 29*, 7540-7548.

Gregory, A. H. & Varney, N. 1996. Cross-cultural comparisons in the affective response to music. *Psychology of Music*, *24*, 47-52.

Grey, J. M. 1977. Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, *61(5)*, 1270-1277.

Grey, J.M. & Gordon, J.W. 1978. Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America*, *63(5)*, 1493-1500.

Gutberlet, I. 2010. Recording EEG signals in side the MRI- In M Ullsperger & S Debener eds., *Simulataneous EEG and fMRI: recording, analysis, and application*, Oxford Univ Press.

Halpern, A. R., Zatorre, R. J., Bouffard, M., Johnson, J. A. 2004. Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia, 42,* 1281–1292.

Handel, S. 1995. Timbre perception and auditory object identification. In B Moore, ed., *Hearing*, pp. 425–462. Academic Press, San Diego, CA.

Harwood, D. L. 1976. Universals in Music: A Perspective From Cognitive Psychology. *Ethnomusicology*, *22*, 521-533.

Hawkins, J., & Blakeslee, S. 2004. *On Intelligence*. New York: Holt.

Helmholtz H. 1954. *On the Sensations of Tone*. Dover, New York.

Holmes, A. P., Blair, R. C., Watson, J. D. G., and Ford, I. 1996. Nonparametric analysis of statistic images from functional mapping experiments. *Journal of Cerebral Blood Flow and Metabolism. 16*, 7–22.

Huron, D. 2001. Is music an evolutionary adaptation? *Annals of the New York Academy of Sciences, 930*, 43–61.

Iverson, P. and Krumhansl, C. L. 1993. Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, *94*, 2595–2603.

Jensen, K. 1999. Timbre models of musical sounds. Unpublished doctoral dissertation, Department of Computer Science, University of Copenhagen.

Jiang, D. N., Lu, L., Zhang, H. J., Tao, J. H., & Cai, L. H. 2002. Music type classification by spectral contrast feature. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, Vol. 1, (pp. 113-116). Lausanne, Switzerland: Los Alamitos, CA: IEEE Computer Society.

Juslin, P. N., & Laukka, P. 2003. Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin*, *129(5)*, 770-814.

Kaiser, H. F. 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20(1)*, 141-151.

Kendall, R.A. & Carterette, E.C. 1993a. Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck adjectives. *Music Perception*, *10(4)*, 445-467.

Kendall, R.A. & Carterette, E.C. 1993b. Verbal attributes of simultaneous wind instrument timbres: II. Adjectives induced from Piston's Orchestration. *Music Perception*, *10(4)*, 469 – 502.

Keysers, C., Kaas, J. H., Gazzola, V. 2010. Somatosensation in Social Perception. *Nature Review Neuroscience, 11*: 417–428.

Koelsch, S., Fritz, T., von Cramon, D. Y., Muller, K., Friederici, A. D. 2006. Investigating emotion with music: an fMRI study. *Human Brain Mapping, 27*: 239- 250.

Kluender, K. R., Coady, J. A., & Kiefte, M. 2003. Sensitivity to change in perception of speech. *Speech Communication, 41*, 59-69.

Krumhansl, C. L. 2010. Plink: "Thin Slices" of music. *Music Perception*, *27(5)*, 337–354.

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., & Louhivuori, J. 2000. Cross-cultural Music Cognition: Cognitive Methodology Applied to North Sami Yoiks, *Cognition*, *76(1)*, 13–58.

Lakatos, S. 2000. A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics*, *62(7)*, 426–1439.

Lartillot, O., & Toiviainen, P. 2007. MIR in Matlab (II): A toolbox for musical feature extraction from audio, In S. Dixon, D. Bainbridge & Rainer Typke (Eds.), *Proceedings of the 8th International Conference on Music Information Retrieval*, (pp. 237-244). Vienna, Austria.

Lazar, N. A. 2008. *The Statistical Analysis of Functional MRI Data*. New York: Springer.

Laukka, P., Juslin, P. N., & Bresin, R. 2005. A dimensional approach to vocal expression of emotion. *Cognition & Emotion, 19(5)*, 633–653.

Laurier, C., Sordo, M., Bozzon, A., Brambilla, M., & Fraternali, P. 2009a. Pharos: An Audiovisual Search Platform using Music Information Retrieval Techniques. In *Conference of the International Society for Music Information Research*, Demo session. Kobe, Japan.

Ledberg, A., Akerman, S., Roland, P. E. 1998. Estimation of the probabilities of 3D clusters in functional brain images. *Neuroimage*, *8*, 113–128.

Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D., Lesaffre, M. 2005. Prediction of Musical Affect Using a Combination of Acoustic Structural Cues, *Journal of New Music Research, 34(1)*, 39-67.

Levitin, D. J., Menon, V. 2003. Musical Structure Is Processed In "Language" Areas Of The Brain: A Possible Role for Brodmann Area 47 in Temporal Coherence. *Neuroimage, 20*, 2142–52.

Li, X. and Stern, R, M. 2004. Feature generation based on maximum normalized acoustic likelihood for improved speech recognition, In *Proceedings of IEEE Transactions on Speech and Audio Processing* (pp. 545-548). Piscataway, NJ: IEEE.

Liu, D., Lu, L., and Zhang, H. J. 2003. Automatic mood detection from acoustic music data. In Holger H. Hoos & David Bainbridge (Eds.), *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR)*, (pp. 81-87). Baltimore, Maryland, USA.

Logan, B. and Salomon, A. 2001. A music similarity function based on signal analysis. In *proceedings IEEE International Conference on Multimedia and Expo* (pp. 745-748). Tokyo, IEEE Computer Spciety, Japan.

Loureiro, M. A., de Paula, H. B. and Yehia, H. C. 2004. Timbre classification of a single musical instrument. In *Proceedings of the 5th International Conference on Music Information Retrieval*, (pp. 546–549). Barcelona: Universitat Pompeu Fabra.

Lu, L., Liu, D., Zhang, H. J. 2006. Automatic Mood Detection and Tracking of Music Audio Signals, *IEEE Transactions on Audio, Speech, and Language Processing, 14(1)*, 5-18.

Lukasik, E., 2005. Towards timbre-driven semantic retrieval of violins, In *Proceedings of the fifth International Conference on Intelligent Systems Design and Applications* (pp. 55-60). Wroclaw, Poland.

McAdams, S. 1984. The auditory image: a metaphor for musical and psychological research on auditory organization. In W. R. Crozier & A.J. Chapman (Eds.), *Cognitive processes in the perception of art.* Amsterdam: North-Holland, 289-323.

McAdams, S. 1993. Recognition of sound sources and events. In S McAdams and E Bigand, eds, *Thinking in sound: the cognitive psychology of human audition*, pp. 146–198. Oxford University Press, Oxford.

McAdams, S. 1999. Perspectives on the contribution of timbre to musical structure. *Computer Music Journal*, *23(3)*, 85-102.

McAdams S. and Giordano B. L. 2009. The perception of musical timbre. In S. Hallam, I. Cross, and M. Thaut (eds.), *Oxford Handbook of Music Psychology* (pp.72-80). Oxford: Oxford University Press.

McAdams, S., Winsberg, S., de Soete, G., and Krimphoff, J. 1995. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes, *Psychological Research*, *58(3)*, 177-192.

Mcavoy, M., Larson-Prior, L., Nolan, T. S., Vaishnavi, S. N., Raichle, M. E., d'Avossa, G. 2008. Resting States Affect Spontaneous BOLD Oscillations In Sensory and Paralimbic Cortex. *Journal of Neurophysiology*, *100(2)*, 922-931.

McAvoy, M.P., Ollinger, J.M., and Buckner, R.L. 2001. Cluster size thresholds for assessment of significant activation in fMRI. *Neuroimage* 13, S198.

McKinney, M. F., and J. Breebaart. 2003. Features for audio and music classification. In Holger H. Hoos & David Bainbridge (Eds.), *Proceedings of the 4th International Conference on Music Information Retrieval* (pp. 151–158). Paris: IRCAM.

Menon, V., Levitin, D., Smith, B., Lembke, A., Krasnow, B., Glazer, D., et al. 2002. Neural correlates of timbre change in harmonic sounds. *Neuroimage*, *17(4)*, 1742-1754.

Merriam, A. 1964. *The Anthropology of Music*. Bloomington: Northwestern University Press.

Meyer, L. B. 1956. *Emotion and Meaning In Music*. (University of Chicago Press, Chicago.

Meyer, R. K., Palmer, C., & Mazo, M. 1998. Affective and coherence responses to Russian laments. *Music Perception*, *16(1)*, 135-150.

Miller, G. 2000. Evolution of human music through sexual selection. In *The origins of music*, N. Wallin, B. Merker, and S. Brown, Eds. Cambridge, MA: MIT Press. 329-360.

Mion, L., & De Poli, G. 2008. Score-independent audio features for description of music expression. *IEEE Transactions on Audio Speech and Language Processing, 16(2)*, 458-466.

Moravec, O., Stepanek, J. 2003. Verbal description of musical sound timbre in Czech language. In A. Askenfelt, S. Felicette, E. Jansson, & J. Sundberg (Eds.), *In Proceedings of the Stockholm Music Acoustics Conference* (pp. 643-645). Stockholm: University of Stockholm.

Morrison S. J., Demorest, S. M., Aylward E. H., Cramer S. C., & Maravillab, K.R. 2003. fMRI investigation of cross-cultural music comprehension, *NeuroImage*, *20*, 378–384.

Nan, Y., Knösche, T. R., & Friederici, A.D. 2006. The perception of musical phrase structure: a cross-cultural ERP study. *Brain Research*, *1094(1)*, 179-191.

Nan, Y., Knosche, T.R., Zysset, S. & Friederici, A.D. 2008. Cross-cultural music phrase processing: an fMRI study, *Human Brain Mapping*, *29 (3)*, 312–328.

Nichols, T.E., Holmes, A.P., 2001. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, *15*, 1–25.

Nykänen, A. & Johansson, Ö. 2003. Development of a language for specifying saxophone timbre. In A. Askenfelt, S. Felicette, E. Jansson, & J. Sundberg (Eds.), *Proceedings of the Stockholm Music Acoustic Conference (*pp. 647-650). Stockholm: University of Stockholm.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. 1957. *The Measurement of Meaning*. Urbana, Ill: University of Illinois Press.

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Gjedde, A. 2009. Cognitive and emotional modulation of brain default operation. *Journal of Cognitive Neuroscience, 21(6),* 1065-1080.

Pampalk, E., Flexer, A. and Widmer. G. 2005. Improvements of audio-based music similarity and genre classification. In *Proceedings of Sixth International Symposium on Music Information Retrieval*, (pp. 628–633). Queen Mary, University of London, UK.

Pettis, K., Bailey, T., Jain, A.K., & Dubes, R. 1979. An intrinsic dimensionality estimator from near-neighbor information. *Institute of Electrical & Electronics Engineers. Transactions on Pattern Analysis and Machine Intelligence*, *1(1)*, 25–36.

Peretz, I., Gagnon, L., & Bouchard, B. 1998. Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68(2)*, 111–141.

Pinker, S. 1997. *How the Mind Works.* New York: W.W. Norton & Co.

Platel, H., Price, C., Baron, J.-C., Wise, R., Lambert, J., Frackowiak, R. S. J., Lechevalier, B., & Eustache, F. 1997. The structural components of music perception: A func- tional anatomical study. *Brain*, *120(2)*, 229-243.

Pratt, R. L. & Doak P. E. 1976. A subjective rating scale for timbre, *Journal of Sound and Vibration*, *43(3)*, 317-328.

Pye, D. 2000. Content-based methods for the management of digital music. In *Proceedings of the 2000 IEEE International Conference on Acoustic Speech and Signal Processing* (pp. 2437-2440). Los Alamitos, CA: IEEE Computer Society.

Sarkar, M., Vercoe, B., & Yang, Y. 2007. Words that describe timbre: A study of auditory perception through language. In J. Cross, J. Hawkins, P. Rebuschat, & M. Rohrmeier (Ed.), *Language and Music as Cognitive Systems Conference*, Conference Booklet (pp. 37-38). Cambridge: University of Cambridge.

Pyper, B. J., Peterman, R. M. 1998. Comparison of methods to account for autocorrelation in correlation analyses of fish data. *Canadian Journal of Fish Aquatic Sciences, 55,* 2127–2140.

Redondo, J., Fraga, I., Padr´on, I., & Pineiro, A. 2008. Affective ratings of sound stimuli. *Behavior Research Methods*, *40(3)*, 784-790.

Salimpoor, V., Benovoy, M., Larcher, K., Dagher, A., Zatorre, R. J. 2011. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*. 14 (2), 257-62.

Salmi, J., Pallesen, K. J., Neuvonen, T., Brattico, E., Korvenoja, A., Salonen, O., et al. 2010. Cognitive and Motor Loops of the Human Cerebro-Cerebellar System. *Journal of Cognitive Neurosciences, 22,* 2663-76.

Sandell, G. J. 1995. Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration. *Music Perception. 13(2)*, 209-246.

Sarkar, M., Vercoe, B., & Yang, Y. 2007. Words that describe timbre: A study of auditory perception through language. In J. Cross, J. Hawkins, P. Rebuschat, & M. Rohrmeier (Eds.), *Language and Music as Cognitive Systems Conference* (pp. 37-38). Cambridge, UK: University of Cambridge

Schellenberg, E., Iverson, P., & McKinnon, M. 1999. Name that tune: identifying popular recordings from brief excerpts. *Psychonomic bulletin & review*, *6(4)*, 641-646.

Scherer, K. R. & Oshinsky, J. S. 1977. Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1(4)*, 331–346.

Schuller, B., Wallhoff, F., Arsic, D and Rigoll, G. 2006. Musical signal type discrimination based on large open feature sets. In Proceedings of IEEE International Conference on Multimedia and Expo (pp. 1089–1092). Los Alamitos, CA: IEEE Computer Society.

Schmidt, E., M. & Kim, Y. E. 2010. Prediction of time-varying musical mood distributions from audio. In Proc. of the *International Society for Music Information Conference*. Utrecht, Netherlands.

Sethares, W. A. 1998. *Tuning, Timbre, Spectrum*, Scale, Springer-Verlag.

Shannon, C. E. 1948. A Mathematical Theory of Communication. *Bell Systems Technical Journal, 27*, 379–423.

Smith, L. D. & Williams, R. N. 1999. Children's artistic responses to musical intervals. *American Journal of Psychology*, *112*, 383– 410.

Terasawa, H., Slaney, M. and Berger, J. 2005. Perceptual distance in timbre space. In *Proceedings of International Conference on Auditory Display*, (pp. 61 – 68). University of Limerick, Ireland.

Toiviainen, P. & Eerola, T. 2003 'Where Is the Beat? : Comparison of Finnish and South African Listeners', in R. Kopiez, A.C. Lehmann, I. Wolther and C. Wolf (eds) *Proceedings of the 5th Triennial Escom Conference*, pp. 501–4. Hanover: Hanover University of Music and Drama.

Toiviainen, P., Kaipainen, M. & Louhivuori, J. 1995. Musical timbre: similarity ratings correlate with computational feature space distances. *Journal of New Music Research*, *24(3),* 282–298.

Toiviainen, P., Tervaniemi, M., Louhivuori, J., Saher, M., Huotilainen, M., and Näätänen, R. 1998. Timbre Similarity: Convergence of Neural, Behavioral, and Computational Approaches. *Music Perception, 16 (2),* 223-242.

Trehub, S. 2000. Human Processing Predispositions and Musical Universals. In N. L. Wallin, B. Merker & S.Brown (Eds.), *The Origins of Music* (pp. 427-448). Cambridge, Massachusetts: MIT Press.

Turnbull, D., Barrington, L., & Lanckriet, G. 2006. Modeling music and words using a multi-class naive bayes approach. In *Proceedings of the International Symposium on Music Information Retrieval* (pp. 254-259). Victoria, Canada: University of Victoria.

Turnbull, D., Barrington, L., Torres, D. and Lanckriet, G. 2007. Towards musical query-by-semantic-description using the cal500 data set. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 439–446). Amsterdam: ACM SIGIR.

Turnbull, D., Barrington, L., Torres, D. and Lanckriet, G. 2008. Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech and Language Processing*, *16 (2)*, 467-476.

Tzanetakis, G. & Cook, P. 2002. Music genre classification of audio signals. *Institute of Electrical & Electronics Engineers. Transactions on Speech and Audio Processing*, *10*, 293–302.

Uddin, L. Q., Kelly, A. M. C., Biswal, B. B., Xavier Castellanos, F., Milham, M. P. 2009. Functional connectivity of default mode network components: correlation, anticorrelation, and causality. *Human Brain Mapping, 30,* 625–637.

Von Bismarck, G. 1974. Sharpness as an attribute of the timbre of steady sounds, *Acustica, 30(3)*, 159-172.

Whitman, B. and Ellis, D. 2004. Automatic record reviews, In *Proceedings of the Fifth International Symposium on Music Information Retrieval*, (pp. 470-477). Barcelona: Universitat Pompeu Fabra.

Worsley, K.J., 2003. Developments in random field theory, In: Frackowiak, R.S.J., Friston, K.J., Frith, C., Dolan, R., Friston, K.J., Price, C.J., Zeki, S., Ashburner, J., Penny, W.D.(Eds.), *Human Brain Function*, 2nd edition. Academic Press.

Worsley, K.J., Evans, A.C., Marrett, S., and Neelin, P. 1992. A three dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, *12*, 900-918.

Yang, Y. H., Lin, Y. C., Su, Y. F., & Chen, H. H. 2008. A regression approach to music emotion recognition. *IEEE Transactions on Audio, Speech and Language Processing (TASLP), 16(2)*, 448–457.

Zacharopoulou, K & Kyriakidou, A. 2009. A Cross-Cultural Comparative Study of the Role of Musical Structural Features in the Perception of Emotion in Greek Traditional Music. *Journal of Interdisciplinary Music Studies*, *3(1)*, 1-15.

Zentner, M. R., Eerola T. 2010. Rhythmic engagement with music in infancy. *Proceeding of the National Academy of Sciences*. *107(13)*, 5768-5773.

Zhu, J., Xue, X. and Lu, H. 2004. Musical genre classification by instrumental features. *In Proceedings of the International Computer Music Conference*, (pp. 580-583). Miami: University of Michigan.

JYVÄSKYLÄ STUDIES IN HUMANITIES

1 Kostiainen, Emma, Viestintä ammattiosaamisen ulottuvuutena. - Communication as a dimension of vocational competence. 305 p. Summary 4 p. 2003.

2 Seppälä, Antti, Todellisuutta kuvaamassa – todellisuutta tuottamassa. Työ ja koti television ja vähän radionkin uutisissa. - Describing reality – producing reality. Discourses of work and home in television and on a small scale in radio news. 211 p. Summary 3 p. 2003.

3 Gerlander, Maija, Jännitteet lääkärin ja potilaan välisessä viestintäsuhteessa. - Tensions in the doctor-patient communication and relationship. 228 p. Summary 6 p. 2003.

4 Lehikoinen, Taisto, Religious media theory - Understanding mediated faith and christian applications of modern media. - Uskonnollinen mediateoria: Modernin median kristilliset sovellukset. 341 p. Summary 5 p. 2003.

5 Jarva, Vesa, Venäläisperäisyys ja ekspressiivisyys suomen murteiden sanastossa. - Russian influence and expressivity in the lexicon of Finnish dialects. 215 p. 6 p. 2003.

6 Uskali, Turo, "Älä kirjoita itseäsi ulos" Suomalaisen Moskovan-kirjeenvaihtajuuden alkutaival 1957–1975. - "Do not write yourself out" The beginning of the Finnish Moscow-correspondency in 1957–1975. 484 p. Summary 4 p. 2003.

7 Valkonen, Tarja, Puheviestintätaitojen arviointi. Näkökulmia lukioikäisten esiintymis- ja ryhmätaitoihin. - Assessing speech communication skills. Perspectives on presentation and group communication skills among upper secondary school students. 310 p. Summary 7 p. 2003.

8 Tampere, Kaja, Public relations in a transition society 1989-2002. Using a stakeholder approach in organisational communications and relation analyses. 137 p. 2003.

9 Eerola, Tuomas, The dynamics of musical expectancy. Cross-cultural and statistical approaches to melodic expectations. - Musiikillisten odotusten tarkastelu kulttuurienvälisten vertailujen ja tilastollisten mallien avulla. 84 p. (277 p.) Yhteenveto 2 p. 2003.

10 Paananen, Pirkko, Monta polkua musiikkiin. Tonaalisen musiikin perusrakenteiden kehittyminen musiikin tuottamis- ja improvisaatiotehtävissä ikävuosina 6–11. - Many paths to music. The development of basic structures of tonal music in music production and improvisation at the age of 6–11 years. 235 p. Summary 4 p. 2003.

11 Laaksamo, Jouko, Musiikillisten karakterien metamorfoosi. Transformaatio- ja metamorfoosiprosessit Usko Meriläisen tuotannossa vuosina 1963–86. - "Metamorphosis of musical characters". Transformation and metamorphosis processes in the works of Usko Meriläinen during 1963–86. 307 p. Summary 3 p. 2004.

12 Rautio, Riitta, *Fortspinnungstypus* Revisited. Schemata and prototypical features in J. S. Bach's Minor-Key Cantata Aria Introductions. - Uusi katsaus kehitysmuotoon. Skeemat ja prototyyppiset piirteet J. S. Bachin kantaattien molliaarioiden alkusoitoissa. 238 p. Yhteenveto 3 p. 2004.

13 Mäntylä, Katja, Idioms and language users: the effect of the characteristics of idioms on their recognition and interpretation by native and non-native speakers of English. - Idiomien ominaisuuksien vaikutus englannin idiomien ymmärtämiseen ja tulkintaan syntyperäisten ja suomea äidinkielenään puhuvien näkökulmasta. 239 p. Yhteenveto 3 p. 2004.

14 Mikkonen, Yrjö, On conceptualization of music. Applying systemic approach to musicological concepts, with practical examples of music theory and analysis. - Musiikin käsitteellistämisestä. Systeemisen tarkastelutavan soveltaminen musikologisiin käsitteisiin sekä käytännön esimerkkejä musiikin teoriasta ja analyysistä. 294 p. Yhteenveto 10 p. 2004.

15 Holm, Jan-Markus, Virtual violin in the digital domain. Physical modeling and model-based sound synthesis of violin and its interactive application in virtual environment. - Virtuaalinen viulu digitaalisella alueella. Viulun fysikaalinen mallintaminen ja mallipohjainen äänisynteesi sekä sen vuorovaikutteinen soveltaminen virtuaalitodellisuus ympäristössä. 74 p. (123 p.) Yhteenveto 1 p. 2004.

16 Kemp, Chris, Towards the holistic interpretation of musical genre classification. - Kohti musiikin genreluokituksen kokonaisvaltaista tulkintaa. 302 p. Yhteenveto 1 p. 2004.

17 Leinonen, Kari, Finlandssvenskt sje-, tje- och s-ljud i kontrastiv belysning. 274 p. Yhteenveto 4 p. 2004.

18 Mäkinen, Eeva, Pianisti cembalistina. Cembalotekniikka cembalonsoittoa aloittavan pianistin ongelmana. - The Pianist as cembalist. Adapting to harpsichord technique as a problem for pianists beginning on the harpsichord. 189 p. Summary 4 p. 2004.

19 Kinnunen, Mauri, Herätysliike kahden kulttuurin rajalla. Lestadiolaisuus Karjalassa 1870-1939. - The Conviction on the boundary of two cultures. Laestadianism in Karelia in 1870-1939. 591 p. Summary 9 p. 2004.

20 Лилия Сиберг, "БЕЛЫЕ ЛИЛИИ". ГЕНЕЗИС ФИНСКОГО МИФА В БОЛГАРИИ. РОЛЬ РУССКОГО ФЕННОИЛЬСТВА. ФИНСКО-БОЛГАРСКИЕ КОНТАКТЫ И ПОСРЕДНИКИ С КОНЦА XIX ДО КОНЦА XX ВЕКА. 284 с. - "Belye lilii". Genezis finskogo mifa v Bolgarii. Rol' russkogo fennoil'stva. Finsko-bolgarskie kontakty i posredniki s konca XIX do konca XX veka. 284 p. Yhteenveto 2 p. 2004.

21 FUCHS, BERTOLD, Phonetische Aspekte einer Didaktik der Finnischen Gebärdensprache als Fremdsprache. - Suomalainen viittomakieli vieraana kielenä. Didaktinen fonetiikka. 476 p. Yhteenveto 14 p. 2004.

22 JÄÄSKELÄINEN, PETRI, Instrumentatiivisuus ja nykysuomen verbinjohto. Semanttinen tutkimus. - Instrumentality and verb derivation in Finnish. A semantic study. 504 p. Summary 5 p. 2004.

23 MERTANEN TOMI, Kahdentoista markan kapina? Vuoden 1956 yleislakko Suomessa. - A Rebellion for Twelve Marks? – The General Strike of 1956 in Finland. 399 p. Summary 10 p. 2004.

24 MALHERBE, JEAN-YVES, L'œuvre de fiction en prose de Marcel Thiry : une lecture d'inaboutissements. 353 p. Yhteenveto 1 p. 2004.

25 KUHNA, MATTI, Kahden maailman välissä. Marko Tapion *Arktinen hysteria* Väinö Linnan haastajana. - Between two worlds. Marko Tapio's Arktinen hysteria as a challenger to Väinö Linna. 307p. Summary 2 p. 2004.

26 VALTONEN, HELI, Minäkuvat, arvot ja mentaliteetit. Tutkimus 1900-luvun alussa syntyneiden toimihenkilönaisten omaelämäkerroista. - Self-images, values and mentalities. An autobiographical study of white collar women in twentieth century Finland. 272 p. Summary 6 p. 2004.

27 PUSZTAI, BERTALAN, Religious tourists. Constructing authentic experiences in late modern hungarian catholicism. - Uskontoturistit. Autenttisen elämyksen rakentaminen myöhäismodernissa unkarilaisessa katolisuudessa. 256 p. Yhteenveto 9 p. Summary in Hungarian 9 p. 2004.

28 PÄÄJOKI, TARJA, Taide kulttuurisena kohtaamispaikkana taidekavatuksessa. - The arts as a place of cultural encounters in arts education. 125 p. Summary 3 p. 2004.

29 JUPPI, PIRITA, "Keitä me olemme? Mitä me haluamme?" Eläinoikeusliike määrittelykamppailun, marginalisoinnin ja moraalisen paniikin kohteena suomalaisessa sanomalehdistössä. - "Who are we? What do we want?" The animal rights movement as an object of discursive struggle, marginalization and moral panic in Finnish newspapers. 315 p. Summary 6 p. 2004.

30 HOLMBERG, JUKKA, Etusivun politiikkaa. Yhteiskunnallisten toimijoiden representointi suomalaisissa sanomalehtiuutisissa 1987–2003. - Front page politics. Representation of societal actors in Finnish newspapers' news articles in 1987-2003. 291 p. Summary 2 p. 2004.

31 LAGERBLOM, KIMMO, Kaukana Kainuussa, valtaväylän varrella. Etnologinen tutkimus Kontiomäen rautatieläisyhteisön elinkaaresta 1950 – 1972. - Far, far away, nearby a main passage. An ethnological study of the life spans of Kontiomäki railtown 1950 – 1972. 407 p. Summary 2 p. 2004.

32 HAKAMÄKI, LEENA, Scaffolded assistance provided by an EFL teacher during whole-class interaction. - Vieraan kielen opettajan antama oikea-aikainen tuki luokkahuoneessa. 331 p. Yhteenveto 7 p. 2005.

33 VIERGUTZ, GUDRUN, Beiträge zur Geschichte des Musikunterrichts an den Gelehrtenschulen der östlichen Ostseeregion im 16. und 17. Jahrhundert. - Latinankoulujen musiikinopetuksen historiasta itäisen Itämeren rannikkokaupungeissa 1500- ja 1600-luvuilla. 211 p. Yhteenveto 9 p. 2005.

34 NIKULA, KAISU, Zur Umsetzung deutscher Lyrik in finnische Musik am Beispiel Rainer Maria Rilke und Einojuhani Rautavaara. - Saksalainen runous suomalaisessa musiikissa, esimerkkinä Rainer Maria Rilke ja Einojuhani Rautavaara. 304 p. Yhteenveto 6 p. 2005.

35 SYVÄNEN, KARI, Vastatunteiden dynamiikka musiikkiterapiassa. - Counter emotions dynamics in music therapy. 186 p. Summary 4 p. 2005.

36 ELORANTA, JARI & OJALA, JARI (eds), East-West trade and the cold war. 235 p. 2005.

37 HILTUNEN, KAISA, Images of time, thought and emotions: Narration and the spectator's experience in Krzysztof Kieslowski 's late fiction films. - Ajan, ajattelun ja tunteiden kuvia. Kerronta ja katsojan kokemus Krzysztof Kieslowskin myöhäisfiktiossa. 203 p. Yhteenveto 5 p. 2005.

38 AHONEN, KALEVI, From sugar triangle to cotton triangle. Trade and shipping between America and Baltic Russia, 1783-1860. 572 p. Yhteenveto 9 p. 2005.

39 UTRIAINEN, JAANA, A gestalt music analysis. Philosophical theory, method, and analysis of Iegor Reznikoff's compositions. - Hahmoperustainen musiikkianalyysi. Hahmofilosofinen teoria, metodi ja musiikkianalyysi Iégor Reznikoffin sävellyksistä. 222 p. Yhteenveto 3 p. 2005.

40 MURTORINNE, ANNAMARI, *Tuskan hauskaa*! Tavoitteena tiedostava kirjoittaminen. Kirjoittamisprosessi peruskoulun yhdeksännellä luokalla. - Painfully fun! Towards reflective writing process. 338 p. 2005.

41 TUNTURI, ANNA-RIITTA, Der Pikareske Roman als Katalysator in Geschichtlichen Abläufen. Erzählerische Kommunikationsmodelle in *Das Leben des Lazarillo von Tormes*, bei Thomas Mann und in EinigenFinnischen Romanen. 183 p. 2005.

42 LUOMA-AHO, VILMA, Faith-holders as Social Capital of Finnish Public Organisations. - Luottojoukot – Suomalaisten julkisten organisaatioiden sosiaalista pääomaa. 368 p. Yhteenveto 8 p. 2005.

63 RUUSKANEN, ESA, Viholliskuviin ja viranomaisiin vetoamalla vaiennetut työväentalot. Kuinka Pohjois-Savon Lapuan liike sai nimismiehet ja maaherran sulkemaan 59 kommunistista työväentaloa Pohjois-Savossa vuosina 1930–1932. - The workers' halls closed by scare-mongering and the use of special powers by the authorities. 248 p. Summary 5 p. 2006.

64 VARDJA, MERIKE, Tegelaskategooriad ja tegelase kujutamise vahendid Väinö Linna romaanis "Tundmatu sõdur". - Character categories and the means of character representation in Väinö Linna's Novel *The Unknown Soldier*. 208 p. Summary 3 p. 2006.

65 TAKÁTS, JÓZSEF, Módszertani berek. Írások az irodalomtörténet-írásról. - The Grove of Methodology. Writings on Literary Historiography. 164 p. Summary 3 p. 2006.

66 MIKKOLA, LEENA, Tuen merkitykset potilaan ja hoitajan vuorovaikutuksessa. - Meanings of social support in patient-nurse interaction. 260 p. Summary 3 p. 2006.

67 SAARIKALLIO, SUVI, Music as mood regulation in adolescence. - Musiikki nuorten tunteiden säätelynä. 46 p. (119 p.) Yhteenveto 2 p. 2007.

68 HUJANEN, ERKKI, Lukijakunnan rajamailla. Sanomalehden muuttuvat merkitykset arjessa. - On the fringes of readership. The changing meanings of newspaper in everyday life. 296 p. Summary 4 p. 2007.

69 TUOKKO, EEVA, Mille tasolle perusopetuksen englannin opiskelussa päästään? Perusopetuksen päättövaiheen kansallisen arvioinnin 1999 eurooppalaisen viitekehyksen taitotasoihin linkitetyt tulokset. - What level do pupils reach in English at the end of the comprehensive school? National assessment results linked to the common European framework. 338 p. Summary 7 p. Sammanfattning 1 p. Tiivistelmä 1 p. 2007.

70 TUIKKA, TIMO, "Kekkosen konstit". Urho Kekkosen historia- ja politiikkakäsitykset teoriasta käytäntöön 1933–1981. - "Kekkonen´s way". Urho Kekkonen's conceptions of history and politics from theory to practice, 1933–1981 413 p. Summary 3 p. 2007.

71 Humanistista kirjoa. 145 s. 2007.

72 NIEMINEN, LEA, A complex case: a morphosyntactic approach to complexity in early child language. 296 p. Tiivistelmä 7 p. 2007.

73 TORVELAINEN, PÄIVI, Kaksivuotiaiden lasten fonologisen kehityksen variaatio. Puheen ymmärrettävyyden sekä sananmuotojen tavoittelun ja tuottamisen tarkastelu. - Variation in phonological development of two-year-old Finnish children. A study of speech intelligibility and attempting and production of words. 220 p. Summary 10 p. 2007.

74 SIITONEN, MARKO, Social interaction in online multiplayer communities. - Vuorovaikutus verkkopeliyhteisöissä. 235 p. Yhteenveto 5 p. 2007.

75 STJERNVALL-JÄRVI, BIRGITTA, Kartanoarkkitehtuuri osana Tandefelt-suvun elämäntapaa. - Manor house architecture as part of the Tandefelt family´s lifestyle. 231 p. 2007.

76 SULKUNEN, SARI, Text authenticity in international reading literacy assessment. Focusing on PISA 2000. - Tekstien autenttisuus kansainvälisissä lukutaidon arviointitutkimuksissa: PISA 2000. 227 p. Tiivistelmä 6 p. 2007.

77 KŐSZEGHY, PÉTER, Magyar Alkibiadés. Balassi Bálint élete. - The Hungarian Alcibiades. The life of Bálint Balass. 270 p. Summary 6 p. 2007.

78 MIKKONEN, SIMO, State composers and the red courtiers - Music, ideology, and politics in the Soviet 1930s - Valtion säveltäjiä ja punaisia hoviherroja. Musiikki, ideologia ja politiikka 1930-luvun Neuvostoliitossa. 336 p. Yhteenveto 4 p. 2007.

79 SIVUNEN, ANU, Vuorovaikutus, viestintäteknologia ja identifioituminen hajautetuissa tiimeissä. - Social interaction, communication technology and identification in virtual teams. 251 p. Summary 6 p. 2007.

80 LAPPI, TIINA-RIITTA, Neuvottelu tilan tulkinnoista. Etnologinen tutkimus sosiaalisen ja materiaalisen ympäristön vuorovaikutuksesta jyväskyläläisissä kaupunkipuhunnoissa. - Negotiating urban spatiality. An ethnological study on the interplay of social and material environment in urban narrations on Jyväskylä. 231 p. Summary 4 p. 2007.

81 HUHTAMÄKI, ULLA, "Heittäydy vapauteen". Avantgarde ja Kauko Lehtisen taiteen murros 1961–1965. - "Fling yourself into freedom!" The Avant-Garde and the artistic transition of Kauko Lehtinen over the period 1961–1965. 287 p. Summary 4 p. 2007.

82 KELA, MARIA, *Jumalan kasvot* suomeksi. Metaforisaatio ja erään uskonnollisen ilmauksen synty. - God's face in Finnish. Metaphorisation and the emergence of a religious expression. 275 p. Summary 5 p. 2007.

83 SAARINEN, TAINA, Quality on the move. Discursive construction of higher education policy from the perspective of quality. - Laatu liikkeessä. Korkeakoulupolitiikan diskursiivinen rakentuminen laadun näkökulmasta. 90 p. (176 p.) Yhteenveto 4 p. 2007.

84 MÄKILÄ, KIMMO, Tuhoa, tehoa ja tuhlausta. Helsingin Sanomien ja New York Timesin ydinaseuutisoinnin tarkastelua diskurssianalyyttisesta näkökulmasta 1945–1998.

and early twentieth-century narrative literature. 208 p. Summary 3 p. 2008.

106 Päärnilä, Ossi, Hengen hehkusta tietostrategioihin. Jyväskylän yliopiston humanistisen tiedekunnan viisi vuosikymmentä. 110 p. 2008.

107 Kangasniemi, Jukka, Yksinäisyyden kokemisen avainkomponentit Yleisradion tekstitelevision Nuorten palstan kirjoituksissa. - The key components of the experience of loneliness on the Finnish Broadcasting Company's (YLE) teletext forum for adolescents. 388 p. 2008.

108 Gajdó, Tamás, Színháztörténeti metszetek a 19. század végétől a 20. század közepéig. - Segments of theatre history from the end of the 19th century to the middle of the 20th century. 246 p. Summary 2 p. 2008.

109 Catani, Johanna, Yritystapahtuma kontekstina ja kulttuurisena kokemuksena. - Corporate event as context and cultural experience. 140 p. Summary 3 p. 2008.

110 Mahlamäki-Kaistinen, Riikka, Mätänevän velhon taidejulistus. Intertekstuaalisen ja -figuraalisen aineiston asema Apollinairen L'Enchanteur pourrissant teoksen tematiikassa ja symboliikassa. - Pamphlet of the rotten sorcerer. The themes and symbols that intertextuality and interfigurality raise in Apollinaire's prose work L'Enchanteur pourrissant. 235 p. Résumé 4 p. 2008.

111 Pietilä, Jyrki, Kirjoitus, juttu, tekstielementti. Suomalainen sanomalehtijournalismi juttutyyppien kehityksen valossa printtimedian vuosina 1771-2000. - Written Item, Story, Text Element. Finnish print journalism in the light of the development of journalistic genres during the period 1771-2000. 779 p. Summary 2 p. 2008.

112 Saukko, Päivi, Musiikkiterapian tavoitteet lapsen kuntoutusprosessissa. - The goals of music therapy in the child's rehabilitation process. 215 p. Summary 2 p. 2008.

113 Lassila-Merisalo, Maria, Faktan ja fiktion rajamailla. Kaunokirjallisen journalismin poetiikka suomalaisissa aikakauslehdissä. - On the borderline of fact and fiction. The poetics of literary journalism in Finnish magazines. 238 p. Summary 3 p. 2009.

114 Knuutinen, Ulla, Kulttuurihistoriallisten materiaalien menneisyys ja tulevaisuus. Konservoinnin materiaalitutkimuksen heritologiset funktiot. - The heritological functions of materials research of conservation. 157 p. (208 p.) 2009.

115 Niiranen, Susanna, «Miroir de mérite». Valeurs sociales, rôles et image de la femme dans les textes médiévaux des *trobairitz*. - "Arvokkuuden peili". Sosiaaliset arvot, roolit ja naiskuva keskiaikaisissa *trobairitz*-teksteissä. 267 p. Yhteenveto 4 p. 2009.

116 Aro, Mari, Speakers and doers. Polyphony and agency in children's beliefs about language learning. - Puhujat ja tekijät. Polyfonia ja agentiivisuus lasten kielenoppimiskäsityksissä. 184 p. Yhteenveto 5 p. 2009.

117 Jantunen, Tommi, Tavu ja lause. Tutkimuksia kahden sekventiaalisen perusyksikön olemuksesta suomalaisessa viittomakielessä. - Syllable and sentence. Studies on the nature of two sequential basic units in Finnish Sign Language. 64 p. 2009.

118 Särkkä, Timo, Hobson's Imperialism. A Study in Late-Victorian political thought. - J. A. Hobsonin imperialismi. 211 p. Yhteenveto 11 p. 2009.

119 Laihonen, Petteri, Language ideologies in the Romanian Banat. Analysis of interviews and academic writings among the Hungarians and Germans. 51 p. (180 p) Yhteenveto 3 p. 2009.

120 Mátyás, Emese, Sprachlernspiele im DaF-Unterricht. Einblick in die Spielpraxis des finnischen und ungarischen Deutsch-als-Fremdsprache-Unterrichts in der gymnasialen Oberstufe sowie in die subjektiven Theorien der Lehrenden über den Einsatz von Sprachlernspielen. 399 p. 2009.

121 Paraczky, Ágnes, Näkeekö taitava muusikko sen minkä kuulee? Melodiadiktaatin ongelmat suomalaisessa ja unkarilaisessa taidemusiikin ammattikoulutuksessa. - Do accomplished musicians see what they hear? 164 p. Magyar nyelvü összefoglaló 15 p. Summary 4 p. 2009.

122 Elomaa, Eeva, Oppikirja eläköön! Teoreettisia ja käytännön näkökohtia kielten oppimateriaalien uudistamiseen. - Cheers to the textbook! Theoretical and practical considerations on enchancing foreign language textbook design. 307 p. Zusammanfassung 1 p. 2009.

123 Helle, Anna, Jäljet sanoissa. Jälkistrukturalistisen kirjallisuuskäsityksen tulo 1980-luvun Suomeen. - Traces in the words. The advent of the poststructuralist conception of literature to Finland in the 1980s. 272 p. Summary 2 p. 2009.

124 Pimiä, Tenho Ilari, Tähtäin idässä. Suomalainen sukukansojen tutkimus toisessa maailmansodassa. - Setting sights on East Karelia: Finnish ethnology during the Second World War. 275 p. Summary 2 p. 2009.

125 Vuorio, Kaija, Sanoma, lähettäjä, kulttuuri. Lehdistöhistorian tutkimustraditiot Suomessa ja median rakennemuutos. - Message, sender, culture. Traditions of research into the history of the press in Finland and structural change in the media. 107 p. 2009.

126 Bene, Adrián Egyén és közösség. Jean-Paul Sartre *Critique de la raison dialectique* című műve a magyar recepció tükrében. - Individual and community. Jean-Paul Sartre's

Critique of dialectical reason in the mirror of the Hungarian reception. 230 p. Summary 5 p. 2009.
127 DRAKE, MERJA, Terveysviestinnän kipupisteitä. Terveystiedon tuottajat ja hankkijat Internetissä. - At the interstices of health communication. Producers and seekers of health information on the Internet. 206 p. Summary 9 p. 2009.
128 ROUHIAINEN-NEUNHÄUSERER, MAIJASTIINA, Johtajan vuorovaikutusosaaminen ja sen kehittyminen. Johtamisen viestintähaasteet tietoperustaisessa organisaatiossa. - The interpersonal communication competence of leaders and its development. Leadership communication challenges in a knowledge-based organization. 215 p. Summary 9 p. 2009.
129 VAARALA, HEIDI, Oudosta omaksi. Miten suomenoppijat keskustelevat nykynovellista? - From strange to familiar: how do learners of Finnish discuss the modern short story? 317 p. Summary 10 p. 2009.
130 MARJANEN, KAARINA, The Belly-Button Chord. Connections of pre-and postnatal music education with early mother-child interaction. - Napasointu. Pre- ja postnataalin musiikkikasvatuksen ja varhaisen äiti-vauva -vuorovaikutuksen yhteydet. 189 p. Yhteenveto 4 p. 2009.
131 BŐHM, GÁBOR, Önéletírás, emlékezet, elbeszélés. Az emlékező próza hermeneutikai aspektusai az önéletírás-kutatás újabb eredményei tükrében. - Autobiography, remembrance, narrative. The hermeneutical aspects of the literature of remembrance in the mirror of recent research on autobiography. 171 p. Summary 5 p. 2009.
132 LEPPÄNEN, SIRPA, PITKÄNEN-HUHTA, ANNE, NIKULA, TARJA, KYTÖLÄ, SAMU, TÖRMÄKANGAS, TIMO, NISSINEN, KARI, KÄÄNTÄ, LEILA, VIRKKULA, TIINA, LAITINEN, MIKKO, PAHTA, PÄIVI, KOSKELA, HEIDI, LÄHDESMÄKI, SALLA & JOUSMÄKI, HENNA, Kansallinen kyselytutkimus englannin kielestä Suomessa: Käyttö, merkitys ja asenteet. - National survey on the English language in Finland: Uses, meanings and attitudes. 365 p. 2009.
133 HEIKKINEN, OLLI, Äänitemoodi. Äänite musiikillisessa kommunikaatiossa. - Recording Mode. Recordings in Musical Communication. 149 p. 2010.
134 LÄHDESMÄKI, TUULI (ED.), Gender, Nation, Narration. Critical Readings of Cultural Phenomena. 105 p. 2010.
135 MIKKONEN, INKA, "Olen sitä mieltä, että". Lukiolaisten yleisönosastotekstien rakenne ja argumentointi. - "In my opinion…" Structure and argumentation of letters to the editor written by upper secondary school students. 242 p. Summary 7 p. 2010.
136 NIEMINEN, TOMMI, Lajien synty. Tekstilaji kielitieteen semioottisessa metateoriassa. - Origin of genres: Genre in the semiotic metatheory of linguistics. 303 p. Summary 6 p. 2010.
137 KÄÄNTÄ, LEILA, Teacher turn allocation and repair practices in classroom interaction. A multisemiotic perspective. - Opettajan vuoronanto- ja korjauskäytänteet luokkahuonevuorovaikutuksessa: multisemioottinen näkökulma. 295 p. Yhteenveto 4 p. 2010. HUOM: vain verkkoversiona.
138 SAARIMÄKI, PASI, Naimisen normit, käytännöt ja konfliktit. Esiaviollinen ja aviollinen seksuaalisuus 1800-luvun lopun keskisuomalaisella maaseudulla. - The norms, practices and conflicts of sex and marriage. Premarital and marital sexual activity in rural Central Finland in the late nineteenth century. 275 p. Summary 12 p. 2010.
139 KUUVA, SARI, Symbol, Munch and creativity: Metabolism of visual symbols. - Symboli, Munch ja luovuus – Visuaalisten symboleiden metabolismi. 296 p. Yhteenveto 4 p. 2010.
140 SKANIAKOS, TERHI, Discoursing Finnish rock. Articulations of identities in the Saimaa-Ilmiö rock documentary. - Suomi-rockin diskursseja. Identiteettien artikulaatioita Saimaa-ilmiö rockdokumenttielokuvassa. 229 p. 2010.
141 KAUPPINEN, MERJA, Lukemisen linjaukset – lukutaito ja sen opetus perusopetuksen äidinkielen ja kirjallisuuden opetussuunnitelmissa. - Literacy delineated – reading literacy and its instruction in the curricula for the mother tongue in basic education. 338 p. Summary 8 p. 2010.
142 PEKKOLA, MIKA, Prophet of radicalism. Erich Fromm and the figurative constitution of the crisis of modernity. - Radikalismin profeetta. Erich Fromm ja modernisaation kriisin figuratiivinen rakentuminen. 271 p. Yhteenveto 2 p. 2010.
143 KOKKONEN, LOTTA, Pakolaisten vuorovaikutussuhteet. Keski-Suomeen muuttaneiden pakolaisten kokemuksia vuorovaikutussuhteistaan ja kiinnittymisestään uuteen sosiaaliseen ympäristöön. - Interpersonal relationships of refugees in Central Finland: perceptions of relationship development and attachment to a new social environment. 260 p. Summary 8 p. 2010.
144 KANANEN, HELI KAARINA, Kontrolloitu sopeutuminen. Ortodoksinen siirtoväki sotien jälkeisessä Ylä-Savossa (1946-1959). - Controlled integration: Displaced orthodox Finns in postwar upper Savo (1946–1959). 318 p. Summary 4 p. 2010.

145 NISSI, RIIKKA, Totuuden jäljillä. Tekstin tulkinta nuorten aikuisten raamattupiirikeskusteluissa. – In search of the truth. Text interpretation in young adults' Bible study conversations. 351 p. Summary 5 p. 2010.
146 LILJA, NIINA, Ongelmista oppimiseen. Toisen aloittamat korjausjaksot kakkoskielisessä keskustelussa. – Other-initiated repair sequences in Finnish second language interactions. 336 p. Summary 8 p. 2010.
147 VÁRADI, ILDIKÓ, A parasztpolgárosodás „finn útja". Kodolányi János finnországi tevékenysége és finn útirajzai. – The "Finnish Way" of Peasant-Bourgeoization. János Kodolányi's Activity in Finland and His Travelogues on Finland. 182 p. Summary 3 p. 2010.
148 HANKALA, MARI, Sanomalehdellä aktiiviseksi kansalaiseksi? Näkökulmia nuorten sanomalehtien lukijuuteen ja koulun sanomalehtiopetukseen. – Active citizenship through newspapers? Perspectives on young people´s newspaper readership and on the use of newspapers in education. 222 p. Summary 5 p. 2011.
149 SALMINEN, ELINA, Monta kuvaa menneisyydestä. Etnologinen tutkimus museokokoelmien yksityisyydestä ja julkisuudesta. – Images of the Past. An ethnological study of the privacy and publicity of museum collections. 226 p. Summary 5 p. 2011. HUOM: vain verkkoversiona.
150 JÄRVI, ULLA, Media terveyden lähteillä. Miten sairaus ja terveys rakentuvat 2000-luvun mediassa. – Media forces and health sources. Study of sickness and health in the media. 209 p. Summary 3 p. 2011.
151 ULLAKONOJA, RIIKKA, Da. Eto vopros! Prosodic development of Finnish students´ read-aloud Russian during study in Russia. – Suomalaisten opiskelijoiden lukupuhunnan prosodinen kehittyminen vaihto-opiskelujakson aikana Venäjällä. 159 p. (208 p.) Summary 5 p. 2011.
152 MARITA VOS, RAGNHILD LUND, ZVI REICH AND HALLIKI HARRO-LOIT (EDS), Developing a Crisis Communication Scorecard. Outcomes of an International Research Project 2008-2011 (Ref.). 340 p. 2011.
153 PUNKANEN, MARKO, Improvisational music therapy and perception of emotions in music by people with depression. 60 p. (94 p.) Yhteenveto 1 p. 2011.
154 DI ROSARIO, GIOVANNA, Electronic poetry. Understanding poetry in the digital environment. – Elektroninen runous. Miten runous ymmärretään digitaalisessa ympäristössä? 327 p. Tiivistelmä 1 p. 2011.
155 TUURI, KAI, Hearing Gestures: Vocalisations as embodied projections of intentionality in designing non-speech sounds for communicative functions. – Puheakteissa kehollisesti välittyvä intentionaalisuus apuna ei-kielellisesti viestivien käyttöliittymä-äänien suunnittelussa. 50 p. (200 p.) Yhteenveto 2 p. 2011.
156 MARTIKAINEN, JARI, Käsitettävä taidehistoria. Kuvalähtöinen malli taidehistorian opetukseen kuvallisen ilmaisun ammatillisessa perustutkinnossa. – Grasping art history. A picture-based model for teaching art history in the vocational basic degree programme in visual arts. 359 p. Summary 10 p. 2011.
157 HAKANEN, MARKO, Vallan verkostoissa. Per Brahe ja hänen klienttinsä 1600-luvun Ruotsin valtakunnassa. – Networks of Power: Per Brahe and His Clients in the Sixteenth-Century Swedish Empire. 216 p. Summary 6 p. 2011.
158 LINDSTRÖM, TUIJA ELINA, Pedagogisia merkityksiä koulun musiikintunneilla perusopetuksen yläluokkien oppilaiden näkökulmasta. – Pedagogical Meanings in Music Education from the Viewpoint of Students of Junior High Grades 7-9. 215 p. 2011.
159 ANCKAR, JOANNA, Assessing foreign language listening comprehension by means of the multiple-choice format: processes and products. – Vieraan kielen kuullun ymmärtämistaidon mittaaminen monivalintatehtävien avulla: prosesseja ja tuloksia. 308 p. Tiivistelmä 2 p. 2011.
160 EKSTRÖM, NORA, Kirjoittamisen opettajan kertomus. Kirjoittamisen opettamisesta kognitiiviselta pohjalta. – The story of writing teacher. Teaching writing from cognitive base. 272 p. Tiivistelmä 4 p. 2011.
161 HUOVINEN, MIKA, Did the east belong to the SS? The realism of the SS demographic reorganisation policy in the light of the germanisation operation of SS- und Polizeiführer Odilo Globocnik. – Kuuluiko Itä SS-järjestölle? SS-järjestön uudelleenjärjestelypolitiikan realistisuus SS- ja poliisijohtaja Odilo Globocnikin germaanistamisoperaation valossa. 268 p. Tiivistelmä 1 p. 2011.
162 PAKKANEN, IRENE, Käydään juttukauppaa. Freelancerin ja ostajan kohtaamisia journalismin kauppapaikalla. – Let's do story business. Encounters of freelancers and buyers at the marketplace of journalism. 207 p. 2011.
163 KOSKELA, HEIDI, Constructing knowledge: Epistemic practices in three television interview genres. – Tietoa rakentamassa: episteemiset käytänteet kolmessa eri televisiohaastattelugenressä. 68 p. (155 p.) Yhteenveto 3 p. 2011.
164 PÖYHÖNEN, MARKKU O., Muusikon tietämisen tavat. Moniälykkyys, hiljainen tieto ja musiikin esittämisen taito korkeakoulun instrumenttituntien näkökulmasta. – The ways of knowing of a musician: Multiple intelligences, tacit knowledge and the art of performing seen through instrumental lessons of bachelor and post-graduate students. 256 p. Summary 4 p. 2011.

165 RAUTAVUOMA, VEERA, Liberation exhibitions as a commemorative membrane of socialist Hungary. 251 p. Yhteenveto 3 p. 2011.

166 LEHTONEN, KIMMO E., Rhetoric of the visual – metaphor in a still image. – Visuaalisen retoriikka – metafora still-kuvan tarkastelussa. 174 p. Yhteenveto 1 p. 2011.

167 SARKAMO, VILLE, Karoliinien soturiarvot. Kunnian hallitsema maailmankuva Ruotsin valtakunnassa 1700-luvun alussa. – Carolean warrior values: an honour-dominated worldview in early-eighteenth-century Sweden. 216 p. Summary 11 p. 2011.

168 RYNKÄNEN, TATJANA, Русскоязычные молодые иммигранты в Финляндии – интеграция в контексте обучения и овладения языком. – Russian-speaking immigrant adolescents in Finnish society – integration from the perspective of language and education. 258 p. Tiivistelmä 9 p. Summary 9 p. 2011.

169 TIAINEN, VEIKKO, Vähentäjää vähentämässä. Tehdaspuu Oy puunhankkijana Suomessa. – Tehdaspuu Oy in Finnish Wood Procurement. 236 p. Summary 5 p. 2011.

170 STOLP, MARLEENA, Taidetta, vastustusta, leikkiä ja työtä? Lasten toimijuus 6–vuotiaiden teatteriprojektissa. – Art, resistance, play and work? Children’s agency in a six-year-olds’ theatre project. 79 p. (142 p.) 2011.

171 COOLS, CARINE, Relational dialectics in intercultural couples’ relationships. – Kulttuurienvälisten parisuhteiden relationaalinen dialektiikka. 282 p. 2011.

172 SAARIO, JOHANNA, Yhteiskuntaopin kieliympäristö ja käsitteet – toisella kielellä opiskelevan haasteet ja tuen tarpeet. – The language environment and concepts in social studies – challenges and need of support for a second language learner. 290 p. Summary 7 p. 2012.

173 ALLURI, VINOO, Acoustic, neural, and perceptual correlates of polyphonic timbre. – Polyfonisen sointivärin hahmottamisen akustiset ja hermostolliset vastineet. 76 p. (141 p.) Yhteenveto 1 p. 2012.