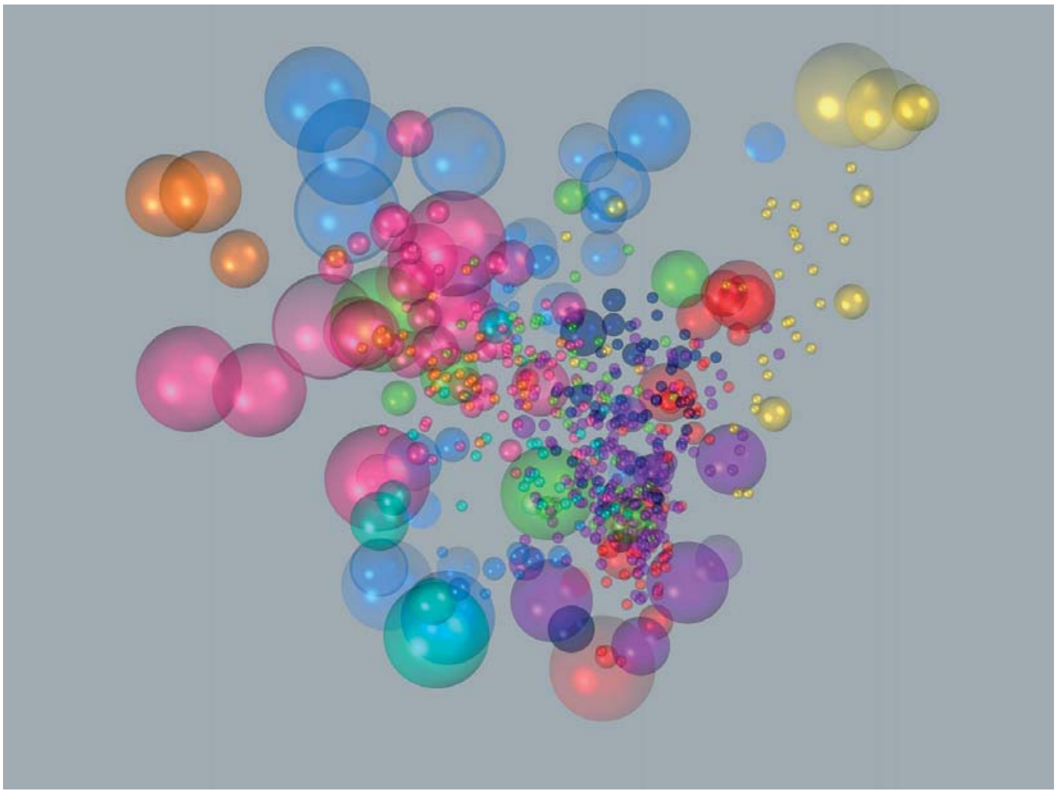
JYVÄSKYLÄ STUDIES IN HUMANITIES

177

Rafael Ferrer

The Socially Distributed Cognition of Musical Timbre

A Convergence of Semantic, Perceptual, and Acoustic Aspects



JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN HUMANITIES 177

Rafael Ferrer

The Socially Distributed Cognition of Musical Timbre

A Convergence of Semantic, Perceptual, and Acoustic Aspects

Esitetään Jyväskylän yliopiston humanistisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston P-rakennuksen Lyhty-salissa
maaliskuun 24. päivänä 2012 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Humanities of the University of Jyväskylä,
in building P, auditorium Lyhty, on March 24, 2012 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2012

The Socially Distributed Cognition of Musical Timbre

A Convergence of Semantic, Perceptual, and Acoustic Aspects

JYVÄSKYLÄ STUDIES IN HUMANITIES 177

Rafael Ferrer

The Socially Distributed Cognition of Musical Timbre

A Convergence of Semantic, Perceptual, and Acoustic Aspects

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2012

Editors
Tuomas Eerola
Department of Music, University of Jyväskylä
Pekka Olsbo
Publishing Unit, University Library of Jyväskylä

Jyväskylä Studies in Humanities
Editorial Board

Editor in Chief Heikki Hanka, Department of Art and Culture Studies, University of Jyväskylä
Petri Karonen, Department of History and Ethnology, University of Jyväskylä
Paula Kalaja, Department of Languages, University of Jyväskylä
Petri Toiviainen, Department of Music, University of Jyväskylä
Tarja Nikula, Centre for Applied Language Studies, University of Jyväskylä
Raimo Salokangas, Department of Communication, University of Jyväskylä

Cover picture by Rafael Ferrer

URN:ISBN:978-951-39-4692-0
ISBN 978-951-39-4692-0 (PDF)

ISBN 978-951-39-4691-3 (nid.)
ISSN 1459-4331

Copyright © 2012, by University of Jyväskylä

Jyväskylä University Printing House, Jyväskylä 2012

ABSTRACT

Ferrer Flores, Rafael
The Socially Distributed Cognition of Musical Timbre: a Convergence of Semantic, Perceptual, and Acoustic Aspects
Jyväskylä: University of Jyväskylä, 2012, 42 p.(+included articles)
(Jyväskylä Studies in Humanities
ISSN 1459-4331;177
ISBN 978-951-39-4691-3 (nid.)
ISBN 978-951-39-4692-0 (PDF)
Finnish summary
Diss.

The cognition of musical timbre is studied from a social perspective. The exploration starts by narrowing the definition of timbre to specific phenomena defined as *timbral environments*, which describe the global sound of musical objects within a social context. For this reason a theoretical framework is provided, to extend the scope of timbre in two aspects: from the perception of short, isolated sonic events, to real-world complex auditory phenomena; and from individual's embodied perception to a socially distributed cognition domain. To test the pragmatic value of such a definition, first a conservative approach relating timbral descriptors to affective dimensions is used to confirm the methodological plausibility of connecting physical entities with high level cognitive functions. Then an empirical approach is employed to analyze the public data, provided by a pioneering online musical social network (Last.fm). In this kind of social network, a form of synthesized natural language has emerged consisting of semantic labels, known as "tags". Specific musical objects such as artists and songs are each given a list of different tags that have been attributed to them by users over time. This language of tags is used by subscribers in such a network to exchange ideas, descriptions, and emotions related with the music (among other kinds of abstractions). The semantic properties of these tags are analyzed, and links with the perceptual and acoustic domains of the objects they describe, are established. The analysis of tags revealed a structure from which specific categories can be extracted and examined more closely, such as musical genres, adjectives or affect related terms. Finally, the practical value of these findings is tested with the implementation of a psychometric tool to assess musical preferences.

Keywords: music, timbre, cognition, social network, semantic analysis, musical preference

Author

Rafael Ferrer Flores
Department of Music
University of Jyväskylä
Finland

Supervisor

Tuomas Eerola
Department of Music
University of Jyväskylä
Finland

Reviewers

Professor Richard Parncutt
Center for Systematic Musicology
University of Graz
Austria

Adjunct Professor Douglas Eck
Department of Computer Science
University of Montreal
Canada

Opponent

Professor Richard Parncutt

PREFACE

Timbre is an umbrella term used to describe the ulterior qualities of sound. It has previously been approached from many angles and disciplines but, because each field in which the word is used has limited the application of its meaning, there is a lack of overall agreement with respect to the physical or perceptual phenomena it refers to. In other words, there is no theory of timbre as there is, say, for color. This parallel that has been present in music since early studies of sound (cf. McClain, 1978) until nowadays, notably in the musical jargon (cf. Bellingham, 2012). While color theorists have created models comprised of few essential parameters (e.g., primary colors) to construct complex tonalities, timbre researchers are striving to provide satisfactory explanations on the multidimensional characteristic of timbre. And there is no consensus on the perceptual and acoustic correspondence of the most salient dimensions of timbre. Therefore, a composition of timbre via theoretical construction – as it is done with colors – seems far to be accomplished. In spite of pioneering steps were researchers have done a careful manipulation of the stimuli (e.g., McAdams and Cunible, 1992), the gap between these monophonic synthetic sounds and actual musical timbre has not been fully addressed.

The problem of defining timbre is therefore attractive and must be solved at some point. There are undoubtedly many seminal works existing in the literature that have attempted this, but it is not within the scope of this dissertation to mention them all here. There is however one personality who, although he did not provide the initial motivation for my work, certainly inspired my efforts once I learned about his legacy. The main reason is because, I suspect he used to listen very deeply and perhaps because of that, he could formulate the notion of the *corps sonore*, showing us that a whole universe exists within a single tone. The person I refer to is Jean-Philippe Rameau.

I think of timbre as being for sound, what color is for things seen with our eyes; indeed it is perhaps so evident, necessary and ever present, that we just take it for granted and do not linger on it very often. It might therefore, just be a matter of simply stopping ourselves - in the spirit of Rameau - to observe and learn more about the quality of musical sounds that bring humanity so much pleasure and perhaps something else from the *meta* domain. Also, I would expect that remarkable ideas concerning timbre exist in many different cultures throughout the history of human civilization, but unfortunately they have not been investigated yet.

For the time being, I therefore present you with but a glimpse into this marvelous universe of timbral phenomena. What I call henceforth *timbral environments* refer to the mental images of prototypical sets of musical sounds that we categorize and name according to our experiences. We developed this term to investigate the possible answers to a fairly common expression “How it sounds like?”. People that is very fond of music and has no formal training (we will call them *non-experts* from now on) are highly skilled listeners, but in contrast with professional musicians (*experts*) they have not learned a specialized vocabulary to refer to the different characteristics of music. By investigating the vocabulary that non-experts use to describe music in social networks, we found that is very rich. It can be grouped in terms referring to genres, stylistic, functional and structural qualities of music, and also, some terms are directly referring to timbral qualities of music. In general, the contents of this dissertation are mostly concerned with the overall sound of music (cf. Prem et al., 2011) and how people digest and communicate its timbral characteristics.

This thesis offers the opportunity to bring to the academic discussion the voices of non-experts towards a redefinition of music. The methods reviewed along this text concerning social networks allows us to speculate and even confront experts traditional assumptions, such as the value of formal aesthetics of music to laymen. In other words, if what we have adopted as truth comes from overspecialized *reductionists*, it may be necessary to incorporate to that truth the views of unintentional *holists*. This bold intellectual challenge could result in a definition of music of broader scopes, where aesthetic hierarchies closely linked with cultural differences are disfavored in order to let perceptually inspired models i.e., akin to human beings, to emerge.

ACKNOWLEDGEMENTS

The invaluable experience of *Tuomas Eerola* resonates in each word of this project. Through intense and kind discussions he taught me exciting paths to make myself to understand what I was looking for. Together with *Petri Toiviainen*, they have built in under a decade, a reputation of producing superior quality research in Finland within the field of music cognition. I humbly and sincerely wish to show them my appreciation and gratitude for sharing their knowledge with me.

Each person involved in the Finnish Centre of Excellence in Interdisciplinary Music Research has had something to do with this thesis. A fabric made of technical, intellectual and emotional support has been the material maintaining myself sane and productive. However, I would like to make a distinction to *Marc*, *Jonna* and *Vinoo*, thanks for the shared space, the hugs, and the stimulating conversations.

To my beloved family, *Atzin*, *Itztli* and *Sandra*, thanks for being around and allowing me to borrow from you the spiritual energy necessary to accomplish this task.

I also wish to thank the *Escalona* family for their support.

CONTENTS

LIST OF INCLUDED ARTICLES

PI Ferrer, R. Timbral Environments: An ecological approach to the cognition of timbre. Empirical Musicology Review, 6(2), 64–74, 2011.

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

PIII Ferrer, R. & Eerola, T. Timbral Qualities of Semantic Structures of Music. In Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR), 571–576, Utrecht, Netherlands, 2010.

PIV Ferrer, R. & Eerola, T. Semantic Structures of Timbre Emerging From Social and Acoustic Descriptions of Music. EURASIP Journal on Audio, Speech, and Music Processing, 2011(1):11, 2011.

PV Ferrer, R. & Eerola, T. Looking Beyond Genres: Identifying Meaningful Semantic Layers from Tags in Online Music Collections. In Proceedings of the Tenth International Conference on Machine Learning and Applications, (2), 112–117, Honolulu, Hawaii, 2011.

PVI Ferrer, R., Eerola, T. & Vuoskoski, J.K. Enhancing Genre-based Measures of Music Preference by User-Defined Liking and Social Tags. Psychology of Music, (in press).

PVII Ferrer, R. & Eerola, T. AMP: Artist-based Musical Preferences Derived from Free Verbal Responses and Social Tags. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), 1–6, Barcelona, Spain, 2011.

Author's contributions to each article:

- In PII, the author designed the interface used to collect the data in the perceptual experiment, while also contributing to empirical data collection and writing.
- In PIII, the author contributed mostly to empirical data collection and the discussions.
- In PIV, the author was mainly responsible for the empirical contribution.
- In PV, the author provided the design of filters and algorithms.
- In PVI, the author took care of the textual processing.
- In PVII, the author was chiefly responsible for the design of the algorithms used.

1 INTRODUCTION

Timbre is to hearing what color is to vision. It is always there, and it helps to define the shapes that surround us while being a shape in itself, a feature of the sonic environment that allows us to make sense of the world about us, while also being "a major structuring force in music and one of the most important and ecologically relevant features of auditory events" (Menon et al., 2002, p. 1742). It is also an issue that has long been worked out by musicians around the world, even though it remains scientifically unexplained on many levels.

The notion of timbre could be regarded as a conceptual space populated by a myriad of definitions for the different qualities of sound. The problem of timbre definition has attracted the attention of researchers from such diverse areas, that they have provided their perspectives in a context of disagreement on a general theory. Timbre has been described as under-theorized (Burgoyne and McAdams, 2008), and as both a partly and poorly understood musical feature (Snyder, 2000; Hajda et al., 1997). The cautious postures that permeate most of the theoretical and experimental (cf. Huron, 2001) approaches to timbre – and therefore the very foundations of this project – have left a shroud of mysticism that obscures any clear definition, making such a quest appear too ambitious.

While conducting the literature reviews for each of the studies that comprise this thesis, I had the opportunity to form a different opinion on this lack of definition for timbre, because I realized that each of the past efforts have actually contributed enormously to defining timbre. Thus, if there is something that the field of timbre research urgently needs, it is a general review that would present an overall taxonomy of the explored fragments so that they might lead towards a unified theory. However, as we are not yet at such a point in timbre research, it is perhaps enough to meticulously specify which aspect of timbre is under scrutiny (e.g., musical timbre, polyphonic timbre, monophonic timbre, environmental timbre, speech timbre, etc.).

It is within this context that the contributions of this thesis fit in, as it explores one particular fragment of timbre studies that has not yet been clearly defined in terms of its conceptual boundaries. For this reason, we created a theoretical framework (see study PI) which would be able to hold key aspects of timbre cognition as it takes place in a social environment. In this respect it was necessary to extend the object of study from the individual to the socially distributed cognition domain, and from single sonic events to complex auditory phenomena. The framework is thus an attempt to maximize the *ecological validity* ideal (cf. Hammond, 1998) in a domain where timbre plays a significant role in defining how non-experts, in their everyday life, describe the contents of their musical collections. Why should we opt for an expert's taxonomy, which is unavoidably biased, when we now have access to thousands of people's opinions and can therefore derive an ontology from it? Furthermore, as there is already a wealth of research concerning monophonic timbre, and refined descriptions of it, why not instead turn our efforts to the polyphonic challenge of the "real" world? Such an aim is obviously quite ambitious, and therefore faces several caveats, but we have tried to put these to right in the best way possible, through careful empirical exploration and the design of a proto-application.

Preceding the exposition of the studies, there is an introductory summary presented in two parts. In the first part I will expose what we call the *tripartite* approach to timbre, consisting of three fundamental aspects: *semantic, perceptual, and acoustic*. The second part highlights key concepts that were used throughout the studies, providing the skeletal framework upon which our approach to musical timbre rests.

A brief exposition of each study's aims and results (chapter 2) follows, highlighting only the most crucial aspects that support the argument of the thesis. Next, there is a general discussion pointing out the main findings, limitations and methodological issues encountered during the development of the project, and finally there is a section on implications for future research (chapter 3).

1.1 Semantic-Perceptual-Acoustic: a Tripartite Approach

The aim is to investigate the relevance of timbre in terms of both culture and cognition. The section below presents a multidisciplinary approach for understanding the concept of timbre better. This consists of better understanding the cultural conventions of timbre through examining more closely the various patterns that connect the language used to describe music, with the perception of music, and with the particular acoustic features that are connected with these perceptions. Therefore timbre will be examined from semantic, perceptual and acoustic perspectives to see just how they relate together. This approach was implemented in studies PIII and PIV.

1.1.1 The Semantic Language of Timbre

The language of timbre has been studied in a number of different contexts previously. There are works that analyze the semantic space derived from vocabularies used to describe timbre in general (e.g., Bismarck, 1974; Disley et al., 2006); as well as in specialists terms: such as pianists (Bernays and Traube, 2011), violinists and organists (Stepánek, 2006), and sound engineers (Porcello, 2004). The list of examples is very long, but these particular studies have been chosen because they share few characteristics with typical studies in the semantics of timbre research. Most research in this field concerns either single tones or single sound sources, and use a well controlled environment specifically designed for the experiment, with the aim being to analyze the jargon used by people closely involved with the specific sound source.

Except for few interesting works considering a non-specialist vocabulary (e.g., Sarkar et al., 2007), most analytic designs ignore the fact they have made an *emic* assumption. These experts are no more external (*etic*) in their observations than laymen or *non-experts*. From that perspective, it seems a bit elitist to discard the sophisticated language developed by non-experts to exchange music from their own collections. This is exactly where social media – and more specifically the tagging of music – offers a new opportunity to revise past methodologies. Its relevance resides in the amount of vocabulary used, the population that uses it, and the large variety of musical sounds they are qualifying. In the past these different kinds of music would number in the hundreds, but now they number in the thousands. The present thesis takes advantage of this opportunity, focusing on how non-experts describe music using the so called *language of tags* (Simons, 2008).

It is noticeable that the most frequently used adjectives in timbre research are taken from other sensory domains than the auditory (cf. Zacharakis et al., 2011). For instance *brightness* and *color* come from the visual domain, and *sharpness* and *roughness* from the haptic. However there is a gap between such terms, which are popular among experts (e.g., researchers, musicians, etc.), and those used by non-experts to describe music. In both expert and non-expert cases "listeners do not perceive the acoustical environment in terms of ′phenomenological descriptions′ but as ′ecological events"′ (Reybrouck, 2005, p. 234). This is, first we listen to our environment as a whole (*ecological event*) until something attracts our attention, then we focus our attention to a fragment of that whole, and sometimes we investigate the laws governing such a fragment (*phenomenological descriptions*). At expense of our own survival, the inverse order in the perception of our acoustical environment seems unlikely. For experts this might be conveyed more subtly in terms taken from other sensory domains, while for non-experts it is less figurative and more literal. Non-expert tags often describe whole *ecological events* in terms of associations to experiences in specific contexts. For instance an indispensable task among teachers and students of music is to exchange subtle aesthetic points of view using a specialized vocabulary, whereas the needs for the non-experts are more general. This thesis therefore proposes that these social tags are an important key to understanding non-specialist vocabulary used to describe music. In other words, characteristics of the vocabulary (e.g., lexical categories, associations with perceptual systems, etc.) can be analyzed, and an ontology derived from it.

The gap mentioned above concerning the differences between verbal descriptors of timbre used by experts and non-experts soon became evident during early analysis of the lexical categories of tags, when we found that the terms used by experts were either never, or scarcely, used to tag music – detailed in Study PV. It could reasonably be argued that when people tag a given song, they are mainly focusing on the structural – most evident – features of music (e.g., rhythm, melody, lyrical content, etc.) but at the same time it seems improbable that they would purposefully exclude – less evident – timbral qualities while tagging. Therefore, in our analysis we assume that the vocabulary consists of a layered network of lexical functions, that with the appropriate filtering (e.g., discarding tags referencing musical genres, lyrical content, individualistic expressions, etc.) can reveal a timbre-focused jargon commonly used by non-experts and expressed either as adjectives, nouns, verbs, references to temporal circumstances or musical instruments.

From a pragmatic point of view, investigating the language of timbre under controlled conditions (e.g., in a laboratory, using standard psychometric tests), might lead to different results than if the information is sought by observing individuals in their natural environment. On the social networks dedicated to music, people is willing to share information, whereas in laboratory conditions, they need an extra motivation and perhaps some entrainment, depending on the task. On the original approach to ecological validity, this dichotomy is unbalanced because it favors the observation in natural environments over the controlled experiments, this is, preferring *idiographic* over *nomothetic* models of behaviour (cf. Hammond, 1998). Our approach to the ecological validity ideal is more realistic in the sense that we have opted for a more balanced approach where both models are combined to take the advantage of collecting data silently and in situ, and by assuming certain generalizations during the analysis. Such an approach is demonstrated in studies PIII, PIV, and PVI.

1.1.2 The Perceptual Psychology of Timbre

The psychology of timbre has received considerable attention from different perspectives: for instance, the psychoacoustic (Helmholtz, 1954; Grey, 1977; Wessel, 1979), the psychophysiological (Bregman, 1990), and the perceptual (Krumhansl and Iverson, 1992; McAdams et al., 1995; Alluri and Toiviainen, 2010). Except for the psychophysiological approach, where the objective is to link the sound with objective sensations, the rest have investigated how to map the acoustic features that characterize a given stimuli with the subjective judgments given by individuals using mainly two paradigms: *semantic differential* (cf. Hajda et al., 1997, p. 259) and *cognitive structuralism* (cf. Leman, 1995, p. 1). The first consists of presenting the participants with sound samples, which they then rate using scales labeled with opposite adjectives at each extreme (e.g., bright–dark, dull–sharp, etc.). The second consists of presenting participants with pairs of sound samples and asking them to rate their similarity on a Likert scale. The analysis methods are straightforward: in the case of the first paradigm, linear correlations between scales and acoustic features are computed; in the second case, the similarities are projected onto a low dimensional space (typically two or three dimensions) and then these are correlated with acoustic descriptors from the stimuli. This thesis makes use of a combination of both these paradigms.

Among the most influential representations of the perceptual characteristics of timbre, Grey's (1977, p. 1272) contribution is perhaps the one that has been most replicated (see Figure 1). In this study, a multidimensional timbral space generated from the perceived similarities between 16 instrumental tones, played by selected wind instruments, was scaled down to three dimensions using Carroll and Chang's (1970) method. Grey interpreted these dimensions as corresponding to the following physical characteristics of the sound signal: *spectral energy distribution* (i.e., description of the intensity of individual frequencies making up the sound); *spectral fluctuation* (i.e., the envelope and deformations of its shape over time); and *inharmonic energy* (i.e., interactions between individual frequencies in the sound over time). Roughly speaking, these three dimensions correspond to the terms used elsewhere in the literature as *"brightness"* (Krimphoff et al., 1994; Iverson and Krumhansl, 1993), *"attack-decay"* (Jensen, 1999), and *"roughness"* (Burgoyne and McAdams, 2008) of sound, respectively.

1.1.3 The Acoustic Description of Timbre

Two decades ago, Iverson and Krumhansl claimed that the "acoustic basis for timbre (were) almost completely undefined." (1993, p. 2595), which has since been one of the reasons for not having a clear definition for the phenomenon. Nevertheless, the scenario has dramatically changed, as at the present time there is a host of methods to perform objective measurements of timbre (e.g., Krimphoff et al., 1994; Sethares, 2005; Jensen, 1999; Juslin, 2000; Tzanetakis and Cook, 2002; Peeters, 2004; Pohle et al., 2005, etc.). Most of them are based on the assumption there is a mental image of sound events, so they have a rationale in common with the works of Helmholtz (1954). The underlying method to investigate the physical properties of timbre consists of transforming the sound signal into a time series representation, then performing measurements on the characteristics of each frame, or time-window, to observe the change between consecutive frames. In this respect it is clear that, in the acoustic domain, timbre can be described by the statistical properties of the sound signal. This means its shape can be described in terms of the distribution and interrelation of frequencies across the audible range that make it up, and how these change over time (i.e., characteristics of the envelope). The chief aim of this thesis was to find connections between semantic descriptions, perceived similarities of sound, and physical characteristics of timbre. For this reason we relied on standard methods to extract the acoustic features, but with a focus on making an accurate selection from the many possibilities, so as to leave us with a compact set that would adequately represent musical timbre and all its semantic and perceptual associations. Existing models aimed at content-based retrieval have studied how to represent timbre using audio features such as, for instance, mel-frequency cepstral coefficients (MFCC), or the spectral flux, and centroid, as these have

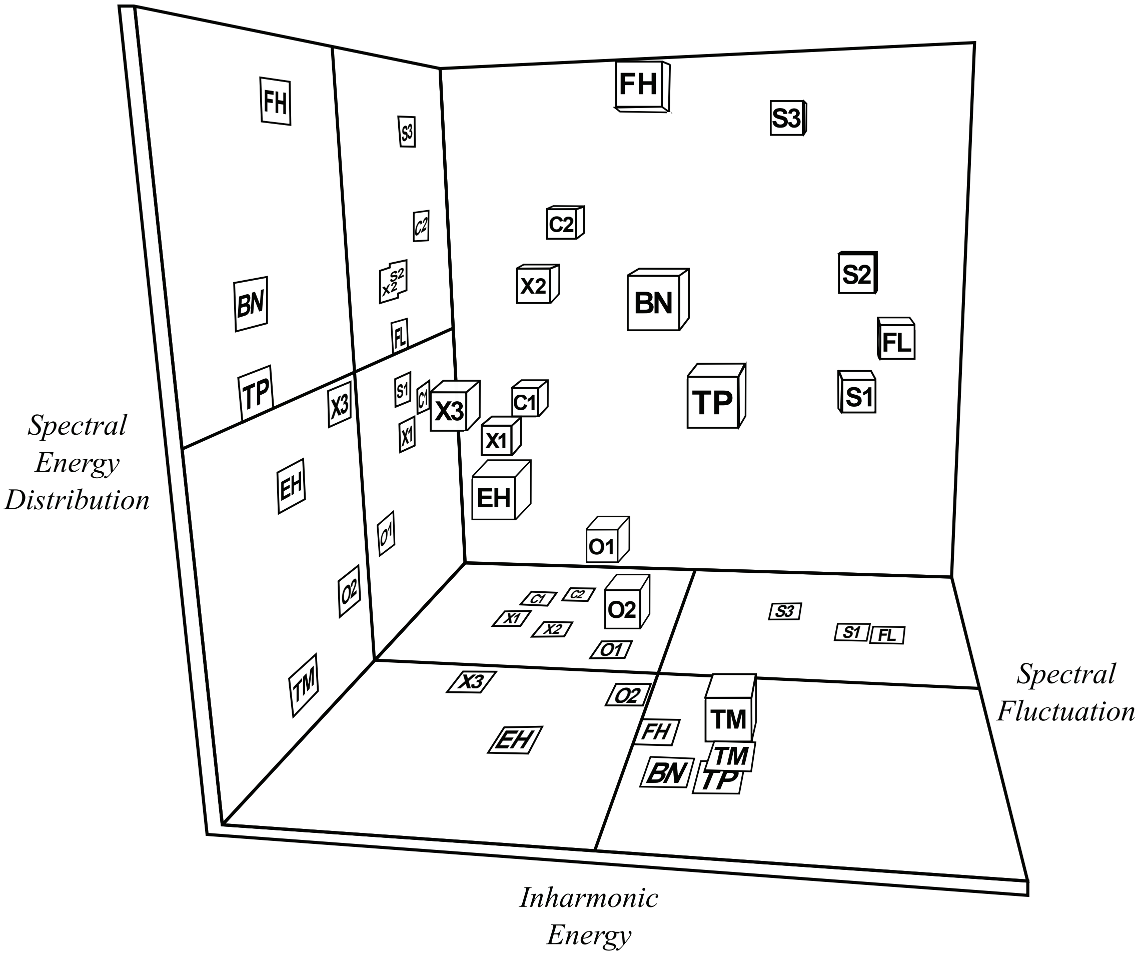


FIGURE 1 A three-dimensional spatial solution for perceptual similarity of wind instruments sounds (adapted from Grey's, 1977. Abbreviations for stimulus points: O1-2 = oboes; C1-2 = clarinets; X1-3 = saxophones; EH = English horn; FH = French horn; S1-3 = strings; TP = trumpet; TM = trombone; FL = flute; BN = bassoon.).

proved effective measures in similarity between sound signals (e.g., Aucouturier and Pachet, 2004; Peeters, 2004; Seyerlehner et al., 2009). However these methods would not satisfy our requirements because, as stated previously, the aims are quite different: while the classical approach to timbre already has names for its physical descriptions of sound qualities, we are applying these physical descriptions to better understand non-expert, but perhaps more ecologically valid, names for timbral qualities.

1.2 Key Issues

While the above mentioned tripartite approach provides a basic foundation for this work, there are relevant aspects beyond this, that are present as implicit or explicit assumptions throughout each of the sub-studies which make up the present dissertation. It should be added at this point that these aspects do not represent an exhaustive list of all the areas in which timbre has tremendous potential for expanding our knowledge of musical communication. For example, the study of brain areas involved in the perception of timbre (e.g., Caclin et al., 2008; Alluri et al., 2012) or the rich vocal manipulations or timbre and their connections to communication of affects (e.g., Tsang and Trainor, 2002) are areas where exciting research on timbre has been conducted. Nevertheless, the key issues described in the section immediately below are crucial to this dissertation.

1.2.1 Timbral Environments as Prototypes of Musical Timbre

Timbral environments is a concept created to study a very specific phenomena of timbre perception. It aims to investigate the perceptual boundaries that allow us to distinguish between different kinds of music grouped according to their most prominent timbral characteristics. It is based on the assumption that different sets of musical sources sharing acoustic and semantic characteristics in the perceptual space, are discernible by empirical means. Timbral environments can also reflect each individual's associations and their own particular reminiscences of the music in a certain context. Studying them might therefore be also relevant to understanding the dynamic exchange of music as information within societies, and give some indication as to whether this exchange can influence music preferences. It is mainly based on the theory of *Perceptual Constancy*, as applied to the concept of timbre (Risset and Wessel, 1999); is also called *Macrotimbre* by Sandell and Chronopoulos (1997); but in addition it adopts some of Schafer's views concerning auditory phenomena, known as *Soundscapes* (1977).

Timbral environments incorporate the ecological aspects from the theory of *distributed cognition* (Barnier et al., 2008) in the social context, by focusing on *transacional memories*. This refers to the context where people do not have to know everything if they can use other people's knowledge (Hesse, 2009). Another ecological aspect they rely on is the *structural coupling* of *autopoietic* systems (Maturana, 2002). These systems, which are self-organised, self-structured and autonomous, are modeled on how our minds are changed by external factors (e.g., brain plasticity). In other words, they are simultaneously transformed by the very environment they are in the process of attempting to transform. The paradigm of *embodied cognition* (Godøy, 2006; Leman, 2007) has served as a common ground to amalgamate such diverse views.

Timbral environments broaden of the scope of research. They also shift this broader focus away from simple correspondences between timbral descriptors and acoustic features, to domains that exist within the social context, in the form of transactional memories. Figure 2 illustrates these domains: *timbre* focus describes perceptual correlates of the physical description of timbre; *macrotimbre* extends the focus to what remains after being exposed to a musical signal – including various musical instruments and timbral variations within a given song – from an individual's perspective. *Timbral environments* encompass both these

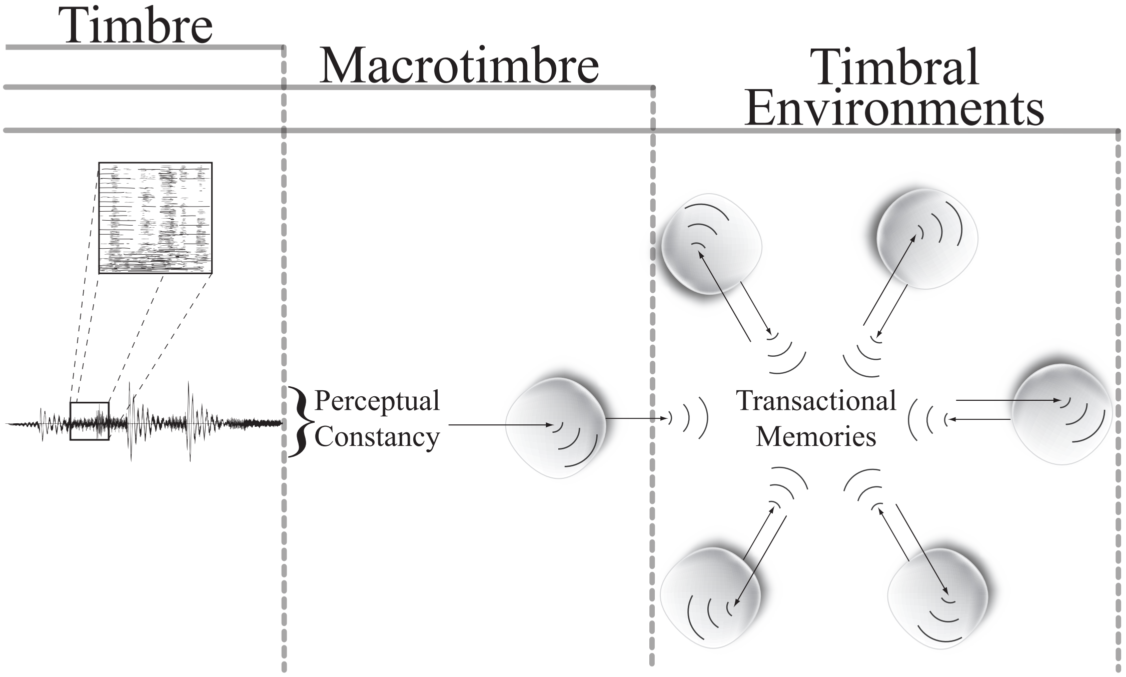


FIGURE 2 Illustration of focus of research from the classical paradigm (timbre) to the notion of timbral environments.

domains, but they also embed them in a social context.

One way of looking at timbral environments would be to draw parallels with the prototype theory (cf. Taylor, 1995). In this respect timbral environments present an alternative to the paradigm that looks simply for binary correspondences between timbral descriptors and acoustic features. The alternative is that timbral environments provide a framework to support fuzzy categorizations, in which multiple words are weighted differently in the perceptual space and their correspondence also sought with a set of acoustic features. Figure 3 depicts this tripartite approach linking the semantic, perceptual and acoustic, as detailed in section 1.1 above.

1.2.2 Social Tags and Distributed Cognition

Within the social ambit of the internet, tags have been regarded as a form of *folksonomy* (Van Damme et al., 2007). This concept combines the idea of *folk* with *taxonomy*, and is indeed relevant because tags may provide a method of grouping people according to aspects of their cultural identity. Such folksonomies provide useful systems of classification, or ontologies (e.g., Mathes, 2004; Van Damme et al., 2007; Lin et al., 2009; Kim et al., 2010) that make it easier to group and retrieve items in collections of the most diverse nature: for example, web pages

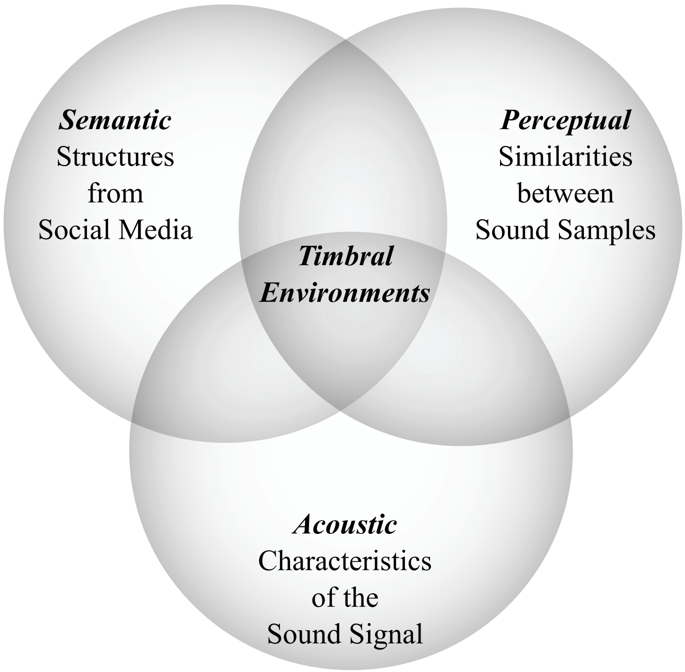


FIGURE 3 A tripartite view on timbre showing the convergence point in the notion of Timbral Environments.

(e.g., del.icio.us[1]), pictures (e.g., flickr[2]) and music (e.g., last.fm[3]). Musical tags in particular rely on a collaborative environment and so are a very good example of a folksonomy. Hence, transactions of musical information that use tags as a kind of "currency", are an ideal example of augmented social cognition (Chi, 2009) in everyday life. Furthermore, within the theory of *information foraging* (Held and Cress, 2009), tagging behaviour is also a good example of a transition from internalized to externalized – and explicitly distributed – form of knowledge. In other words, what escapes one individual's perception can be captured by another's. In this way, tags have become a form of externalized memory or a cue for knowledge because they provide evidence that a process of socially distributed cognition is taking place.

[1] http://delicious.com
[2] http://www.flickr.com
[3] http://www.last.fm

2 STUDIES

This dissertation consists of seven studies, that focus on the relevance of timbre for the everyday musical decisions of non-experts. Overall, the studies follow a traditional theoretical-empirical-practical order, which is described below.

The first study (PI) provides a theoretical framework for the study of timbre from an ecological perspective that is fundamental to the rest of the papers. It was needed for this thesis, because presently there are no such studies that are both as singularly devoted to timbre, and yet multidisciplinary enough to sufficiently cover the ground for such topics (i.e., perception of musical timbre in the social dimension).

The second study (PII) aims to investigate the role of timbre in the perception of affect, which is a major motivational means for human beings not only to listen to music, but to achieve other fundamental goals. This thereby ratifies further explorations for then linking acoustic features to other variables in later studies. Methodologically, and within the present dissertation, it also represents one axis of the aforementioned tripartite approach (see 1.1) that is being explored through behavioural experiments.

Studies PIII to PV are a trio of studies that explore the notion of *timbral environments* from an empirical perspective by searching for links between the semantic acoustic and perceptual spaces. They represent an empirical interpretation of the theoretical ideas introduced in study PI and develop the work of PII by adding the semantic variable. Thus PIII exposes the initial approach to timbral environments, while PIV extends this by refining some of the arguments, and finally PV introduces a model of semantic filtering for musical tags, which was included for its pivotal role in subsequent studies, despite the fact that it was comprised mostly of post-hoc observations.

The last two articles (PVI and PVII) illustrate one of the possible practical applications of the whole project, focusing on the development of a psychometric test aimed to assess musical preferences. It also puts the whole project firmly within the ambit of music psychology, as such tests represent one of the defining characteristics of most studies in the field. PVI first discusses the suitability of genre-based measurements and introduces an alternative method that creates a musical preference profile that requires only minimal input from participants, and social media. The method is called *Artist based Musical Preferences* (AMP), and study PVII goes deeper into the technical details of it. The following parts of chapter 2 contain a brief summary of each of these studies.

2.1 Study PI: Theoretical Background

The opening article introduces the theoretical notion of *timbral environments* as an ecological framework, within which empirical explanations regarding the influence of timbre in music appreciation schemata can be made. Timbral environments are based on the merging of two views concerned with the auditory phenomena: *macrotimbre* (Sandell, 1998) and *soundscape* (Schafer, 1977). However these are additionally supported by an ecological perspective of music cognition, which incorporates concepts from *embodied cognition* (Leman, 2007), *cybernetic theory* (Reybrouck, 2005; Godøy, 2006), *autopoiesis* (Maturana, 2002), and the *representational theory of mind* (Nussbaum, 2007).

The paper proposes that individuals are cognizant entities bound to a network of interactions. These networks have an influence on how individuals further assimilate and create new links, in a constant exchange of information. The interactions in question carry information concerning musical timbre, therefore they are embodied as internal representations that, when externalized, become represented as language. Online social music networks that rely on collaborative tagging of music, such as Last.fm, provide a concrete example of this very process. Users describe their personal reminiscences and associations with the music that they like or dislike, and this has some influence on other users – which implies interactions in a social context for music appreciation. With regard to Last.fm, timbral environments refer to the perceptual boundaries between tagged objects, i.e., they describe the characteristic features of a category into which people lump certain kinds of music. The idea enables us to speculate on whether the semantic space derived from such collaborative tagging is isomorphic with any other aspect of the acoustic experience within embodied cognition.

In sum, this first paper provides a theoretical framework to explain how listeners might use timbre for interactions in their everyday lives. It provides enough space for discussion but also offers a fresh approach to the cognition of timbre from an ecological perspective. Furthermore, it provides the basis for empirically testing embodied aspects of musical timbre.

2.2 Study PII: Timbre and its Affective Connotations

One of the most important aspects of timbre is its capacity to communicate affects and emotions (e.g., Juslin and Laukka, 2003; Laukka et al., 2005), and yet relatively little attention has been paid on just how those acoustic features contribute to emotional expression in music. Understanding the role of timbre in this process could therefore be crucial to learning more about, for instance, how musical preferences are established, or how people use music to regulate their emotions (Saarikallio, 2011).

The aim of this study was to investigate the role of timbre in the perception of affect dimensions when listening to isolated musical sounds. It consisted of three behavioural experiments. In the first, participants were asked to rate samples of instrumental sounds that each lasted only 1 second. Participants rated the samples using five bipolar scales: valence, energy arousal, tension arousal, preference and emotional intensity. In the second experiment an emotional dissimilarity task was applied to a subset of the same instrument sounds to better reveal the underlying affect structure. The third experiment was a replication of the first, but used a different set of sounds. These included systematic manipulations in the dynamics of sound production, articulation and the ratio of high-frequency to low-frequency energy. The affect dimensions resulting from this were then explained in terms of certain acoustic features extracted from the stimuli. The high agreement found among the participants' ratings across the experiments suggested that even isolated instrument sounds contain particular cues which indicate affective expression, and these are recognized as such by the listeners.

This study should be considered as an essential step away from a design that simply links acoustic descriptors of natural sounds to high-level cognitive functions – such as the emotional response to instruments sounds. Rather, it is a step towards an exploration of the semantics of timbre. The stability found among participants' ratings suggests the possibility that there are mental representations of timbre which are beyond cultural conventions.

2.3 Studies PIII-PV: the Semantics of Timbre in Social Media

This set of three studies form the core of the project. Firstly an exploratory scheme makes a case for the tripartite approach to timbre (see 1.1). It focuses on the way people describe timbre in terms of musical tags, from which a semantic structure can be extracted. This semantic structure is then used to reorganize a collection of music according to prototypical sets of semantic descriptors, and these descriptors are next used to create acoustic summaries for each of them in the form of spliced sound samples. In PIV, these acoustic summaries are then looked at in greater detail. Finally, in PV, the perceptual similarity of semantic descriptors and acoustic summaries is studied empirically, to learn whether the semantic space is indeed analogous to the acoustic space.

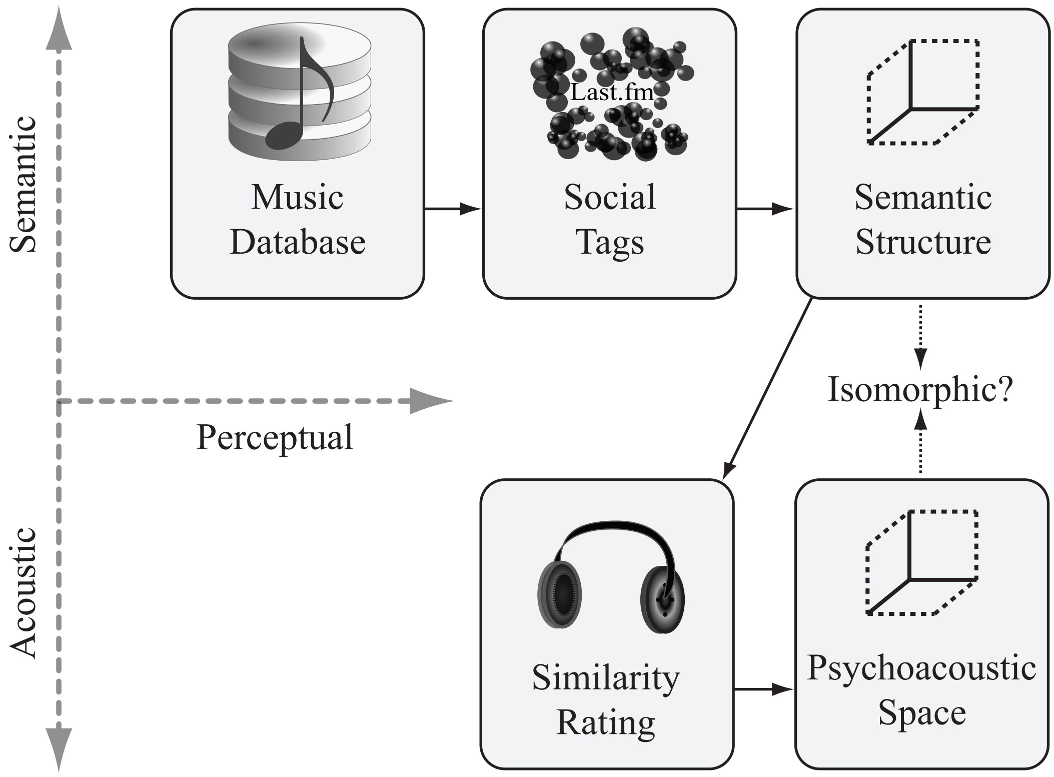


FIGURE 4 Visual summary of the processing of the information from the semantic to the perceptual and acoustic analysis of timbre.

2.3.1 Study PIII: the Timbral Qualities of Semantic Structures I

Figure 4 offers a visual summary of the tripartite approach (semantic, perceptual and acoustic) described in PIII. It shows how this approach forms the theoretical basis for the following process: music was retrieved according to an index created from clusters of musical tags, that indicated perceived similarity. The three boxes at the top of the figure highlight this *semantic* process. From these, the acoustic stimuli were created that were to be used in a *perceptual* similarity test, and these in turn determined a *psychoacoustic* space.

First musical tags were retrieved for a specific collection of six thousand songs (Eerola and Ferrer, 2009), that could also be found on *last.fm*. The tags were free verbal descriptions of musical objects (e.g., songs, artists or albums) that were composed of one or more words written by users of the network. These descriptions are available publicly and can be accessed by automatic means with several scripting languages through a dedicated *application programming interface* (API). The resultant corpus of tags takes the form of ranked lists for each of the songs in the collection.

The tagging system at Last.fm does not impose a specific vocabulary on the users, although it can suggest tags based on previous users' input – see 3.2.4 for an extended discussion on this. For this reason it is imbued with vernacular expressions in multiple languages and often highly specialized terminology from diverse musical cultures. This heterogeneous form of language to describe music ought to be separated in categories relevant for the purposes of timbral environments exploration. So the corpus was filtered to reduce the noise while aiming at preserving the original statistical distribution, and to discern the lexical and musicological categories. This left us with a subset of the original tag vocabulary, composed only of *adjectives*, *nouns*, *instrument names*, *temporal references*, and *verbs*.

After this filtering, the next step was to search for clusters by geometric means. The procedure consisted of making a presence-absence analysis of tags within songs to create a binary matrix that was then used to compute a matrix based on Euclidean distances (from tag to tag). The Euclidean distances were then used to perform cluster analysis with a hybrid method combining agglomerative and partitional techniques (Langfelder et al., 2009). The resulting clusters gave us prototypical timbral environments, i.e., groups of tags that were semantically related, and separated from other groups by conceptual boundaries.

These timbral environments were then used to reorganize the database (i.e., to perform vector quantization) of music by ranking each song according to how well it exemplified each environment. The best representatives of each (i.e., the top ranks in the re-indexed collection) were then used to explore similarities in the perceptual domain with an empirical similarity rating experiment.

This similarity experiment was to see whether the obtained clusters were perceptually meaningful. Participant's responses were projected onto a low-dimensional geometric space and then correlated with selected acoustic descriptors. Three dimensions were eventually found to best represent the original multidimensional space: the first dimension related to the distribution of energy in the spectrum; the second represented the periodic organization of the spectrum; and the third described the temporal fluctuation of the spectrum.

These results were found to agree with earlier findings that have employed similarity ratings (e.g., Grey, 1977; McAdams et al., 1995). But this was particularly interesting as the stimuli used in the above examples were steady monophonic sounds, whereas ours were a spliced signal made up of tiny slices of actual music. Another finding was that there was a clear pattern of correspondence between the perceptual similarity rating of acoustic stimuli and the verbal descriptions that these stimuli represented. For example, the semantic opposition of the musical descriptors "Aggresive" with "Sexy" corresponded to extremes in the *fluctuation centroid* and *roughness* descriptors; while a "Coy - Guitar virtuoso" continuum corresponded with the distribution of energy along the frequency spectrum. So study PIII is a preliminary attempt to show how the semantics of tags, and the timbre of spliced signals relate to perceptual similarity spaces. In some ways however, it opened more new questions than it answered. For instance, are higher level structural aspects of music, such as pitch and melody, needed to explain semantic structures or are low-level, timbral characteristics sufficient? Or could the semantic semantic layers be filtered differently to better correspond to the timbral qualities of music? These two questions were addressed subsequently.

2.3.2 Study PIV: Timbral qualities of semantic structures II

The second paper looks more specifically than PIII at the selection of acoustic features and includes an automatic classification of clusters based exclusively on acoustic descriptors extracted from the sound signal.

Finding the right set of acoustic features to best represent timbre was a major part issue of this study. Therefore we opted for a typical and safe strategy that would create models that are not over-fitted and that do not capitalize on chance. Firstly, acoustic descriptors were extracted via a fairly standard procedure, that would best capture the differences between timbral environments (discriminating power), and secondly, we ensured that such descriptors were not too similar to each other (low collinearity). Accordingly, this left only a compact set of acoustic features for further analysis.

The set of features we were left with offered more possibilities for interpretation than ones we had used in PIII – such as the MFCC, for instance. Furthermore a measure used for structural analysis (chromagram) also emerged as a strong acoustic descriptor of perceptual similarities. Figure 5 (taken directly from study PIV), shows these results, as the tags *Aggressive* and *Chill out* are in opposite corners of the psychological space as well as in terms of chromagram measurements. There is also a clear acoustical organization of the excerpts, as cluster number 5 (*Composer*, *Cello*) is depicted as being high in *roughness* and high in *spectral regularity*, with a well defined set of harmonics. In addition, those clusters that have similar overall descriptors, such as 15 (*Affirming*, *Lyricism*), 7 (*Mellow*, *Sad*), and 11 (*Autumnal*, *Wistful*), are located within close proximity of each other.

The automatic classification of the original sound examples used an implementation of the *Random Forest* algorithm (Breiman, 2001; Pang et al., 2006).The aim was to find the boundaries between clusters based exclusively on the acoustic features describing them. Results revealed that the agglomerative level of clustering is determinant, this is, success in classification is dependent on the number of clusters. Where tags were spread over only a few clusters, the classifier performed better, thus suggesting that assessing boundaries exclusively by automatic means, would reach a glass ceiling when a certain fine level of detail was reached in determining between different prototypical timbral environments. This limitation has also been found in other studies using music classification based only on acoustic features (cf. Music Information Retrieval Evaluation eXchange competition[1]). Therefore it may be possible that the relative success of these classification designs is not so much dependent on which features are used, or how sophisticated the employed system is, but rather on how the approach needs to be complemented with musical metadata as shown by our contribution.

[1] http://www.music-ir.org/mirex

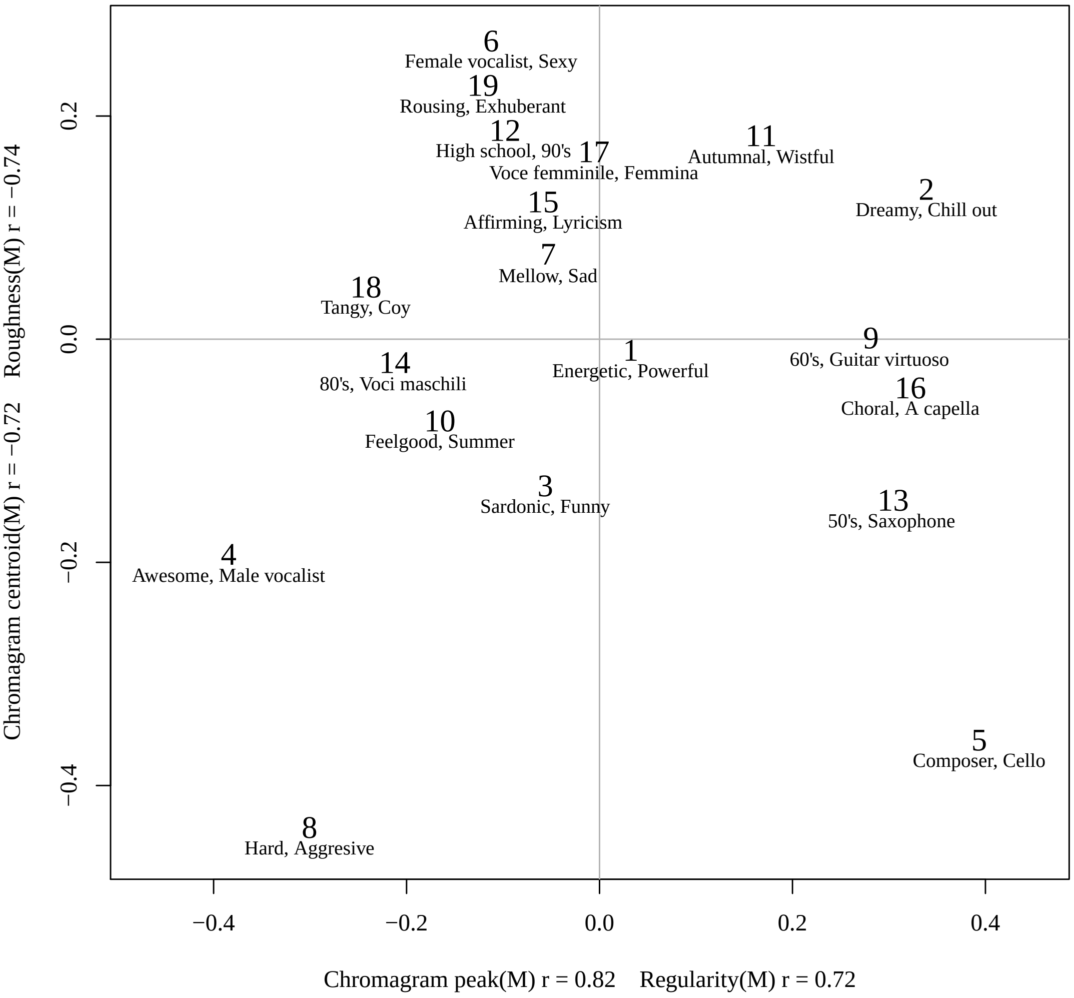


FIGURE 5 Multidimensional scaling of perceived similarity of spliced sound stimuli (from study PIV).

2.3.3 Study PV: the Semantic layers Beyond Musical Genres

This study presents an automatic classifier designed to identify *semantic layers*, which is based on a string matching algorithm. A semantic layer is a set of interrelated concepts extracted from tags that can describe, for instance, genres, affect related words or artists names. The purpose of such a semantic layer, is that it can be used to then qualitatively filter a given corpus of tags. During the past five years, musical tags in collaborative systems have been used for diverse purposes: such as making auto-taggers, discovering semantic structures in music, and music indexing and recommendation (e.g., Baccigalupo et al., 2008; Bertin-Mahieux et al., 2008; Levy and Sandler, 2009; Chen et al., 2009). Understanding the semantic characteristics of tags is crucial to building an appropriate semantic space for any application. For instance, some approaches have performed an ad hoc extraction of semantic layers (e.g., Laurier et al., 2009). Nevertheless, such an approach entails discarding any tags that are not related with the chosen semantic layer, and it therefore means that an important part of the semantic space is neglected, because relevant semantic information embedded within the relations between tags is also discarded. This is because, within the language of tags (Simons, 2008), words acquire a different meaning than in their original language context. Hence it could be argued that the meaning of a tag is, in part, determined by the tag context – or the mesh of relations that it has with other tags. By allowing for fuzzy categorization however, this study is able to show the degree to which each tag belongs to each category in the semantic layer. This then enables a much fuller characterization of the semantic space provided by the tags.

Study PV also includes a revision of the method that was used to filter tags in PIII. In that study (PIII), quantitative filtering was based on the *most frequent* method used in previous research (Bertin-Mahieux et al., 2008; Baccigalupo et al., 2008; Levy and Sandler, 2009). In other words, tags were selected from a given corpus according to the frequency of their use. In study PV, we showed how this filter can in fact severely distort the spectrum of word frequencies, and can actually introduce an artifact that eliminates tags that might in fact be relevant for any study aiming to construct a folksonomy or collaborative taxonomy (Vander Wal, 2007). Consequently, we provide an alternative method that implements in three steps a filter that has the capability of removing as much noise as possible while still keeping the overall shape of the corpus' spectrum as intact as possible.

2.4 Studies PVI-PVII: Assessing Musical Preferences Using Timbre Descriptions

The last two studies offer a practical application for timbral environments. The underlying assumption is linked with the former studies as follows: there are prototypical timbral environments in the (acoustic and semantic) perceptual space that describe different kinds of music and these can be cued with verbal references such as tags. If that holds true, it should then be possible to design a psychometric tool capable of identifying an individual's musical preference, if mapped to such a space based on their verbal cues.

Both studies contribute to the same idea, although with a very different emphasis. While study PVI offers a review of current tools that are used to assess musical preference, including all their various drawbacks from a *psycho-musical* perspective; study PVII stresses the technical characteristics of our possible solution in the form of a new application. This takes the form of an instrument called **AMP** (Artist Based Musical Preferences), that with an input of three liked and three disliked artists returns –with the aid of online sources– a standardized set of items with ratings. The chosen form of the item output for these studies was *musical genres*, and their ratings took the form of Likert scales. Nevertheless, such formalization is optional, as the tool has the possibility of presenting the data in terms of any *semantic layer* that might be of interest for the researcher.

3 GENERAL DISCUSSION

3.1 Main Findings

The theoretical framework outlined in PI provided a novel perspective to timbre, which emphasized its interactive nature. Taking the ideas from embodied cognition and ecological psychology, it was proposed that the semantic aspects of timbre should reflect a natural interaction with sound. For this reason, social media was chosen as the main focus of study, since this is one place where millions of music listeners describe, annotate, and share music in their everyday lives. These rich descriptions of music, sounds in particular, had not yet been used in conjunction with timbre studies. But the decision to focus on a large public also brought on new challenges.

Appropriate analysis methods for obtaining meaningful information from the rich data provided by social media is still in its infancy (cf. Aucouturier and Pampalk, 2008). While the present study utilized some of the main strategies for obtaining such knowledge, such as the clustering of distances provided by tag matrices (Levy and Sandler, 2008, 2009), new ways to handle such data were also used. For example, understanding the meanings encoded in tags is not really possible without discriminating the various semantic layers embedded in such data. For this reason, a part of the studies focused on determining whether the tags used by listeners refer to genres, verbs, bands, adjectives, affects, functions, nouns or other categories. These kind of methods have been used for different purposes, but not explicitly applied to the study of timbre. Finally, to demonstrate that this kind of timbre research has practical applications, we showed how it might be used to measure musical preference.

From a musical perspective, we have been able to map the acoustic characteristics of a musical sound that might be perceived as sexy, coy, dreamy, aggressive, etc. This could be of interest to functional composers, for instance to build up a theatrical scene. Also for musicologists interested in organizing music according to an innovative criteria based on adjectives rather than musical genres. Overall, our findings also suggest that a) timbre is an issue of major concern to non-experts even though they may not be fully aware of it, b) that the heterogeneous vocabulary used to communicate timbre is crucial for exchanging musical impressions, and c) that such vocabulary is evolving in parallel with the social networks.

3.2 Limitations and Methodological Issues

There are several topic areas that proved controversial and in need of discussion. They are covered in this section.

3.2.1 Environmental vs. Musical Sound

We are usually talking about *crisp sets* (i.e., hard cluster assignment) when categorizing things. Being a member of such a set is an all or nothing affair. The opposite of this is a *fuzzy set* (i.e., soft cluster assignment). So there are obvious differences between musical and environmental sound in examples taken from crisp sets – which are all one thing or the other. Nevertheless, in practice it is never this clear cut, as many musical styles actually use environmental sounds – e.g., soundscape compositions, or ambient music. Furthermore, within this overlap of sound categories, the discrimination between them could be the result of a high-level (aesthetic) rather than a low-level (perceptual) process. Indeed, this thesis tends to view music as a very much attached to the environment around it. Hence a crucial assumption is that music overlaps with environmental sounds at the common level of timbre, while distinctions between sound and music only happen at a higher formal level, such as melody or harmony).

We have been careful in not disregarding the work of Aucouturier et al. (2007), who studied the discriminatory power of a summarization method based on spliced signals, when applied to urban recordings and polyphonic music. In our research we use a similar approach (called *bag-of-frames*), but we also introduce subtle technical differences. For instance, the point where the random slices were taken was at the peak of energy in each case (i.e., at the onset of spectral flux) – a very relevant perceptual issue, according to the detailed analysis performed by Seyerlehner et al. (2009). The frame length was randomized within a range that would reduce the artifact produced by the splicing of frames. And the final spliced stimulus was composed of slices taken from different songs pertaining to the same *timbral environment*. This was because the focus was not on the retrieval of a particular song, but on finding the boundaries between different clusters of songs that shared the same timbral characteristics.

3.2.2 Timbral Environments and Auditory Scene Analysis

Of crucial significance to the concept of perceptual memory, and this dissertation, are Bregman's (1990) views on timbre, particularly with respect to auditory scene analysis (ASA). The analysis of a stream of music depends on hearing correctly the many individual components that make it up, and timbre has a significant part to play in this. Segmentation, integration and segregation of the auditory stream depend on physiological, psychological and neurological processes, the traces of which perhaps provide the basis for information regarding the audio stream as a whole. It is this information about music (i.e., semantic references to acoustic phenomena), that individuals share using social media such as Last.fm. We therefore subscribe to Bregman's idea that timbre perception relies on dynamic internal processes and is the "...a perceptual description of a stream, not of an acoustic waveform" (Bregman and Pinker, 1978, p. 24).

3.2.3 Recognizing Spliced Audio with Repetition

A post hoc observation of the similarity experiment performed in study PIV was that, after repeated exposition to spliced stimuli, individuals experienced an expansion in their focus of attention, which allowed them to recognize the origin of some of the excerpts in each spliced sample. This was revealed only in their informal comments after the experiment, and was not considered fully at the experimental design stage. It also had an important implication for the main argument and the overall results, because it showed a direct link with the long term memory. Understandably, when a second of the spliced stimuli is listened to only once, it is more likely to sound like a burst of unrelated sonic events, depending on the individual of course. The scenario changes, however, when the same sample is listened to another 18 times! Then the sample, that first seemed like a burst of unrelated sound is converted into a rich, complex and condensed tool for mining past musical experiences. Complex because it can be composed of several acoustic events and condensed because these events are fitted into a very brief sound sample. The impact of this implication leaves us with an open empirical question, and suggests interesting paths for future research and applications. For instance it would be interesting to learn about the minimum number of repetitions required to harvest such memories, and this could then be taken into account when designing tools to study the development of musical preference within a very concise protocol.

3.2.4 Particularities of Tags Retrieved from Last.fm

Last.fm tags have proved to be a reliable resource for music information retrieval research and applications (e.g., Levy and Sandler, 2007; Aucouturier and Pampalk, 2008; Bertin-Mahieux et al., 2008; Chen et al., 2009; Laurier et al., 2009; Levy and Sandler, 2009). It does have a few limitations however, as Lamere (2008) has pointed out – hacker attacks, inherent noise and cultural bias, for example). A further criticism could arise from the fact that the precise demographics of the sample are unclear, just as the technical details pertaining to the accretion and summarization of tags is itself also obscure. For example, do tags placed earlier induce a bias for latter taggers? And is the ranking of tags a democratic or consensual process? We would add to the discussion at this point that one way of looking at these inherent *flaws* in the database, could be to instead consider them as properties of this particular language, as Simons (2008) did when he proposed the term "tag-elese". In addition, we explored the spectrum of ranked frequencies of tags within the corpus using Popescu and Altmann's (2009) geometric assumptions, and this endorsed Simons's views and opened new filtering possibilities, such as the ones discussed in study PV.

3.2.5 Tags as Rule Based Representations in Memory

Tags are synthetic descriptions given to objects of interest. They are useful for categorizing items so that they are easier to retrieve from large collections. Musical genres are one good example of tags because they have historically helped to organize ever growing music collections with the ever-changing development of new forms of recording technology. The internet too has made tags a useful tool for exchanging music and it has actually challenged the original taxonomy designed by experts (e.g., musical genres), with there being so many more kinds of music, and from so many different sources, readily available. Within this context –where the vocabulary of tags has experienced a huge expansion and has changed from the single entity based (e.g., musical genres) to a multilayered space (see Study PV) – a common view has been to consider tags as semantic units "mirroring" existing perceptual objects, as in the *picture theory* (Wittgenstein, 1922). In other words, while it is assumed that tags are written in plain English, they function more like a pidgin language without grammar, derived from English and other major natural languages (Simons, 2008). So, we need to decide whether tags are to be considered as mentally represented as such or if they are derived from a rule based representation (cf. Hare et al., 2001). In this project tags are treated as abstract objects, that is, as mnemonic units (i.e., mental images of an external object) characterizing only a minimal characteristic of the musical object they refer to, not as representations of the object itself. This implies that tags should not be cut in pieces (i.e., tokenized, lemmatized or stemed) to further study them. Nevertheless, an empirical validation of such an assumption is still pending.

3.3 Future Directions

The initial convergence of the semantic, perceptual and acoustic domains found in the present series of studies is but an early exploration, that should to be followed up with a sharper focus on specific topics. For instance we could use a method that would allow a more direct comparison between these domains. Although using two dimensional representations to superimpose the semantic distances onto the perceptual distances have worked to some degree and certainly demonstrate the feasibility of such a comparison, it would be desirable to quantify the actual fitness for the convergence of all three domains. The whole procedure would in this way be optimized, and all methodological variants tested.

How linguistic representations of timbre are cognitively embodied remains an open empirical question. For instance some tags refer specifically to body parts, or perceptual experiences closely related to other (non-auditory) senses, such as smells or colors. It may be that tags are not particularly best suited to unveil such phenomena, but it could be approached by adapting the same tripartite approach to study timbre imagery, particularly that reflected, for instance, in the onomatopoeic vocabulary (i.e., words sounding like the sounds they refer to).

This convergence approach (semantic, acoustic and perceptual) could also be shifted from the low-level features domain (timbre) to the structural domain (e.g., melody, rhythm, harmony, etc.) in further studies. It might be worth exploring, for example, how these are embodied and codified in the semantic space for deeper insights into the psychology of music or for the improvement of music retrieval systems.

It would be desirable that each of these suggestions for future studies acknowledge that timbre needs to be studied within the social context, and from an ecological perspective.

4 CONCLUSIONS

Timbral phenomena have been explored from many different perspectives previously. The present dissertation has employed an ecological framework to support empirical explorations using social media that is readily available on Internet. It has also been based on a multidisciplinary approach that has aimed to find convergences between semantic structures, acoustic descriptions and perceptual similarities associated with timbre. In other words, the common language that people use to describe the qualities of music was compared with the physical description of the sounds they describe. This allowed the perceptual and semantic characteristics of timbre to be mapped using a bottom-up approach.

The paradigms of socially distributed cognition and embodied cognition were used together to support arguments which propose that the social and individual levels are equally important in the exchange of information about music. This allowed us to investigate a large amount of public data in situ, without fragmentation and in concordance with the ideal of ecological validity.

AMP was based on these findings, and showed that an individual's music preference profile could be created with minimal input from the participant. In that respect, this application is a potential contribution to research as it groups people according to their musical taste in a more practical fashion than by musical genre alone. Musical Genres have proved to be transient and context dependent, whereas the connections between timbre and verbal descriptors have proved to be more robust in a variety of contexts, as revealed by the behavioural experiments in this thesis.

The proposed organization of tags in semantic layers which are fuzzy, as they are in social media, rather than in simple binary or crisp sets (i.e., belonging to or not) allows for a richer variety in description. In other words, an affect related term can still be used as a tag, even if the existing tag vocabulary describes different musical genres. Although the fuzzy classification proposed might pose methodological challenges compared to relatively straightforward crisp sets, the direction is worthwhile as it has the power of portraying a more natural classification of objects according to prototypes. Besides, it might well be one of the few acceptable approaches to study corpora where the words carry only a part of the information, if we assume that tags in isolation contain non-explicit meaning (e.g., “rock” is meaningless withouth a context). All these things should be considered if we are to organize music collections or to construct music recommendation systems based on tagging, among other applications.

The epistemological foundation and empirical evidence presented in this dissertation could have important implications for timbre and music research in general. First of all, it questions the validity of contemporary or future studies that intend to use only monophonic or artificial sounds. This is because the increased availability of musical material should be considered as a very important factor in reshaping our music-related behaviour at the social and individual levels. Secondly, it subscribes to the pragmatism of multidisciplinarity in music research, suggesting that monodisciplinary endeavors have perhaps reached their maximum explanatory potential already. The growth of systematic musicology (Parncutt, 2007; Leman, 2008), which is founded on multidisciplinarity within music research, is no doubt due to this.

It might be possible that timbre research is now at a point of sufficient maturity for an all-inclusive review to be written: particularly of the methods applied in the field in the last decades. Aesthetic views from a multicultural perspective may lead in a natural way to a preliminary draft of such a general theory of timbre. Such an ambitious endeavor is inexorably becoming a requisite for promoting further advancement in the field.

REFERENCES

Alluri, V. and Toiviainen, P. (2010). Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception*, 27(3):223–242.

Alluri, V., Toiviainen, P., Jääskeläinen, I., Glerean, E., Sams, M., and Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, 59(4):3677 – 3689.

Aucouturier, J.-J., Defreville, B., and Pachet, F. (2007). The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music. *The Journal of the Acoustical Society of America*, 122(2):881–891.

Aucouturier, J.-J. and Pachet, F. (2004). Improving timbre similarity: How high is the sky. *Journal of Negative Results in Speech and Audio Sciences*, 1(1):1–13.

Aucouturier, J.-J. and Pampalk, E. (2008). Introduction-from genres to tags: A little epistemology of music information retrieval research. *Journal of New Music Research*, 37(2):87–92.

Baccigalupo, C., Plaza, E., and Donaldson, J. (2008). Uncovering affinity of artists to multiple genres from social behaviour data. In *Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR)*, pages 275–280, Philadelphia, USA.

Barnier, A., Sutton, J., Harris, C., and Wilson, R. (2008). A conceptual and empirical framework for the social distribution of cognition: the case of memory. *Cognitive Systems Research*, 9(1-2):33–51.

Bellingham, J. (2012). Tone-colour. In Latham, A. editor, *The Oxford Companion to Music* Retrieved from http://www.oxfordmusiconline.com/subscriber/article/opr/t114/e6836.

Bernays, M. and Traube, C. (2011). Verbal expression of piano timbre: Multidimensional semantic space of adjectival descriptors. In *International Symposium on Performance Science*, pages 299–304.

Bertin-Mahieux, T., Eck, D., Maillet, F., and Lamere, P. (2008). Autotagger: A model for predicting social tags from acoustic features on large music databases. *Journal of New Music Research*, 37(2):115–135.

Bismarck, G. (1974). Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acustica*, 30(3):146–159.

Bregman, A. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge.

Bregman, A. and Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 32(1):19.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Burgoyne, J. and McAdams, S. (2008). A meta-analysis of timbre perception using nonlinear extensions to clascal. In *Computer Music Modeling and Retrieval. Sense of Sounds*, pages 181–202. Springer.

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

Carroll, J. and Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35(3):283–319.

Chen, L., Wright, P., and Nejdl, W. (2009). Improving music genre classification using collaborative tagging data. In *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining*, pages 84–93. ACM.

Chi, E. (2009). Augmented social cognition: Using social web technology to enhance the ability of groups to remember, think, and reason. In *Proceedings of the 35th SIGMOD International Conference on Management of Data*, pages 973–984, Providence, Rhode Island, USA.

Disley, A., Howard, D., and Hunt, A. (2006). Timbral description of musical instruments. In *International Conference on Music Perception and Cognition*, pages 61–68.

Eerola, T. and Ferrer, R. (2009). Setting the standards: Normative data on audio-based musical features for musical genres. Poster presented at *The 7th Triennial Conference of European Society for the Cognitive Sciences of Music*.

Godøy, R. (2006). Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound*, 11(02):149–157.

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

Hajda, J., Kendall, R., Carterette, E., and Harschberger, M. (1997). Methodological issues in timbre research. In Deliège, I. and Sloboda, J., editors, *Perception and Cognition of Music*, pages 253–306. Psychology Press.

Hammond, K. (1998). Ecological validity: Then and now. Retrieved from http://www.albany.edu/cpr/brunswik/notes/essay2.html.

Hare, M., Ford, M., and Marselen-Wilson, W. (2001). Ambiguity and frequency effects in regulr verb inflection. In Bybee, J. and Hopper, P., editors, *Frequency and the emergence of linguistic structure*, volume 45, pages 181–200. John Benjamins Publishing Company.

Held, C. and Cress, U. (2009). Learning by foraging: The impact of social tags on knowledge acquisition. In Cress, U., Dimitrova, V., and Specht, M., editors, *Learning in the Synergy of Multiple Disciplines*, volume 5794 of *Lecture Notes in Computer Science*, pages 254–266. Springer Berlin / Heidelberg.

Helmholtz, H. v. (1954). *On the sensations of tone as a physiological basis for the theory of music*. Dover Publications, New York.

Hesse, F. (2009). Use and acquisition of externalized knowledge. In Cress, U., Dimitrova, V., and Specht, M., editors, *Learning in the Synergy of Multiple Disciplines*, volume 5794 of *Lecture Notes in Computer Science*, pages 5–6. Springer Berlin / Heidelberg.

Huron, D. (2001). Toward a theory of timbre. Retrieved from http://musicog.ohio-state.edu/Huron/Talks/SMTmidwest.2001/talk.01.html.

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

Jensen, K. (1999). *Timbre models of musical sounds*. Department of Computer Science, University of Copenhagen.

Juslin, P. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human perception and performance*, 26(6):1797–1812.

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

Kim, H., Decker, S., and Breslin, J. (2010). Representing and sharing folksonomies with semantics. *Journal of Information Science*, 36(1):57–72.

Krimphoff, J., McAdams, S., and Winsberg, S. (1994). Caractérisation du timbre des sons complexes. ii. analyses acoustiques et quantification psychophysique. *Journal de Physique*, 4(C5):625–628.

Krumhansl, C. and Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, 18(3):739–751.

Lamere, P. (2008). Social tagging and music information retrieval. *Journal of New Music Research*, 37(2):101–114.

Langfelder, P., Zhang, B., and Horvath, S. (2009). *dynamicTreeCut: Methods for detection of clusters in hierarchical clustering dendrograms.* R package version 1.20.

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

Laurier, C., Sordo, M., Serrà, J., and Herrera, P. (2009). Music mood representation from social tags. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan.

Leman, M. (1995). *Music and Schema Theory: Cognitive foundations of systematic musicology*. Springer, Berlin Heidelberg.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. MIT Press, Cambridge.

Leman, M. (2008). Systematic musicology at the crossroads of modern music research. In Schneider, A., editor, *Systematic and comparative musicology: concepts, methods, findings*, pages 89–115. Peter Lang, Frankfurt am Main.

Levy, M. and Sandler, M. (2007). A semantic space for music derived from social tags. In Dixon, S., Bainbridge, D., and Typke, R., editors, *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR)*, volume 1, page 12, Vienna, Austria. Österreichische Computer Gesellschaft.

Levy, M. and Sandler, M. (2008). Learning latent semantic models for music from social tags. *Journal of New Music Research*, 37(2):137–150.

Levy, M. and Sandler, M. (2009). Music information retrieval using social tags and audio. *IEEE Transactions on Multimedia*, 11(3):383–395.

Lin, H., Davis, J., and Zhou, Y. (2009). An integrated approach to extracting ontological structures from folksonomies. In *Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications*, page 668, Heraklion, Greece. Springer.

Mathes, A. (2004). Folksonomies-cooperative classification and communication through shared metadata. Retrieved from http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html.

Maturana, H. (2002). Autopoiesis, structural coupling and cognition: A history of these and other notions in the biology of cognition. *Cybernetics & Human Knowing*, 9(3-4):5–34.

McAdams, S. and Cunible, J.-C. (1992). Perception of timbral analogies. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 336(1278):383–389.

McAdams, S., Winsberg, S., Donnadieu, 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.

McClain, G. (1978). *The Pythagorean Plato: Prelude to the song itself*. Nicolas-Hays.

Menon, V., Levitin, D., Smith, B., Lembke, A., Krasnow, B., Glazer, D., Glover, G., and McAdams, S. (2002). Neural correlates of timbre change in harmonic sounds. *Neuroimage*, 17(4):1742–1754.

Nussbaum, C. (2007). *The Musical Representation: Meaning, ontology, and emotion*. MIT Press, Cambridge.

Pang, H., Lin, A., Holford, M., Enerson, B., Lu, B., Lawton, M., Floyd, E., and Zhao, H. (2006). Pathway analysis using random forests classification and regression. *Bioinformatics*, 22(16):2028.

Parncutt, R. (2007). Systematic musicology and the history and future of western musical scholarship. *Journal of interdisciplinary music studies*, 1(1):1–32.

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the cuidado project. Cuidado ist project report, IRCAM, Paris, France.

Pohle, T., Pampalk, E., and Widmer, G. (2005). Evaluation of frequently used audio features for classification of music into perceptual categories. In *Proceedings of the Fourth International Workshop on Content-Based Multimedia Indexing*, Riga, Latvia. Tampere University of Technology.

Popescu, I. and Altmann, G. (2009). *Word frequency studies*. Walter de Gruyter, Berlin.

Porcello, T. (2004). Speaking of sound. *Social studies of science*, 34(5):733.

Prem, D., Parncutt, R., Giesriegl, A., and Stiegler, H. (2011). Jazz vocal sound: A timbre knowledgebase for research and practice. Retrieved from http://www.n-ism.org/CIM2011/Abstracts/Abstract_PremParncutt.pdf.

Reybrouck, M. (2005). A biosemiotic and ecological approach to music cognition: Event perception between auditory listening and cognitive economy. *Axiomathes*, 15(2):229–266.

Risset, J. and Wessel, D. (1999). Exploration of timbre by analysis and synthesis. In Deutsch, D., editor, *The psychology of music*, pages 113–169. Orlando, FL: Academic Press.

Saarikallio, S. (2011). Music as emotional self-regulation throughout adulthood. *Psychology of Music*, 39(3):307.

Sandell, G. (1998). Macrotimbre: Contribution of attack, steady state, and verbal attributes. *The Journal of the Acoustical Society of America*, 103:2966.

Sandell, G. and Chronopoulos, M. (1997). Perceptual constancy of musical instrument timbres; generalizing timbre knowledge across registers. In Gabrielsson, A., editor, *Proceedings of the Third Triennial (ESCOM) Conference*, pages 222–227, Uppsala. Uppsala Unversity.

Sarkar, M., Vercoe, B., and Yang, Y. (2007). Words that describe timbre: A study of auditory perception through language. In *Language and Music as Cognitive Systems Conference (LMCS-2007), Cambridge, UK*, pages 11–13.

Schafer, R. M. (1977). *The Tuning of the World*. McClelland & Stewart, Toronto, Canada.

Sethares, W. (2005). *Tuning, Timbre, Spectrum, Scale*. Springer Verlag.

Seyerlehner, K., Pohle, T., Widmer, G., and Schnitzer, D. (2009). Informed selection of frames for music similarity computation. In *Proceedings of the 12th International Conference on Digital Audio Effects (DAFx)*, Como, Italy.

Simons, J. (2008). Tag-elese or the language of tags. *Fibreculture Journal*, 12.

Snyder, B. (2000). *Music and memory: an introduction*. The MIT Press.

Stepánek, J. (2006). Musical sound timbre: Verbal description and dimensions. In *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx06), Montreal, Canada, September 18*, volume 20.

Taylor, J. (1995). *Linguistic categorization: Prototypes in linguistic theory*. Clarendon press Oxford.

Tsang, C. and Trainor, L. (2002). Spectral slope discrimination in infancy: Sensitivity to socially important timbres. *Infant Behavior and Development*, 25(2):183–194.

Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302.

Van Damme, C., Hepp, M., and Siorpaes, K. (2007). Folksontology: An integrated approach for turning folksonomies into ontologies. In *Proceedings of the ESWC Workshop "Bridging the Gap between Semantic Web and Web 2.0*, volume 2, pages 57–70. Springer.

Vander Wal, T. (2007). Folksonomy. Retrieved from http://vanderwal.net/folksonomy.html.

Wessel, D. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 3(2):45–52.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. International Library of Psychology Philosophy and Scientific Method. Kegan Paul, Trench, Traubner & Co., LTD., London.

Zacharakis, A., Pastiadis, K., Papadelis, G., and Reiss, J. (2011). An investigation of musical timbre: uncovering salient semantic descriptors and perceptual dimensions. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, pages 807–812.

YHTEENVETO (FINNISH SUMMARY)

Musiikillisen sointivärin jakautunut kognitio

Tämä väitöskirja tutkii musiikillisen sointivärin kognitiota sosiaalisesta näkökulmasta. Ilmiön tarkastelu aloitetaan rajaamalla sointivärin käsite sointiväriympäristöihin, joilla kuvataan musiikillisten objektien kokonaisvaltaista ääntä sosiaalisessa kontekstissa. Tämän työn teoreettinen viitekehys laajentaa sointivärin määritelmää kahdessa eri merkityksessä: yksittäisten äänten havaitsemisesta reaalimaailman monitahoisiin ääni-ilmiöihin, sekä yksilön kehollisesta kognitiosta sosiaalisesti jakautuneeseen kognitioon. Tämän määritelmän pragmaattista arvoa testattiin eri näkökulmista. Ensin tarkasteltiin sointiväripiirteiden yhteyttä affektiivisiin dimensioihin, tavoitteena testata korkean tason kognitiivisten toimintojen ja fysikaalisten objektien kuten äänen yhdistämisen metodologista uskottavuutta. Tämän jälkeen analysoitiin erään johtavan musiikillisen sosiaalisen median (Last.fm) käyttäjien tuottamaa tietoa, jota voidaan kuvata äänlaiseksi keinotekoiseksi, semanttisista "tägeistä" muodostuvasi kieleksi. Last.fm – palvelussa käyttäjät ovat antaneet musiikillisille objekteille kuten kappaleille ja artisteille listan erilaisia tägejä. Palvelun käyttäjät käyttävät tägejä välittämään ideoita, kuvauksia ja emootioita musiikkikappaleisiin tai – artisteihin liittyen. Näiden tägien semanttisia ominaisuuksia analysoitiin, ja yhdistettiin niiden kuvaamien objektien akustisiin ja havaittuihin ominaisuuksiin. Tägien analyysi paljasti rakenteen josta on erotettavissa tarkempia kategorioita kuten musiikkigenrejä, adjektiiveja ja tunnesanoja. Lopuksi tulosten käytännön arvoa testattiin kehittämällä musiikkimakua mittaava psykometrinen testi.

Avainsanat: musiikki, sointiväri, kognitio, sosiaalinen media, semanttinen analyysi, musiikkimaku

ORIGINAL PAPERS

PI

TIMBRAL ENVIRONMENTS: AN ECOLOGICAL APPROACH TO THE COGNITION OF TIMBRE

by

Ferrer, R. 2011

Empirical Musicology Review, 6(2), 64–74

Timbral Environments: An Ecological Approach to the Cognition of Timbre

RAFAEL FERRER
Finnish Centre of Excellence in Interdisciplinary Music Research,
Department of Music, University of Jyväskylä

ABSTRACT: This study formulates an ecological framework that links the environment and human systems, to support further arguments on the influence of timbre in the music appreciation schemata. At the core of the framework is the notion of *timbral environments,* which is introduced as an epistemological foundation to characterize perceptual cues of internalized representations of music, and to explore how these are expressed in the dynamics of diverse external environments. The proposed notion merges the concepts of *macrotimbre* (Sandell, 1998) and *soundscape* (Schafer, 1977) to distinguish between the formulated framework and traditional approaches to timbre, which are mainly concerned with short-term temporal auditory events. The notion of timbral environments enables the focus of timbre research to be shifted from isolated events to socially relevant sounding objects, hence facilitating the identification of connections between semantic descriptors and the physical properties of sounds. [1]

Submitted 2009 January 26; accepted 2011 April 7.

KEYWORDS: *timbral environments, macrotimbre, soundscape*

THE word timbre is arbitrarily used to refer to multiple qualities of sound, which is an important part of the problem to define it. Despite the fact that categorisation of specific qualities of music, in terms of timbre, began almost two hundred years ago, psychologically inspired descriptions of timbre and existent research have not attained sufficient strength for generating a general theory of timbre (Huron, 2001).

Introductory paragraphs of this paper present selected ideas conceiving entities in a continuum of interactions (taken from different fields e.g., biology, cybernetics, music cognition) as epistemological basis to explore the cognition of timbre from an ecological perspective. Then, a review of major concerns with respect to timbre, such as history and categorization issues, precedes a description of how the ideas presented in the introduction could be used in the particular case of timbre. Here a distinction is made between the classical empirical approach to timbre that has been concerned with short, isolated sound events and a concept of global timbre, which covers a longer time-span, such as macrotimbre and soundscape (Sandell, 1998; Schafer, 1977). This exposition ends with the introduction of a new term: *timbral environments*, which merges the concepts of macrotimbre and soundscape to shift the focus away from the traditional approach to timbre. The closing part of the paper includes pragmatisms concerning empirical possibilities for the introduced term.

AN ECOLOGICAL PERSPECTIVE

Music perception and cognition could be modelled as an *autopoietic* (i.e., self-organised, self-structured and autonomous) system if music is considered a psychological construct and the sonic environment a continuum of information in which the individual exists. Such a framework serves to focus on the dynamic interactions between the components of that system rather than in the components themselves. Autopoietic theory has been used before to explain the cognition of polyphonic music (Chagas, 2005). In this paper, autopoiesis is used as a framework to present a theoretical model of timbre cognition by considering an individual and her musical schemata (i.e., mental structures reflecting the relations between the perceived objects, see Leman, 1995) as an autonomous unit defined by its participative *interaction* with the environment.

An individual must intentionally project herself into the environment in order to internally represent specific fragments of the environment. According to this view, the human body is an autonomous unit interacting with the environment through its *sensors* and *effectors* (Godøy, 2006), which is in concordance with the paradigm of *embodied cognition* (Leman, 2007). Analogically, music listeners can be considered as *adaptive devices* in that they organize their sensors and effectors to adapt themselves to the world, while simultaneously modifying it (Reybrouck, 2005). According to the representational theory of mind (Nussbaum, 2007), individuals use their bodies, throughout their lives, to develop a consciousness about themselves and the complexities of the surrounding environment. A body has perceptual capabilities that allow it to internally represent the outside environment. This interaction with the environment, which includes both objects and other individuals, is the point at which the interplay between action and perception begins.

Interactionism maintains that mental and physical events "…causally influence each other" (Tye, 2008), and proposes a solution to the mind-body problem (Popper & Eccles, 1984) if we agree to extend the boundaries of the mind as an object of study beyond the individual into society in terms of *augmented* (Chi, 2009) *or distributed* (Barnier, Sutton, Harris, & Wilson, 2008) *social cognition*. However, the word *interaction* evokes a neutral relation, and for music perception, a term reflecting a more active role should be used. For that purpose, Kaipainen (1996) proposes the use of *participation*, arguing that through a conscious and participative interaction with the environment, we generate fluctuations in the system and at the same time promote changes in our internal structures (e.g., *neural plasticity)*. The dynamic interplay where structures and their internal organisations are mutually deformed is termed *structural coupling* (Maturana, 2002), (see Figure 1). In addition to the body acting as a mechanical medium, language - in a multimodal sense - is considered by Maturana (1988) as the subject matter of reciprocal coupling, hence, a social dimension is implied. In the views presented in this paper, the social domain is the fabric composing the environment, as necessary as unavoidable in a musical context.

By participating with the environment human beings, develop a categorization of musical phenomena (Dura, 2006), including finer variations of sounding qualities (Bregman, 1990). An example is the ability of a one month old baby to distinguish its mother's voice (Mehler, Bertoncini, Barriere, & Jassik-Gerschenfeld, 1978), which later in life develops into the ability to discriminate subtle timbral variations such as phonemes (Hinton, Nichols, & Ohala, 1995; McMullen & Saffran, 2004; Patel & Iversen, 2003). Clarke (2005) identifies this process of recurrent categorization as *perceptual learning*, which in an ecological context and from the perspective of *information foraging* (Held & Cress, 2009) theory, might be an expression of an externalized rather than internalized form of knowledge. At the core of the participation with the environment could be the *transactional memory*, i.e., people not having to know everything if they can use other people's knowledge (Hesse, 2009); what escapes one individual's perceptual capabilities, is captured by another, thus allowing the exchange of memory cues used to make transactions (Chi, 2009). The purpose of these transactions could be the *adaptation* (Maturana, 2002) of the individual to a given environment.

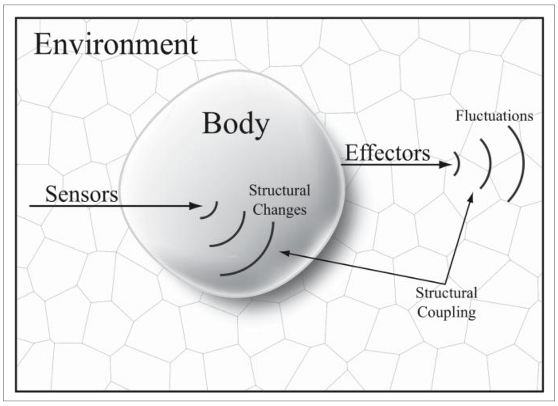


Fig 1. Diagram of the embodied-participative model

This model might well aid in the exploration of a wide variety of musical phenomena. For instance, individuals' musical preferences (Delsing, ter Bogt, Engels, & Meeus, 2008), developed in the dynamics of a social context (Gregory, 1999; Rentfrow & Gosling, 2007), contribute in the construction of the self (DeNora, 1999), by means of social identity (Bakagiannis & Tarrant, 2006) or interpersonal perception (Rentfrow & Gosling, 2006). In this example, the social dimension corresponds to the environment. The dynamic organisation, as the social identity and interpersonal perception, are fluctuations in the environment, and the self could be considered as the embodied entity experiencing constant changes in its structure.

I will now turn to the more difficult concept, namely timbre, and provide an exposition into the history of the concept before applying the interactional/embodied/autopoietic view to it. Finally, an attempt will be made to advance such embodied framework for timbre studies and provide the reader with lists of benefits associated with this perspective applied to timbre.

TIMBRE AS A CATEGORY OF SOUND

Many kinds of sound phenomena fall into the category of *timbre*, which remains ill defined for several reasons (cf. Donnadieu, 2007). One of the most problematic is that timbre is an umbrella term that has been used to describe many categories of sound, which have since been differentiated by empirical methods. To overcome this polysemic conflict I propose to use the term *timbral environments* to refer to a very specific category of the general sounding phenomena. Timbral environments are concerned with reminiscences of music pieces that roughly share the same range of acoustic features but have the property of being meaningfully grouped by noticeable distances in the perceptual space. An example of this meaningful grouping is the taxonomy of musical genres (Pachet & Cazaly, 2000), however, the term timbral environments is created with the purpose of generating alternative taxonomies that extend beyond musical genre into different layers of the music ontology. Ideally, timbral environments will aid the investigation of the emergence and functionality of musical schemata, particularly in relation to the music preferences. An elaboration of past ideas relating to timbre, the current proposal and connections with the general framework of the paper are presented next.

Highlights in the History of Timbre

Words tend to acquire new meanings from time to time as the concept they refer becomes more complex. For example, the conceptual shift provoked by the invention of perspective (in the visual arts) or polyphony (in music); the same elements in a given representational space but organised in a different way led to a whole new idea about depth in the visual and auditory domain respectively. After being exposed to these inventions a change in our minds emerges, and consequently, the way we use our bodies to perceive new characteristics about the things we already know is also transformed.

The invention of timbre as a novel category of sound wrought similar changes. According to Fales (2005), the modern meaning for the term timbre can be traced back to the Age of Enlightenment. Through a historical review of the concept, Fales argues that one of the first Westerners that became aware of timbre, in the sense that we use it today, was Jean Phillipe Rameau (1683-1764). He proposed that the difference between "hearing and listening"[2] posed problems for an effective understanding of the *corps sonore*. The distinction between the two tasks of hearing and listening, suggests that he intuitively recognised the need to make a conscious effort to grasp the particular qualities of sound he was capable of perceiving. For him, the *corps sonore* was an agglomerate of sound attributes that needed to be dissected - and perhaps this is the closest analogy to our present understanding of timbre. Nevertheless, scholars of his epoch failed to grasp the idea and nobody was able to give an empirical explanation for the phenomenon. This remained the state of affairs for a century, until Hermann von Helmholtz (1821-1894) started to relate the perceptual attributes of sound with its physical properties (Helmholtz, 1954). After him, the music psychologist Carl Seashore (1866-1949) proposed timbre as the most important and complex aspect of tone, over pitch, loudness and duration (Seashore, 1967).

A new tradition in timbre research started with the advent of *cognitive structuralism*, which was largely based on similarity tests (i.e., assessment of subjective similarity relations between audible stimuli). It led to an understanding of the multidimensional nature of timbre (Grey, 1977; McAdams, Winsberg, Donnadieu, Soete, & Krimphoff, 1995; Wessel, 1979), and has remained the classical approach in timbre research for the last 40 years. It has also been used in the development of computational models, with the goal being to find representations of timbre that are "*isomorphic* with human perception" (Terasawa, Slaney, & Berger, 2006). Nevertheless, cognitive structuralism has a major drawback, as explained by Leman (1995): it cannot capture the dynamics of the perceptual system. Therefore, recent efforts have been aimed at understanding timbre in both dynamic (Hajda, 2007) and complex settings (Donnadieu, 2007).

An embodied view of timbre embedded in an autopoietic framework might be able to provide a solution to the explanatory deficiency of cognitive structuralism because it extends the focus of research to an ecological context, where the dynamic relations between the environment and the elements coexisting within are at the core of the model.

Embodiment of Timbre

Direct and inferential theories of perception are often presented as antagonists (Chemero, 2003); the difference between the two views is related with the localization of *meaning*, whether it is in the environment (direct) or in the individuals (inferential). In this work, these views are rather presented as complementary, but in order to do that we need a common ground, which can be the theory of affordances (Gibson, 1986).

The capability of an individual to *afford* a participative interaction with the environment depends on how aware she is about the contents of that environment. This awareness is constructed by linking three different kinds of reality: the first comprises physical entities that exist in the environment; the second kind is the mental state or states of consciousness associated with thinking and perceiving; and a third reality is composed of abstractions and ideas, or intuitions in the old Platonic sense (Popper & Eccles, 1984). The same three realities also translate into musicological research, albeit with alternative terminology. The first sees music as a morphology that consists of physical entities; the second is an internal and isomorphic representation of those morphologies (Dura, 2006; Terasawa et al., 2006), and the third consists of *isomorphisms of second order*, which are abstractions that control the emergence and functionality of perception (Leman, 1995). These three realities are linked by loops of action-perception (i.e., structural coupling) that bring closure to the system in the autopoietic sense. However there is still a question regarding the nature of the isomorphisms. The paradigm of embodied cognition sheds light on a possible explanation, which maintains that such isomorphisms are encoded by, and in the body. In other words, if internalised representations of external objects use the body as a medium, they are most likely to be anthropomorphic projections (Godøy, 2006). It is probable that these projections reflect a unique part of the individual's self, as well as a fingerprint of the cultural environment embedded on her self as a result of her development. Such a reflection can be identified as identity, at an individual and at a social level. Identity is what remains after the individual participates with the environment and reorganises itself, preserving its unity, structure, and autonomy as a closed system. If this holds true, perceptual schemata are an ontological expression of the adaptive self, which potentially afford any information contained in the environment. Affordances can be viewed as learning methods, developed to apprehend specific characteristics of the environment, distributed among individuals in the environment and possibly taking the form of transactional memories.

Timbre can be explained in these three hypothetical worlds. For instance, the first where the physical attributes of sound exist (e.g., acoustic descriptors such as Mel Frequency Cepstral Coefficients and other statistical descriptions of the sound spectrum), the second, where individuals perceive (e.g., as in the studies of John M. Grey, David Wessel and Stephen McAdams among others), and the third comprising all the possible descriptions, hypotheses and theories about it. In this third world, it is safe (from an epistemological point of view) to speculate about the existence of an *unembodied* timbre; created and reserved only to be empirically tested. It is also on this third world where internalized representations of sound, imagery, and words to describe the sound experience, converge (e.g., internalized experiences uttered as onomatopoeias).

The specific embodiment of timbre remains unexplained due the particularities of the phenomena such as its multidimensionality, and perhaps because of a failure in the way we conceive the abstraction of our bodies in projection onto different categories of sound. For example, as timbre is a quality of sound, we could assert that the auditory system is the main sensory apparatus involved in the acquisition of an isomorphic representation. Nevertheless that can only be part of the truth, because by regarding two of the most used verbal descriptors of timbre such as *colourful* - *colourless* and *dull* - *sharp* (Sethares, 1999), it is evident that individuals' embodiment of timbre is mostly visual and tactile. Efforts in the vein of this example, where free verbal descriptions are paired with acoustic descriptors have shed light on this issue (Sarkar, Lan, Diaz, & Vercoe, 2009), nevertheless, the vocabulary has never been filtered and processed to obtain an anthropomorphic ontology. An example of this is provided at the end of the following section.

Defining Timbral Environments

The idea of timbral environments can be regarded as an extension of the work of Sandell (1998), who proposed the term *Macrotimbre.* This term challenges the traditional concept of timbre by referring to the set of qualities that remain invariant across several pitches at different loudness levels. Sandell's notion differs from the classical ANSI definition (American National Standards Institute, 1973), which also alludes to the characteristic of sound that allow us distinguish one source from another, but conceptually separated from loudness and pitch. The difference—apart from the inclusion/exclusion of obvious dimensions such as loudness and pitch—resides in how broad the scope is in terms of time. The ANSI view—that has permeated most of the corpus of research—presents a fragmented auditory object of a short duration in the order of seconds and milliseconds. In contrast, macrotimbre refers to events beyond such time restrictions, in the order of minutes or hours. Therefore it provides a better approach in terms of how timbre is internally represented in a holistic way, closer to the popular expression "it sounds like.…" While the classical empirical approach is mostly concerned with short and monophonic isolated sounds (Grey, 1977; Krumhansl & Iverson, 1992; McAdams et al., 1995; Terasawa et al., 2006; Wessel, 1979), Sandell's view is concerned with a summary of characteristics that makes us able to differentiate one source from the other even if they are performed at different loudness levels and pitches. The key to temporal span considerations resides in memory, which plays a central role in the form of *perceptual constancy* (Sandell & Chronopoulos, 1997).

The whole idea of microtimbre suggests that the schemata controling the perception of timbre enables us to understand that across pitches, loudness levels, attack types and articulations (i.e., *sul ponticello,* muted, *staccato*), the sound source remains the same. Such an interpretation has a high ecological quality and validity, since "…listeners do not perceive the acoustical environment in terms of 'phenomenological descriptions' but as 'ecological events'…" (Reybrouck, 2005, p. 234); reminiscences of musical events represent a global impression of past events. For instance, it can be argued that musical genres are characterized by their prototypical macrotimbres. Furthermore, the addition of the prefix "macro" to the word timbre is useful to make an epistemological distinction between the classical studies investigating the perceptual correlates of short excerpts of isolated sounds and further explorations that extend beyond such conceptual and methodological constraints.

The novel approach proposed here is aimed at characterizing longer temporal (i.e., beyond the lifespan of an individual's reminiscence about auditory events and complex timbral events such as *soundscapes* (Schafer, 1977). Schafer's term was constructed by substituting the prefix of the word 'landscape' with 'sound' to transpose the concept from the visual to the auditory domain. The term has inspired a host of publications within the field of acoustic ecology, where for example the sonic environment of two geographical locations is analyzed by contrasting their salient acoustic characteristics (Ge & Hokao, 2005).

What I propose is to merge the two notions of macrotimbre and soundscapes into *timbral environments*. To take advantage of the different perspective that macrotimbre affords with respect to the classical interpretations of timbre, and apply the methodologies and experiences that have been developed over the past forty years of soundscape research into different temporal domains and levels of complexity. Within the notion of timbral environments, the principle of perceptual constancy supporting macrotimbre, could be used to discriminate among prototypical soundscapes (e.g., predominant sounds surrounding a house in a city, in contrast to the predominant sounds surrounding a house near the sea, or distinguishing the differences between salient perceptual characteristics of musical genres). The perceptual validity for musicological research would be to provide a means to better explain our evident ability to discriminate not only between sources, but also between prototypical mixtures of sources (i.e., the overall sound of a rock ensemble versus the overall sound of a big-band orchestra). Listeners are able to recognize that such sources belong to well-defined categories of sound despite the huge numerical variance in objectively measured timbral descriptors. If these categories can be empirically characterized and systematically differentiated, we could call them timbral environments. This would make them distinct from the classical approach to timbre and *timbral spaces*, and stress the ecological approach (Godøy, 2006; Leman, 2007; Reybrouck, 2005). For a visual comparison between the three different definitions, see Figure 2. As illustrated in the figure, the role of memory is one of the crucial differences between the definitions, since the first one (classical timbre) does not have any direct relation, the second one (macrotimbre) is based on individuals' memory capacities for handling perceptual constancy and recognition and the third definition, timbral environments, is based on social, collective memory.

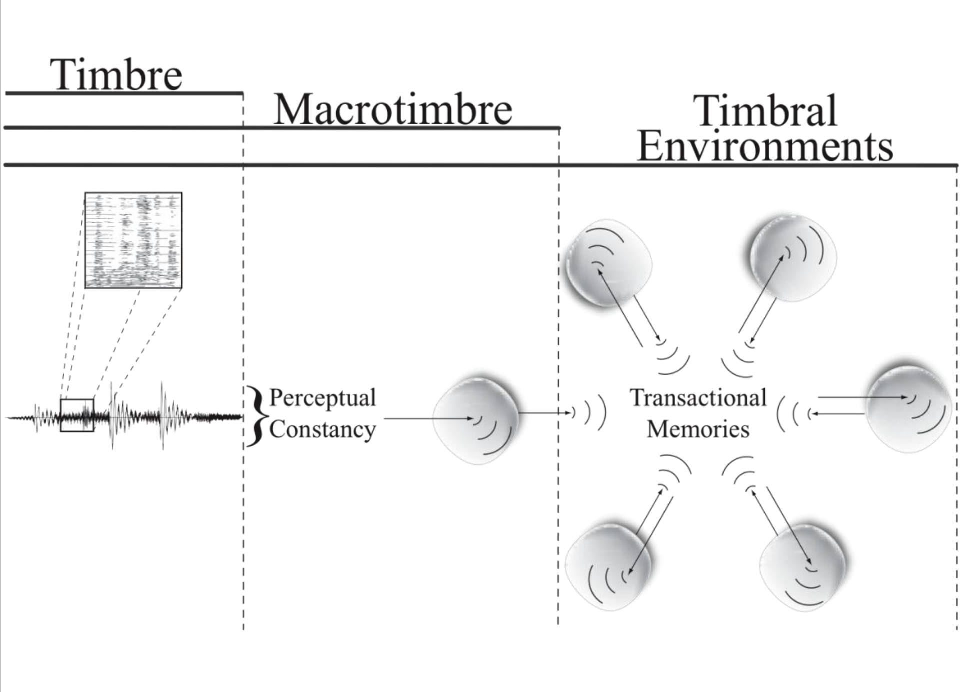


Fig 2. Comparison of the focus of research between the different definitions of timbre.

The notion of timbral environments could be used to represent a convergence between semantic and acoustic spaces empirically, for instance by filtering verbal descriptions of music to an anthropomorphic ontology and correlating such structure with the acoustic descriptors of the described music. Take for example the emotional attributes of a piece of music. A song is said to be sad or happy, although it can be argued that there is no such a thing contained in the song, or that there is no consensus about it. What is certain is the interpretation of the listener, or in the context of this paper, sadness or happiness are projections of the listener's self, who judges the piece and attaches a particular label to it. Such labelling does exist in the everyday chain of consumption-distribution of music, and it is called tagging (Lamere, 2008); users of social media (e.g., Last.fm) tag their music according to their own projections in the most varied semantic categories. The corpus of verbal descriptions can be classified and filtered according to categories related to the body or attributes inherently human such as the emotions [3]. Such an analysis has been carried out by Laurier, Sordo, Serrà and Herrera (2009), who derived a mood space by filtering 6,814,068 tags attributed to 575,149 pieces of music. Moreover, such a semantic structure can be connected with the music to establish a correlation between the semantic and acoustic domains (Ferrer & Eerola, 2010). The set of qualities describing the connections between the semantic and the acoustic domain could be considered as timbral environments, thus allowing us to refer to the set of acoustic descriptors that define, for example, sadness or happiness, or any other category related with the perception and cognition of auditory events.

CONCLUSIONS

The term *timbral environments*, is presented as a theoretical solution to further investigate the perceived general sounding quality of music in an ecologically valid fashion. However, it remains to be validitated by empirical means. Therefore, this theoretical foundation will be used in future research that is targeted at investigating the existing forces that shape the emergence and functionality of perceptual schemata of timbre.

Perception of timbre involves a complex system of interactions between listeners and their environment. Therefore, in order to extend our knowledge to reach an effective ontology of musical timbre, it would be advisable to learn about which parts of our body (or sensory systems) are involved when we attempt to grasp an internal representation of it. Objects in the environment cannot be represented as static morphologies, but as fluctuating anthropomorphic projections of the self. The ideas presented in this paper represent an effort to promote the awareness of timbre as an environmental issue that, for example, may have a possible influence on our daily decisions about what to listen to.

If the notion of timbral environments survives the process of empirical validation, it will have various implications for future studies by extending the focus of research beyond the traditional views displayed for instance in monophonic and polyphonic timbre research, or by contributing with empirical evidence to the definition of timbre as an aeshetic resource in Western and non-Western traditions. It will also be useful to derive the sounding objects and their features from conceptual units and sources that are meaningful and common for the listeners (everyday sounds, speech, typical instrument combinations), allowing for a better connection between semantic descriptors and acoustic features. Timbral environments could be studied using an array of behavioural methods (similarity ratings, priming tasks, semantic rating scales) as has been done in the past, but perhaps using richer sets of sound categories to keep the comparisons at a meaningful level. This will result in the creation of sets of stimuli in a bottom-up fashion, in which listeners' natural sound categories (e.g., musical genres, associations of sounds) are taken as the meaningful units.

NOTES

[1] Part of this work was presented in the SysMus08 conference in Graz, Austria, and was selected for publication in the British Postgraduate Musicology on-line under the title of "Embodied Cognition Applied to Timbre and Musical Appreciation: Theoretical Foundation."

[2] In *Observations sur notre instinct pour la musique* (1754).

[3] Note that in the interpretation made here, emotions elicited by music are considered as anthropomorphic attributions of music with the purpose of extending Godøy's (2006) term, anthropomorphic projection, beyond physical appearance.

REFERENCES

American National Standards Institute. (1973). Psychoacoustical terminology. In *S3.20-1973.* New York: American National Standards Institute.

Bakagiannis, S., & Tarrant, M. (2006). Can music bring people together? Effects of shared musical preference on intergroup bias in adolescence. *Scandinavian Journal of Psychology*, Vol. 47, pp. 129-136.

Barnier, A., Sutton, J., Harris, C., & Wilson, R. (2008). A conceptual and empirical framework for the social distribution of cognition: the case of memory. *Cognitive Systems Research*, Vol. 9, No. 1-2, pp. 33-51.

Bregman, A. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge: MIT Press.

Chagas, P. (2005). Polyphony and embodiment: a critical approach to the theory of autopoiesis. *TRANS-Transcultural Music Review*, Vol. 9, Article 15. Retrieved September 6, 2010, from http://www.sibetrans.com/trans/trans9/chagas.htm

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology*, Vol. 15, No. 2, pp. 181-195.

Chi, E. (2009). Augmented social cognition: using social web technology to enhance the ability of groups to remember, think, and reason. In: C. Binnig & B. Dageville (Eds.), *Proceedings of the 35th SIGMOD International Conference on Management of Data*, Providence, Rhode Island, USA, pp. 973-984.

Clarke, E. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. USA: Oxford University Press.

Delsing, M., ter Bogt, T., Engels, R., & Meeus, W. (2008). Adolescents' music preferences and personality characteristics. *European Journal of Personality*, Vol. 22, No. 2, pp. 109-130.

DeNora, T. (1999). Music as a technology of the self. *Poetics*, Vol. 27, No. 1, pp. 31-56.

Donnadieu, S. (2007). Mental representation of the timbre of complex sounds. In: J. W. Beauchamp (Ed.), *Analysis, synthesis, and perception of musical sounds: The sound of music.* New York: Springer.

Dura, M. (2006). The phenomenology of the music-listening experience. *Arts Education Policy Review*, Vol. 107, No. 3, pp. 25-32.

Eerola, T., Alluri, V., & Ferrer, R. (2008). Emotional connotations of isolated instruments sounds. In: *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC)*. Sapporo, Japan: University of Hokkaido, pp. 483-489.

Fales, C. (2002). The paradox of timbre. *Ethnomusicology*, Vol. 46, No. 1, pp. 56-95.

Fales, C. (2005). Listening to timbre during the French enlightenment. In: C. Traube & S. Lacasse (Eds.), *Proceedings of the 2005 conference on interdisciplinary musicology (CIM).* Montréal, Québec, Canada: Centre for Interdisciplinary Research in Music Media and Technology.
Available from http://www.oicm.umontreal.ca/doc/cim05/articles/FALES C CIM05.pdf

Ferrer, R. & Eerola, T. (2010) Timbral qualities of semantic structures of music. In: *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*. Utrecht, Netherlands, pp. 571-576.

Ge, J., & Hokao, K. (2005). Applying the methods of image evaluation and spatial analysis to study the sound environment of urban street areas. *Journal of Environmental Psychology*, Vol. 25, No. 4, pp. 455-466.

Gibson, J. (1986). *The ecological approach to visual perception*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Godøy, R. (2006). Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound*, Vol. 11, No. 2, pp. 149-157.

Gregory, C. (1999). Stereotypes and personalities of musicians. *Journal of Psychology*, Vol. 133, No. 1, pp. 104-114.

Grey, J. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, Vol. 61, No. 5, pp. 1270-1277.

Hajda, J. (2007). The effect of dynamic acoustical features on musical timbre. In J. W. Beauchamp (Ed.), *Analysis, synthesis, and perception of musical sounds: The sound of music.* New York: Springer.

Held, C. & Cress, U. (2009). Learning by foraging: The impact of social tags on knowledge acquisition. In: *Learning in the Synergy of Multiple Disciplines: 4th European Conference on Technology Enhanced Learning (EC-TEL*). Nice, France: Springer, pp. 254-266.

Helmholtz, H. (1954). *On the sensations of tone as a physiological basis for the theory of music*. New York: Dover Publications.

Hesse, F. (2009). Use and acquisition of externalized knowledge. In: *Learning in the Synergy of Multiple Disciplines: 4th European Conference on Technology Enhanced Learning (EC-TEL).* Nice, France. Springer, p. 5.

Hinton, L., Nichols, J., & Ohala, J. J. (Eds.). (1995). *Sound symbolism.* Cambridge: Cambridge University Press.

Huron, D. (2001). Toward a theory of timbre. Paper presented at the 12th Annual Conference of Music Theory Midwest. Abstract retrieved September 6, 2010, from http://www.musiccog.ohio-state.edu/Huron/Talks/SMTmidwest.2001/talk.01.html

Kaipainen, M. (1996). Prospects for ecomusicology: Inner and outer loops of the musical mind-environment system. In: P. Pylkkänen, P. Pylkkö, & A. Hautamäki (Eds.), *Brain, Mind and Physics.* Netherlands: IOS Press, pp. 266-277.

Krumhansl, C., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 18, No. 3, pp. 739-751.

Laurier, C., Sordo, M., Serrà, J., & Herrera, P. (2009). Music mood representations from social tags. In: *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR),* Kobe, Japan, pp. 381-386.

Lamere, P. (2008). Social tagging and music information retrieval. *Journal of New Music Research*, Vol. 37, No. 2, pp. 101-114.

Leman, M. (1995). *Music and schema theory: Cognitive foundations of systematic musicology*. Berlin, Heidelberg: Springer.

Leman, M. (2007). *Embodied music cognition and mediation technology*. Cambridge: MIT Press.

Maturana, H. (1988). Ontología del conversar. *Revista Terapia Psicológica*, Vol. 10, pp. 15-23.

Maturana, H. (2002). Autopoiesis, structural coupling and cognition: A history of these and other notions in the biology of cognition. *Cybernetics & Human Knowing*, Vol. 9, No. 3-4, pp. 5-34.

McAdams, S., Winsberg, S., Donnadieu, S., Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, Vol. 58, No. 3, pp. 177-192.

McMullen, E., & Saffran, J. (2004). Music and language: A developmental comparison. *Music Perception*, Vol. 21, No. 3, pp. 289-311.

Mehler, J., Bertoncini, J., Barriere, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, Vol. 7, No. 5, pp. 491-497.

Nussbaum, C. (2007). *The musical representation: Meaning, ontology, and emotion*. Cambridge: MIT Press.

Pachet, F. & Cazaly, D. (2000). A taxonomy of musical genres. In: *Proc. Content-Based Multimedia Information Access* (RIAO), pp. 1238-1245.

Patel, A., & Iversen, J. (2003). Acoustic and perceptual comparison of speech and drum sounds in the north indian tabla tradition: An empirical study of sound symbolism. In: M. J. Solé, D. Recasens, & J. Romero (Eds.), In *Proceedings of the 15th international congress of phonetic sciences (ICPhS).* Barcelona, Spain: ICPhS, pp. *925-928.*

Popper, K., & Eccles, J. (1984). *The self and its brain*. New York, USA: Routledge.

Rentfrow, P., & Gosling, S. (2003). The do re mi's of everyday life: Examining the structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, Vol. 84, pp. 1236-56.

Rentfrow, P., & Gosling, S. (2006). Message in a ballad: The role of music preferences in interpersonal perception. *Psychological Science*, Vol. 17, No. 3, pp. 236-242.

Rentfrow, P., & Gosling, S. (2007). The content and validity of music-genre stereotypes among college students. *Psychology of Music*, Vol. 35, No. 2, pp. 306-326.

Reybrouck, M. (2005). A biosemiotic and ecological approach to music cognition: Event perception between auditory listening and cognitive economy. *Axiomathes*, Vol. 15, No. 2, pp. 229-266.

Sandell, G. (1998). Macrotimbre: Contribution of attack, steady state, and verbal attributes. *The Journal of the Acoustical Society of America*, Vol. 103, p. 2966.

Sandell, G., & Chronopoulos, M. (1997). Perceptual constancy of musical instrument timbres; generalizing timbre knowledge across registers. In: A. Gabrielsson (Ed.), *Proceedings of the third triennial ESCOM conference*. Uppsala: Uppsala Unversity, pp. 222-227.

Sarkar, M., Lan, C., Diaz, J., & Vercoe, B. (2009). The effect of musical experience on describing sounds with everyday words. *The Journal of the Acoustical Society of America*, Vol. 125, p. 2683.

Schafer, R. M. (1977). *The tuning of the world*. Toronto, Canada: McClelland & Stewart.

Seashore, C. (1967). *Psychology of music*. New York: Dover Publications.

Sethares, W. (1999). *Tuning, timbre, spectrum, scale*. Great Britain: Springer.

Terasawa, H., Slaney, M., & Berger, J. (2006). Determining the euclidean distance between two steady state sounds. In: M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the 9th international conference on music perception & cognition, international conference on music perception &*

cognition. Bologna, Italy: ICMPC-ESCOM. Retrieved September 6, 2010, from https://www-ccrma.stanford.edu/~hiroko/timbre/Terasawa2006_ICMPC9.pdf

Tye, M. (2008). Dualism. In: *Stanford Encyclopedia of Philosophy*. Stanford, CA: The Metaphysics Research Lab. Retrieved October 1, 2008, from http://plato.stanford.edu/entries/dualism/

Wessel, D. (1979). Timbre space as a musical control structure. *Computer Music Journal*, Vol. 3, No. 2, pp. 45-52.

PII

TIMBRE AND AFFECT DIMENSIONS: EVIDENCE FROM AFFECT AND SIMILARITY RATINGS AND ACOUSTIC CORRELATES OF ISOLATED INSTRUMENT SOUNDS

by

Eerola, T., Ferrer, R. & Alluri, V. (in press)

Music Perception

PIII

TIMBRAL QUALITIES OF SEMANTIC STRUCTURES OF MUSIC

by

Ferrer, R. & Eerola, T. 2010

In Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR), 571–576, Utrecht, Netherlands

TIMBRAL QUALITIES OF SEMANTIC STRUCTURES OF MUSIC

Rafael Ferrer and Tuomas Eerola
Finnish Centre of Excellence in Interdisciplinary Music Research
rafael.ferrer-flores@jyu.fi; tuomas.eerola@jyu.fi

ABSTRACT

The rapid expansion of social media in music has provided the field with impressive datasets that offer insights into the semantic structures underlying everyday uses and classification of music. We hypothesize that the organization of these structures are rather directly linked with the ”qualia” of the music as sound. To explore the ways in which these structures are connected with the qualities of sounds, a semantic space was extracted from a large collection of musical tags with latent semantic and cluster analysis. The perceptual and musical properties of 19 clusters were investigated by a similarity rating task that used spliced musical excerpts representing each cluster. The resulting perceptual space denoting the clusters correlated high with selected acoustical features extracted from the stimuli. The first dimension related to the high-frequency energy content, the second to the regularity of the spectrum, and the third to the fluctuations within the spectrum. These findings imply that meaningful organization of music may be derived from low-level descriptions of the excerpts. Novel links with the functions of music embedded into the tagging information included within the social media are proposed.

1. INTRODUCTION

Attempts to craft a bridge between acoustic features and the subjective sensation they provoke [3] have usually started with concepts describing instrument sounds, using adjectives or bipolar scales (e.g., bright-dark, static-dynamic) and matching these with acoustic descriptors (such as shape of the envelope and energy distribution) [11, 20].

In this study, we present a purely bottom-up approach to the conceptual mapping between sound qualities and emerging meanings. We utilized social media to obtain a wide sample of music and extract an underlying semantic structure of this sample. Next, we evaluated the validity of the obtained mapping by investigating the acoustic features underlying the semantic structures. This was done by an analyzing of the examples representing the semantic space, and by having participants to rate the similarity of random spliced sound examples representing the semantic space.

Social tagging is an activity, where descriptive verbal characterizations are given to items of interest, such as songs, images, or links as a part of the normal use of the popular online services. Tags can be considered as semantic representations of abstract concepts created essentially for mnemonic purposes and used typically to organize items [14]. Tagging music is not a novel idea, as any labeling scheme such as musical genres may be considered as tags themselves, but in recent years in the context of social networks, tagging has acquired a new relevance and meaning [1].

Despite all the possibilities offered by large databases containing tags, a central problem remains on how to derive an ontology from them [19]. Starting with the assumption of an underlying structure existing in an apparently unstructured set, we consider a sample of tags to extract a semantic structure, explained next.

2. ANALYSIS OF TAGS

2.1 Material

A collection of 6372 songs [7] representing 15 musical genres (Alternative, Folk, Finnish Iskelmä, Pop, World, Blues, Gospel, Jazz, Rock, Classical, Heavy, Soul, Electronic, Hip-Hop, Soundtrack) served as the initial database of music. Musical genres were used in establishing the sample in order to maximize musical variety in the collection and to be compatible with a host of music preference studies (e.g., [6, 22]) that have provided lists of 13 to 15 broad musical genres relevant for most Western adult listeners. The tags related to the songs in this collection were retrieved from an online music service (*last.fm*[1]) with a dedicated API (Application programming interface) named *Pylast*[2].

2.2 Description of the corpus

The retrieved *corpus* consists of 5,825 lists of tags (mean length of 62.27 tags), each list (*document* in this context) is associated with a piece of music. The number of times each tag had been used in the system until the time of the retrieval was also obtained, representing a measure of “popularity”.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.
© 2010 International Society for Music Information Retrieval.

[1] http://www.last.fm

[2] http://code.google.com/p/pylast/

In total, the corpus contains 362,732 tags, from which 77,537 are distinct. Each tag is formed by one or more words (M=2.48, SD=1.86), a small proportion of the distinct tags in the corpus contain long expressions (e.g. 6% of the distinct tags are formed by 5 words or more). In this study a tag is considered as a unit representing an element of the *vocabulary*, disregarding the number of words that compose it. Treating tags as *collocations* (i.e. frequent juxtaposition of words) shifts the focus from data processing to concept processing [2], also allowing the tags to function as conceptual expressions [23] instead of words or phrases.

2.3 Lexical layers of the vocabulary

Preprocessing is necessary in any text mining application because retrieved data does not follow any particular set of rules, and there are not standard steps to follow [13].

Three filtering rules where applied to the corpus in the quantitative domain. First, *hapax legomena* (i.e. tags used only once in the corpus), are removed under the rationale of discarding unrelated data. To capture the most prevalent and relevant tags, a second filter uses the associated popularity measure of each tag to eliminate the tags below the mean popularity index of the vocabulary. The third step eliminates tags with three or more words to prune short sentence-like descriptions from the corpus. The subset resulting from such reductions represents 46.6% of the corpus (N=169,052, Vocabulary=2,029 tags).

At this point, data has been de-noised but for the extraction of a meaningful semantic ontology from the tags, a semantic analysis and qualitative filtering is necessary. To categorize the tags at a functional level [24] (e.g. musicological and lexicological), an analysis was performed by using the Brown Corpus [9] as parts-of-speech (POS) tagger, Wordnet database [8] for word sense disambiguation, and Urban Dictionary online[3] and Last.fm database for general reference. Tags are looked-up in these sources and the selection of a category is decided by reviewing each case. The criteria applied in this process favors categories closely related to music, such as genre, artist, instrument, form and company, then adjectives, and finally other types. For instance, "Acid" is a noun but it is also a term extensively used to describe certain musical genres, so it was classified according to its musical function. Proposed categories, percentage of the vocabulary, definition and examples are shown in Table 1. The resulting layers were used to make a finer discrimination of the tags to uncover the semantic structure. Since one of the main motivations of this project was to obtain prototypical timbral descriptions, we focused on tags related to adjectives, nouns, instruments, temporal and verbs.

2.4 Semantic structure

Tag structure (or folksonomy) is obtained by using *latent semantic analysis* (LSA) as a framework [5], a method that has been used before in the domain of musical tags [17, 18]. In this study, detection of semantic structure has three stages: 1) construction of a *Term-Document Matrix*, 2) calculation of similarity coefficients, and 3) cluster analysis. First, a Term-Document Matrix $\mathbf{X} = \{x_{ij}\}$ is constructed. Where each song i, corresponds to a "Document" and each unique tag (or item of the vocabulary) j, to a "Term". The result is a binary matrix $\mathbf{X}(0,1)$ containing information about the presence or absence of a particular tag to describe a given song. Second, a similarity matrix $n \times n$ $\mathbf{D}$ with elements d_{ij} where $d_{ii} = 0$ for all i, is created by computing similarity indexes between tag vectors x_{i*j} of $\mathbf{X}$ with:

$$d_{ij} = \frac{ad}{\sqrt{(a+b)(a+c)(d+b)(d+c)}} \tag{1}$$

where *a* is the number of (1,1) matches, *b* for (1,0), *c* for (0,1) and *d* for (0,0).

There are several methods to compute similarity coefficients between binary vectors (c.f., [10]). This coefficient was selected because of its *symmetric* quality, which considers the double absence (0,0) as important as (1,1), that presumably has positive impact on ecologic applications [10]. A hierarchical clustering algorithm was used to transform the similarity matrix into a sequence of nested partitions. The method used in the hierarchical clustering was Ward's minimum variance, to find compact, spherical clusters [21] and because it has demonstrated its proficiency in comparison to other methods [12].

After obtaining a hierarchical structure, the clusters are derived from the resulting dendrogram by "pruning" the branches with an algorithm that uses a partitioning around medioids (PAM) clustering method in combination with the height of the branches [15]. Figure 1 shows a two dimensional projection (obtained with multidimensional scaling) of the similarity matrix used in the hierarchical clustering. Each dot represents a tag, and the numbers show the centers of their corresponding clusters. Each number is enclosed in a circle that shows the relative size of the cluster in terms of the number of tags contained in it. A more detailed reference on the content of the clusters can be consulted in Table 2.

2.5 Ranking of musical examples in the clusters

In order to explore any acoustic or musical aspects of the clusters, we need to link the clusters with the specific songs represented by the tags. For this, a $m \times n$ *Term Document Matrix* (TDM) $\mathbf{X} = \{x_{ij}\}$ is constructed, where lists of tags attributed to a particular song are represented as m, and preselected tags as n. A list of tags is a finite set $\{1, ..., k\}$, where $1 \leq k \leq 96$. Each element of the matrix contains a value of the normalized rank of a tag if found on a list, and it is defined by:

$$x_{ij} = \left(\frac{r_k}{k}\right)^{-1} \tag{2}$$

Where r_k is the cardinal rank of the tag j if found in i, and k is the total length of the list. To obtain a cluster profile,

[3] http://www.urbandictionary.com

Categories	%	Definition	Examples
Genre	36.72%	Musical genre or style	Rock, Alternative, Pop
Adjective	12.17%	General category of adjectives	Beautiful, Mellow, Awesome
Noun	9.41%	General category of nouns	Love, Melancholy, Memories
Artist	8.67%	Artists or group names	Coldplay, Radiohead, Queen
Locale	8.03%	Geographic situation or locality	British, American, Finnish
Personal	6.80%	Words used to manage personal collections	Seen Live, Favourites, My Radio
Instrument	4.83%	Sound source	Female vocalists, Piano, Guitar
Unknown	3.79%	Unclassifiable gibberish	aitch, prda, <3
Temporal	2.41%	Temporal circumstance	80's, 2000, Late Romantic
Form	2.22%	Musical form or compositional technique	Ballad, Cover, Fusion
Company	1.72%	Record label, radio station, etc.	Motown, Guitar Hero, Disney
Verb	1.63%	General category of verbs	Chillout, Relax, Wake up
Content	1.03%	Emphasis in the message or literary content	Political, Great lyrics, Love song
Expression	0.54%	Exclamations	Wow, Yeah, lol

Table 1. Main categories of tags, their prevalence, definition and examples.

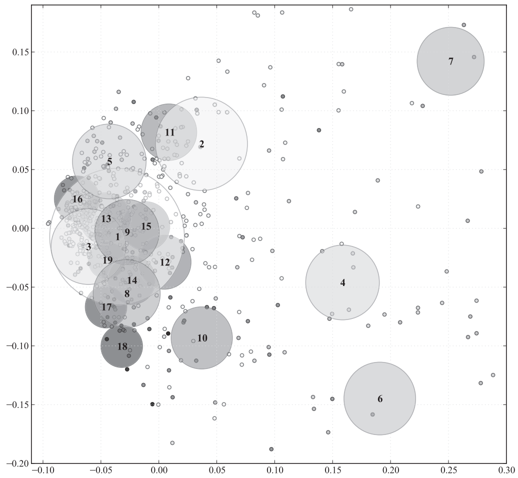


Figure 1. 19 clusters obtained with hierarchical clustering and hybrid pruning.

mean rank of the tag across the TDM is calculated with:

$$\bar{r}_j = \frac{\sum_{i=1}^{m} x_{ij}}{m} \quad (3)$$

Thus the cluster profile or mean ranks vector is defined as:

$$\mathbf{p}_l = \bar{r}_{j \in C_l} \quad (4)$$

C_l denotes a given cluster l for $1 \leq l \leq 19$ (optimal number of clusters for this dataset), and $\mathbf{p}$ is a vector $\{5, ..., k\}$, where $5 \leq k \leq 334$.

Last step aims to obtain ranked lists of songs ordered in terms of its closeness to each cluster profile. This is carried out by calculating the euclidean distance between each song rank vector $x_{i,j \in C_l}$ and the cluster profile $\mathbf{p}_l$:

$$d_i = \sqrt{\sum_{j \in C_l} (x_{ij} - \mathbf{p}_l)^2} \quad (5)$$

The examples of the results can be seen in Table 2, where top artists of each cluster are displayed below central tags of the cluster.

3. EXPERIMENT

In order to explore whether the obtained clusters are perceptually meaningful and to further understand what kinds of acoustic and musical attributes they consist of, empirical data unrelated to the existing structures about the clusters is needed. A similarity rating experiment was designed to assess the timbral qualities of songs pertaining to each of the clusters. We chose to emphasize the low-level, non-structural qualities of music since we wanted to minimize the confounding factors caused by recognition of songs, artists and the subsequent associations with these as well as the lyrical contents of the music. To this end, the stimuli for the experiment consisted of semi-randomly spliced, brief excerpts, explained in detail below.

3.1 Experiment details

3.1.1 Stimuli

Initially, 5-second audio samples were taken from a random middle part (25% after the beginning and 25% before the end) of the 25 top ranked songs (see ranking procedure in section 2.5) from each cluster. For each sample, the temporal position of notes onsets were estimated based on *spectral flux* using MIRToolbox [16]. The highest onset was selected as a reference point from which slices of random length ($150ms \leq t \leq 250ms$) were taken from $10ms$ before the peak onset of each sample, then equalized in loudness, and finally mixed together using a fade in-out of $50ms$ with an overlap window of $100ms$ This resulted in 19 excerpts (each representing a cluster) of variable length, that were finally trimmed to $1750ms$, with a fade in-out of $100ms$ To prepare these 19 excerpts for a similarity rating, the 171 paired combinations were mixed with a silence of $600ms$. between them.

3.1.2 Participants

12 females and 9 males (age M=26.8, SD=4.15) participated to the experiment. 9 of them possessed least one year of musical training. 12 reported listening to music attentively between one and 10 hours per week.

Cluster ID	*Tags proximate to cluster centroids*	Top artists in the cluster
1	*Energetic, Female vocal, Powerful, Hot, Sex*	Amy Adams, Fred Astaire, Kelly Clarkson
2	*Dreamy, Chill out, Haunting, Sleep, Moody*	Nick Drake, Radiohead, Massive Attack
3	*Sardonic, Sarcastic, Cynical, Humorous, Funny*	Alabama 3, Yann Tiersen, Tom Waits
4	*Awesome, Amazing, Male vocalist, Loved, Great*	Guns N' Roses, U2, Metallica
5	*Composer, Cello, Piano, Cello rock, Violin*	Camille Saint-Saëns, Tarja Turunen, Franz Schubert
6	*Female vocalist, Female vocalists, Female, 00s, Sexy*	Fergie, Lily Allen, Amy Winehouse
7	*Mellow, Beautiful, Chillout, Chill, Sad*	Katie Melua, Phil Collins, Coldplay
8	*Hard, Angry, Loud, Aggressive, Rock out*	System of a Down, Black Sabbath, Metallica
9	*60s, 70s, Guitar virtuoso, Sixties, Guitar solo*	Simon & Garfunkel, Janis Joplin, The Four Tops
10	*Feelgood, Summer, Feel good, Cheerful, Gute laune*	Mika, Goo Goo Dolls, Shekinah Glory Ministry
11	*Autumnal, Wistful, Intimate, Sophisticated, Reflective*	Soulsavers, Feist, Leonard Cohen
12	*High school, 90's, 1990s, 1995, 1996*	Fool's Garden, The Cardigans, No Doubt
13	*50s, Saxophone, Trumpet, Tenor sax, Sax*	Miles Davis, Thelonious Monk, Charles Mingus
14	*1980s, 80's, Eighties, 80er, Voci maschili*	Ray Parker Jr., Alphaville, Michael Jackson
15	*Affirming, Lyricism, Life song, Vocalization*	Lisa Stansfield, KT Tunstall, Katie Melua
16	*Choral, A capella, Acapella, Choir, A cappella*	Mediæval Bæbes, Alison Krauss, Blackmore's Night
17	*Voce femminile, Femmina, Voci femminili, Femmine*	Avril Lavigne, The Cranberries, Diana Krall
18	*Tangy, Coy, Sleek, Attitude, Flirty*	Kylie Minogue, Ace of Base, Solange
19	*Rousing, Exuberant, Confident, Playful, Passionate*	James Brown, Does It Offend You, Yeah?, Tchaikovsky

Table 2. Most representative tags and typical artists of each of the 19 clusters.

3.1.3 Procedure

Participants were presented with pairs of sound excerpts in random order using a computer interface and high-quality headphones. Their task was to rate the similarity of sounds on a 9-level Likert scale, whose extremes were labeled as *dissimilar* and *similar*. Before the actual experimental trials, they were given instructions and practice trials to familiarize themselves with the task.

3.1.4 Audio features

To explore the acoustic and musical features underlying the perceptual similarities of the clusters, 41 audio features (listed on Table 3) were extracted from each spliced stimuli using MIR toolbox [16]. The choice of features was restricted to those which would be applicable to spliced examples and would not require high-level feature analysis such as structural repetition or tonality. The extraction was carried out using frame-based approach with $50ms$ analysis frame using 50% overlap.

Category	No.	Feature
Dynamics	1-2	RMS energy
	3-4	Attack time (M, SD)
Rhythm	5-6	Fluctuation peak pos. (M, SD)
	7	Fluctuation centroid (M, SD)
Pitch	8-9	Pitch (M, SD)
	10-11	Chromagram (unwr.) centr. (M, SD)
Harmony	12	Entropy (oct. collap. spectr.) (M)
	13	Roughness (M)
	14	Inharmonicity (M, SD)
Timbre	15-16	Brightness (cut-off 110 Hz) (M, SD)
	17-18	Spectral centroid (M, SD)
	19-20	Zerocross (M, SD)
	20-21	Spread (M)
	22	Spectral entropy (M)
	23	Spectral flux (M)
	24	Flatness (M)
	25	Kurtosis (M)
	26-27	Regularity (M, SD)
	28-29	1st MFCC (M, SD)
	⋮	⋮
	30-41	7th MFCC (M, SD)

Table 3. List of extracted audio features (M= mean, SD= standard deviation)

3.2 Results

Highly consistent pattern of similarities between the 21 participants were obtained (Cronbach $\alpha = 0.94$). For this reason, a mean similarity matrix of the individual ratings was subjected to metric multidimensional scaling (MDS) analysis based on stress minimization by means of majorization (SMACOF) [4]. This yielded adequate low - dimensional projections of the data, from which we focus on 2 - dimensional (stress=0.065) and 3 - dimensional (stress=0.027) solutions.

The organization of the clusters (represented with sliced samples) illustrates a clear organization in terms of the semantic qualities of the clusters (see Figure 2), showing the *Awesome* and *Hard* examples on the left uppermost corner, and the semantically distant, *Autumnal* and *Dreamy* in the lower right-hand corner.

To investigate the perceived organization of the semantic clusters in terms of the acoustic qualities, the 3 dimensions were correlated with the extracted audio features. Highly significant correlations, top five shown in Table 4, were observed for dimensions 1 and 2. We may interpret these correlations in terms of the qualities of the sound spectrum: The first dimension is related to the distribution of energy along the frequency (spectral centroid, flatness, brightness, MFCC1, etc.), where the items in the MDS solution are arranged from the high-frequency energy content in the left to the prevalence of low-frequency energy content in the right. The second dimension may be interpreted as the periodic organization of the spectrum, i.e., whether the spectrum is harmonic (roughness,skewness, spread and fluctuation centroid). The clusters represented by the items in the lower part of the MDS solution possess clearer organization of the spectrum in comparison with the the items high on the MDS solution. The third dimension seem to be related the temporal fluctuation of the spectrum (MFCC6 [SD], Fluctuation position [M], MFCC22 [M]).

Dimension 1			**Dimension 2**			**Dimension 3**		
Acoustic feature	***r***		**Acoustic feature**	***r***		**Acoustic feature**	***r***	
MFCC 1 (M)	0.94	***	Fluctuation centroid (M)	-0.72	***	MFCC 6 (SD)	0.51	*
Flatness (M)	-0.86	***	Roughness (M)	0.68	**	Fluctuation position (M)	-0.50	*
Centroid (M)	-0.83	***	Skewness (M)	0.67	**	MFCC 2 (M)	-0.46	*
Brightness (M)	-0.81	***	Spread (M)	-0.65	**	Fluctuation peak (M)	0.45	
Spectral entropy (M)	-0.80	***	Kurtosis (M)	0.57	*	Irregularity (SD)	0.44	
$* * * = p < .001$, $** = p < .01$, $* = p < .05$								

Table 4. Correlations between the dimensions of the multidimensional scaling solution and acoustic descriptors.

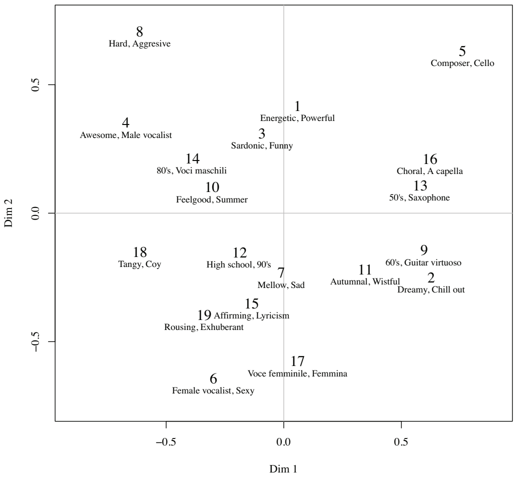


Figure 2. Dimensions 1 and 2 of the MDS with behavioural responses and associated tags

3.3 Discussion

In sum, when brief and spliced excerpts taken from the clusters representing semantic structures of the music descriptions are presented to listeners, they are able to form coherent distances between them. An acoustic analysis of the excerpts was used to label the dimensions embedded in the cluster similarities. This analysis showed clear correlations between the dimensional and timbral qualities of music. However, it should be emphasized that the high relevance of many timbral features is only natural since the timbral characteristics of the excerpts were preserved and structural aspects were masked by the semi-random splicing.

We are careful in not taking these early results to mean literally that the semantic structure of the initial sample would be explainable by means of the same timbral features. This is of course another question which is easily empirically approached using feature extraction of the typical examples representing each cluster and either classify the clusters based on features, or predict the coordinates of the clusters within a low dimensional space by means of regression using a larger set of acoustic features (including those that are relevant for full excerpts such as tonality and structure). However, we are positively surprised at the level of coherence from the part of the listener ratings and their explanations in terms of the acoustic features despite the limitations we imposed on the setting (i.e. discarding tags connected with musical genres), splicing and having a large number of clusters to test. Our intention is to follow this analysis with more rigorous selection of acoustic features (PCA and other data reduction techniques) and use multiple regression to assess whether linear combinations of the features would be necessary for explaining the perceptual dimensions.

4. CONCLUSIONS

The present work provided a bottom-up approach to semantic qualities of music descriptions, which capitalized social media, natural language processing, similarity ratings and acoustic analysis. Semantic structures of music descriptions have been extracted from the social media previously [18] but the main difference here was the careful filtering of such data. We used natural language processing to focus on categories of tags that are meaningful but do not afford immediate categorization of music in a way that, for example, musical genre does.

Although considerable effort was spent on finding the optimal way of teasing out reliable and robust structures of the tag occurrences using cluster analysis, several other techniques and parameters within clustering could also have been employed. We realize that other techniques would probably have led to different structures but it is an open empirical question whether the connections between the similarities of the tested items and their acoustic features would have been entirely different. A natural continuation of the current study would be to predict the typical examples of the clusters with the acoustic features by using either classification algorithms or mapping of the cluster locations within a low dimensional space using correlation and multiple regression. However, the issue at stake here was the connection of timbral qualities with semantic structures.

The implications of the present findings are related to several open issues. The first one is the question whether structural aspects of music are required in explaining the semantic structures or whether the low-level, timbral characteristics are sufficient, as was indicated by the present findings. Secondly, what new semantic layers (as indicated by categories of tags) can be meaningfully connected with the acoustic properties of the music? Finally, if the timbral characteristics are indeed strongly connected with such semantic layers as *adjectives*, *nouns* and *verbs*, do these arise by means of learning and associations, or are the underlying regularities connected with emotional, functional and gestural cues of the sounds?

5. REFERENCES

[1] J.J. Aucouturier and E. Pampalk. Introduction-from genres to tags: A little epistemology of music information retrieval research. *Journal of New Music Research*, 37(2):87–92, 2008.

[2] J. Brank, M. Grobelnik, and D. Mladenic. Automatic evaluation of ontologies. In Anne Kao and Stephen R.Poteet, editors, *Natural Language Processing and Text Mining*. Springer, USA, 2007.

[3] O. Celma and X. Serra. Foafing the music: Bridging the semantic gap in music recommendation. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):250–256, 2008.

[4] J. de Leeuw and P. Mair. Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, 31(3):1–30, 2009.

[5] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407, 1990.

[6] M.J. Delsing, T.F. ter Bogt, R.C. Engels, and W.H. Meeus. Adolescents music preferences and personality characteristics. *European Journal of Personality*, 22(2):109–130, 2008.

[7] T. Eerola and R. Ferrer. Setting the standards: Normative data on audio-based musical features for musical genres. In *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music, ESCOM*, 2009.

[8] Christiane Fellbaum, editor. *WordNet: An electronic lexical database.* Language, speech, and communication. MIT Press, Cambridge, Mass, 1998.

[9] W.N. Francis and H. Kucera. *Brown corpus. A Standard Corpus of Present-Day Edited American English, for use with Digital Computers*. Department of Linguistics, Brown University, Providence, Rhode Island, USA, 1979.

[10] J.C. Gower and P. Legendre. Metric and euclidean properties of dissimilarity coefficients. *Journal of classification*, 3(1):5–48, 1986.

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

[12] A.K. Jain and R.C. Dubes. *Algorithms for clustering data*. Prentice Hall, Englewood Cliffs, NJ, 1988.

[13] Anne Kao and Stephen R. Poteet, editors. *Natural Language Processing and Text Mining*. Springer Verlag, 2006.

[14] P. Lamere. Social tagging and music information retrieval. *Journal of New Music Research*, 37(2):101–114, 2008.

[15] P. Langfelder, B. Zhang, and S. Horvath. *dynamicTreeCut: Methods for detection of clusters in hierarchical clustering dendrograms.*, 2009. R package version 1.20.

[16] O. Lartillot, P. Toiviainen, and T. Eerola. A matlab toolbox for music information retrieval. *Data Aalysis, Machine Learning and Applications*, pages 261–8, 2008.

[17] C. Laurier, M. Sordo, J. Serra, and P. Herrera. Music mood representation from social tags. In *Proceedings of the 10th International Society for Music Information Conference, Kobe, Japan*, 2009.

[18] M. Levy and M. Sandler. Learning latent semantic models for music from social tags. *Journal of New Music Research*, 37(2):137–150, 2008.

[19] H. Lin, J. Davis, and Y. Zhou. An integrated approach to extracting ontological structures from folksonomies. In *Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications*, page 668. Springer, 2009.

[20] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff. Perceptual scaling of synthesized musical timbres: Common dimensions, specificities and latent subject classes. *Psychological Research*, 58(3):177–192, 1995.

[21] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0.

[22] P.J. Rentfrow and S.D. Gosling. Message in a ballad: the role of music preferences in interpersonal perception. *Psychol Sci*, 17(3):236–242, 2006.

[23] J.M. Siskind. Learning word-to-meaning mappings. *Models of language acquisition: inductive and deductive approaches*, pages 121–153, 2000.

[24] B. Zhang, Q. Xiang, H. Lu, J. Shen, and Y. Wang. Comprehensive query-dependent fusion using regression-on-folksonomies: a case study of multimodal music search. In *Proceedings of the seventeen ACM international conference on Multimedia*, pages 213–222. ACM, 2009.

PIV

SEMANTIC STRUCTURES OF TIMBRE EMERGING FROM SOCIAL AND ACOUSTIC DESCRIPTIONS OF MUSIC

by

Ferrer, R. & Eerola, T. 2011

EURASIP Journal on Audio, Speech, and Music Processing, 2011(1):11

Ferrer and Eerola *EURASIP Journal on Audio, Speech, and Music Processing* 2011, **2011**:11
http://asmp.eurasipjournals.com/content/2011/1/11

EURASIP Journal on
Audio, Speech, and Music Processing
a SpringerOpen Journal

RESEARCH Open Access

Semantic structures of timbre emerging from social and acoustic descriptions of music

Rafael Ferrer* and Tuomas Eerola

Abstract

The perceptual attributes of timbre have inspired a considerable amount of multidisciplinary research, but because of the complexity of the phenomena, the approach has traditionally been confined to laboratory conditions, much to the detriment of its ecological validity. In this study, we present a purely bottom-up approach for mapping the concepts that emerge from sound qualities. A social media (http://www.last.fm) is used to obtain a wide sample of verbal descriptions of music (in the form of tags) that go beyond the commonly studied concept of genre, and from this the underlying semantic structure of this sample is extracted. The structure that is thereby obtained is then evaluated through a careful investigation of the acoustic features that characterize it. The results outline the degree to which such structures in music (connected to affects, instrumentation and performance characteristics) have particular timbral characteristics. Samples representing these semantic structures were then submitted to a similarity rating experiment to validate the findings. The outcome of this experiment strengthened the discovered links between the semantic structures and their perceived timbral qualities. The findings of both the computational and behavioural parts of the experiment imply that it is therefore possible to derive useful and meaningful structures from free verbal descriptions of music, that transcend musical genres, and that such descriptions can be linked to a set of acoustic features. This approach not only provides insights into the definition of timbre from an ecological perspective, but could also be implemented to develop applications in music information research that organize music collections according to both semantic and sound qualities.

Keywords: timbre, natural language processing, vector-based semantic analysis, music information retrieval, social media

1 Introduction

In this study, we have taken a purely bottom-up approach for mapping sound qualities to the conceptual meanings that emerge. We have used a social media (http://www.last.fm) for obtaining as wide a sample of music as possible, together with the free verbal descriptions made of music in this sample, to determine an underlying semantic structure. We then empirically evaluated the validity of the structure obtained, by investigating the acoustic features that corresponded to the semantic categories that had emerged. This was done through an experiment where participants were asked to rate the perceived similarity between acoustic examples of prototypical semantic categories. In this way, we were attempting to recover the correspondences between semantic and acoustic features that are ecologically relevant in the perceptual domain. This aim also meant that the study was designed to be more exploratory than confirmative. We applied the appropriate and recommended techniques for clustering, acoustic feature extraction and comparisons of similarities; but this was only after assessing the alternatives. But, the main focus of this study has been to demonstrate the elusive link that exists between the semantic, perceptual and physical properties of timbre.

1.1 The perception of timbre

Even short bursts of sound are enough to evoke mental imagery, memories and emotions, and thus provoke immediate reactions, such as the sensation of pleasure or fear. Attempts to craft a bridge between such acoustic features and the subjective sensations they provoke [1] have usually started with describing instrument

* Correspondence: rafael.ferrer-flores@jyu.fi
Finnish Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Jyväskylä, Finland

© 2011 Ferrer and Eerola; licensee Springer. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

sounds via adjectives on a bipolar scale (e.g. bright-dark, static-dynamic) and matching these with more precise acoustic descriptors (such as the *envelope* shape, or high-frequency energy content) [2,3]. However, it has been difficult to compare these studies when such different patterns between acoustic features and listeners' evaluations have emerged [4]. These differences may be attributed to the cross-study variations in context effects, as well as the choice of terms, stimuli and rating scales used. It has also been challenging to link the findings of such studies to the context of actual music [5], when one considers that real music consists of a complex combination of sound. A promising approach has been obtained to evaluate short excerpts of recorded music with a combination of bipolar scales and acoustic analysis [6]. However, even this approach may well omit certain sounds and concepts that are important for the majority of people, since the music and scales have usually been chosen by the researcher, not the listeners.

1.2 Social tagging

Social tagging is a way of labelling items of interest, such as songs, images or links as a part of the normal use of popular online services, so that the tags then become a form of categorization in themselves. Tags are usually semantic representations of abstract concepts created essentially for mnemonic purposes and used typically to organize items [7,8]. Within the theory of *information foraging* [9], tagging behaviour is one example of a transition from internalized to externalized forms of knowledge where, using *transactional memory*, people no longer have to know everything, but can use other people's knowledge [10]. What is most evident in the social context is that what escapes one individual's perception can be captured by another, thus transforming tags into memory or knowledge cues for the undisclosed transaction [11].

Social tags are usually thought to have an underlying *ontology* [12] defined simply by people interested in the matter, but with no institutional or uniform direction. These characteristics make the vocabulary and implicit relations among the terms considerably richer and more complex than in formal taxonomies where a hierarchical structure and set of rules are designed *apriori* (cf. folksonomy versus taxonomy in [13]). When comparing ontologies based on social tagging and the classification by experts, it is presumed that there is an underlying organization of musical knowledge hidden among the tags. But, as raised by Celma and Serra [1]), this should perhaps not to be taken for granted. For this reason, Section 2 addresses the uncovering of an ontology from the tags [14] in an unsupervised form, to investigate whether such an ontology is not an imposed construction. Because a latent structure has been assumed, we use a technique called *vector-based semantic analysis*, which is a generalization of *Latent Semantic Analysis* [15] and similar to the methods used in latent semantic mapping [16] and latent perceptual indexing [17]. Thus, although some of the terminology is borrowed from these areas, our method is also different in several crucial respects. While ours is designed to explore emergent structures in the semantic space (i.e. clusters of musical descriptions), the other methods are designed primarily to improve information retrieval by reducing the dimensionality of the space [18]. In our method, the reduction is not part of the analytical step, but rather implemented as a pre-filtering stage (see Appendix sections A.1 and A.2). The indexing of documents (songs in our case) is also treated separately in Section 2.2 which presents our solution based on the Euclidean distances of clusters profiles in a vector space. The reasons outlined above show that tags, and the structures that can be derived from them, impart crucial cues about how people organize and make sense of their experiences, which in this case is music and in particular its timbre.

2 Emergent structure of timbre from social tags

To find a semantic structure for timbre analysis based on social tags, a sample of music and its associated tags were taken. The tags were then filtered, first in terms of their statistical relevancy and then according to their semantic categories. This filtering left us with five such categories, namely *adjectives*, *nouns*, *instruments*, *temporal references* and *verbs* (see Appendix A for a detailed explanation of the filtering process). Finally, the relations between different combinations of tags were analysed by means of distance calculations and hybrid clustering.

The initial database of music consisted of a collection of 6372 songs [19], from a total of 15 musical genres (with approximately 400 examples for each genre), namely, *Alternative*, *Blues*, *Classical*, *Electronic*, *Folk*, *Gospel*, *Heavy*, *Hip-Hop*, *Iskelmä*, *Jazz*, *Pop*, *Rock*, *Soul*, *Soundtrack* and *World*. Except for some songs in the *Iskelmä* and *World* genres (which were taken from another corpus of music), all of the songs that were eventually chosen in November 2008 from each of these genres could already be found on the musical social network (http://www.last.fm), and they were usually among the "top tracks" for each genre (i.e. the most played songs tagged with that genre on the Internet radio). Although larger sample sizes exist in the literature (e.g. [20,21]), this kind of sample ensured that (1) typicality and diversity were optimized; while (2) the sample could still be carefully examined and manually verified. These musical genres were used to maximize musical variety in the collection, and to ensure that the sample was compatible with a host of other music preference studies (e.g. [22,23]), as these studies have also provided lists of between 13 and 15 broad musical genres that are relevant to most Western adult listeners.

All the tags related to each of the songs in the sample were then retrieved in March 2009 from the millions of users of the mentioned social media using a dedicated *application programming interface* called *Pylast* (http://code.google.com/p/pylast/). As expected, not quite all (91.41%) of the songs in the collection could be found; those not found were probably culturally less familiar songs for the average Western listener (e.g., from the *Iskelmä* and *World* music genres). The retrieved *corpus* now consisted of 5825 lists of tags, with a mean length of 62.27 tags. As each list referred to a particular song, the song's title was also used as a label, and together these were considered as a document in the *Natural Language Processing* (NLP) context (see the preprocessing section of Appendix A). In addition to this textual data, numerical data for each list were obtained that showed the number of times a tag had been used (*index of usage*) up to the point when the tags were retrieved. The corpus contained a total of 362,732 tags, of which 77,537 were distinct and distributed over 323 frequency classes (in other words, the shape of the spectrum of rank frequencies), and this is reported here to illustrate the prevalence of hapax legomena–tags that appear only once in the corpus–in Table 1 (cf. [24]). The tags usually consisted of one or more words (M = 2.48, SD = 1.86), with only a small proportion containing long sentences (6% with five words or more). Previous studies have *tokenized* [20,25] and *stemmed* [26] the tags to remove common words and normalize the data. In this study however, a tag is considered as a holistic unit representing an element of the *vocabulary* (cf. [27]), disregarding the number of words that compose it. Treating tags as *collocations* (i.e. words that are frequently placed together for a combined effect)–rather than as separate, single keywords–has the advantage of keeping the link between the music and its description a priority, rather than the words themselves. This approach shifts the focus from data processing to concept processing [28], where the tags function as conceptual expressions [29] instead of purely words or phrases. Furthermore, this treatment (collocated versus separated) does not distort the underlying nature of the corpus, given that the distribution of the sorted frequencies of the vocabulary still exhibits a Zipfian curve. Such a distribution suggests that tagging behaviour is also governed by the principle of least effort [30], which is an essential underlying feature of human languages in general [27].

Table 1 Frequency classes of tags

Class	*N*	Cumulative (%)
1 (hapaxes)	46 727	60.26
2	11 724	75.38
3	5512	82.49
4	2938	86.28
5	2020	88.89
6	1420	90.72
7	1055	92.08
8	838	93.16
9	674	94.03
10+	4094	100

2.1 Exposing the structure via cluster analysis

The tag structure was obtained via a vector-based semantic analysis that consisted of three stages: (1) the construction of a Term-Document Matrix, (2) the calculation of similarity coefficients and (3) cluster analysis.

The *Term Document Matrix* $\mathbf{X} = \{x_{ij}\}$ was constructed so that each song i corresponded to a "Document" and each unique tag (or item of the vocabulary) j to a "Term". The result was a binary matrix $\mathbf{X}(0, 1)$ containing information about the presence or absence of a particular tag to describe a given song.

$$x_{ij} = \begin{cases} 1, & \text{if } j \in i \\ 0, & \text{if } j \notin i \end{cases} \tag{1}$$

The similarity matrix $n \times n$ $\mathbf{D}$ with elements $d_{ij}$where d_{ii}= 0 was created by computing similarity indices between tag vectors $x_{i*j}$of $\mathbf{X}$ with:

$$d_{ij} = \frac{ad}{\sqrt{(a+b)(a+c)(d+b)(d+c)}} \tag{2}$$

where a is the number of (1,1) matches, b = (1,0), c = (0,1) and d = (0,0). A choice then had to be made between the several methods available to compute similarity coefficients between binary vectors [31]. The coefficient (2) corresponding to the 13th coefficient of Gower and Legendre was selected because of its *symmetric* quality. This effectively means that it considers double absence (0,0) as equally important as double presence (1,1), which is a feature that has been observed to have a positive impact in ecological applications [31]. Using Walesiak and Dudek algorithm [32], we then compared its performance with nine alternative similarity measures used for binary vectors, in conjunction with five distinct clustering methods. The outcome of this comparison was that the coefficient we had originally chosen was indeed best suited to create an intuitive and visually appealing result in terms of *dendrograms* (i.e. visualizations of hierarchical clustering).

The last step was to find meaningful clusters of tags. This was done using a hierarchical clustering algorithm that transformed the similarity matrix into a sequence of nested partitions. The aim was to find the most compact, spherical clusters, hence Ward's minimum variance method [33] was chosen due to its advantages in general [34], but also in this particular respect, when compared to other methods (i.e. single, centroid, median, McQuitty and complete linkage).

After obtaining a hierarchical structure in the form of a dendrogram, the clusters were then extracted by "pruning" the branches with another algorithm that combines a "partitioning around medioids" clustering method with the height of the branches [35]. The result of this first hybrid operation can be seen in the 19 clusters shown in Figure 1, shown as vertical-coloured stripes in the top section of the bottom panel. In addition, the typical tags related to each of these cluster medioids are shown in Table 2.

To increase the interpretability of these 19 clusters, a second operation was performed, consisted of repeating the hybrid pruning to increase the minimum amount of items per cluster (from 5 to 25), which thereby decreased the overall number of actual clusters. It resulted in five meta-clusters, shown in the lower section of stripes in Figure 1. These were labelled according to their contents as *Energetic* (I), *Intimate* (II), *Classical* (III), *Mellow* (IV) and *Cheerful* (V).

In both the above operations, the size of the clusters varied considerably. This was most noticeable for the first cluster in both, which was significantly larger than the rest. We interpreted this to be due to the fact that these first clusters might be capturing tags with weak relations. Indeed, for practical purposes, the first in both solutions was not as well defined and clean-cut in the semantic domain as the rest of the clusters. This was probably due to the fact that the majority of tags used in them was highly polysemic (i.e. using words that have different, and sometimes unrelated senses).

2.2 From clustered tags to music

This section explains how the original database, of 6372 songs, was then reorganized according to their closeness to each tag cluster in the semantic space. In other words, the 19 clusters from the analysis were now considered as prototypical descriptions of 19 ways that music shares similar characteristics. These prototypical descriptions were referred to as "clusters profiles" in the vector space, containing sets of between 5 and 334 tags in common (to a particular concept). Songs were then described in terms of a comparable ranked list of tags, varying in length from 1 to 96. The aim was then to measure (in terms of Euclidean distance) how close each song's ranked list of tags was to each prototypical description's set of tags. The result of this would tell us how similar each song was to each prototypical description.

An $m \times n$ *Term Document Matrix* $\mathbf{Y} = \{y_{ij}\}$ was therefore constructed to define the cluster profiles in the vector space. In this matrix, the lists of tags

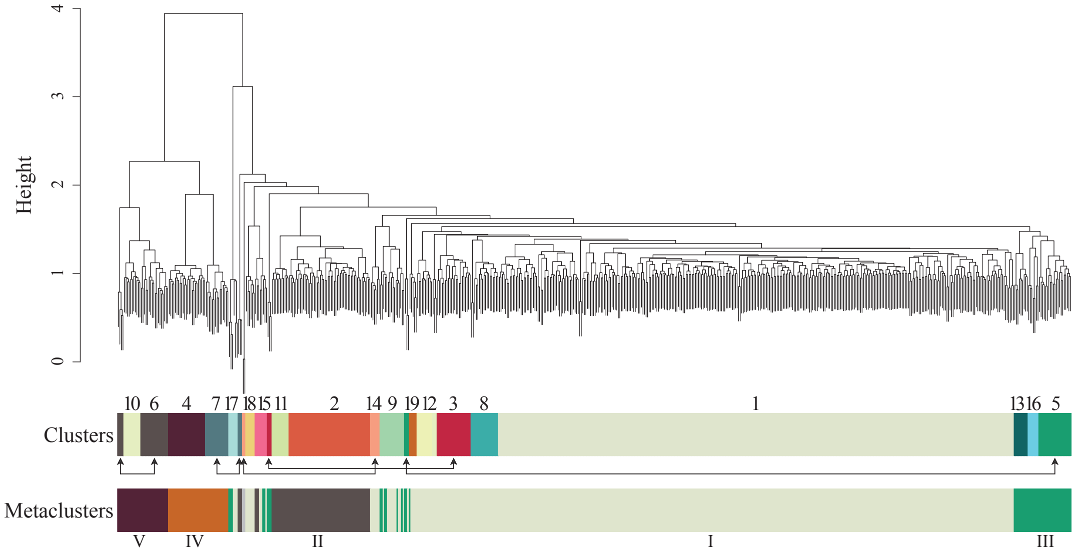


Figure 1 Hierarchical dendrogram and hybrid pruning showing 19 cluster solution (upper stripe) and 5 cluster solution (lower stripe).

Table 2 Most representative tags and corresponding artists for each of the 19 clusters

ID	Tags closest to cluster centroids	Top artists in the cluster
1	*energetic, powerful, hot*	Amy Adams, Fred Astaire, Kelly Clarkson
2	*dreamy, chill out, sleep*	Nick Drake, Radiohead, Massive Attack
3	*sardonic, sarcastic, cynical*	Alabama 3, Yann Tiersen, Tom Waits
4	*awesome, amazing, great*	Guns N' Roses, U2, Metallica
5	*cello, piano, cello rock*	Camille Saint-Saëns, Tarja Turunen, Franz Schubert
6	*00s, sexy, catchy*	Fergie, Lily Allen, Amy Winehouse
7	*mellow, beautiful, sad*	Katie Melua, Phil Collins, Coldplay
8	*hard, angry, aggressive*	System of a Down, Black Sabbath, Metallica
9	*60s, 70s, legendary*	Simon & Garfunkel, Janis Joplin, The Four Tops
10	*feelgood, summer, cheerful*	Mika, Goo Goo Dolls, Shekinah Glory Ministry
11	*wistful, intimate, reflective*	Soulsavers, Feist, Leonard Cohen
12	*high school, 90's, essential*	Fool's Garden, The Cardigans, No Doubt
13	*50s, saxophone, trumpet*	Miles Davis, Thelonious Monk, Charles Mingus
14	*1980s, eighties, voci maschili*	Ray Parker Jr., Alphaville, Michael Jackson
15	*affirming, lyricism, life song*	Lisa Stansfield, KT Tunstall, Katie Melua
16	*choral, a capella, medieval*	Mediæval Bæbes, Alison Krauss, Blackmore's Night
17	*voce femminile, donna, bella topolina*	Avril Lavigne, The Cranberries, Diana Krall
18	*tangy, coy, sleek*	Kylie Minogue, Ace of Base, Solange
19	*rousing, exuberant, passionate*	James Brown, Does It Offend You, Yeah?, Tchaikovsky

attributed to a particular song (i.e. the song descriptions) are represented as *m*, and *n* represents the 618 tags left after the filtering stage (i.e. the preselected tags). Each list of tags (*i*) is represented as a finite set {1, ..., *k*}, where 1 ≤ *k* ≤ 96 (with a mean of 29 tags per song). Finally, each element of the matrix contains a value of the normalized rank of a tag if found on a list, and it is defined by:

$$\gamma_{ij} = \left(\frac{r_k}{k}\right)^{-1} \tag{3}$$

where $r_k$is the cardinal rank of the tag *j* if found in *i*, and *k* is the total length of the list. Next, the mean rank of the tag across **Y** is calculated with:

$$\bar{r}_j = \frac{\sum_{i=1}^{m} \gamma_{ij}}{m} \tag{4}$$

And the cluster profile or *mean ranks vector* is defined by:

$$\mathrm{p}_l = \bar{r}_{j\in C_l} \tag{5}$$

$C_l$denotes a given cluster *l* where 1 ≤ *l* ≤ 19, and **p** is a vector {5, ..., *k*}, where 5 ≤ *k* ≤ 334 (5 is the minimum number of tags in one cluster, and 334 is the maximum in another).

The next step was to obtain, for each cluster profile, a list of songs ranked in order according to their closeness to the profile. This consisted in calculating the Euclidean distance $d_i$between each song's rank vector $\gamma_{i,j\in C_l}$ and each cluster profile **p***l* with:

$$d_i = \sqrt{\sum_{j\in C_l} (\gamma_{ij} - \mathrm{p}_l)^2} \tag{6}$$

Examples of the results can be seen in Table 2, where top artists are displayed beside the central tags for each cluster, while Figure 2 shows more graphically how the closeness to cluster profiles was calculated for this ranking scheme. In it are shown three artificial and partly overlapping clusters (I, II and III). In each cluster, the centroid $\mathbf{p}_l$has been calculated, together with the Euclidean distance from it to each song, as formally explained in Equations 3-6. This distance is graphically represented by the length of each line from centroid to the songs (*a*, *b*, *c*, ...), and the boxes next to each cluster

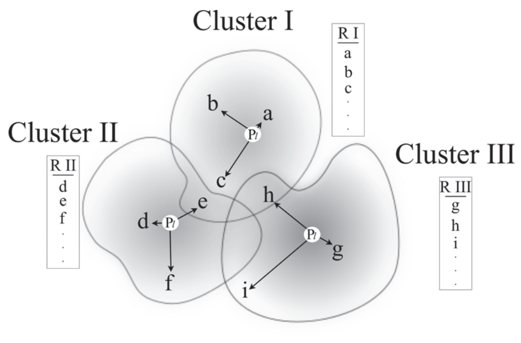


Figure 2 Visual example of the ranking of the songs based on their closeness to each cluster profile.

show their ranking (the boxes with R I, R II, R III) accordingly. Furthermore, this method allows for systematic comparisons of the clusters to be made when sampling and analysing the musical material in different ways, which is the topic of the following section.

3 Determining the acoustic qualities of each cluster

Previous research on explaining the semantic qualities of music in terms of its acoustic features has taken many forms: genre discrimination tasks [36,37], the description of soundscapes [5], bipolar ratings encompassing a set of musical examples [6] and the prediction of musical tags from acoustic features [21,38-40]. A common approach in these studies has been to extract a range of features, often low-level ones such as timbre, dynamics, articulation, Mel-frequency cepstral coefficients (MFCC) and subject them to further analysis. The parameters of the actual feature extraction are dependent on the goals of the particular study; some focus on shorter musical elements, particularly the MFCC and its derivatives [21,39,40]; while others utilize more high-level concepts, such as harmonic progression [41-43].

In this study, the aim was to characterize the semantic structures with a combined set of non-redundant, robust low-level acoustic and musical features suitable for this particular set of data. These requirements meant that we employed various data reduction operations to provide a stable and compact list of acoustic features suitable for this particular dataset [44]. Initially, we considered a large number of acoustic and musical features divided into the following categories: dynamics (e.g. root mean square energy); rhythm (e.g. fluctuation [45] and attack slope [46]); spectral (e.g. brightness, roll-off [47,48], spectral regularity [49] and roughness [50]); spectro-temporal (e.g. spectral flux [51]) and tonal features (e.g. key clarity [52] and harmonic change [53]). By considering the mean and variance of these features across 5-s samples of the excerpts (details given in the following section), we were initially presented with 50 possible features. However, these features contained significant redundancy, which limits the feasibility of constructing predictive classification or regression models and also hinders the interpretation of the results [54]. For this reason, we did not include MFCC, since they are particularly problematic in terms of redundancy and interpretation [6].

The features were extracted with the *MIRtoolbox* [52] using a frame-based approach [55] with analysis frames of 50-ms using a 50% overlap for the dynamic, rhythmic, spectral and spectro-temporal features and 100-ms with an overlap of 87.5% for the remaining tonal features.

The original list of 50 features was then reduced by applying two criteria. Firstly, the most stable features were selected by computing the Pearson's correlation between two random sets taken from the 19 clusters. For each set, 5-s sound examples were extracted randomly from each one of the top 25 ranked songs representing each of the 19 clusters. More precisely: $P(t)$ for $0.25T \leq t \leq 0.75T$, where T represents the total duration of a song. This amounted to 475 samples in each set, which were then tested for correlations between sets. Those features correlating above r = 0.5 between two sets were retained, leaving 36 features at this stage. Secondly, highly collinear features were discarded using a *variance inflation factor* $(\hat{\beta}_i < 10)$ [56]. This reduction procedure resulted in a final list of 20 features, which are listed in Table 3.

3.1 Classification of the clusters based on acoustic features

To investigate whether they differed in their acoustic qualities, four test sets were prepared to represent the clusters. For each cluster, the 50 most representative songs were selected using the ranking operation defined in Section 2.2. This number was chosen because an analysis of the rankings within clusters showed that the top 50 songs per cluster remained predominantly within the target cluster alone (89%), whereas this discriminative property became less clear with larger sets (100 songs at 80%, 150 songs at 71% and so on). From these

Table 3 Selected 20 acoustic features

Domain	Name	Σ	*MDA*
Rhythm	Attack time	M	0.23
		SD	0.08
	Fluctuation centr.	M	0.63
	Fluctuation peak	M	0.58
Spectral	Brightness	SD	0.39
	Entropy	SD	0.66
	Flatness	SD	0.60
	Regularity	M	0.33
		SD	0.26
	Roll-off	SD	0.06
	Roughness	M	0.75
	Spread	M	0.54
Spectro-Temporal	Spectral flux	M	1.20
		SD	0.44
Tonal	Chromagram centr.	M	0.98
		SD	0.35
	Chromagram peak	M	0.60
	Harmonic change	M	0.50
		SD	0.61
	Key clarity	M	0.07

Σ stands for the summary measure, where M = mean and SD = standard deviation. MDA is the *Mean Decrease Accuracy* in classification of the five meta-clusters by the acoustic features using RF.

candidates, two random 5-s excerpts were then extracted to establish two sets, to train and test each clustering, respectively. For 19 clusters, this resulted in 950 excerpts per set; and for the 5 meta-clusters, it resulted in 250 excerpts per set. After this, classification was carried out using Random Forest (RF) analysis [57]. RF is a recent variant of the regression tree approach, which constructs classification rules by recursively partitioning the observations into smaller groups based on a single variable at a time. These splits are created to maximize the between groups sum of squares. Being a non-parametric method, regression trees are thereby able to uncover structures in observations which are hierarchical, and yet allow interactions and nonlinearity between the predictors [58]. RF is designed to overcome the problem of overfitting; bootstrapped samples are drawn to construct multiple trees (typically 500 to 1000), which have randomized subsets of predictors. Out-of-bag samples are used to estimate error rate and variable importance, hence, eliminating the need for cross-validation, although in this particular case we still resorted to validation with a test set. Another advantage of RF is that the output is dependent only on one input variable, namely, the number of predictors chosen randomly at each node, heuristically set to 4 in this study. Most applications of RF have demonstrated that this technique has improved accuracy in comparison to other supervised learning methods.

For 19 clusters, a mere 9.1% of the test set could correctly be classified using all 20 acoustic features. Although this is nearly twice the chance level (5.2%), clearly the large number of target categories and their apparent acoustic similarities degrade the classification accuracy. For the meta-clusters however, the task was more feasible and the classification accuracy was significantly higher: 54.8% for the prediction per test set (with a chance level of 20%). Interestingly, the meta-clusters were found to differ quite widely in their classification accuracy: Energetic (I, 34%), Intimate (II, 66%), Classical (III, 52%), Mellow (IV, 50%) and Cheerful (V, 72%). As mentioned in Section 2.1, the poor classification accuracy of meta-cluster I is understandable, since that cluster contained the largest number of tags and was also considered to contain the weakest links between the tags (see Figure 1). However, the main confusions for meta-cluster I were with clusters III and IV, suggesting that labelling it as "Energetic" may have been premature (see Table 4). The advantage of the RF approach is the identification of critical features for classification using the Mean Decrease Accuracy [59].

Another reason for RF classification chosen was that it uses relatively unbiased estimates based on out-of-bag samples and the permutation of classification trees. The mean decrease in accuracy (MDA) is the average of such estimates (for equations and a fuller explanation, see [57,60]). These are reported in Table 3, and the normalized distributions of the three most critical features are shown in Figure 3. Spectral flux clearly distinguishes the meta-clusters II from III and IV from V, in terms of the amount of change within the spectra of the sounds used. Differences in the dominant registers also distinguish meta-clusters I from II and III from V, and these are reflected in differences in the estimated mean centroid of the chromagram for each, and roughness, the remaining critical feature, partially isolates cluster IV (Mellow, Awesome, Great) from the other clusters.

Table 4 Confusion matrix for five meta-clusters (showing 54.8% success in RF classification)

		Predicted				
		I Energetic	II Intimate	III Classical	IV Mellow	V Cheerful
Actual	I Energetic	17	5	3	2	5
	II Intimate	9	33	10	11	2
	III Classical	8	4	26	5	3
	IV Mellow	13	5	3	25	4
	V Cheerful	3	3	8	7	36

The classification results imply that the acoustic correlates of the clusters can be established if we are looking only at the broadest semantic level (meta-clusters). Even then, however, some of the meta-clusters were not adequately discriminated by their acoustical properties. This and the analysis with all 19 clusters suggest that many of the pairs of clusters have similar acoustic contents and are thus indistinguishable in terms of classification analysis. However, there remains the possibility that the overall structure of the cluster solution is nevertheless distributed in terms of the acoustic features along dimensions of the cluster space. The cluster space itself will therefore be explored in more detail next.

3.2 Acoustic characteristics of the cluster space

As classifying the clusters according to their acoustic features was not hugely accurate at the most detailed cluster level, another approach was taken to define the differences between the clusters in terms of their mutual distances. This approach examined in more detail their underlying acoustic properties; in other words, whether there were any salient acoustic markers delineating the concepts of cluster 19 ("Rousing, Exuberant, Confident, Playful, Passionate") from the "Mellow, Beautiful, Chill-out, Chill, Sad" tags of cluster 7, even though the actual boundaries between the clusters were blurred.

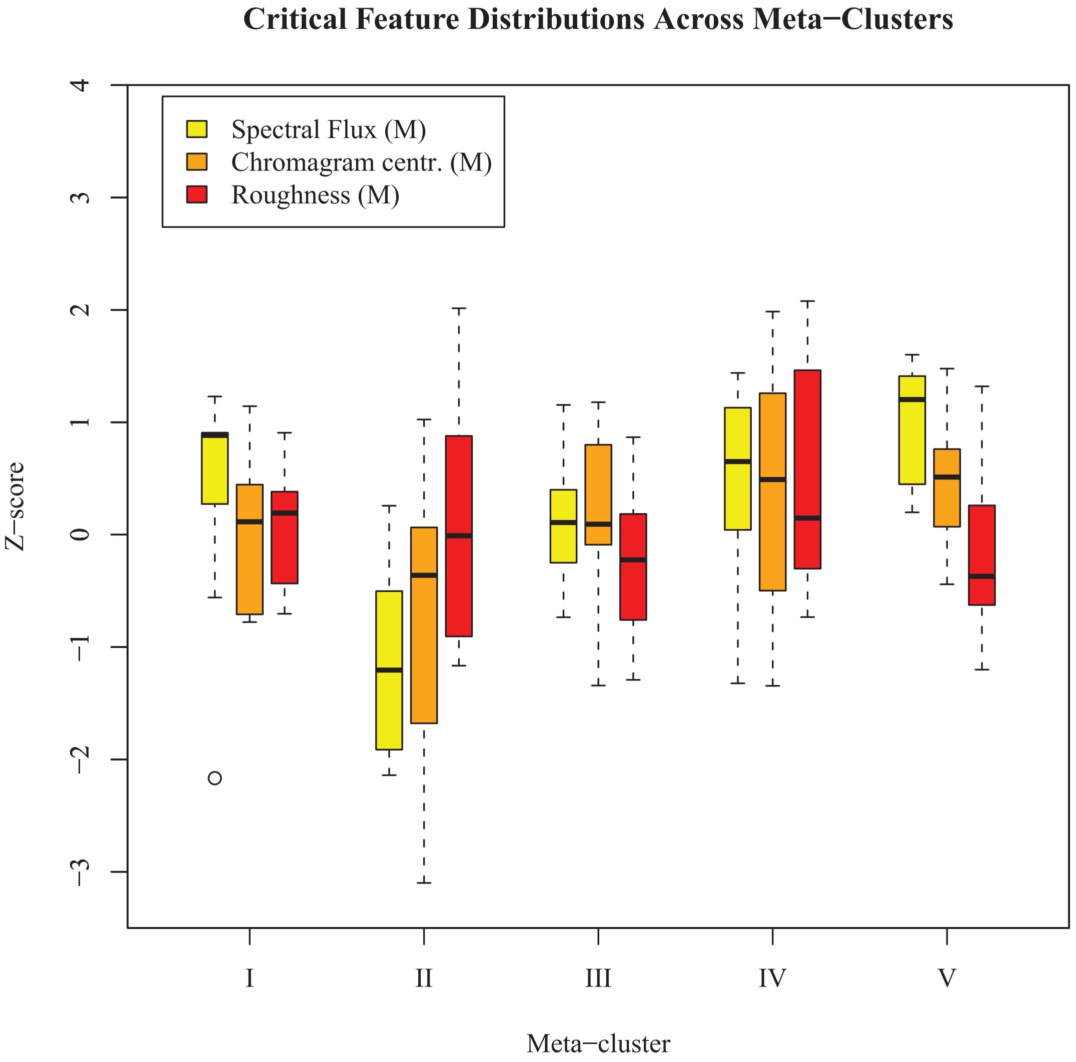


Figure 3 **Normalized distribution of the three most important features for classification of the five meta-clusters by means of RF analysis**.

To explore this idea fully, the intercluster distances were first obtained by computing the closest Euclidean distance between two tags belonging to two separate clusters [61]:

$$\mathrm{dist}(C_i, C_j) = \min\{d(x, y) : x \in C_i, y \in C_j\} \qquad (7)$$

where $C_i$and $C_j$represent a pair of clusters and x and y two different tags.

Nevertheless, before settling on this method of single linkage, we checked three other intercluster distance measures (Hausdorff, complete and average) for the purposes of comparison. Single linkage was finally chosen due to its intuitive and discriminative performance in this material and in general (cf. [61]).

The resulting distance matrix was then processed with classical metrical *Multidimensional Scaling* (MDS) analysis [62]. We then wanted to calculate the minimum number of dimensions that were required to approximate the original distances in a lower dimensional space. One way to do this is to estimate the proportion of variation explained:

$$\frac{\sum_{i=1}^{p}\lambda_i}{\sum(\text{positive eigenvalues})} \tag{8}$$

where p is the number of dimensions and $\lambda_i$represents the eigenvalues sorted in decreasing order [63].

However, the results of this procedure suggested that considering only a reduced number of dimensions would not satisfactorily reflect the original space, so we instead opted for an exploratory approach (cf. [64]). An exploration of the space meant that we could investigate whether any of the 18 dimensions correlated with the previously selected set of acoustic features, which had been extracted from the top 25 ranked examples of the 19 clusters. This analysis yielded statistically significant correlations for dimensions 1, 3 and 14 of the MDS solution with the acoustic features that are shown in Table 5. For the purpose of illustration, Figure 4 shows the relationship, in the inter-cluster space, between four of these acoustic features (shown in the labels for each axis) and two of these dimensions (1 and 3 in this case). If we look at clusters 14 and 16, we can see that they both contain tags related with the human voice (*Voci maschili* and *Choral*, respectively), and they are situated around the mean of the X-axis. However, this is in spite of a large difference in sound character, which can best be described in terms of their perceptual dissonance (e.g. spectral roughness), hence their positions at either end of the Y -axis. Another example of tags relating to the human voice, concerns clusters 17 and 4 (*Voce femminile* and *Male Vocalist*, respectively), but this time they are situated around the mean of the Y -axis, and it is in terms of the shape of the spectrum (e.g. spectral spread) that they differ most, hence their positions at the end of the X-axis. In sum, despite the modest classification accuracy of the clusters according to their acoustic features, the underlying semantic structure embedded into tags could nonetheless be more clearly explained in terms of their relative positions to each other within the cluster space. The dimensions yielded intuitively interpretable patterns of correlation, which seem to adequately pinpoint the essence of what musically characterize the concepts under investigation in this study (i.e. adjectives, nouns, instruments, temporal references and verbs). However, although these semantic structures could be distinguished sufficiently by their acoustic profiles at the generic, meta-cluster level; this was not the case at the level of the 19 individual clusters. Nevertheless, the organization of the individual clusters across the semantic space could be connected by their acoustic features. Whether the acoustic substrates that musically characterize these tags is what truly distinguishes them for a listener is an open question that will be explored more fully next.

4 Similarity rating experiment

In order to explore whether the obtained clusters were perceptually meaningful, and to further understand what kinds of acoustic and musical attributes they actually consisted of, new empirical data about the clusters needed to be gathered. For this purpose, a similarity rating experiment was designed, which assessed the timbral qualities of songs from each of the tag clusters. We chose to focus on the low-level, non-structural qualities of music, since we wanted to minimize the possible confounding factor of association, caused by recognition of lyrics, songs or artists. The stimuli for the experiment therefore consisted of semi-randomly spliced [37,65], brief excerpts. These stimuli, together with other details of the experiment, will be explained more fully in the remaining parts of this section.

4.1 Experiment details

4.1.1 Stimuli

Five-second excerpts were randomly taken from a middle part ($P(t)$ for $0.25T \leq t \leq \cdot 0.75T$, where T represents the total duration of a song) of each of the 25 top ranked songs from each cluster (see the ranking procedure detailed in Section 2.2). However, when splicing the excerpts together for similarity rating, we wanted to minimize the confounds that were caused by disrupting the onsets (i.e. bursts of energy). Therefore, the exact temporal position of the onsets for each excerpt was detected with the aid of the MIRToolbox [52]. This

Table 5 Correlations between acoustic features and the inter-item distances between the clusters

Dimension 1		Dimension 3		Dimension 14	
Acoustic feature	***r***	**Acoustic feature**	***r***	**Acoustic feature**	***r***
Fluctuation centroid (M)	0.53*	Regularity (SD)	-0.51*	Chromagram centroid (M)	0.60**
Spread (M)	0.51*	Harmonic change (SD)	-0.50*	Flatness (SD)	0.54*
Entropy (SD)	0.50*	Roughness (M)	0.50*	Attack time (M)	-0.51*
Brightness (SD)	0.49*	Harmonic change (M)	-0.50*	Regularity (M)	-0.51*
Flatness (SD)	0.49*	Chromagram centroid (SD)	-0.45*	Attack time (SD)	-0.48*
Flux (SD)	0.49*	Flux (SD)	-0.45*	Chromagram peak (M)	-0.46*

* p <0.05, ** p <0.01, df = 17

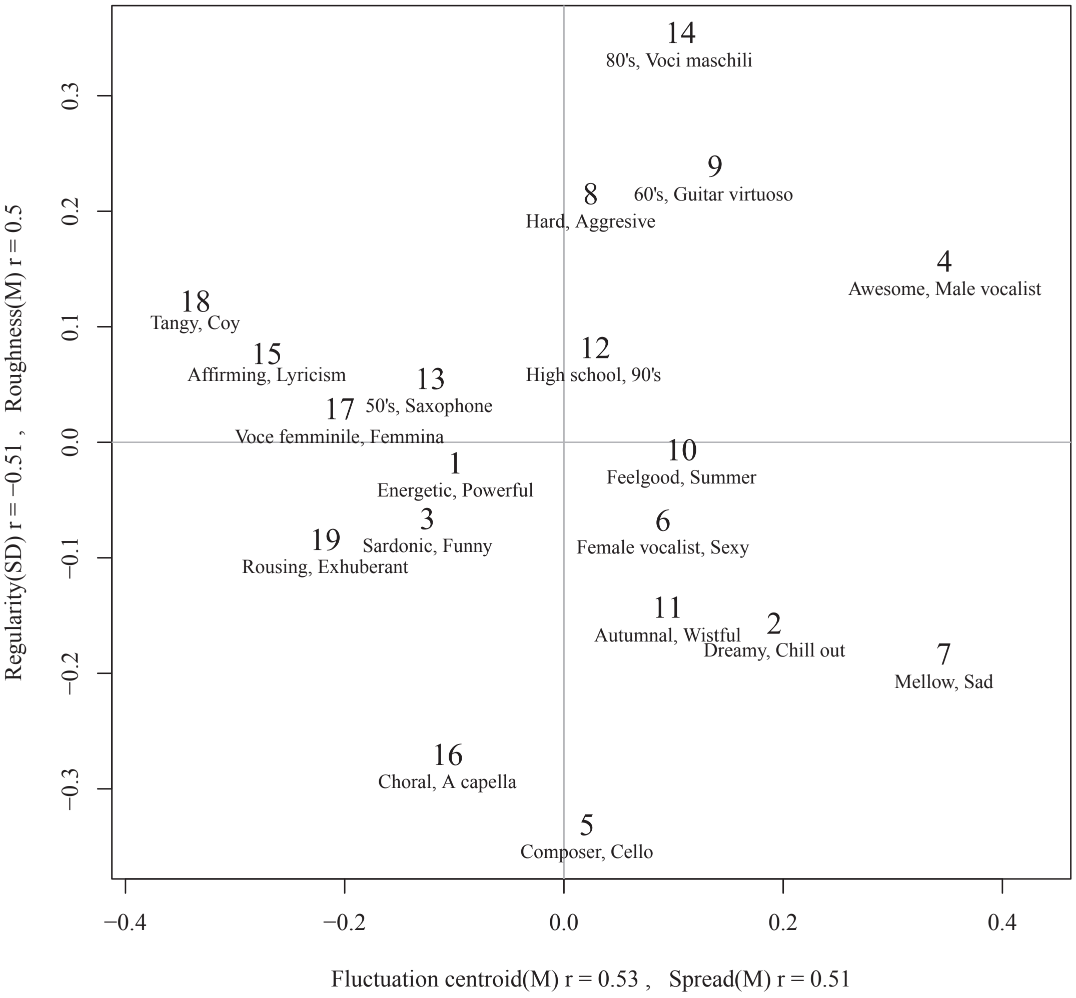


Figure 4 MDS (dimensions 1, 3) of intercluster distances.

process consisted of computing the *spectral flux* within each excerpt by focussing on the increase in energy in successive frames. It produced a temporal curve from which the highest peak was selected as the reference point for taking a slice, providing that this point was not too close to the end of the signal ($t \leq 4500$ ms).

Slices of random length ($150 \leq t \leq 250$ ms) were then taken from a point that was 10 ms before the peak onset for each excerpt that was being used to represent a tag cluster. The slices were then equalized in loudness, and finally mixed together using a fade in/out of 50 ms and an overlap window of 100 ms. This resulted in 19 stimuli (examples of the spliced stimuli can be found at http://www.jyu.fi/music/coe/materials/splicedstimuli) of variable length, each corresponding to a cluster, and each of which was finally trimmed to 1750 ms (with a fade in/out of 100 ms). To finally prepare these 19 stimuli for a similarity rating experiment, the resulting 171 paired combinations were mixed with a silence of 600 ms between them.

4.1.2 Participants

Twelve females and nine males were participated in this experiment (age M = 26.8, SD = 4.15). Nine of them had at least 1 year of musical training. Twelve reported listening to music attentively between 1 and 10 h/week, and 19 of the subjects listened to music while doing another activity (63% $1 \leq t \leq 10$, 26% $11 \cdot \leq t \leq 20$, 11% $t \leq 21$ h/week).

4.1.3 Procedure

Participants were presented with pairs of sound excerpts in random order using a computer interface and high-quality headphones. Their task was to rate the similarity of sounds on a 9-level Likert scale, the extremes of which were labelled as *dissimilar* and *similar*. Before the actual experimental trials, the participants were also given instructions and some practice to familiarize themselves with the task.

4.2 Results of experiment

The level at which participants' ratings agreed with each other was estimated with Cronbach's method (α = 0.94), and the similarity matrices derived from their ratings were used to make a representation of the perceptual space. Individual responses were thus aggregated by computing a mean similarity matrix, and this was subjected to a classical metric MDS analysis. With Cox and Cox's [63] method (8) we estimated that four dimensions were enough to represent the original space since these can explain 70% of the variance.

4.2.1 Perceptual distances

As would be hoped, the arrangement of clusters, as represented by their spliced acoustic samples, illustrates a clear organization according to an underlying semantic structure. This perceptual distance can be seen in Figure 5 where, for example, *Aggressive* and *Chill out* are in opposite corners of the psychological space. There is also a clear acoustical organization of the excerpts, as cluster number 5 (*Composer*, *Cello*) is depicted as being high in roughness and high in spectral regularity, with a well-defined set of harmonics, and those clusters that have similar overall descriptors, such as 15 (*Affirming*, *Lyricism*), 7 (*Mellow*, *Sad*) and 11 (*Autumnal*, *Wistful*), are located within proximity of each other. Noticeably, cluster number 1 is located at the centre of the MDS solution, which could be expected from a cluster that worked as a trap for tags with weak relations.

4.2.2 Acoustic attributes of the similarities between stimuli

Acoustic features were extracted from the stimuli in a similar fashion to that described in Section 3, but the list of features was consolidated again by trimming it down to a robust minimal set. Trimming consisted of creating another random set of stimuli and correlating their acoustic features with the stimuli used in the experiment. Those features which performed poorly (r <0.5, df = 17) were removed from the list. After this, the coordinates of the resulting 4-dimensional space were correlated with the set of acoustic features extracted from the stimuli to show the perceptual distances of the stimuli from one another. Only dimensions 1 and 2 had statistically significant linear correlations with the acoustic features, the other dimensions having only low correlations ($|r| \leq 0.5$, or p >0.05, df = 17). The final selection of both acoustic features and dimensions is displayed in Table 6.

The first dimension correlates with features related to the organization of pitch and harmonics, as revealed by the mean chromagram peaks (r = 0.82) and the degree of variation between successive peaks in the spectrum (mean spectral regularity r = 0.72). There is also correlation with the variance of the energy distribution (standard deviation of the spectral roll-off at 95% r = 0.7); the distance between the spectrum of successive frames (mean spectral flux r = -0.7); and to a lesser degree with the shape of the spectrum in terms of its "width" (mean spectral spread r = -0.61). The second dimension correlates significantly with the perceived dissonance (mean roughness r = -0.74); pitch salience (chromagram centroid r = -0.72); and also captures the mean spectral spread (r = 0.65), although in an inverse fashion. Table 6 provides a more detailed summary of this.

4.2.3 Comparing a semantic structure based on social tags, to one based on perceptual similarities

As we have now explored the emergent structure from tags using a direct acoustic analysis of the best exemplars in each cluster, and probed this semantic space further in a perceptual experiment, the question remains as to whether the two approaches bear any similarities. The most direct way to examine this is to look at the pattern of correlations between both: i.e. to compare tables 5 and 6. Although the lists of features vary slightly, due to the difference in redundancy and robustness criteria applied to each set of data, convergent patterns can still be found. An important shared feature is the variation in brightness, which is both present in dimension 1 of the direct cluster analysis, and in the perceptual space depicting the spliced stimuli (from the same 19 clusters). In the first case, it takes the form of "brigthenss SD", and in the second, it is "roll-off SD" (virtually identical). In addition, the second dimension in both solutions is characterized by roughness, although the underlying polarities of the space have been flipped in each. Of course, one major reason for differences between the two sets of data must be due to the effects of splicing, conducted in the perceptual experiment but not in the other. However, there were nevertheless analogies between the two perspectives of the semantic structure that could be detected in the acoustic substrates. They have been used here to highlight such features that are little affected by form, harmony, lyrics and other high-level musical (and extra-musical) characteristics. From this perspective, a tentative convergence between the two approaches was successfully obtained.

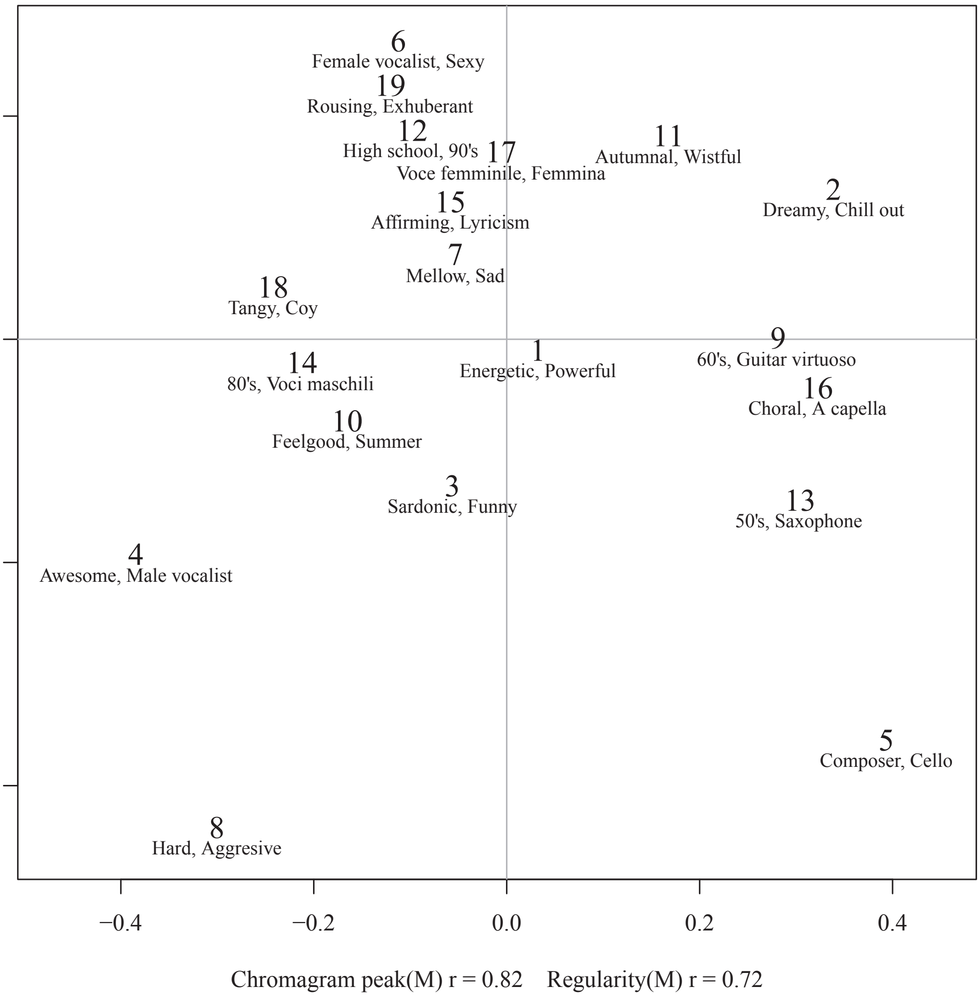


Figure 5 MDS (dimensions 1, 2) of perceived distances of the stimulus material representing the 19 clusters using semi-randomly spliced excerpts.

5 Discussion and conclusions

Semantic structures within music have been extracted from the social media previously [20,25] but the main difference between the prior genre-based studies and this study is that we focussed more on the way people describe music in terms of how it sounds in conceptual expressions. We argue that these expressions are more stable than musical genres, which have previously proven to be of a transient nature and a source of disagreement (cf. [37]), despite important arguments vindicating their value for classification systems [66]. Perhaps the biggest problem with expert classifications (such as genre) is that the result may not reach the same level of ecological validity in describing how music sounds, as a semantic structure derived from social tags. This is a very important reason to examine tag-based semantic structures further, in spite of their inherent weaknesses as pointed out by Lamere [7].

A second way in which this study differs from those previous lies in the careful filtering of the retrieved tags using manual and automatic methods before the actual analysis of the semantic structures was conducted. Not only that, but a prudent trimming of the acoustic features was done to avoid overfitting and any possible increases in model complexity. Finally, a perceptual exploration of the semantic structure found was carried

Table 6 Correlations between MDS solutions (dimensions 1 and 2) and acoustic features for the experiment

Domain	Name	Σ	Dim 1	Dim 2
Spectral	Entropy	SD	0.36	0.46*
	Flatness	SD	-0.13	0.32
	Regularity	M	0.72***	0.10
	Roll-Off	SD	0.70***	0.14
	Roughness	M	-0.35	-0.74***
	Spread	M	-0.61**	0.65**
Spectro-temporal	Spectral flux	M	-0.70***	-0.16
Tonal	Chromagram centroid	M	-0.23	-0.72***
	Chromagram peak	M	0.82***	-0.28

* *p* <0.05, ** *p* <0.01, *** *p* <0.001, *df* = 17

out to assess whether the sound qualities alone would be sufficient to uncover the tag-based structure.

The whole design of this study offers a preliminary approach to the cognition of timbre in semantic terms. In other words, it uses verbal descriptions of music, expressed by the general population (in the form of social tags), as a window to study how a critical feature of music (timbre) is represented in the *semantic memory* [67]. It is however evident that if each major step of this study was treated separately, there would be plenty of room for refining their respective methodologies, namely, tag filtering, uncovering the semantic structure, acoustic summarization and conducting a perceptual experiment to examine the two empirical perspectives. This being said, we *did* consider some of the alternatives for these steps to avoid methodological pitfalls (particularly in the clustering and the distance measures). But even if each analytical step was optimized to enhance the solution to an isolated part of the problem, this would inevitably come at the expense of unbalancing the overall picture. Since this study relies on an exploratory approach, we chose mainly conventional techniques for each step, with the expectation that further research will be conducted to corroborate the findings and improve the techniques used here.

The usefulness of signal summarization based on the random spliced method [37] has been assessed for audio pattern recognition [65]. Our findings in the perceptual domain seem to vindicate the method where listeners rate sounds differing in timbral qualities, especially if the scope is the long-term non-structural qualities of music [68]. Such a focus is attained by cutting the slices in a way that preserves important aspects of music (onsets and sample lengths), while ensuring that they are from a wide cross section of timbrically related songs (i.e. belonging to the same semantic region or *timbral environment* [69] in the perceptual space).

In conclusion, this study provided a bottom-up approach for finding the semantic qualities of music descriptions, while capitalizing on the benefits of social media, NLP, similarity ratings and acoustic analysis to do so. We learned that when listeners are presented with brief and spliced excerpts taken from the clusters representing a tag-based categorization of the music, they are able to form coherent distinctions between them. Through an acoustic analysis of the excerpts, clear correlations between the dimensional and timbral qualities of music emerged. However, it should be emphasized that the high relevance of many timbral features is only natural since the timbral characteristics of the excerpts were preserved and structural aspects were masked by the semi-random splicing. Nevertheless, we are positively surprised at the level of coherence in regard to the listener ratings and their explanations in terms of the acoustic features; in spite of the limitations we imposed on the setting using a random splicing method, and the fact that we tested a large number of clusters.

The implications of the present findings relate to several open issues. The first is whether structural aspects of music are required to explain the semantic structures or whether low-level, timbral characteristics are sufficient, as was suggested by the present findings. Secondly, what new semantic layers (as indicated by the categories of tags) can meaningfully be connected with the acoustic properties of the music? Finally, if the timbral characteristics are indeed strongly connected with such semantic layers as *adjectives*, *nouns* and *verbs*, do these arise by means of learning and associations or are the underlying regularities connected with the emotional, functional and gestural cues of the sounds?

A natural continuation of this study would be to go deeper into the different layers of tags to explore which layers are more amenable to direct mapping by acoustic qualities, and which are mostly dependent on the functional associations and cultural conventions of the music.

A Preprocessing

Preprocessing is necessary in any text mining application because the retrieved data do not follow any particular set of rules, and there are no standard steps to follow [70]. Moreover, with the aid of *Natural Language Processing* (NLP) [71,72] methods, it is possible to explore the nature of the tags from statistical and lexicological perspectives. In the following sections, the rationale and explanation for each preprocessing step is given.

A.1 Filtering

Three filtering rules were applied to the corpus.

Remove *Hapax legomena* (i.e. tags that appear only once in the corpus), under the rationale of discarding unrelated data (see Table 1).

Capture the most prevalent tags by eliminating from the vocabulary those whose *index of usage* (see Section 2) is below the mean.

Discard tags composed by three or more words in order to prune short sentence-like descriptions from the corpus.

The subset resulting from such reductions represents 46.6% of the corpus (N = 169, 052, Vocabulary = 2029 tags).

A.2 Lexical categories for tags

At this point, the data had been de-noised but only in the quantitative domain. To extract a meaningful ontology from the tags, not only filtering, but semantic analysis of the tags was necessary. To do so in an effective fashion, a qualitative analysis was performed using a number of sources: the Brown Corpus [73] to identify parts of speech; the Wordnet database [74] to disambiguate words; and the online Urban Dictionary (http://www.urbandictionary.com) and http://www.Last.fm database for general reference. We were thus aiming for a balanced set of references; two sources were technical (the Brown and Wordnet), one vernacular (the Urban Dictionary) and one highly specialized in musical jargon (Last.fm's wiki pages). An underlying motivation for relying on this broad set of references, rather than exclusively on an English dictionary, was to recognize the multilingual nature of musical tags. Tag meanings were thus looked up and the selection of a category was decided case by case. The criteria applied in this process favoured categories of meaning closely related to music and the music industry, such as the genre, artist, instrument, form of music, and commercial entity. The next most important type of meaning looked for was adjectival, and finally other types of descriptor were considered. For instance, "Acid" is well known to be a corrosive substance, but it is also a term used extensively to describe certain musical genres, so this latter meaning took priority. Table 7 shows the aforementioned tag categories, examples of each, a definition of each, and their percentage of distribution in the sample.

The greatest percentage of tags refer to *musical genres*, but there are significant percentages in other categories. For instance, the second most commonly found tags are *adjectives*, followed by *nouns* which except for some particular contextual connotations, are used for the most part adjectivally to describe the general sound of a song (e.g. *mellow*, *beautiful* for adjectives and *memories* and *melancholy* for nouns).

The rest of the categories suggest that music is often tagged in terms of association, whether it be to known auditory objects (e.g. instruments and band names), specific circumstances (e.g. geographical locations and time of the day or season) or idiosyncratic things that only make sense at a personal level. This classification is mainly consistent with past efforts [7], although the vocabulary analysed is larger, and there are consequently more categories.

The result allowed for a finer discrimination of tags to be made, that might better uncover the semantic structure. Since one of the main motivations of this project was to obtain prototypical timbral descriptions, we focused on only a few of the categories: *adjectives*, *nouns*, *instruments*, *temporal references* and *verbs*, and this resulted in a vocabulary of 618 tags.

The rest of the tag categories were left for future analysis. Note that this meant discarding such commonly used descriptors as musical genres, which on the one hand provide an easy way to discriminate music [36] in

Table 7 Main categories of tags

Categories	%	Definition	Examples
Genre	36.72	Musical genre or style	Rock, Alternative, Pop
Adjective	12.17	General category of adjectives	Beautiful, Mellow, Awesome
Noun	9.41	General category of nouns	Love, Melancholy, Memories
Artist	8.67	Artists or group names	Coldplay, Radiohead, Queen
Locale	8.03	Geographic situation or locality	British, American, Finnish
Personal	6.80	Words used to manage personal collections	Seen Live, Favourites, My Radio
Instrument	4.83	Sound source	Female vocalists, Piano, Guitar
Unknown	3.79	Unclassifiable	aitch, prda, < 3
Temporal	2.41	Temporal circumstance	80's, 2000, Late Romantic
Form	2.22	Musical form or compositional technique	Ballad, Cover, Fusion
Commercial	1.72	Record label, radio station, etc.	Motown, Guitar Hero, Disney
Verb	1.63	General category of verbs	Chillout, Relax, Wake up
Content	1.03	Emphasis in the message or literary content	Political, Great lyrics, Love song
Expression	0.54	Exclamations	Wow, Yeah, lol

terms of fairly broad categories, but on the other hand makes them hard to adequately define by virtue of this very same quality [37]. This manuscript is devoted to exploring timbre and by extension the way people describe the general sound of a piece of music, hence the idea has been to explore the concepts that lie underneath the genre descriptions. For this reason, *genre* was utilized as the most significant semantic filter. The other discarded categories had their own reasons, for instance *Personal* and *Locale* contents are strongly centered in the individual's perspective, *Artist* contents are redundantly referring to the creator/performer of the music. The rest of the omissions concerned rare categories (e.g. *unknown terms*, *expressions*, *commercial branches or recording companies*) or not explicitly related with timbre (e.g. *musical form*, description of the *lyrics*); these were left out to simplify the results.

Acknowledgements
This study was supported by the Finnish Centre of Excellence in Interdisciplinary Music Research.

Competing interests
The authors declare that they have no competing interests.

Received: 25 March 2011 Accepted: 7 December 2011
Published: 7 December 2011

References

1. O Celma, X Serra, FOAFing the Music: Bridging the semantic gap in music recommendation. *Web Semantics. Science, Services and Agents on the World Wide Web*. **6**(4), 250–256 (2008). [Semantic Web Challenge 2006/2007]. doi:10.1016/j.websem.2008.09.004
2. J Grey, Multidimensional Perceptual Scaling of Musical Timbres. The Journal of the Acoustical Society of America. **61**(5), 1270–1277 (1977). doi:10.1121/1.381428
3. S McAdams, S Winsberg, S Donnadieu, G De Soete, J Krimphoff, Perceptual Scaling of Synthesized Musical Timbres: Common dimensions, specificities and latent subject classes. Psychological Research. **58**(3), 177–192 (1995). doi:10.1007/BF00419633
4. J Burgoyne, S McAdams, A Meta-analysis of Timbre Perception Using Nonlinear Extensions to CLAS-CAL. Computer Music Modeling and Retrieval. Sense of Sounds, 181–202 (2009)
5. JJ Aucouturier, F Pachet, M Sandler, The Way it Sounds: Timbre models for analysis and retrieval of music signals. Multimedia, IEEE Transactions on. **7**(6), 1028–1035 (2005)
6. V Alluri, P Toiviainen, Exploring Perceptual and Acoustical Correlates of Polyphonic Timbre. Music Perception. **27**(3), 223–242 (2010). doi:10.1525/mp.2010.27.3.223
7. P Lamere, Social Tagging and Music Information Retrieval. Journal of New Music Research. **37**(2), 101–114 (2008). doi:10.1080/09298210802479284
8. JJ Aucouturier, E Pampalk, Introduction-From Genres to Tags: A little epistemology of music information retrieval research. Journal of New Music Research. **37**(2), 87–92 (2008). doi:10.1080/09298210802479318
9. C Held, U Cress, Learning by Foraging: The impact of social tags on knowledge acquisition, in *Learning in the Synergy of Multiple Disciplines. 4th European Conference on Technology Enhanced Learning*, Nice, France, (2009)
10. F Hesse, Use and Acquisition of Externalized Knowledge, in *Learning in the Synergy of Multiple Disciplines. 4th European Conference on Technology Enhanced Learning*, (Nice, France: Springer, 2009), p. 5
11. E Chi, Augmented Social Cognition: Using social web technology to enhance the ability of groups to remember, think, and reason, in *Proceedings of the 35th SIGMOD International Conference on Management of Data, Providence*, Rhode Island, USA, (2009)
12. H Kim, S Decker, J Breslin, Representing and Sharing Folksonomies with Semantics. Journal of Information Science. **36**, 57–72 (2010). doi:10.1177/0165551509346785
13. A Mathes, Folksonomies-cooperative Classification and Communication through Shared Metadata. http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html (2004)
14. H Lin, J Davis, Y Zhou, An Integrated Approach to Extracting Ontological Structures from Folksonomies, in *Proceedings of the 6th European Semantic Web Conference on The Semantic Web. Research and Applications*, (Heraklion, Greece: Springer, 2009), p. 668
15. S Deerwester, S Dumais, G Furnas, T Landauer, R Harshman, Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science. **41**(6), 391–407 (1990). doi:10.1002/(SICI)1097-4571(199009)41:63.0.CO;2-9
16. J Bellegarda, Latent Semantic Mapping: Principles & applications. Synthesis Lectures on Speech and Audio Processing. **3**, 1–101 (2007). doi:10.2200/S00048ED1V01Y200609SAP003
17. S Sundaram, S Narayanan, Audio Retrieval by Latent Perceptual Indexing, in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, (IEEE, 2008), pp. 49–52
18. S Dumais, Latent Semantic Analysis. Annual Review of Information Science and Technology (ARIST). **38**, 189–230 (2004)
19. T Eerola, R Ferrer, Setting the Standards: Normative data on audio-based musical features for musical genres, in *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music*, Jyväskylä, Finland, (2009)
20. M Levy, M Sandler, Learning Latent Semantic Models for Music from Social Tags. Journal of New Music Research. **37**(2), 137–150 (2008). doi:10.1080/09298210802479292
21. T Bertin-Mahieux, D Eck, F Maillet, P Lamere, Autotagger: A model for predicting social tags from acoustic features on large music databases. Journal of New Music Research. **37**(2), 115–135 (2008). doi:10.1080/09298210802479250
22. P Rentfrow, S Gosling, Message in a Ballad. Psychological Science. **17**(3), 236–242 (2006). doi:10.1111/j.1467-9280.2006.01691.x
23. M Delsing, T ter Bogt, R Engels, W Meeus, Adolescents' Music Preferences and Personality Characteristics. European Journal of Personality. **22**(2), 109–130 (2008). doi:10.1002/per.665
24. I Popescu, G Altmann, *Word Frequency Studies* (Berlin: Walter de Gruyter, 2009)
25. M Levy, M Sandler, A Semantic Space for Music Derived from Social Tags, in *Proceedings of the 8th International Society for Music Information Retrieval Conference*, vol. 1, ed. by Dixon S, Bainbridge D, Typke R (Vienna, Austria: Österreichische Computer Gesellschaft, 2007), p. 12
26. B Zhang, Q Xiang, H Lu, J Shen, Y Wang, Comprehensive Query-dependent Fusion Using Regression-on-folksonomies: A case study of multimodal music search, in *Proceedings of the seventeen ACM international conference on Multimedia*, Beijing, China: ACM, pp. 213–222 (2009)
27. H Halpin, V Robu, H Shepherd, The Complex Dynamics of Collaborative Tagging, in *Proceedings of the 16th international conference on World Wide Webal Conference on World Wide Web*, Banff, Alberta, Canada: ACM, p. 220 (2007)
28. J Brank, M Grobelnik, D Mladenic, Automatic Evaluation of Ontologies, in *Natural Language Processing and Text Mining*, ed. by Kao A, RPoteet S (USA: Springer, 2007)
29. J Siskind, Learning Word-to-meaning Mappings, in *Models of Language Acquisition. Inductive and deductive approaches*, USA: Oxford University Press, pp. 121–153 (2000)
30. G Zipf, *Human Behavior and the Principle of Least Effort, An introduction to human ecology* (addison-wesley press, 1949)
31. J Gower, P Legendre, Metric and Euclidean Properties of Dissimilarity Coefficients. Journal of Classification. **3**, 5–48 (1986). doi:10.1007/BF01896809
32. M Walesiak, A Dudek, *ClusterSim. Searching for optimal clustering procedure for a data set* (2011). http://CRAN.R-project.org/package=clusterSim [R package version 0.39-2]
33. R Development Core Team, *R A language and environment for statistical computing*, (R Foundation for Statistical Computing, Vienna, Austria, 2009). http://www.R-project.org [ISBN 3-900051-07-0]
34. A Jain, R Dubes, *Algorithms for Clustering Data* (Englewood Cliffs, NJ: Prentice Hall, 1988)

35. P Langfelder, B Zhang, S Horvath, *DynamicTreeCut. Methods for detection of clusters in hierarchical clustering dendrograms* (2009). http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting/ [R package version 1.20]
36. G Tzanetakis, P Cook, Musical Genre Classification of Audio Signals. IEEE Transactions on Speech and Audio Processing. **10**(5), 293–302 (2002). doi:10.1109/TSA.2002.800560
37. R Gjerdingen, D Perrott, Scanning the Dial: The rapid recognition of music genres. Journal of New Music Research. **37**(2), 93–100 (2008). doi:10.1080/09298210802479268
38. M Hoffman, D Blei, P Cook, Easy as CBA: A simple probabilistic model for tagging music, in *Proceedings of the 10th International Society for Music Information Retrieval Conference*, Kobe, Japan, (2009)
39. MI Mandel, DPW Ellis, A Web-based Game for Collecting Music Metadata. Journal of New Music Research. **37**(2), 151–165 (2008). doi:10.1080/09298210802479300
40. D Turnbull, L Barrington, D Torres, G Lanckriet, 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*, SIGIR '07, New York, NY, USA: ACM, pp. 439–446 (2007)
41. K Jacobson, M Sandler, B Fields, Using Audio Analysis and Network Structure to Identify Communities in On-line Social Networks of Artists, in *Proceedings of the 9th International Society for Music Information Retrieval Conference*, ed. by Bello JP, Chew E, Turnbull D (Philadelphia, USA, 2008), pp. 269–274
42. C Laurier, O Meyers, J Serrà, M Blech, P Herrera, X Serra, Indexing Music by Mood: Design and integration of an automatic content-based annotator. *Multimedia Tools and Applications*. **48**, 161–184 (2010). [Springerlink link: http://www.springerlink.com/content/jj01750u20267426]. doi:10.1007/s11042-009-0360-2
43. J Bello, J Pickens, A Robust Mid-level Representation for Harmonic Content in Music Signals, in *Proceedings of the 6th International Society for Music Information Retrieval Conference*, London, UK, pp. 304–311 (2005)
44. S Chu, S Narayanan, CC Kuo, Environmental Sound Recognition With Time-frequency Audio Features. Audio, Speech, and Language Processing, IEEE Transactions on. **17**(6), 1142–1158 (2009)
45. E Pampalk, A Rauber, D Merkl, Content-based Organization and Visualization of Music Archives, in *Proceedings of the tenth ACM international conference on Multimedia*, Juan les Pins, France: ACM, p. 579 (2002)
46. G Peeters, A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project. *CUIDADO IST Project Report* 1–25 (2004)
47. P Juslin, Cue Utilization in Communication of Emotion in Music Performance: Relating performance to perception. Journal of Experimental Psychology. Human perception and performance., 6 **26**, 1797–1812 (2000)
48. P Laukka, P Juslin, R Bresin, A Dimensional Approach to Vocal Expression of Emotion. Cognition & Emotion. **19**(5), 633–653 (2005). doi:10.1080/02699930441000445
49. K Jensen, *Timbre Models of Musical Sounds* (Department of Computer Science, University of Copenhagen, 1999)
50. W Sethares, *Tuning, Timbre, Spectrum, Scale* (Springer Verlag, 2005)
51. J Bello, C Duxbury, M Davies, M Sandler, On the use of Phase and Energy for Musical Onset Detection in the Complex Domain. Signal Processing Letters, IEEE. **11**(6), 553–556 (2004). doi:10.1109/LSP.2004.827951
52. O Lartillot, P Toiviainen, T Eerola, A Matlab Toolbox for Music Information Retrieval, in *Data Aalysis, Machine Learning and Applications*, ed. by Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R (Berlin, Germany: Springer, 2008), pp. 261–268. Studies in Classification, Data Analysis, and Knowledge Organization
53. C Harte, M Sandler, M Gasser, Detecting Harmonic Change in Musical Audio, in *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia*, Santa Barbara, CA, USA: ACM, p. 26 (2006)
54. I Guyon, A Elisseeff, An Introduction to Variable and Feature Selection. Journal of Machine Learning Research. **3**, 1157–1182 (2003)
55. G Tzanetakis, P Cook, Manipulation, Analysis and Retrieval Systems for Audio Signals, PhD thesis, (Princeton University, Princeton, NJ, 2002)
56. J Fox, G Monette, Generalized Collinearity Diagnostics. Journal of the American Statistical Association. **87**(417), 178–183 (1992). doi:10.2307/2290467
57. L Breiman, Random Forests. Machine Learning. **45**, 5–32 (2001). doi:10.1023/A:1010933404324
58. B Ripley, *Pattern Recognition and Neural Networks*, (Cambridge: Cambridge University Press, 1996)
59. K Archer, R Kimes, Empirical Characterization of Random Forest Variable Importance Measures. Computational Statistics & Data analysis. **52**(4), 2249–2260 (2008). doi:10.1016/j.csda.2007.08.015
60. H Pang, A Lin, M Holford, B Enerson, B Lu, M Lawton, E Floyd, H Zhao, Pathway Analysis Using Random Forests Classification and Regression. Bioinformatics. **22**(16), 2028 (2006). doi:10.1093/bioinformatics/btl344
61. L Nieweglowski, *CLV: Cluster validation techniques* (2009). http://CRAN.R-project.org/package=clv [R package version 0.3-2]
62. J Gower, Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. Biometrika. **53**(3-4), 325 (1966). doi:10.1093/biomet/53.3-4.325
63. M Cox, T Cox, *Multidimensional Scaling: Handbook of data visualization*, (USA: Chapman & Hall, 2001)
64. I Borg, P Groenen, *Modern Multidimensional Scaling: Theory and applications*, (Springer Verlag, 2005)
65. JJ Aucouturier, B Defreville, F Pachet, The Bag-of-frames Approach to Audio Pattern Recognition: A sufficient model for urban soundscapes but not for polyphonic music. The Journal of the Acoustical Society of America. **122**(2), 881–891 (2007). doi:10.1121/1.2750160
66. C McKay, I Fujinaga, Musical Genre Classification: Is it worth pursuing and how can it be improved. in *Proceedings of the 7th International Society for Music Information Retrieval Conference* 101–6 (2006)
67. D Balota, J Cohane, Semantic Memory, in *Learing and Memory: A comprehensive reference, Volume 2 Cognitive Pyshology of Memory*, ed. by Byrne JH, III HLR (Oxford, UK: Academic Press, 2008), pp. 511–534
68. G Sandell, Macrotimbre: Contribution of attack, steady state, and verbal attributes. The Journal of the Acoustical Society of America. **103**, 2966 (1998)
69. R Ferrer, Embodied Cognition Applied to Timbre and Musical Appreciation: Theoretical Foundation. British Postgraduate Musicology. **X,** http://www.bpmonline.org.uk/bpm10/ferrer_rafael-embodied_cognition_applied_to_timbre_and_musical_appreciation_theoretical_foundation.pdf (2009)
70. Kao A, Poteet SR (eds.), *Natural Language Processing and Text Mining* (Springer Verlag, 2006)
71. C Manning, H Schütze, *Foundations of Statistical Natural Language Processing* (MIT Press, 2002)
72. S Bird, E Klein, E Loper, *Natural Language Processing with Python* (Oreilly & Associates Inc, 2009)
73. W Francis, H Kucera, *Brown Corpus. A Standard Corpus of Present-Day Edited American English, for use with Digital Computers* (Department of Linguistics, Brown University, Providence, Rhode Island, USA, 1979)
74. Fellbaum C (ed.), *WordNet: An electronic lexical database* (Language, speech, and communication, Cambridge, Mass: MIT Press, 1998)

doi:10.1186/1687-4722-2011-11
Cite this article as: Ferrer and Eerola: **Semantic structures of timbre emerging from social and acoustic descriptions of music.** *EURASIP Journal on Audio, Speech, and Music Processing* 2011 **2011**:11.

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Immediate publication on acceptance
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com

PV

LOOKING BEYOND GENRES: IDENTIFYING MEANINGFUL SEMANTIC LAYERS FROM TAGS IN ONLINE MUSIC COLLECTIONS

by

Ferrer, R. & Eerola, T. 2011

In Proceedings of the Tenth International Conference on Machine Learning and
Applications, (2), 112–117, Honolulu, Hawaii

2011 10th International Conference on Machine Learning and Applications

Looking Beyond Genres: Identifying Meaningful Semantic Layers from Tags in Online Music Collections

Rafael Ferrer and Tuomas Eerola
Finnish Centre of Excellence in Interdisciplinary Music Research
University of Jyväskylä, Finland
rafael.ferrer-flores@jyu.fi, tuomas.eerola@jyu.fi

***Abstract*—A scheme for identifying the semantic layers of music-related tags is presented. Arguments are provided why the applications of the tags cannot be effectively pursued without a reasonable understanding of their semantic qualities. The identification scheme consists of a set of filters. The first is related with social consensus, user-count ratio, and n-gram properties of tags. The next relies on look-up functions across multiple databases to determine the probable semantic layer of each tag. Examples of the semantic layers with prevalence rates are given based on application of the scheme to a subset of the Million Song Dataset. Finally, a validation of the results was carried out with an independent, smaller hand-annotated dataset, in which high agreement between the identification provided by the scheme and annotations was found.**

***Index Terms*—musical genre; social tags; semantic layers; music information retrieval**

I. Introduction

During the past five years in music information research, social tags of music have been capitalized in a variety of purposes, such as construction of auto-taggers, semantic structures of music, and as building blocks for music indexing and recommendation [1], [2], [3], [4]. Typical approaches include analyzing the tags with natural language processing methods to build semantic spaces that are then linked with the acoustic properties of the objects (songs or artists). One of the most prominent sources of social tags for the last decade has been *Last.fm* [5], which gives users the possibility of tagging musical objects in an unsupervised fashion. This sort of repositories have a high ecological value because they represent crowds knowledge in a form that has overwhelmed the empirical possibilities existing in controlled settings, given the massive amount of participants. However, this rather positive aspect of the available data does not come at any cost, as it also contains elements that reflect the diversity of human creativity. For this reason, researchers dealing with the problem of deriving any reasonably clean semantic structure from tags have relied in quantitative heuristics (typically taking the most frequent elements of a given corpus), or filtered out specific materials with the aid of controlled databases (e.g., finding the emotional layer with the support of emotional terms databases, [6]).

Understanding the implicit meaning embedded in tags is crucial to build an appropriate semantic space for any application. Outside the musical domain, systems such as *ClasTag* [7] have emerged to propose a solution to such need. Within the specific case of musical tags, no general classification system has been presented. Instead, ad hoc extraction of semantic layers has been performed (e.g., [6]) and tags not related with the chosen semantic layer have been discarded. This ad hoc approach might shadow an important part of the semantic space because it eliminates not only "untargeted" tags, but also relevant information concerned with the implicit semantic content embedded within the relations between the tags. A semantic layer is thus a set of interrelated concepts extracted from tags, such as genres, words related with emotions, artists names, etc. that can be used to filter qualitatively a given corpus of tags. Preliminary approaches in discriminating semantic layers of tags have been proposed by [8] in a set of 500 tags and extended by [9] in a set of 2 029 tags. We chose the term *semantic layer*, as it has the power to reflect that tags are embedded in a network of entities. An alternative terminology to semantic layers would be "semantic categories", however we wanted to emphasize that the output of our algorithm is a fuzzy classification, not a sharp one. Large scale databases contain considerable more distinct tags than the ones that are manually checked in these two example studies, therefore an automatic approach is required. These two examples of previous categorization of tags share few attributes, for instance *genre* as the most prevalent, *locale*, *instrument* in the same rank order.

It is worth mentioning that in the example of the 2 029 tags [9] there are significantly more suggested semantic categories than in the study carried out with 500 set [8]. This may indicate the need of more categories as the number of tags increases. For this reason, a hierarchy of expected categories should be approximated beforehand. Nevertheless, by doing that we should first answer the question about which kind of categories are the most relevant to characterize music at a meta-acoustical level. It is thus pertinent to question the value of musical genres for such a task, particularly since musical genres are, among other things, imposed constructs created by businessman to trade with music [10]. The same applies to the case of references to locales, these can have a high organizational value for an individual or a group within a social network, but might not be musically informative as the instrumentation.

978-0-7695-4607-0/11 $26.00 © 2011 IEEE
DOI 10.1109/ICMLA.2011.89

IEEE computer society

In general, it can be argued that non directly descriptive tags should not be filtered out in order to strengthen the statistical bond within tags, nevertheless it is difficult to determine whether such a strength is a source of noise when the objective is to learn about the music itself and not merely the contextual or social issues associated with it. Another issue of major concern is the practical impossibility of attributing polysemic tags to a single class, for example the tags *beat* and *swing* that can refer to a genre, artist name, verb, etc. Hence a distinction must be made between *synsemantic* and *autosemantic* tags i.e., between those that require a context to acquire a meaning and those capable of being self-explanatory.

In the present paper, we implement and illustrate strategies to work around the issues mentioned above in the form of an identification process. More specifically, we propose a set of rules dedicated to reach an automatic classification of social tags with the aid of online resources which will help to discern among the different semantic layers.

First we will introduce the experimental dataset (II-A), then explain a set of filters and attribution processes that hierarchically attempt to ascribe various semantic layers to each tag (II-B). After this, different strategies involved in aggregating the data to model a weighting system is presented (II-C). Finally, examples of the semantic layers with prevalence rates is shown (II-D), and the fitness of the classification is assessed with a dataset containing a manual classification [9] (II-D1).

II. Identification of Semantic Layers

A. Experimental Dataset

As a starting point, we chose the *musiXmatch* dataset, a subset of the *Million Song Dataset* (MSD) that includes lyrics [11], because it is the largest dataset currently available and also to maximize comparability with other ongoing studies. This dataset contained a list of 237 662 songs that served to retrieve a corpus of 3 621 778 tags (352 472 distinct) from *last.fm*. From these distinct tags 212 107 were hapax legomenon (i.e., tags that appeared only once in the corpus), which means that more than the half of the vocabulary of tags (60%) was truly unique. The number of users utilizing these distinct tags and the total use count has also been retrieved.

B. Filtering

Tags related to music, as available from collaborative non-supervised systems such as Last.fm, require an efficient form of noise reduction. Most approaches apply the *most frequent* rule (cf. [4], [1], [3]). These studies impose a threshold based on the number of times tags exist in a given corpus. The procedure cuts the long tail of the distribution, removing not only the assumed noise proceeding from what can be called accidents such as misspellings, mistaken attributions, individual's expressions, etc., but also putatively relevant tags. Furthermore, there is no evidence supporting that such a procedure will cut such noise in data which results from semantic ambiguity, often called *polysemy*.

The aim of the present work was to keep as many tags as possible and remove any incidental noise without relying solely on their frequencies. We assume that 237 662 songs are enough to retrieve an heterogeneous collection of tags, and that such a corpus can sufficiently represent the language of tags (or *tag-elese* [12]) (e.g., tags frequencies on the sample have a high correlation r=0.86 with the frequencies of the last.fm system). In this study, a tag is considered a single entry of the vocabulary disregarding the number of words that compose it (cf. [13]). Treating tags as units rather than as separate, single keywords has the advantage of shifting the focus from data processing to concept processing [14], thus allowing to study tags as *conceptual expressions* [15] instead of purely words or phrases.

To work around the limitations of the *most frequent* approach, we propose a multi-stage set of three novel filters that are designed to eliminate noise related to lack of majority opinion (*consensus* and *used-users* filters) and complex compound tags (*n-grams* filter).

1) Consensus filter: This filter is based on an loose interpretation of the cultural consensus theory [16], that to our knowledge has not been explicitly used before in the context of musical tags. The reference to this theory is made to support the assumption of an underlying cultural knowledge reflected in the agreement between taggers (or *informants* as introduced by Boster, see [17]). This is, tags that reach consensus are more than spontaneous expressions, therefore we need to make a distinction between those that function as memes and those that only have representational power in the individual's domain. The filter consists of setting a threshold of a minimum number of users applying a given tag. Interpreted as "supporters of the tag", this information is available as *built by n people* in the web page devoted to each tag in last.fm web site. As we tried to be rather liberal in our inclusion criterium, we set the threshold at 30 users (at the time of the retrieval of this info during March 2011) following Weller's (2007) recommendation. This filter reduced dramatically the number of distinct tags to be analyzed, from 352 472 to 28 873.

2) Used-users filter: A second filter focused on the ratio of *number of times used* and *built by* quantities. This process allows to detect such deviant tags that may be the result of an artificial application of tags. For instance a given tag used one time by one user (ratio of 1) is the starting point and it seems a natural tagging behavior. In contrast, a ratio of 5 103 -which is the case of the tag *wdzh-FM*, that had been used 35 726 times by only 7 users- seems the outcome of a small group of individuals or even an individual with seven accounts using some sort of automatic tagging. The ratios for the output of the consensus filter were typically low; this is, 97% of the computed ratios were below 20 (M=4.73, SD=14.7). For instance, one of the most popular tags is *rock*, the ratio of $times\ used/built\ by$ is 9.99. This is twice our reported mean, so we thought that a maximum ratio of 10 would be enough to remove the tags attributed by automatic means. While a minimum ratio of 2 would ensure a degree of consistence in the use of the tag (i.e., any tag having a ratio > 10 or < 2 was eliminated). After this filter, the number of distinct tags was reduced to 19 460.

3) N-grams filter: A third filter based on the distribution of n-grams (i.e., number of words composing the tags) was applied to reduce the complexity of the sample. A proportion of 95 % of the remaining (output from the second filter) tags were composed by one, two or three words. We set this filter to remove the tags composed by four words or more. This eliminated tags such as 'I need to remember to check this band out' and 'If you dont like this song theres just plain and simple something wrong'. This operation reduced the sample to 18 494 tags.

Figure 1 offers a visualization on how these three filters have affected the corpus and also how the *most frequent* approach would trim the data. The figure depicts a spectrum of tags frequencies within the corpus, associating them by classes. Hence the tags present only once in the corpus (212 107 *hapax legomena*) are grouped into class 1, the number of tags repeated twice correspond to class 2, and so on. The right end of the spectrum depicts high frequency classes, such as class 51 981, a frequency class possessed only by one tag which is also the most popular (i.e., *rock*). Our filter cuts some tags from the lowest frequency classes but without severely distorting the spectrum. On the contrary, when *the most frequent* approach is applied, radically different spectrum of tags would be obtained (shown in boxed area of Figure 1).

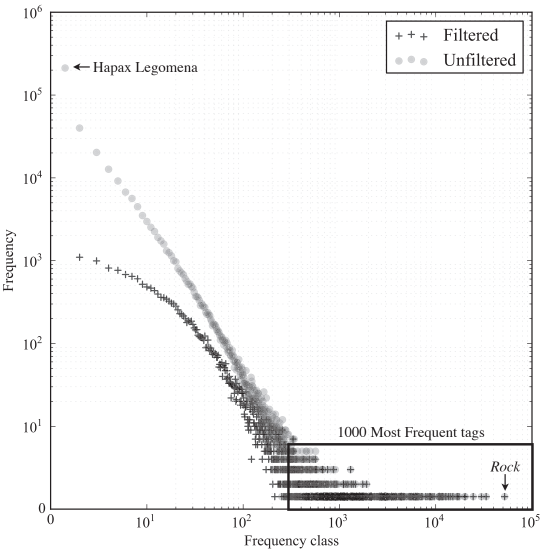


Fig. 1. Frequency spectrum of the corpus showing pre and post filter stages and the area of tags belonging to the pruning approach called *most frequent*.

With our multi-stage approach we have been careful to avoid the generalized assumption that unpopularity is a synonym of low relevance. We expected, however, that this procedure results in an efficient removal of incidental noise although the selected filter thresholds aimed at maximizing the heterogeneity on the sample. Note that the thresholds used on each of these filters could be more strict if what is needed is to derive a more compact list of tags.

C. Classifier

Distinct tags were extracted from the corpus and this set was de-noised based on a multi-stage filter focused on social relevance, frequency of use/number of users ratio, and number of words composing the tags. Noise, however, also has a semantic origin, so no matter how efficient a quantitative filter might be, a semantic filter is necessary. The difficulty is where to draw the line between what is noise and what is not noise semantically in the social context. For instance, words and their meanings are created and extensively used at the core of the cultures before they are finally codified in dictionaries. Therefore classification of this kind -and perhaps any kind- of corpus is challenged by the assumption of a predefined ontology, that is, we can not assert about the noisiness without the aid of external references. With respect to these points, the classification presented in this project is divided in two parts. The first part looks for exact matches of the tags in external repositories of (presumably) musically relevant entities, and the second part parses the tags not found in such repositories through a set of rules to discern their most probable classification.

1) External databases: Following the top semantic layers found in previous tag classifications [8], [9] as a model, we utilized a set of databases containing monotonic entities such as musical genres (here n=1 264 items), instruments (n=667 items), artists (n=72 665 items), geographical locations (n=14 214 items), affect terms (n=131 items). Additionally a multipurpose lexical database named *WordNet* [18], *Wikipedia* and Last.fm were used for reference. Genres and instrument databases contains lists of popular musical genres and instruments extracted from Wikipedia. The list of geographical locations is provided by *GeoDataSource* [19] and *MaxMind* [20] databases, and the list of affect terms is the by-product of empirical studies of the most common emotions words in the context of music [21], [22]. In the case of each external database, a match between the tag and the database entries was carried out. A portion (43%) of the tags to be classified (n=18 494) was found in at least one of the databases. The rest of the tags needed to be treated in a more refined fashion using different resources to discern their meaning.

2) Rules: To discern the attributes of the tag (T) we used the following set of rules.

- if T has a high Levenshtein similarity (ratio > 0.85) with any of the pre-classified tags, T inherits the attributes of the preclassified tag. The aim is to find tags close to already identified items (e.g., heavy metal, heavymetal, heavy/metal, etc.).
- if T has been defined in the last.fm wiki and the keywords "genre" or "style" are found, T acquires the attribute of *genre*. These tags that does not reach the formal definition of a genre in an expert's taxonomy, nevertheless are defined as such in the last.fm wiki. Typical examples include terms such as "abstract hip-hop" and "sophsti-

pop".

- if T has returned a page when searched in Wikipedia, the title of that page has a high similarity with T (Levenshtein ratio > 0.80) and the keywords "genre" or "style", "artist", "band" or "dance" are found, T acquires the attribute named as the found keyword. (e.g., "boogie", "headbanging")
- if T is composed by four digits or two digits followed by aphostrophe "s", T acquires the attribute of "Temporal reference" (e.g., 1989, or 90s).
- if T is found on the title of the track/song, artist, title of the album containing the track/song or the bag of words of the lyrics of the songs concerned with the tag, T is attributed with the "song title", "artist name", "album name" or "lyrics reference" category. (e.g., "Glashaus" is an artist name, song title, and also a reference to the album containing that song. Common words such as "death", "sky" or "war" can be typically included in the lyrics)
- if T is a bigram or trigram and contains a monogram very similar (Levenshtein similarity ratio > 0.75) to any of the monograms in the genres database, the tag is attributed as "genre reference". (e.g., "neo swing", "rock argento")
- if T is less than four characters, contains only punctuation symbols or repeats any character more than three times, the tag is marked as a candidate for deprecation. (e.g., "<3", "niiice", "vgm")
- if T is a monogram and can be found in Wordnet database, it acquires all the lexical categories derived from all the possible senses that the monogram can adopt in different contexts.
- T is parsed with the Brill's [23] part of speech (POS) tagger.

3) Post Filtering: When the set of rules (described in II-C2) is applied to the vocabulary of previously filtered tags (n=18 494), 90% of the tags were attributed with one or more of the different types (i.e.,*entities*) derived from the rules. Raw results were organized in a binary matrix $\mathbf{X} = \{x_{ij}\}$ (where i represents the set of distinct tags and j the distinct types of attributes) containing information about the linkage between tags and attributes. Next, this matrix was simplified by manually grouping and discarding certain attributes by focusing on their qualities. These were divided in two broad categories, namely: *musical* if the attribute was directly referencing a quality contained in the musical object, and *extra-musical* when the attribute was non directly related to the musical object. The musical category contains genres/styles, artists/band names, song and album titles and instruments names, and the extra-musical category comprises adjectives, affect terms, references to geographic locations, time and place, nouns referring to body parts, nature events and objects, food, shapes, etc. and verbs depicting motion, change, perception, communication, etc. Noticeable, the taxonomies of nouns and verbs are so diverse and profuse (more than 20% of the vocabulary), that they deserve a special attention. For this reason, attributes nouns and verbs were filtered out to prevent them from masking the musically relevant information. Only those attributes that contributed directly to represent the musical object or its context in a very concrete manner were kept. Hence, from the extra-musical attributes only adjectives, geographical locations, time and place, and affect were considered for further processing. This finer selection of attributes reduced the number of tags (n=12 386) to be included in the final output.

The semantic layers were defined as the grouping of selected attributes, which is another way to refer to the simplification operated on the raw binary matrix. The reduction (or simplification) of $\mathbf{X}$ is expressed as a transformation to $\mathbf{Y} = \{y_{ik}\}$ (where i represents the tags and k the layers) and defined with:

$$\mathbf{Y} = x_{\sum i,j \in G} \tag{1}$$

where G refers to the groups of attributes defining the layers listed below:

- *Adjective:* Wordnet relational adjective + Wordnet noun denoting attribute of people and objects + Wordnet adverb + Brill POS adjective for monogram.
- *Affect:* affect term + Wordnet noun denoting feelings and emotions + Wordnet verb of feeling.
- *Album:* T is he name of the album containing the track that was tagged with T.
- *Artist:* artist, band or group as defined in Last.fm + Wikipedia + MSD unique artists list + T is the name of the artist that produced the track from which the tag was retrieved.
- *Genre:* genre, style and keywords found in definitions in Last.fm + Wikipedia + Wikipedia musical genres + any term from Wikipedia musical genres found within T.
- *Geographic:* Geodata + Maxmind + Wikipedia popular cities + Wordnet noun spatial position.
- *Instrument:* MUMS subfamily + Wikipedia list of instruments.
- *Lyrics:* T is a term found in the lyrics of the song that is tagging.
- *Time-place:* Wordnet noun denoting stable states of affairs + Wordnet noun describing time and temporal relations + Wordnet noun denoting acts or actions + years and decades found with *regular expressions*.
- *Songtitle:* T is the title of the song that is tagging.

Each of the attributes had attached a binary value, so after the reduction, layers had different maximums. For this reason, the values had to be normalized with $y_{i,k}/max(y_{i*k})$. This is, each value of the matrix divided by the maximum per row. The result is a matrix containing classification scores for each layer and tag. For example, for the tag *Swing*, the weights for the semantic layer are as follows: genre=1.00, time-place=0.66, artist=0.66, adjective= 0.33, lyrics=0.33 and songtitle = 0.33, and the rest of the categories have a score of zero. Note that from this representation, it is also easy to derive the most probable layer of a tag, if needed.

The whole scheme is summarized in Figure 2.

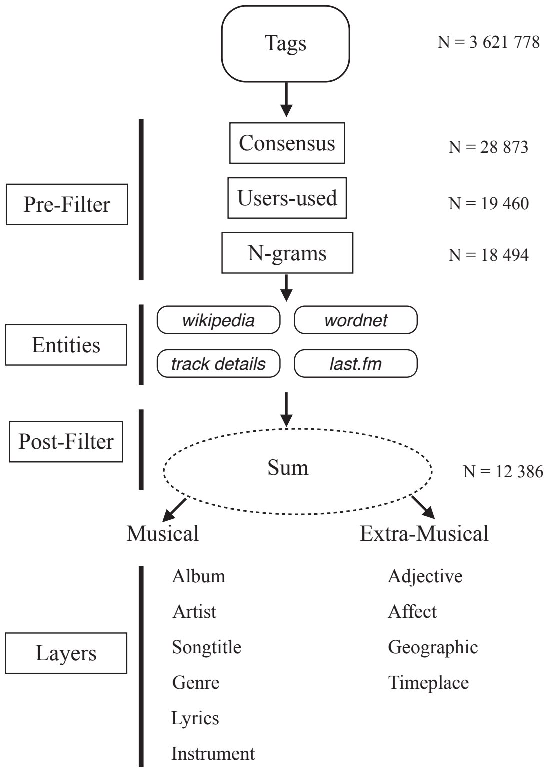


Fig. 2. Diagram of the automatic layer detection process.

Category	Layer	Percentage
Musical	Genre	31.8%
	Artist	24.8%
	Songtitle	9.2%
	Lyrics	4.0%
	Album	3.8%
	Instrument	0.6%
Extra-musical	Adjective	10.1%
	Time-place	9.2%
	Affect	2.2%
	Geographic	4.2%

TABLE I
PROPORTION OF TAGS IN THE SEMANTIC LAYERS (N=12 386)

D. Results

12 386 tags acquired a fuzzy semantic definition by means of a score that linked them to the proposed, main semantic layers. This lead to a 83.6% of the sample being unambiguously assigned to a single layer, 11.9% to two layer, 3.4% to three and the remaining 1.1% to four to seven layers. Some applications of tags might benefit from the fuzzy classification scores and some might require a strict taxonomic classification with no overlapping layers. Where required, the proportion of weights derived from the layers (see Table I) can be used as a model to discern the class of a given tag if this has been found as a primary choice in two or more layers. The method consists in multiplying the tied scores by their corresponding proportions and then normalizing them to one. Through this method we could attribute a strict taxonomic classification to the ambiguous tags.

We offer the resulting semantic layer of musical tags as freely available[1]. This dataset will be useful for research capitalizing on semantic qualities of the tags, such as attempts to predict artist, genre, mood, or geographical location of songs from audio or from other forms of data. As such, the dataset represents a benchmark for further refinements of the semantic layer identification, a task that still has a considerable room for improvement.

1) Validation: To assess the validity of the results, a manually annotated dataset was used (see [9]). It consists of a list of 2 029 tags classified according to 14 categories. Making the comparison was not straightforward because only 7 categories were comparable at a conceptual level, an issue that also reduced the number of tags available for comparison to 1 288. The procedure consisted in using the manual classification as a ground truth and testing the accuracy of our system to predict the right classification. After assigning a single category to each of the 1 288 compatible tags by using the proportions model (explained in II-D), our system was able to predict with an acceptable accuracy the manual annotations. The confusion matrix with the detailed results is displayed in Table II.

	Adj.	Aff.	Art.	Gen.	Geo.	Ins.	Tim.
Adjective	107	19	10	9	3	0	2
Affect	0	15	1	3	0	0	2
Artist	0	0	129	5	1	0	2
Genre	20	0	18	719	0	0	12
Geographic	48	0	4	14	55	0	0
Instrument	3	0	2	18	1	20	1
Timeplace	1	0	1	6	0	0	37

TABLE II
CONFUSION MATRIX FOR MANUAL VERSUS AUTOMATIC CLASSIFICATION (N=1 288)

III. CONCLUSIONS

In our opinion, no significant advances in the analysis of social metadata and tags can be made without a closer scrutiny of their underlying semantic qualities. For this reason, we present a way to organize tags according to entities relevant to the musical domain – such as genre, artist, mood, verb, adjective, verb or location to name the most important ones– that we call *semantic layers*. We argue that a plausible discrimination of these layers embedded into tagging activities is sorely needed in order to build more reliable and sophisticated applications based on tags.

The implementation consists of a multi-stage model that attributes most tags to a meaningful semantic layer. This model has a filtering stage, attribution stage, and a summarization

[1] https://www.jyu.fi/music/coe/materials/tags/byondg

stage. In the first stage, new filters were proposed based on consensus, user-count and word combination statistics that will eliminate some of the inherent noise in tagged data. However, we do not encourage to filter tags purely on their their frequency, since the result will often reflect the most common tags, which have the least novel and discriminative information to offer. In other words, this strategy will stress the tags that are so common that they will have little discrimination power (such as "Rock"). Instead, we promote the idea of relying on tags that are neither too popular nor too rare.

We also eliminate the basic descriptors of the track details (artist, title, lyrics), which will help to focus on those aspects of tags that are the actual semantic contributions rather mere duplications of the lyrics and title. To our knowledge, this has not been done before.

Our proposed analysis procedure handles the tags either by assigning them to classes directly or disambiguating them by means of scores. When a tag is tied to one or more categories, a weighting provided by the combination of entities seems to offer a balanced solution. That is, the grouping of different entities results in the reinforcement of each group as a super-group by aggregating weight in terms of the sources given to each group.

In the single-category classification few controversies might emerge, as it is evident in the validation of the results. This should not be immediately considered as an error because of the multiple meanings a tag can acquire. Within the language of tags, words acquire a meaning slightly different than in their original languages. Thus, for instance when taggers use the word "rock", what they actually mean is far from the English meaning of the word referring to a solid mineral material. In the musical context, "rock" often refers to a musical (timbre, rhythm, melody, etc.) or extra musical (style, cultural background, historical period, etc.) quality, rather than concrete things outside the musical domain. We could argue that the meaning of the tags can only be found in the mesh of relations with other tags. This is, their meaning is implicit and could not be explicit because the multidimensional characteristic of the phenomenon they are trying to represent. Our approach provides a formalization of such characteristic because it shows the degree of belonging to each entity in the semantic layer, and we believe this has an obvious advantage over the single categorization approach.

Although further studies about the reliability of the proposed scheme is needed, a first validation with an unrelated, hand-annotated dataset provided highly similar pattern of results. This suggest that the filtering and the identification scheme proposed here is a useful tool for making sense of rich social data provided online music collections. Furthermore, understanding the semantic characteristics of musical tags is a crucial -and often oversimplified- aspect that is at the foundation of any machine learning application in the music information retrieval context.

Acknowledgment

Finnish Centre of Excellence in Interdisciplinary Music Research

References

[1] M. Levy and M. Sandler, "Music information retrieval using social tags and audio," *IEEE Transactions on Multimedia*, vol. 11, no. 3, pp. 383–395, 2009.

[2] L. Chen, P. Wright, and W. Nejdl, "Improving music genre classification using collaborative tagging data," in *Proc. 2nd ACM Int. Conf. on Web Search and Data Mining (WSDM)*. ACM, 2009, pp. 84–93.

[3] C. Baccigalupo, E. Plaza, and J. Donaldson, "Uncovering Affinity of Artists to Multiple Genres From Social Behaviour Data," in *Proc. 9th Int. Conf. on Music Information Retrieval (ISMIR)*, 2008, pp. 275–280.

[4] T. Bertin-Mahieux, D. Eck, F. Maillet, and P. Lamere, "Autotagger: A model for predicting social tags from acoustic features on large music databases," *Journal of New Music Research*, vol. 37, no. 2, pp. 115–135, 2008.

[5] "Api – last.fm," http://www.last.fm/api.

[6] C. Laurier, M. Sordo, J. Serrà, and P. Herrera, "Music mood representation from social tags," in *Proceedings of the 10th International Society for Music Information Conference*, Kobe, Japan, October 2009.

[7] S. Overell, B. Sigurbjörnsson, and R. Van Zwol, "Classifying tags using open content resources," in *Proceedings of the Second ACM International Conference on Web Search and Data Mining*. ACM, 2009, pp. 64–73.

[8] P. Lamere, "Social tagging and music information retrieval," *Journal of New Music Research*, vol. 37, no. 2, pp. 101–114, 2008.

[9] R. Ferrer and T. Eerola, "Timbral qualities of semantic structures of music," in *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, Netherlands, August 2010, pp. 571–576.

[10] J. Lena and R. Peterson, "Classification as culture: Types and trajectories of music genres," *American Sociological Review*, vol. 73, no. 5, pp. 697–718, 2008.

[11] T. Bertin-Mahieux, D. Ellis, W. B., and P. Lamere, "The million song dataset," in *Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR 2011)*, 2011, (submitted).

[12] J. Simons, "Tag-elese or the language of tags," *Fibreculture Journal*, vol. 12, 2008. [Online]. Available: http://www.journal.fibreculture.org/issue12/issue12_simons.html

[13] H. Halpin, V. Robu, and H. Shepherd, "The complex dynamics of collaborative tagging," in *Proceedings of the 16th international conference on World Wide Webal Conference on World Wide Web*. Banff, Alberta, Canada: ACM, May 2007, p. 220.

[14] J. Brank, M. Grobelnik, and D. Mladenic, "Automatic evaluation of ontologies," in *Natural Language Processing and Text Mining*, A. Kao and S. R.Poteet, Eds. USA: Springer, 2007.

[15] J. Siskind, "Learning word-to-meaning mappings," *Models of Language Acquisition: Inductive and deductive approaches*, pp. 121–153, 2000.

[16] S. Weller, "Cultural consensus theory: Applications and frequently asked questions," *Field Methods*, vol. 19, no. 4, pp. 339–368, 2007.

[17] A. Romney, S. Weller, and W. Batchelder, "Culture as consensus: A theory of culture and informant accuracy," *American anthropologist*, pp. 313–338, 1986.

[18] C. Fellbaum, Ed., *WordNet: An electronic lexical database*, ser. Language, speech, and communication. Cambridge, Mass: MIT Press, 1998.

[19] "Geodatasource," http://www.geodatasource.com.

[20] "Maxmind - free world cities database," http://www.maxmind.com/app/worldcities.

[21] P. Juslin and P. Laukka, "Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening," *Journal of New Music Research*, vol. 33, no. 3, pp. 217–238, 2004.

[22] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: Differentiation, classification, and measurement," *Emotion*, vol. 8, no. 4, pp. 494–521, 2008.

[23] E. Brill, "A simple rule-based part of speech tagger," in *Proceedings of the third conference on Applied natural language processing*. Association for Computational Linguistics, 1992, pp. 152–155.

PVI

ENHANCING GENRE-BASED MEASURES OF MUSIC PREFERENCE BY USER-DEFINED LIKING AND SOCIAL TAGS

by

Ferrer, R., Eerola, T. & Vuoskoski, J.K. (in press)

Psychology of Music

PVII

AMP: ARTIST-BASED MUSICAL PREFERENCES DERIVED FROM FREE VERBAL RESPONSES AND SOCIAL TAGS

by

Ferrer, R. & Eerola, T. 2011

In Proceedings of the IEEE International Conference on Multimedia and Expo
(ICME), 1–6, Barcelona, Spain

AMP: ARTIST-BASED MUSICAL PREFERENCES DERIVED FROM FREE VERBAL RESPONSES AND SOCIAL TAGS

Rafael Ferrer and Tuomas Eerola

Finnish Centre of Excellence in Interdisciplinary Music Research
University of Jyväskylä, Finland
rafael.ferrer-flores@jyu.fi tuomas.eerola@jyu.fi

ABSTRACT

Operational definitions of music preferences are at the core of psychological research exploring individual differences (personality, expertise) and their relation to a variety of musical behaviors. However, the measurement instruments for music preferences mostly rely on subjective likings for genres, a notion known to be problematic in several ways. We present a framework to derive music preferences based on free responses about liked and disliked artists. The framework utilizes social tags and online databases to aggregate comparable data to the existing genre-based measures. This framework was tested in a sample of 408 participants, who indicated their musical preferences using a genre-based measure and free textual responses. A comparison of both forms of data suggested that a genre-based measure can be reliably recovered from the free responses using the framework. The framework has the advantage of being ecologically valid and flexible in terms of the possible inputs and outputs.

Index Terms— music preference, social tags, genres, artists, language processing, instrument development

1. INTRODUCTION

Instruments for measuring musical preferences are required for a host of applications in the most diverse fields related to music, ranging from industrial (i.e. music recommendation) to psychological and social research. Despite the urge to improve the accuracy for such instruments, most of the existing ones are based on musical genres [1, 2], which are known to be problematic [3, 4]. Some of the problems are related to the (a) hierarchical nature of musical genres, (b) prevalence of artists and songs which resist such categorization, (c) lack of knowledge for many individuals about the exact genre classification of their preferred music, and (d), cultural dependency and constant redefinition of musical genres. Despite these different problems, an instrument for quickly capturing the essence of musical preferences is critically important for many branches of music research. For instance, the widely cited results explaining the musical choices in terms of listener personalities [5, 1] rely on rating 15 musical genres in terms liking for each one of them. Each subsequent research [2, 6, 7] has modified the list of genres to better reflect the pool of participants in question, leading rather different patterns of results and underlying preference dimensions.

There are also other ways of obtaining information about the musical preferences of individuals. For instance, the actual music listening behavior provides highly valid perspective to musical preferences, and this can take the form of analyzing the digital playlists of individuals or using experiments where the selection is done specifically (using the Operant Music Listening Recorder, see [8]). One can also ask the listeners to describe freely what kinds of music they like, which is a natural task and comes effortlessly from most individuals. We will focus on this last method, since it provides the richest and the most natural way of describing ones own preferences.

In this paper we present a framework to recover an individual's musical preference profile from free verbal responses about liked and disliked items (i.e., artists, genres, adjectives, etc.). The method utilizes social tags and online databases in hierarchical manner in inferring the appropriate genres for the particular liked and disliked items.

Previous research about the analysis of social tags has rapidly accumulated and spawned into different directions, from methodological considerations [9, 10], to descriptions of the semantic space underlying tags [11], enhancing genre classification by utilizing the metadata provided by the tags [12], to the definitions of affinities between genres and artists [13] and to predictions of tags from the acoustic characteristics of songs [14, 15]. These studies have produced considerable insights into the possibilities offered by these vast and dynamic datasets. However, the social tags and the different analysis techniques and approaches applied to them have not yet been connected to such research that already sought to establish links between the actual preferences, individual differences and other behaviour related to music [1, 7]. To take such studies into a larger context of music consumption and to advance in the understanding about the role of musical preferences for these issues, we see a direct need to improve the way music preferences are measured. We will especially focus on overcoming the need for people to know particular

978-1-61284-350-6/11/$26.00 ©2011 IEEE

genres in order to define what they actually like by allowing responses in a more natural form, that is, by writing and listing favorite artists, songs and bands. We also propose that descriptions of disliked music in a similar form will also be valuable information about the music preferences of the individuals.

2. ANALYSIS OF FREE RESPONSES

We propose a framework that measures music preference by allowing free textual input in the form of description of liked and disliked terms (typically artists and genres) separated by commas. Consequently, these data are not constrained to a particular set of types (artists, songs, or genres), nor homogeneously formatted. For this reason it is, however, necessary to implement a system to retrieve and classify the free responses. This system is implemented in three major modules: pre-processor for breaking up the strings and homogenizing them, classifier including a lookup methods from separate online databases, and post-processor to aggregate the responses. These modules will be explained next.

2.1. Pre-processor

Pre-processor performs an initial cleaning and filtering of the raw input for each participant's (P) disliked (D) and liked (L) responses. Thus the purpose of the pre-processor is to define a survey S as a set of tuples $P(D, L)$, where D and L are lists of strings.

Strings of S are mostly artist names and musical genres separated by commas, but also sentences expressing personal opinions and uses of music. Some of these have objects enclosed in quotation marks and parenthesis or additional objects separated by dots, colons and semicolons.

Initially, the module splits long strings where it find commas. The output is passed to a filter that removes items of less than two characters in length, transforms all the input into lowercase and splits the string again where it finds a dot, colon, semicolon, or an expression enclosed in parenthesis and quotation marks. Next, the strings pass through a filter for remaining long strings (i.e., composed by 5 words or more); this looks for artists existing in a local cache (specifications are explained in the next section) within the string, if found, it is removed from the long string and added as an additional item. It is worth noting that the system favors artists names over other types of objects.

The pre-processor works at a very basic level, for instance the raw input "iskelmämusiikki (anne mattilat ja kumppanit jne.)" is transformed into "iskelmämusiikki, anne mattila".

2.2. Classifier

A classification of the strings is needed for a finer discrimination of the material in qualitative terms and to reduce the noise inherent to social tags [9] and free responses. S contains unique tuples (i.e., combinations of participant and liked and disliked items). Within the tuples several items of D and L are shared among the different P. To avoid redundancy, the contents of D and L of all P are combined in a single set that we treat as a corpus c. From c it is possible to extract a vocabulary V of unique collocations (i.e., combinations of words such as artists names) and compute their frequencies F (i.e., the number of times each item is repeated through the corpus).

The classification is performed $\forall x \in V$ according to a hierarchical list of conditions that simultaneously refer to the certainty of the classification. If the condition is not fulfilled in a given level, then it is passed to the next level. If the condition is true, x is tagged according to the level name. The level name is composed by two characters depicting the type ('G', 'A', 'P' or 'M' for Genre, Artist, Popular unknown term, Mixture of known terms, respectively) and an index of certainty (0 is extremely certain, 8 is less certain). The hierarchy is listed next:

G0: $x \in$ local Genres db
A0: $x \in$ local Artist db
A1: $x \in 3$ online Artist db
A2: $x \in 2$ online Artist db
A3: x not found. Top search result is identical in 3 online Artist db, match $\in$ local Artist db
A4: x not found. Top search result is identical in 2 online Artist db, match $\in$ local Artist db
A5: x not found. Top search result is identical in 3 online Artist db, match $\notin$ local Artist db
A6: x not found. Top search result is identical in 2 online Artist db, match $\notin$ local Artist db
A7: $x \in 1$ online Artist db, if length $x > 6$ characters $\Rightarrow$ correct spelling, corrected $\in$ local Artist db
P0: $x \in 1$ online Artist db $\wedge F \geq 2$
P1: $x \in 1$ online Artist db $\wedge$ listeners > 10000
P2: $x \in 1$ online Artist db
A8: x contains a known artist (plus noise)
G1: x contains a known genre (plus noise)
M0: x contains a mixture of known items (plus noise)

Local Genres db is a set of 746 popular musical genres discerned manually from last.fm tags. Local Artist db is a set of 2141 artists/band names. The online Artist databases correspond to three different public services: Last.fm [16], Echo Nest [17], and MusicBrainZ [18]. An additional index of popularity named "listeners" is used in P1; it refers to the total number of listeners for a given artist within the last.fm system.

The output of the classifier is a mapping of V as $x : \{class, y\}$, where y is a corrected x (e.g., if the class of a given x is A3, y is to the top search result; if x is classified as A1 $y = x$; if x could not be classified, y is empty; etc.). In other words, the classifier also gives information about the level of certainty of the classification that may be used to define the desired level of fidelity of the module.

2.3. Post-processor

With the contents of V classified, the resulting mapping is used to return the data to the original form of S (i.e., $P(D, L)$). Survey contents are typified with different levels of certainty and unclassifiable strings are removed. In this way the data becomes homogeneous and is ready for further processing. Note that at this stage one can also constrain the data according to the desired level of certainty by stipulating, for example, that only level 6+ in certainty, for instance, will be considered in constructing the final list of artist attributions. In a similar vein, the processing may be constrained to one particular class of data (artists or genres).

3. ARTIST-BASED MUSICAL PREFERENCES (AMP)

Deriving a musical preference measure that is compatible with the previous notions of preferences (genre-based measures) from the homogeneous lists of liked and disliked artists, requires this information to be projected into the social tag space and be formalized as genre-based measures similar to STOMP [1]. This construction, which we will call *Artist-based Musical Preferences* (AMP) measure, is comparable to any existing genre-based measure, but would be created from social tags and the free listing of liked and disliked artists. Such a measure has the benefits of the traditional genre-based measure, namely, (1) fixed number of dimensions ready for comparison with other variables, and (2) the possibility of performing comparisons with a large number of important previous studies using such measures, but also two significant advantages, namely, (1) avoiding the problem of predefining the level of description in musical preferences, and (2) offering a faster and more natural way of describing ones music preferences.

AMP comprises of the following processes: 1) project $P(D, L)$ into the social tags space $P_{DL}(T)$, 2) project the intended AMP model $P(I, r)$ where I is a set of concepts (e.g., musical genres, adjectives, etc.) and r is their corresponding weight (as emulating ratings in a Likert scale) into the social tags space $P_{AMP}(T)$, 3) assign the weights of $P_{DL}(T)$ to $P_{AMP}(T)$, 4) transform $P_{AMP}(T)$ into $P(I, r)$. Each step will be explained next.

To obtain $P_{DL}(T)$, the top tags given to the artists a listed on D and L are retrieved through the Last.fm *artist.getTopTags* method of the public API. The output $\forall a \in D \cup L$, are sets of unique tags "...ordered by popularity" [16] $a(t_W)$ of variable length n. The objective is to derive a single list of unique tags with weights T, so all $a(t_w)$ need to be aggregated. To this end, the inverse ordinal rank $(n + 1) - w$ is used as a weight w', and the concepts of liked and disliked are processed here; if $a \in D$, w' is multiplied by -1. Some tags may be present in two or more sets so their weights are summed.

For the projection of the intended model $P(I, r)$, the tags similar to each *concept* I (concepts are genres in this particular example of AMP application) are retrieved using the Last.fm *tag.getSimilar* method of the public API. The output is a set of 50 unique tags "...ranked by similarity, based on listening data" [16] per concept $i(t_w)$. A weight w' is assigned to all tags, hence the projection can be expressed as $P_{AMP}(T) = I(t_{w'})$. In the third part of the process, the weights of $P_{DL}(T)$ are multiplied with the weights of $P_{AMP}(T)$ at the intersection $P_{DL}(T) \cap P_{AMP}(T)$. The final step in AMP collapses $P_{AMP}(T)$ into the model $P(I, r)$. The predicted rating r for each concept of I is obtained with $\sum i(t_{w'})$, hence producing a standard and comparable genre-based measure of musical preferences.

A diagrammatic summary of the elements in AMP and the process of comparison are given in Figure 1.

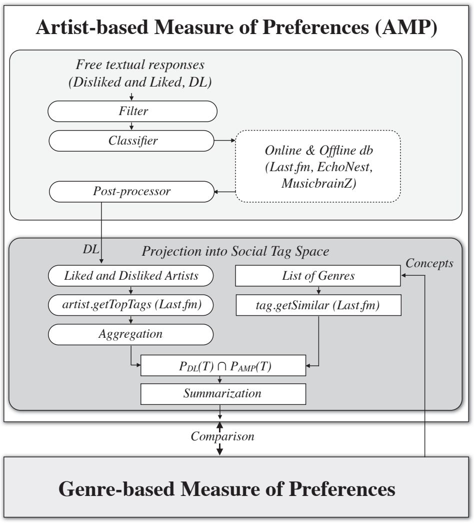

Artist-based Measure of Preferences (AMP)
Free textual responses (Disliked and Liked, DL)
Filter
Classifier
Online & Offline db (Last.fm, EchoNest, MusicbrainZ)
Post-processor
DL
Projection into Social Tag Space
Concepts
Liked and Disliked Artists
List of Genres
artist.getTopTags (Last.fm)
tag.getSimilar (Last.fm)
Aggregation
$P_{DL}(T) \cap P_{AMP}(T)$
Summarization
Comparison
Genre-based Measure of Preferences

Fig. 1. Summary of the processes involved in AMP.

4. EXPERIMENT

An experiment was designed to compare an existing genre-based measure with the results obtained with AMP, which relies on free responses. To this end, we used a survey containing a genre-based measure and free responses from the same pool of participants. The free responses were transformed by the AMP to emulate the results of a genre-based measure of musical preferences. To assess the fitness of the method, the correlations between these two measures (free-responses fed into AMP and genre-based measure) were calculated.

Table 1. Correlations (r) between AMP and genre-based measure with different minimal number of artists required (*a*).

a	N	*r*
1	397	0.47*
2	251	0.48*
3	156	0.49*
4	94	0.52*
5	54	0.54*
6	34	0.53*

* p<0.01, *df*=18

4.1. Materials

A web-based survey was administered to 408 university students (82.1% female, 17.9% male), their mean age was 23.87 years (SD 4.35). The survey contained a modified version of the STOMP [1], in which *Country & Western* was replaced with *Iskelmä* (a particularly popular Finnish national genre), *Dance* for *Electronic*, *Heavy Metal* for *Heavy*, *Gospel* was removed and *Hip-Hop*, *Indie*, *Punk*, *Reggae*, *RnB*, *Singer-Songwriter*, and *World* were added, yielding 20 genres altogether. The additions to the original STOMP were motivated by a pilot study where the most frequently mentioned genres were obtained from 346 students of this particular population. In the actual survey, the respondents were asked to indicate their liking for each genre using a Likert scale from 1 to 7 (where 1 = 'Dislike strongly' and 7 = 'Like strongly'). In addition, the survey contained open-ended questions about liked and disliked artists, bands and genres, to explore aspects of musical preferences not captured by the genre-based items.

4.2. Results

All respondents completed the genre-based measure and free responses of liked and disliked artists. From these, 397 included at least one liked and one disliked artist (discernible by the classifier at 'A0' level). However, as the minimum number of artists increases, the number of participants fulfilling that criteria is reduced considerably (e.g., $l(a) \geq \{2, 3, 4\}, N = \{251, 156, 94\}$, see Table 1). To find an optimal criteria, we replicated the comparison of the genre-based measure and AMP with different $l(a)$ values. The results, summarized in Table 1, suggest that even 1 liked and 1 disliked artists may be enough to give reliable results. Beyond 5 artists, the overall results are not dramatically changed, so we conclude that less than 5 is more than enough to be fed into the AMP method to provide an estimation of the genre-based preferences based on the items.

In sum, the overall match between the two forms of data is moderately strong (*r*=0.50, p<0.01) and suggests that they indeed tap into the same underlying construct. However, it is easy to find individuals in the data who receive low fit between these two measures. In order to understand whether these mismatches stem from poor comparability or processing errors, these were explored in detail. It seems that in such cases the information provided by genres and individual items simply do not match well.

We will illustrate this with an individual whose correlation between AMP and genre-based measure was negative (*r=-0.26*).This person has reported liking 'Habakuk', 'dc Talk' and 'Think Twice' and disliking 'Lordi', 'Michael Jackson' and 'Iron Maiden'. Her ratings of the 20 genres are as follows (from 1 to 7, where the latter is strongly liked): *rock* (7), *classical* (7), *soundtrack* (7), *pop* (6), *jazz* (5), *world* (5), *blues* (5), *rnb* (4), *singer-songwriter* (4), *indie* (4), *soul* (3), *hip-hop* (3), *rap* (3), *iskelmä* (3), *alternative* (3), *folk* (2), *punk* (2), *reggae* (2), *metal* (1), *electronic* (1). By comparing the two forms of the data, a clear mismatch arises in the case of the disliked item 'Michael Jackson' and the high rating given for *pop*. The same happens with 'Habakuk' and the low rating given for reggae. Michael Jackson has often been called the prince of pop, and 'Habakuk' has been one of the most famous Polish reggae groups. So even if the rest of the artists are in concordance with the genre ratings, these two examples give rise to distortion between the measures.

Another example, close to the mean correlation (*r*=0.46), mentions liking 'Nightwish', 'Kotiteollisuus', 'Children of Bodom', 'Tarja Turunen' and 'Sarah Brightman', and disliking 'Elastinen', 'Common', 'Movetron'. The genre ratings for the individual are: *metal* (7), *rock* (6), *classical* (6), *soundtrack* (6), *jazz* (5), *iskelmä* (4), *pop* (4), *rnb* (4), *singer-songwriter* (4), *indie* (4), *punk* (4), *reggae* (4), *folk* (3), *world* (3), *blues* (3), *electronic* (3), *soul* (2), *alternative* (2), *hip-hop* (1), *rap* (1). This example shows a clear concordance between both measurements because 'Elastinen' is a typical representative of rap in Finland, and 'Common' has won a Grammy award as a rapper. 'Movetron' has been tagged as *eurodance*, so it is close to the *electronic* genre. The list of liked artists demonstrates high concordance with the likings for genres, since 'Tarja Turunen', the ex-vocalist of 'Nightwish' are identified with the *symphonic metal* style, 'Children of Bodom' and 'Kotiteollisus' with *metal*, and 'Sarah Brightman' as a representative of a *classical* music, all receiving high genre-based ratings.

Finally, an example of the highest match (*r*=0.90) between the measures is displayed in Table 2 to demonstrate the coherence between the measures. It is worth noting that this level of high correspondence was obtained by an individual indicating liking 'HIM', 'Disturbed', 'Kotiteollisuus', 'Lapko', 'Disco Ensemble', 'In Flames', 'Muse', and 'Metallica' and disliking 'Enrique Iglesias', 'Britney Spears', 'Bob Marley', 'Finlanders', and 'Janne Tulkki'.

5. DISCUSSION

A method for constructing a genre-based measure from free responses was proposed. This method utilized a series of

Table 2. An example of the high correspondence (r=0.90) between Genre-based Measure and AMP.

Genre	Score	AMP	Genre	Score	AMP
Alternative	7	6	*Pop*	3	4
Blues	1	2	*Punk*	6	4
Classical	2	1	*Rap*	1	1
Electronic	2	3	*Reggae*	1	2
Folk	1	2	*Rnb*	2	2
Hip-Hop	1	1	*Rock*	7	6
Indie	5	5	*Songwriter*	2	3
Iskelmä	1	2	*Soul*	2	2
Jazz	1	2	*Soundtrack*	3	2
Metal	7	7	*World*	1	1

data processing modules and drew on social media to attribute artists and their popularity into the queries from the free responses. A direct comparison of the constructed method and an existing genre-based measure yielded promising results despite the possible drawbacks associated with both measures.

Although the performance of the AMP was adequate in most cases, the comparison did reveal some of the problems inherent in genre-based measures [3, 13]. It is perfectly possible for people to like songs and artists and dislike the genres these artists are commonly thought to represent. This concern is a more serious one for genre-based measures, since it forces people to choose from rather broad categories whereas picking particular favorite artists may more easily carry distinctive information about the subtle aspects of an individual's musical preferences. Since this is an issue where concept (here genres) frequencies obtained from online databases may be useful for further development, this issue will be discussed next from a perspective of analysis of language properties.

5.1. Optimal concepts for AMP or genre-based measures

The results of AMP and genre-based measure also give insights into the optimal ways of constructing novel genre-based measures since the choice of genres is a property that can be optimized in terms of their discriminatory power; genres that are too specific or too generic will not have large impact on the definition of music preferences. This aspect becomes evident when the comparative information about genres is obtained from social media. The social tags are not only providing fuzzy sets of terms describing each musical artist or song, but as a whole, they represent also a kind of language used by millions of people with the purpose of describing their musical experience. As such, the language of tags is susceptible to an analysis of language properties [19]. This brings new implications, for example, for the possibility of discerning between tags that are not very useful without a context (e.g., 'rock') or tags which are not practical because of their rarity (e.g., 'Northern Finnish Goth Metal with Arctic Angst').

To illustrate this idea, we took the corpus of tags from the survey and estimated various statistical properties of the corpus. The highest frequency corresponded to 'rock', which is present 819 times, followed by 'alternative' with 756. These two examples might be the equivalent of the most frequent words in the English language: 'the' and 'of'. Indeed, 'rock' as a musical descriptor is many things and used in many contexts, just as the word 'the'. Hence, without a proper context, these words are not very informative, and thus are not particularly useful in a genre-based measure of music preferences. Words appearing only once in the spectrum are named as *hapaxes*, this corpus has 13,250 of them, representing 70% of the vocabulary. This is an indication of the richness of the vocabulary, but also shows the individuality of musical descriptions that are not useful in terms of an instrument measuring musical preferences.

How to establish the optimal musical descriptor then? From the spectrum of the corpus, we can estimate the *k-point* [20], which is a fuzzy referent dividing the words that are *autosemantic* (semantically independent) from those that are *synsemantic* (words that need a context to acquire meaning). In this material, the *k-point* is localized in the class 8, meaning that tags whose frequency is close to 8 might be good candidates for informative descriptors because of their balance between the need of a context and self-definition.

In this corpus, there were 119 terms whose frequency are 8. These could be further filtered to suit particular research interests. Here we find genre-related tags such as *Suomi pop*, *chamber music*, *progressive jazz*, which refer to more subtle characterization of musical genres than those found in a common genre-based instrument which collapses all jazz into one genre and similarly with the classical music. Any of these terms near the *k-point*, a selected subset, or all, could be used to derive an AMP.

5.2. Other concepts and input for AMP

It is worth pointing out that the AMP could also be taken to represent the underlying dimensions of music preferences. For instance, in previous studies of music preferences [2, 1], the following four factors identified, namely (1) reflective and complex, (2) intense and rebellious, (3) upbeat and conventional, and (4) energetic and rhythmic. In these previous studies, the musical genres have been mapped into these four factors (e.g., 'classical', 'blues', 'folk', and 'jazz' are considered to represent reflective and complex, and 'alternative', 'rock', and 'heavy' represent intense and rebellious) and, subsequently, the results of such a reduction are used in interpreting the preferences in a more generic way than merely reflecting the individual preferences for musical genres. Such a reduction is also easily imposed to AMP since it involves a linear mapping of genres to the underlying dimensions. It is also possible to focus on entirely other concepts of preferences by focusing on adjectives given in the free responses since these are also prevalent as tags in online databases. Another interesting idea is to use the actual user data aggregated from playlists (iTunes, iPods, Online services, Spotify) as input to AMP. In such a case the preprocessing can be largely eliminated but still be useful to define a homogeneous output according to the needs of the measurement.

6. CONCLUSIONS

The AMP framework provides a method for measuring an individuals musical preference in an effortless yet reliable way, thus offering an alternative to the common genre-based measures of musical preferences. This alternative measure avoids many of the pitfalls associated with the predefined lists of genres yet produces comparable standardized information usable for comparison with other psychometric constructs (personality, empathy, discrimination ability, etc.). It must be noted, however, that the AMP measure is still under development. For instance, a more careful evaluation of the need for keeping in both the liked and disliked items, or how the popularity of the artists or the social context contribute to the results, is needed. One fruitful avenue for future work could also entail implementation of AMP with song names instead of artists.

7. REFERENCES

[1] P.J. Rentfrow and S.D. Gosling, "The do re mi's of everyday life: the structure and personality correlates of music preferences," *Journal of Personality and Social Psychology*, vol. 84, no. 6, pp. 1236–1256, 2003.

[2] M.J. Delsing, T.F.M. ter Bogt, R.C. Engels, and W.H. Meeus, "Adolescents music preferences and personality characteristics," *European Journal of Personality*, vol. 22, no. 2, pp. 109–130, 2008.

[3] J.J. Aucouturier and E. Pampalk, "Introduction-From Genres to Tags: A Little Epistemology of Music Information Retrieval Research," *Journal of New Music Research*, vol. 37, no. 2, pp. 87–92, 2008.

[4] M. Sordo, O. Celma, M. Blech, and E. Guaus, "The Quest for Musical Genres: Do the Experts and the Wisdom of Crowds Agree?," in *Proc. 9th Int. Conf. on Music Information Retrieval (ISMIR)*, 2008, pp. 255–260.

[5] P.J. Rentfrow and S.D. Gosling, "Message in a ballad: The role of music preferences in interpersonal perception," *Psychological Science*, vol. 17, no. 3, pp. 236–242, 2006.

[6] K.D. Schwartz and G.T. Fouts, "Music preferences, personality style, and developmental issues of adolescents," *Journal of Youth and Adolescence*, vol. 32, no. 3, pp. 205–213, 2003.

[7] A.C. North, L. Desborough, and L. Skarstein, "Musical preference, deviance, and attitudes towards music celebrities," *Personality and individual differences*, vol. 38, no. 8, pp. 1903–1914, 2005.

[8] A.C. North and D.J. Hargreaves, "Musical preferences during and after relaxation and exercise," *The American Journal of Psychology*, vol. 113, no. 1, pp. 43–67, 2000.

[9] P. Lamere, "Social tagging and music information retrieval," *Journal of New Music Research*, vol. 37, no. 2, pp. 101–114, 2008.

[10] D. Turnbull, L. Barrington, and G. Lanckriet, "Five approaches to collecting tags for music," in *Proc. 8th Int. Conf. on Music Information Retrieval (ISMIR)*, 2008, pp. 225–230.

[11] M. Levy and M. Sandler, "Music information retrieval using social tags and audio," *IEEE Transactions on Multimedia*, vol. 11, no. 3, pp. 383–395, 2009.

[12] L. Chen, P. Wright, and W. Nejdl, "Improving music genre classification using collaborative tagging data," in *Proc. 2nd ACM Int. Conf. on Web Search and Data Mining (WSDM)*. ACM, 2009, pp. 84–93.

[13] C. Baccigalupo, E. Plaza, and J. Donaldson, "Uncovering Affinity of Artists to Multiple Genres From Social Behaviour Data," in *Proc. 9th Int. Conf. on Music Information Retrieval (ISMIR)*, 2008, pp. 275–280.

[14] T. Bertin-Mahieux, D. Eck, F. Maillet, and P. Lamere, "Autotagger: A model for predicting social tags from acoustic features on large music databases," *Journal of New Music Research*, vol. 37, no. 2, pp. 115–135, 2008.

[15] D. Eck, P. Lamere, T. Bertin-Mahieux, and S. Green, "Automatic generation of social tags for music recommendation," *Advances in neural information processing systems*, vol. 20, pp. 385–392, 2007.

[16] "Api – last.fm," `http://www.last.fm/api`.

[17] "Echo nest api overview - the echo nest v4.2 documentation," `http://developer.echonest.com/docs/v4/`.

[18] "Xml web service - musicbrainz," `http://musicbrainz.org/doc/XML_Web_Service`.

[19] J. Simons, "Tag-elese or the language of tags," *Fibreculture Journal*, vol. 12, 2008.

[20] I.I. Popescu and G. Altmann, *Word frequency studies*, Walter de Gruyter, Berlin, 2009.

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. Rolʹ russkogo fennoilʹstva. 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.

43 Penttinen, Esa Martti, Kielioppi virheiden varjossa. Kielitiedon merkitys lukion saksan kieliopin opetuksessa. - Grammar in the shadow of mistakes. The role of linguistic knowledge in general upper secondary school German grammar instruction. 153 p. Summary 2 p. Zusammenfassung 3 p. 2005.

44 Kaivapalu, Annekatrin, Lähdekieli kielenoppimisen apuna. - Contribution of L1 to foreign language acquisition. 348 p. Summary 7 p. 2005.

45 Salavuo, Miikka,Verkkoavusteinen opiskelu yliopiston musiikkikasvatuksen opiskelukulttuurissa - Network-assisted learning in the learning culture of university music education. 317 p. Summary 5 p. 2005.

46 Maijala, Juha, Maaseutuyhteisön kriisi-1930-luvun pula ja pakkohuutokaupat paikallisena ilmiönä Kalajokilaaksossa. - Agricultural society in crisis – the depression of the 1930s and compulsory sales as a local phenomenon in the basin of the Kalajoki-river. 242 p. Summary 4 p. 2005.

47 Jouhki, Jukka, Imagining the Other. Orientalism and occidentalism in Tamil-European relations in South India. -Tulkintoja Toiseudesta. Orientalismi ja oksidentalismi tamileiden ja eurooppalaisten välisissä suhteissa Etelä-Intiassa. 233 p. Yhteenveto 2 p. 2006.

48 Lehto, Keijo, Aatteista arkeen. Suomalaisten seitsenpäiväisten sanomalehtien linjapapereiden synty ja muutos 1971–2005. - From ideologies to everyday life. Editorial principles of Finnish newspapers, 1971–2005. 499 p. Summary 3 p. 2006.

49 Valtonen, Hannu, Tavallisesta kuriositeetiksi. Kahden Keski-Suomen Ilmailumuseon Messerschmitt Bf 109 -lentokoneen museoarvo. - From Commonplace to curiosity – The Museum value of two Messerschmitt Bf 109 -aircraft at the Central Finland Aviation Museum. 104 p. 2006.

50 Kallinen, Kari, Towards a comprehensive theory of musical emotions. A multi-dimensional research approach and some empirical findings. - Kohti kokonaisvaltaista teoriaa musiikillisista emootioista. Moniulotteinen tutkimuslähestymistapa ja empiirisiä havaintoja. 71 p. (200 p.) Yhteenveto 2 p. 2006.

51 Iskanius, Sanna, Venäjänkielisten maahanmuuttajaopiskelijoiden kieli-identiteetti. - Language and identity of Russian-speaking students in Finland. 264 p. Summary 5 p. Реферат 6 с. 2006.

52 Heinänen, Seija, Käsityö – taide – teollisuus. Näkemyksiä käsityöstä taideteollisuuteen 1900-luvun alun ammatti- ja aikakauslehdissä. - Craft – Art – Industry: From craft to industrial art in the views of magazines and trade publications of the early 20th Century. 403 p. Summary 7 p. 2006.

53 Kaivapalu, Annekatrin & Pruuli, Külvi (eds), Lähivertailuja 17. - Close comparisons. 254 p. 2006.

54 Alatalo, Pirjo, Directive functions in intra-corporate cross-border email interaction. - Direktiiviset funktiot monikansallisen yrityksen englanninkielisessä sisäisessä sähköpostiviestinnässä. 471 p. Yhteenveto 3 p. 2006.

55 Kisantal, Tamás, „…egy tömegmészárlásról mi értelmes dolgot lehetne elmondani?" Az ábrázolásmód mint történelemkoncepció a holokauszt-irodalomban. - "...there is nothing intelligent to say about a massacre". The representational method as a conception of history in the holocaust-literature. 203 p. Summary 4 p. 2006.

56 Matikainen, Satu, Great Britain, British Jews, and the international protection of Romanian Jews, 1900-1914: A study of Jewish diplomacy and minority rights. - Britannia, Britannian juutalaiset ja Romanian juutalaisten kansainvälinen suojelu, 1900–1914: Tutkimus juutalaisesta diplomatiasta ja vähemmistöoikeuksista. 237 p. Yhteenveto 7 p. 2006.

57 Hänninen, Kirsi, Visiosta toimintaan. Museoiden ympäristökasvatus sosiokulttuurisena jatkumona, säätelymekanismina ja innovatiivisena viestintänä. - From vision to action. Environmental education in museums as a socio-cultural continuum, regulating mechanism, and as innovative communication 278 p. Summary 6 p. 2006.

58 Joensuu, Sanna, Kaksi kuvaa työntekijästä. Sisäisen viestinnän opit ja postmoderni näkökulma. - Two images of an employee; internal communication doctrines from a postmodern perspective. 225 p. Summary 9 p. 2006.

59 Koskimäki, Jouni, Happiness is… a good transcription - Reconsidering the Beatles sheet music publications. - Onni on… hyvä transkriptio – Beatles-nuottijulkaisut uudelleen arvioituna. 55 p. (320 p. + CD). Yhteenveto 2 p. 2006.

60 Hietaharju, Mikko, Valokuvan voi repiä. Valokuvan rakenne-elementit, käyttöympäristöt sekä valokuvatulkinnan syntyminen. - Tearing a photograph. Compositional elements, contexts and the birth of the interpretation. 255 p. Summary 5 p. 2006.

61 Jämsänen, auli, Matrikkelitaiteilijaksi valikoituminen. Suomen Kuvaamataiteilijat -hakuteoksen (1943) kriteerit. - Prerequisites for being listed in a biographical encyclopedia criteria for the Finnish Artists Encyclopedia of 1943. 285 p. Summary 4 p. 2006.

62 Hokkanen, Markku, Quests for Health in Colonial Society. Scottish missionaries and medical culture in the Northern Malawi region, 1875-1930. 519 p. Yhteenveto 9 p. 2006.

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.

- ”Powerful, Useful and Wasteful”. Discourses of Nuclear Weapons in the New York Times and Helsingin Sanomat 1945–1998. 337 p. Summary 7 p. 2007.

85 Kantanen, Helena, Stakeholder dialogue and regional engagement in the context of higher education. - Yliopistojen sidosryhmävuoropuhelu ja alueellinen sitoutuminen. 209 p. Yhteenveto 8 p. 2007.

86 Almonkari, Merja, Jännittäminen opiskelun puheviestintätilanteissa. - Social anxiety in study-related communication situations. 204 p. Summary 4 p. 2007.

87 Valentini, Chiara, Promoting the European Union. Comparative analysis of EU communication strategies in Finland and in Italy. 159 p. (282 p.) 2008.

88 Pulkkinen, Hannu, Uutisten arkkitehtuuri - Sanomalehden ulkoasun rakenteiden järjestys ja jousto. - The Architecture of news. Order and flexibility of newspaper design structures. 280 p. Yhteenveto 5 p. 2008.

89 Meriläinen, Merja, Monenlaiset oppijat englanninkielisessä kielikylpyopetuksessa - rakennusaineita opetusjärjestelyjen tueksi. - Diverse Children in English Immersion: Tools for Supporting Teaching Arrangements. 197 p. 2008.

90 Vares, Mari, The question of Western Hungary/Burgenland, 1918-1923. A territorial question in the context of national and international policy. - Länsi-Unkarin/Burgenlandin kysymys 1918–1923. Aluekysymys kansallisen ja kansainvälisen politiikan kontekstissa. 328 p. Yhteenveto 8 p. 2008.

91 Ala-Ruona, Esa, Alkuarviointi kliinisenä käytäntönä psyykkisesti oireilevien asiakkaiden musiikkiterapiassa – strategioita, menetelmiä ja apukeinoja. – Initial assessment as a clinical procedure in music therapy of clients with mental health problems – strategies, methods and tools. 155 p. 2008.

92 Oravala, Juha, Kohti elokuvallista ajattelua. Virtuaalisen todellisen ontologia Gilles Deleuzen ja Jean-Luc Godardin elokuvakäsityksissä. - Towards cinematic thinking. The ontology of the virtually real in Gilles Deleuze’s and Jean-Luc Godard’s conceptions of cinema. 184 p. Summary 6 p. 2008.

93 Kecskeméti, István, Papyruksesta megabitteihin. Arkisto- ja valokuvakokoelmien konservoinnin prosessin hallinta. - From papyrus to megabytes: Conservation management of archival and photographic collections. 277 p. 2008.

94 Suni, Minna, Toista kieltä vuorovaikutuksessa. Kielellisten resurssien jakaminen toisen kielen omaksumisen alkuvaiheessa. - Second language in interaction: sharing linguistic resources in the early stage of second language acquisition. 251 p. Summary 9 p. 2008.

95 N. Pál, József, Modernség, progresszió, Ady Endre és az Ady–Rákosi vita. Egy konfliktusos eszmetörténeti pozíció természete és következményei. 203 p. Summary 3 p. 2008.

96 Bartis, Imre, „Az igazság ismérve az, hogy igaz”. Etika és nemzeti identitás Sütő András Anyám könnyű álmot ígér című művében és annak recepciójában. 173 p. Summary 4 p. 2008.

97 Ranta-Meyer, Tuire, Nulla dies sine linea. Avauksia Erkki Melartinin vaikutteisiin, verkostoihin ja vastaanottoon henkilö- ja reseptiohistoriallisena tutkimuksena. - *Nulla dies sine linea*: A biographical and reception-historical approach to Finnish composer Erkki Melartin. 68 p. Summary 6 p. 2008.

98 Koivisto, Keijo, Itsenäisen Suomen kanta-aliupseeriston synty, koulutus, rekrytointitausta ja palvelusehdot. - The rise, education, the background of recruitment and conditions of service of the non-commissioned officers in independent Finland. 300 p. Summary 7 p. 2008.

99 Kiss, Miklós, Between narrative and cognitive approaches. Film theory of non-linearity applied to Hungarian movies. 198 p. 2008.

100 Ruusunen, Aimo, Todeksi uskottua. Kansandemokraattinen Neuvostoliitto-journalismi rajapinnan tulkkina vuosina1964–1973. - Believed to be true. Reporting on the USSR as interpretation of a boundary surface in pro-communist partisan journalism 1964–1973. 311 p. Summary 4 p. 2008.

101 Härmälä, Marita, Riittääkö *Ett ögonblick* näytöksi merkonomilta edellytetystä kielitaidosta? Kielitaidon arviointi aikuisten näyttötutkinnoissa. – Is *Ett ögonblick* a sufficient demonstration of the language skills required in the qualification of business and administration? Language assessment in competence-based qualifications for adults. 318 p. Summary 4 p. 2008.

102 Coelho, Jacques, The vision of the cyclops. From painting to video ways of seeing in the 20th century and through the eyes of Man Ray. 538 p. 2008.

103 Brewis, Kielo, Stress in the multi-ethnic customer contacts of the Finnish civil servants: Developing critical pragmatic intercultural professionals. – Stressin kokemus suomalaisten viranomaisten monietnisissä asiakaskontakteissa: kriittis-pragmaattisen kulttuurienvälisen ammattitaidon kehittäminen. 299 p. Yhteenveto 4 p. 2008.

104 Belik, Zhanna, The Peshekhonovs’ Workshop: The Heritage in Icon Painting. 239 p. [Russian]. Summary 7 p. 2008.

105 Moilanen, Laura-Kristiina, Talonpoikaisuus, säädyllisyys ja suomalaisuus 1800- ja 1900-lukujen vaihteen suomenkielisen proosan kertomana. – Peasant values, estate society and the Finnish in late nineteenth- and early

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.

174 Vuoskoski, Jonna Katariina, Emotions represented and induced by music: The role of individual differences. – Yksilöllisten erojen merkitys musiikillisten emootioiden havaitsemisessa ja kokemisessa. 58 p. (132 p.) Tiivistelmä 1 p. 2012.

175 Leinonen, Jukka, The beginning of the cold war as a phenomenon of realpolitik – U.S. secretary of state James F. Byrnes in the field of power politics 1945–1947. – Kylmän sodan synty reaalipoliittisena ilmiönä – James F. Byrnes suurvaltapolitiikan pelikentällä Jaltasta Stuttgartiin 1945–1947. 393 p. Yhteenveto 8 p. 2012.

176 Thompson, Marc, The application of motion capture to embodied music cognition research. - Liikkeenkaappausteknologian soveltaminen kehollisen musiikkikognition tutkimuksessa. 86 p. (165 p.) Yhteenveto 1 p. 2012.

177 Ferrer, rafael, The socially distributed cognition of musical timbre: a convergence of semantic, perceptual, and acoustic aspects. - Musiikillisen sointivärin jakautunut kognitio. 42 p. (156 p.) Yhteenveto 1 p. 2012.