TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Jukka Holm

Visualizing Music Collections Based on Metadata: Concepts, User Studies and Design Implications

Julkaisu 1068 • Publication 1068

Tampere 2012

Tampereen teknillinen yliopisto. Julkaisu 1068
Tampere University of Technology. Publication 1068

Jukka Holm

Visualizing Music Collections Based on Metadata: Concepts, User Studies and Design Implications

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB109, at Tampere University of Technology, on the 5th of October 2012, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2012

ISBN 978-952-15-2897-2 (printed)
ISBN 978-952-15-2908-5 (PDF)
ISSN 1459-2045

Abstract

Modern digital music services and applications enable easy access to vast online and local music collections. To differentiate from their competitors, software developers should aim to design novel, interesting, entertaining, and easy-to-use user interfaces (UIs) and interaction methods for accessing the music collections. One potential approach is to replace or complement the textual lists with static, dynamic, adaptive, and/or interactive visualizations of selected musical attributes. A well-designed visualization has the potential to make interaction with a service or an application an entertaining and intuitive experience, and it can also improve the usability and efficiency of the system.

This doctoral thesis belongs to the intersection of the fields of human-computer interaction (HCI), music information retrieval (MIR), and information visualization (Infovis). HCI studies the design, implementation and evaluation of interactive computing systems; MIR focuses on the different strategies for helping users seek music or music-related information; and Infovis studies the use of visual representations of abstract data to amplify cognition.

The purpose of the thesis is to explore the feasibility of visualizing music collections based on three types of musical metadata: musical genre, tempo, and the release year of the music. More specifically, the research goal is to study which visual variables and structures are best suitable for representing the metadata, and how the visualizations can be used in the design of novel UIs for music player applications, including music recommendation systems. The research takes a user-centered and constructive design-science approach, and covers all the different aspects of interaction design: understanding the users, the prototype design, and the evaluation.

The performance of the different visualizations from the user perspective was studied in a series of online surveys with 51-104 (mostly Finnish) participants. In addition to tempo and release year, five different visualization methods (colors, icons, fonts, emoticons and avatars) for representing musical genres were investigated. Based on the results, promising ways to represent tempo include the number of objects, shapes with a varying number of corners, and *y*-axis location combined with some other visual variable or clear labeling. Promising ways to represent the release year include lightness and the perceived location on the *z*- or *x*-axis. In the case of genres, the most successful method was the avatars, which used elements from the other methods and required the most screen estate.

In the second part of the thesis, three interactive prototype applications (avatars, potentiometers and a virtual world) focusing on visualizing musical genres were designed and evaluated with 40-41 Finnish participants. While the concepts had great potential for complementing traditional text-based music applications, they were too simple and restricted to replace them in longer-term use. Especially the lack of textual search functionality was seen as a major shortcoming.

Based on the results of the thesis, it is possible to design recognizable, acceptable, entertaining, and easy-to-use (especially genre) visualizations with certain limitations. Important factors include, e.g., the used metadata vocabulary (e.g., set of musical genres) and visual variables/structures; preferred music discovery mode; available screen estate; and the target culture of the visualizations.

Preface

This work was started at Nokia Research Center (NRC), Tampere, in 2007-2009 and finalized at the Unit of Human-Centered Technology (IHTE) at Tampere University of Technology (TUT) during 2009-2012.

First and foremost, I would like to thank my thesis supervisors Prof. Kaisa Väänänen-Vainio-Mattila (TUT) and Dr. Harri Siirtola (University of Tampere). Kaisa accepted me as her post-graduate and believed in my work when the value of the research was not that clear to some other institutes. I also want to thank Harri for co-authoring three publications with me and for his experience and help on statistics and information visualization. I would like to thank the pre-examiners of this thesis, Associate Professor Sally Jo Cunningham and Professor Tuomas Eerola, for their review of the manuscript, and Professor Emeritus Yngve Sundblad and Professor Petri Toiviainen for being my opponents at the public examination.

This thesis would not have been possible without the help of several Nokia colleagues and friends. First, I want to thank Arto Lehtiniemi for the possibility to do user studies on the SuperMusic prototypes and for co-authoring the related publications. Dr. Antti Eronen co-authored the avatar paper and checked the MIR part of this thesis, and Jarno Seppänen participated in the writing of the emoticons paper. All three also did a great job in our SuperMusic team. I also want to express my gratitude to the rest of the SuperMusic team / Dublin graveyard crew (Mikko Heikkinen, Timo Kosonen and Marko Takanen) for developing the best music player in the world.

In addition, I would like to thank Mr. Kz Havukainen for all those EJS years and our performances at NRC parties. Dr. Antti Aaltonen helped me to start my work on visualizing music collections and co-authored the first papers. I am grateful to Dr. Juha Arrasvuori for his general support and especially all the great MCG work that we did together during the years. I also thank my brother Harri Holm for the icon and emoticon graphics that he did for free (surfing trips to Hawaii do not count) outside office hours.

Lauri Laaksonen from Idean Enterprises Oy designed the avatar graphics that played an important role in my thesis, and his work is gratefully acknowledged.

I wish to thank my ex-bosses Matti Hämäläinen, Kari Laurila, Mauri Väänänen, and Jyri Huopaniemi for the encouraging and rewarding working atmosphere at NRC. Matti let me work on what I love the most (music) for many years, and Kari gave me permission to use the avatar designs and other important documents in my post-graduate studies after I had already left NRC. Last but not least, I want to express my gratitude to the upper Nokia management for letting me leave the company voluntarily – without you I would never have started my post-graduate studies.

Funding and financial support by UCIT (Poika Isokoski and Jenni Anttonen), Nokia Foundation Scholarship (2009 and 2011), and TeliaSonera Finland Oyj Research and Education Foundation Scholarship 2011 is gratefully acknowledged. UCIT paid for my three conference trips, and the TeliaSonera and Nokia grants helped me to cover my basic living costs during 2009-2011.

Finally, my biggest thanks go to my three ladies (Katja, Siiri and Maija) for their love and support, for being there when needed, and especially for not being there when I needed time to write this thesis.

Tampere, 11.1.2012 Jukka Holm

Supervisors:	Professor Kaisa Väänänen-Vainio-Mattila Department of Software Systems – Human-Centered Technology Tampere University of Technology
	Ph.D. Harri Siirtola School of Information Sciences – Tampere Unit for Computer-Human Interaction University of Tampere
Pre-examiners:	Associate Professor Sally Jo Cunningham Department of Computer Science University of Waikato, New Zealand
	Professor Tuomas Eerola Department of Music University of Jyväskylä
Opponents:	Professor Emeritus Yngve Sundblad School of Computer Science and Communication KTH Royal Institute of Technology, Sweden
	Professor Petri Toiviainen Department of Music University of Jyväskylä

Contents

List of original publications

This thesis is based on the following publications, which are reproduced here with permission. The publications are referred to in the text by **P1**, **P2**, and so forth.

P1 Holm, J., & Aaltonen, A. (2007). Associating graphical objects with musical tempo. *Proceedings of the 2nd Audio Mostly Conference* (pp. 174-179). Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany, 27-28 September. (online proceedings http://www.audiomostly.com/amc2007)

P2 Holm, J., & Aaltonen, A. (2008). Associating graphical objects with release year of music. *Proceedings of the 3rd IASTED International Conference on Human-Computer Interaction* (pp. 218-223). ACTA Press.

P3 Holm, J., Aaltonen, A., & Siirtola, H. (2009). Associating colours with musical genres. *Journal of New Music Research, 38(1)*, 87-100. Routledge.

P4 Holm, J., Holm, H., & Seppänen, J. (2010). Associating emoticons with musical genres. *Proceedings of the International Conference on New Interfaces for Musical Expression*. University of Technology Sydney, Australia, 7-11 June, 2010. (online proceedings http://www.educ.dab.uts.edu.au/nime/PROCEEDINGS/)

P5 Holm, J., Siirtola, H., & Laaksonen, L. (2010). Associating avatars with musical genres. *Proceedings of the 14th International Conference on Information Visualisation* (pp. 186-193). IEEE Computer Society.

P6 Holm, J., Lehtiniemi, A., & Eronen, A. (2010). Evaluating an avatar-based user interface for discovering new music. *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*. ACM Press.

P7 Holm, J., & Lehtiniemi, A. (2011). A virtual world prototype for interacting with a music collection. *Proceedings of the 4th International Conference on Online Communities and Social Computing, held as Part of HCI International 2011* (pp. 326-335). Springer-Verlag.

P8 Lehtiniemi, A., & Holm, J. (2011). Evaluating a potentiometer-based graphical user interface for discovering new music. *Proceedings of the 15th International Conference on Information Visualisation* (pp. 110-118). IEEE Computer Society.

P9 Holm, J., & Siirtola, H. (2012). A comparison of methods for visualizing musical genres. *Proceedings of the 16th International Conference on Information Visualisation* (pp. 636-645). IEEE Computer Society.

The publications are not included in the electronic version of the dissertation.

1. Introduction

This chapter establishes the background and motivation of the research, presents the scope and research questions of the thesis, summarizes the included original publications and the main contribution of the thesis, and gives an overview of the structure of the thesis.

1.1 Background and motivation

Music has become an important and ubiquitous element in our daily lives. The consumption of music can vary from active listening to selected songs to exposure to background music and sounds in, e.g., a shopping mall or a busy city street. Music can be listened to alone or in a social context, and it can be used for various purposes including relaxation, motivation, mood enhancement, and moderating arousal levels. For several decades, the music industry took benefit of all of this, grew bigger and bigger, and in the end most of the money was collected by a few multinational corporations and mega-artists. Then came the "laptop" or "wired generation" (Kot 2009) which was familiar with the Internet and peer-to-peer networks, and everything started to change.

After the introduction of Napster in 1999, the sales of physical music recordings started to decrease rapidly. According to RIAA reports, in 1999-2009 US album sales were falling on the average 8% per year (Goldmann 2010). In the end of 2009, the sales per capita were down 64% from their peak in 1999; Americans were now buying only slightly more than one album per person per year (DeGusta 2011). A similar phenomenon can also be noticed in other countries, including Finland, where the sales of recorded music (excluding physical video recordings) decreased by 13% from 2008 to 2009 (Liikkanen 2010). However, by the end of 2011, US sales had risen by 3%, indicating that the music business may have finally "hit the bottom" (Pulley 2011).

While music consumption on the web is increasing, so far it has been difficult to monetize. Only a few companies, such as Apple, have been successful with downloads. In the US, the sales of downloaded albums and singles have increased greatly since the introduction of the iTunes Store in 2003, but the numbers are not large enough to offset the losses of physical products (DeGusta 2011). In Finland, the trade value of digital downloads increased 24.5% from 2008 to 2009, but only 7% of the 2009 sales came from digital products (Liikkanen 2010). One recent phenomenon has been cloud-based "storage lockers" (Kirn 2011a), which can be used for storing one's (at least in theory legally purchased) music content, from which it can be downloaded to any Internet-connected device. While several companies such as Apple, Google, and Amazon have already released such services, their commercial success still remains to be seen.

Downloading is not the optimal solution for every person or every age group. During the last few years, there has been a significant change in consumer behavior, and the current generation of adolescents is not as interested in collecting and owning audio-visual content as their parents once were. For these consumers, streamed and cloud-based content is the way to go. Streaming music services have become immensely popular in both Europe and the US; it has also been shown that such services lead to both increased music consumption and an increase in the range of music consumed (Music Industry Report 2010). In the end of 2011, one of the most successful services, Spotify (www.spotify.com), had already approx. 20 million free and paid users in Europe and the United States (Van Buskirk 2011). This number was likely to increase further due to Spotify's recently published collaboration with Facebook. According to a forecast by the analyst firm Gartner (Andrews 2011), *"unlimited-access music subscription services will take almost a third of consumers' digital music spending by 2015."*

Modern music consumers have easy access to vast digital music collections containing millions of songs. In the end of 2010, record companies had already licensed 13 million tracks to online music services, and there were over 400 licensed digital music services worldwide (IFPI 2011). As making the content available to users is no longer an issue, the emphasis is now moving more and more to inventing new service concepts and improving the user experience of existing products. To differentiate from their competitors and reduce the risk of mixing brands, music service developers should aim to design novel, interesting, entertaining, and easy-to-use user interfaces (UIs) and interaction methods for exploring the vast music collections. Still, most companies are "playing it safe" and relying on textual lists and pictures of album front covers. In such UIs, users may browse, navigate, filter, and search the music collection according to selected metadata attributes such as artist and song names, musical genre, tempo, mood, and the release year of the music.

However, a huge volume of music easily overwhelms the user if he/she does not know where to start from or what to look for. The traditional browsing and searching methods may be insufficient to maintain an overview of the collection (Goussevskaia, Kuhn, & Lorenzi 2008). Especially in the case of mobile devices, browsing through long text lists may become a boring and time-consuming task, and the process does not encourage the user in any way to (re)discover "forgotten" songs in the collection. Several studies such as (Seyerlehner et al. 2009) and (Celma 2008) have shown that many songs of a music recommendation system are not reachable by browsing; they stay hidden in the so-called Long Tail (Anderson 2006). The same also applies to local (personal) music collections. For example, in a study of 5000 iPod users (Celma 2007) it was found that 23% of songs were played 80% of time and 64% of songs were never played. One solution to this is to use random/shuffle play or so-called smart playlists that are created automatically and contain only songs that match certain criteria, but it has been shown that they are not very popular in everyday use (Holm 2008).

One potential alternative for representing and accessing digital music collections is to replace or complement the textual lists with static, dynamic, adaptive, and/or interactive visualizations. In this type of graphical user interfaces (GUIs), selected visual attributes are mapped to the musical metadata of a user's personal music collection, an online music store, or a music recommendation service. By clicking on the desired part(s) of the visualization, the user can quickly start exploring or browsing the music collection, select what kind of music to listen to, or generate smart playlists or receive new recommendations; or else the visualizations can simply be used to show additional information on the screen in an aesthetic and entertaining manner.

A well-designed visualization can make the interaction with a music collection an entertaining, enjoyable, playful, and intuitive experience, and greatly help the user in finding similar songs from the collection. If designed properly, such visualizations can minimize the need to remember things and the need for explicit user input, increase the amount of useful information on the screen, and improve the usability and efficiency of the system (Hoashi et al. 2009). By relying on clear associations between visual and musical attributes, the user's cognitive load can be reduced; thus the users do not have to be musical experts to access the music collections. In the case of mobile phones and other small devices with limited screen estate, visualizations can be used to complement textual lists and show additional 1-*N* dimensional information on the screen. An aesthetic design can also result in improved usability; as shown by Norman (2004), "attractive things work better."

1.2 Scope and research questions of the thesis

This thesis studies different methods for visualizing large online or local music collections, subsets of the collections, and single music tracks. The topic is related to the fields of human-computer interaction (HCI), music information retrieval (MIR), and information visualization (Infovis).

Human-computer interaction (HCI) is *"a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them"* (Hewett et al. 1992). The usefulness, usability, and user experience (UX) of such systems can be improved by utilizing proper interaction design methods, which consist of understanding users, designing prototypes, and evaluating the designs.

Music information retrieval (MIR) is an interdisciplinary research field studying different strategies for enabling access to music collections (Casey et al. 2008). Instead of using textual lists, an interesting alternative for accessing the collections is to transform the data into some type of visual form, i.e., rely on information visualization (Infovis). In the MIR literature, music collections have been visualized in diverse ways including networks, "geographical" maps, rainbows, 3D spirals, blobs surrounding genre-label magnets, discs, and so forth. However, so far the research has been largely systems-focused (Weigl & Gustavino 2011). Most papers have concentrated on describing new algorithms and reporting their (automatically calculated) performance, instead of testing the GUIs with real users. Understandably, focusing on the users has been identified as one of the five most important MIR challenges for the next ten years (Downie, Byrd, & Crawford 2009).

The work presented in this thesis takes a user-centered and constructive approach to visualizing music collections; it involves all the different aspects of interaction design (understanding users, designing prototypes and evaluating the designs). The goal was to study the feasibility of visualizing musical metadata ("data about data") especially in the context of music recommendation services; design novel visualizations and GUIs; and evaluate the prototypes with real users. All the proposed visualization methods can be implemented based on existing MP3 ID3v2 (http://id3.org) or other musical metadata, and thus the studied methods do not require the development of new MIR signal processing algorithms.

The research topic was approached through three research questions:

RQ1: Can musical genres and other musical metadata be visualized in a recognizable and acceptable way?

RQ2: Which visual variables and structures are best suited to represent the metadata, and what is their performance from the user perspective?

RQ3: How can the genre visualizations be used in the design of novel graphical user interfaces for music player applications, and which visualization techniques are acceptable to the end users?

RQ1 is a general question that is not addressed specifically in any particular publication(s); instead, the question is more general and related to the publications as a whole. In a sense, RQ1 was the starting point of the whole research project. Due to the scarcity of related MIR user studies, it was unclear how our ideas would be perceived and accepted by the users, and thus we wanted to study how far we could proceed with the ideas.

The different aspects of RQ2 were studied in a series of surveys (online questionnaires) with 51-104 participants per survey. Taking the surveys presented in **P9** into account, the total number of participants was 543. The surveys concentrated on associating graphical objects with the tempo and release year of the music, and associating colors, icons, fonts, emoticons, and avatars with musical genres. For each survey, a collection of visual variables or structures was first developed, and their performance was then measured by asking the survey participants to map them to musical metadata.

To answer RQ3, findings from the surveys were used to design UIs for three interactive prototype applications (avatars, potentiometers and a virtual world) for the latest version of Nokia Research Center's SuperMusic music recommendation system (Lehtiniemi 2008). The prototypes were evaluated in user studies with 40-41 participants, the used research methods being observation, semi-structured interviews, and surveys. Both first impressions and longer-term user experience were studied.

1.3 Summary of publications

The results of the surveys and the user studies were published in several conference papers and journals, of which nine (eight conference papers and the journal) are included in this thesis. Publications **P1**-**P5** and **P9** are mostly related to RQ2, while **P6**-**P8** address mainly RQ3. This section summarizes the main results and contributions of the publications, and describes the author's contributions to them.

Publication **P1** focuses on mapping musical tempo to various visual attributes such as the number of objects, shape, size, orientation, color, and blur. Based on the results of the conducted survey, promising ways to present tempo include the number of objects, shapes with a varying number of corners, and *y*-axis location combined with some other graphical attribute. In all these cases, high numbers map to faster tempi than low numbers. In this publication, the author was the originator of the concept, arranged the survey, conducted the statistical analysis, and wrote the largest part of the paper. The concept was developed jointly with Dr. Antti Aaltonen.

Publication **P2** discusses mapping the release year of the music to various visual attributes such as lightness, location, color, and shape. Based on the results, promising ways to present release year include lightness (lighter colors represent newer music) and perceived location on the z-axis (foreground objects represent newer music). The publication also suggests using the *x*-axis location for presenting release year as it is commonly associated with time. In this publication, the author was the originator of the concept, arranged the survey, conducted the statistical analysis, and wrote the largest part of the paper. The concept was developed jointly with Dr. Antti Aaltonen.

Publication **P3** considers the concept of associating colors with musical genres. Based on the results of the conducted survey, it is not possible to design a globally accepted color-genre mapping. Instead, the publication proposes that a music player application could be set-up with a default mapping that is most suitable for the given country or region. The publication also presents a compromise mapping for use in Finland. In this publication, the author was the originator of the concept, arranged the survey, did part of the statistical analysis, and was the main author of the paper. Dr. Harri Siirtola implemented the contingency table and the association plot.

Publication **P4** studies visualizing musical genres in the context of context-aware music recommendation systems. In this type of recommender, the user's current emotional state also plays an important role. The aim of the study was to study which musical genres people listen to in different emotional states, and whether these genres and states can be represented with descriptive emoticons. Based on the results of the conducted survey, the publication presents a list of genres that could be used as a starting point for making recommendations fitting the current mood of the user. In this publication, the author was the originator of the concept, arranged the survey, conducted the statistical analysis, and wrote the largest part of the paper. Harri Holm implemented the graphics, and Jarno Seppänen helped to improve the presentation of the publication.

Publication **P5** considers the concept of mapping musical genres to stereotypical avatars consisting of head, body, and background elements. Each avatar element was studied separately in an online survey. Based on the results, it is possible to design representative avatars that are recognized accurately. For example, the different elements of the classical music avatar were recognized correctly by ≥93% of the participants. In this publication, the author was the originator of the concept, arranged the survey, did part of the statistical analysis, and wrote most of the paper. Dr. Harri Siirtola implemented the heatmaps, and Lauri Laaksonen designed the avatars based on instructions from the author.

Publication **P6** describes an avatar-based application for discovering new music. Each of the three avatar parts represents a single musical genre, and the application generates a new playlist of music recommendations based on the selected combination of avatar parts. Based on the results of the qualitative and quantitative user study, the concept has great potential as a complementary user interface for traditional text-based music applications. In the longer-term use (in this study, up to three weeks), the prototype was too restricted and lacked some important functionality (e.g., the possibility to do textual searches) to suffice as the only music player application. In this publication, the author was a co-inventor of the concept, designed and implemented the user study in collaboration with Arto Lehtiniemi, did the statistical analysis, and wrote the largest part of the publication.

Publication **P7** studies the concept of using 3D virtual worlds as a music player interface. Each element in the world can function as an interface to a song, artist, or album in the music collection or be associated with a specific music player function. To study the concept in practice, a simplified off-line version of such an application was developed, and evaluated in a quantitative and qualitative user study. Based on the results, the concept does not fit well to consumers' current music listening habits. To have a chance to succeed, such an application should contain more activity and interactive elements, and thus one possible solution for this would be to modify an existing game or virtual world application by augmenting it with music recommendation capabilities. In this publication, the author was a co-inventor of the concept, designed and implemented the user study in collaboration with Arto Lehtiniemi, did the statistical analysis, and wrote the largest part of the publication.

Publication **P8** presents the idea of using tempo and energy level potentiometers as a graphical user interface for interacting with a music recommendation service. The potentiometers are used to fine-tune the playlist of new music recommendations, and the look of the interface is changed to reflect the currently selected genre. Based on the results of a qualitative and quantitative user study, it is shown that also this concept has potential for complementing traditional music player applications, and that it possible to design recognizable music genre icons and skins. In this publication, the author was a co-inventor of the concept, designed and implemented the user study in collaboration with Arto Lehtiniemi, did the statistical analysis, and co-authored the publication.

P9 summarizes four different methods (colors, icons, fonts and avatars) for visualizing musical genres and compares them against each other. To allow comparisons to the compromise mapping proposed in **P3**, the publication concentrates only on the Finnish sample of the different surveys. Based on the results, the best performance can be achieved by combining different visualization methods. The best performing method was the avatars, followed by icons, fonts, and colors. The easiest genre to visualize was metal; it performed best in both surveys and the user tests of the prototypes. In this publication, the author wrote most of the text, while Dr. Harri Siirtola implemented the balloon plot and the boxplots.

1.4 Main contribution of the thesis

This thesis provides contributions to both academic researchers and developers of digital music services, with the emphasis being on Western views and design implications.

The first contribution comprises a set of suggested visual variables and structures for visualizing tempo, release year, and genre metadata. The performance of the encodings was studied in a series of online surveys, and selected genre designs were further user-tested with three interactive prototypes. To the best of our knowledge, such systematic studies on visualizing musical metadata have not yet been conducted elsewhere. Based on the results, promising ways to represent tempo include the number of objects, shapes with a varying number of corners, and *y*-axis location combined with some other visual variable or clear labeling. Promising ways to represent the release year include lightness and the perceived location on the z- or x-axis. In the case of genres, the most successful method was avatars, which used elements from the other methods and also required the most screen estate.

The second contribution of the thesis includes a set of design implications for digital music service development. Based on the results of the conducted user tests, it is possible to design recognizable, acceptable, entertaining, and easy-to-use music visualizations for selected (especially genre) metadata with certain limitations. Regardless of how entertaining or easy-to-use the music visualization might be, it cannot meet all the needs of a modern music consumer without the support for textual searches, and thus visualizations are more suitable for complementing than replacing traditional text-based applications. Furthermore, the performance, suitability, and acceptability of visualizations depends on factors such the type of available musical metadata, metadata vocabulary (e.g., the set of musical genres), used visual variables and/or structures, available screen estate (size of the visualization), the user's preferred music discovery type, the target culture of the visualizations, and so forth. For example, while the avatar concept was received well in Finland, the concept of using stereotypical avatars may be less acceptable in some other cultures.

1.5 Outline of the thesis

The thesis is divided into the following parts: Chapter 2 presents an overview of the field of music information retrieval (MIR) and discusses selected subtopics relevant to the contents of this thesis: metadata, music discovery, and interaction design. Chapter 3 provides a short introduction to the field of information visualization (Infovis), and Chapter 4 gives several examples of visualizing tempo, release year, and genre metadata in both commercial and academic applications. Chapter 5 describes the used research process, approach, and methods within the field of HCI, and discusses in more detail the conducted surveys, the evaluated prototypes, and the selection of musical genres for the surveys. Chapter 6 presents the results of the research in terms of research questions, and Chapter 7 discusses the results from various viewpoints including reliability and validity, product usage and music discovery modes, genre selection, racism, and the number of metadata dimensions. Finally, Chapter 8 concludes the thesis and suggests some directions for future work.

Appendix A contains a summary of genres used by selected music services in 2007, and Appendix B explains the definitions and history of genres used in the thesis. In addition, the thesis contains the nine original articles published in international conference proceedings and journals.

2. Music Information Retrieval

This chapter provides relevant background information on Music Information Retrieval (MIR), including metadata, music discovery and music recommendation systems, and interaction design.

Music information retrieval is an interdisciplinary research field studying different strategies for enabling access to both new and historical music collections (Casey et al. 2008). MIR includes topics as diverse as computational methods for music analysis, musicology and music theory, software development, HCI, and UI development. Casey et al. (2008) identify three main audiences benefiting from MIR: industry bodies engaged in recording, aggregating and disseminating music, music professionals (performers, teachers, etc.), and the end-users who want to find and use music in a personalized way. Common MIR use cases include, e.g., providing metadata about an unknown track, finding similar-sounding music based on a given seed song, recommending new songs matching the user's profile, automatic transcription of music, and finding music from a certain genre.

There are three strategies for solving the different use cases: 1) low-level audio features, 2) high-level music content description, and 3) conceptual metadata (Casey et al. 2008). The strategies do not have to be separate; they can also support each other. For example, musical tempo (a high-level music content descriptor) can be stored as textual metadata inside the song.

Low-level audio features are *"measurements of audio signals that contain information about a musical work and musical performance"* (Casey et al. 2008, p. 672). The features are typically segmented based on frames, beats, or statistical measures. Techniques for extracting the features include, e.g., short-time magnitude and mel spectrums, chromagram, onset detection, and spectral flux. The results of the low-level analysis are often used to obtain a high-level description of the musical content. (Casey et al. 2008)

The high-level descriptions embody *"the types of knowledge that a sophisticated listener would have about a piece of music, whether or not they know they have that knowledge"* (Casey et al. 2008, p. 671). Examples of such musical features include timbre, melody, rhythm and tempo, key, structure, and lyrics, and they are typically created by the means of automatic audio content analysis. Some of the high-level descriptions (e.g., musical tempo) are commonly stored as textual metadata inside the individual songs.

Currently, the most common way to access music collections is through textual metadata (Casey et al., 2008). This approach has also been taken in this thesis.

2.1 Metadata in MIR

Metadata is data representing the actual content, i.e., "data about data." Metadata makes content easier to manage, and it can be used for various tasks including searching, organizing, summarizing, constructing views, and recommendations (Lehikoinen et al. 2007, pp. 94-100).

Most current music services are accessed through textual metadata describing the contents of the music collection. Such metadata may include several keywords or attributes (e.g., artist name, song name, musical genre, or release year of music) from a controlled vocabulary and their values (e.g., pop or 1973). A common MIR task in home environments has been to use online services such as Gracenote (www.gracenote.com) to seek this metadata for ripped audio CDs.

The best-known metadata standard for music is ID3 (http://id3.org). ID3v2 metadata is stored in the same files with the actual content, and it is supported by the major music player applications such as Windows Media Player and Apple iTunes. Other audio tagging formats include, e.g., those used by Windows Media Audio (WMA) and AAC (Advanced Audio Coding) music files (Lehikoinen et al. 2007, p. 105). A comparison of existing music metadata standards can be found from Corthaut et al. (2008).

Metadata can be either factual or cultural. Factual metadata contains *"objective truths about a track"* (Casey et al. 2008); it includes data such as artist, album and track names, release year, and duration. Cultural metadata, on the other hand, contains subjective information such as mood and musical genre (Casey et al. 2008); this type of information is implicitly present in huge amounts of data (Celma 2008, p. 66). Most current music services utilize both types of metadata.

Metadata may be created manually by users, content creators or companies (contextual data), or automatically by computers (content-based data). A common form of user-generated metadata is folksonomies, which consist of *"collaboratively generated, open-ended labels that categorize content"* (Lehikoinen et al. 2007, p. 83). The labels are also knows as tags. Many MIR related services such as Last.fm (www.last.fm) and Qloud (2011) use folksonomies, i.e., the process of "social tagging" or "collaborative tagging" (Celma 2008, p. 40), as a part of their recommendation system. In 2008, Last.fm had already collected over 40 million social tags, most of which were related to musical genre (Lamere 2008). Users' motivations and incentives for tagging may vary from social signaling and contribution to assisting personal retrieval of songs from the collection (Lamere 2008). The tags can also be the end result of annotation games such as TagaTune (Law et al., 2007) or MoodSwings (Kim et al., 2008).

In addition to users themselves, the metadata can be entered by an expert or a group of experts. One well-known example of such a commercial music service is Pandora (www.pandora.com), which generates a personalized online radio channel based on user preferences. Pandora uses manually annotated metadata for estimating track and artist similarity, and it has been estimated that entering the metadata for a single music track takes 20-30 minutes of one expert's time (Casey et al. 2008). As the process is costly and time-consuming, there were "only" 800,000 analyzed songs in the collection in June 2011 (www.pandora.com). Another service based on expert opinions is the AllMusic (originally All Music Guide) website (www.allmusic.com), which contains a large database with genre and mood annotations.

While manually generated metadata can be very beneficial, it also has some problems, which limit its usage in music services (Casey et al. 2008). When several different people (either experts or users) create the keywords and their values, it can be difficult to maintain consistency; there is always some noise in the data. The process is always subjective and prone to errors such as typos, and the tagging of vast, ever-increasing music collections is an extremely tedious job (cf. Pandora). Due to these problems, the automatic extraction of metadata (and high-level music descriptions) has become a popular research topic within the MIR community. For example, Tzanetakis et al. (2001) and Chang et al. (2010) describe algorithms for the automatic genre categorization of songs in a music collection.

All visualization methods presented in this thesis are based on the use of already existing metadata. From the methods' point of view, it is not important when and how the metadata was generated – as long as it is there in a usable format.

2.2 Music discovery

One important and common MIR use case is the problem of music discovery, which can be exploratory, active, or passive (Lillie 2008, p. 24). In the case of *exploratory* discovery, the user is browsing through the music collection without knowing exactly what he/she is looking for. Still, the user can get satisfaction from the actual searching experience. In the case of *active* discovery, the user is searching for something and has at least a rough idea of what he/she wants. For example, the user could write some lyrics to Google's search box or search Spotify for the oldest album of his/her favorite artist.

As we are gradually reaching the "celestial jukebox" *("the mass of all recorded music, from all of history, that'll one day live freely in cyberspace"* [Dahlen 2006]), the modern music consumer is also faced with a new type of discovery problem: how to find personally relevant and interesting music from the vast collections containing millions of songs? Instead of the basic exploratory and active music discovery methods, the problem has to be tackled with *passive* discovery. The term "passive discovery" is related to the use of music recommendation systems, which are able to select new tracks for the user based on a personal preference profile or one or more example (seed) songs. In the first case, the system automatically observes the personal preferences of users (and thus builds the user profiles), models the properties of a music catalog, and suggests music based on the model. In the latter case, the user either selects the seed song by him/herself or modifies some parameters (e.g., genre or tempo), based on which the seed songs and the recommendations are then generated. In either case, the recommender system helps the user to discover songs they might not find on their own (Lillie 2008, p. 24).

2.2.1 Music recommendation systems

Modern music recommendation systems can be divided into five categories based on the used information filtering method. The categories are 1) demographic filtering, 2) collaborative filtering, 3) context-based filtering, 4) content-based filtering, and 5) hybrid methods. (Celma 2008, p. 29)

Systems based on demographic filtering try to identify what type of users like certain items, and then classify users into profiles based on demographic data such as age, city, and interests. The main problem with this approach is that the recommendations are very general. One example of a (non music-related) system based on demographic filtering is the Grundy (Rich 1979), where information gathered from an interactive dialogue was used to recommend books. (Celma 2008, p. 30)

In collaborative-filtering based methods, the system predicts user preferences based on users' listening habits and previously given user feedback, such as ratings for songs. The similarity between songs is measured in terms of the similarity of their listener populations. One problem with collaborative filtering is the so-called "cold-start problem", which means that new songs (that have not yet been listened to) and new users (who have not yet given many ratings) are difficult to recommend and categorize. Due to this, songs in the "Long Tail" are often neglected in the recommendations (Celma 2008, pp. 34-35). Examples of music recommenders using collaborate filtering include Last.fm and the now discontinued MyStrands social music service (Lillie 2008, p. 25).

Context-based filtering systems use cultural metadata (e.g., mood or genre) to compute artist or song similarities. The cultural metadata is gathered using collaborative tagging or web mining techniques (Celma 2008, pp. 85-86). Examples of such music recommendation systems include Pandora, AllMusic, Last.fm, and Qloud (2011). Context-based filtering should not be confused with context-aware music recommendation, which means that the system suggests music to match the current situation (location, time, mood, etc.) of the listener (**P4**).

Content-based filtering systems help us *"get at what the music actually sounds like"* (Lillie 2008, p. 27). In these systems, similarity is considered in terms of low-level and/or high-level musical attributes, i.e., audio similarity. Examples of such attributes include loudness (low-level), timbre (low-level), rhythm (high-level), and musical form (high-level) (Lillie 2008, p. 27). Content-based music recommendation systems include, e.g., MusicSurfer (Cano, Koppenberg, & Wack 2005) and PlaysSOM (Neumayer, Lidy, & Rauber 2005).

Finally, the fifth category of music recommendation systems consists of hybrid methods. While Magno and Sable (2008) have shown that the performance of content-based filtering is approaching that of context-based and collaborative filtering, many authors – such as West, Cox, and Lamere (2006) – believe that optimal performance can be achieved by combining content-based and contextual features, i.e., by using a hybrid approach. For example, Yoshii et al. (2008) improved the performance of their system by combining content-based filtering with collaborative filtering, and Bogdanov & Herrera (2011) refined their content-based approach by using genre metadata.

2.2.2 SuperMusic

Another example of the hybrid method is SuperMusic (Lehtiniemi 2008), which is a context-aware music recommendation system developed at Nokia Research Center between 2006 and 2009. SuperMusic is a dedicated client-server system, where the Symbian or Windows client interacts with a server built with Python. In addition to providing the basic Symbian Series 60 music player functionality, the mobile version of SuperMusic supports both active and passive music discovery. Users can perform textual searches, look for similar-sounding music for the selected seed song, and get personalized music recommendations based on their current context (e.g., location) and listening history. One can also recommend tracks to his/her friends and rate the recommendations. The songs are played either locally or streamed from a relatively large online music collection (approx. half a million tracks).

SuperMusic is a hybrid system utilizing both textual metadata and audio similarity. The similarity of metadata tags is calculated at the artist level, and to limit the range of values, the audio similarity is calculated only between tracks of artists whose similarity exceeds a certain threshold. The final similarity value is a weighted sum of the artist tag similarity and song similarity. To make online searches faster, song similarities (*N* per track) are calculated offline and stored in a database. For more details on the used algorithms, see **P6**.

2.3 Interaction design in MIR

To be successful, a product (in this case, a music application) has to be useful, usable, and provide an enlightening overall user experience (Jones & Marsden 2006, p. 40). User experience (UX) refers to *"a person's perceptions and responses that result from the use and/or anticipated use of a product, system or service"* (ISO 2009). UX involves both pragmatic and hedonic quality aspects of the product, and it is highly personal. UX also changes over time (Karapanos et al. 2009); longer-term user experience is especially important for business success (Roto 2007).

The usefulness, usability, and UX of products can be improved by utilizing proper interaction design methods. Interaction design refers to *"designing interactive products to support people in their everyday and working lives"* (Sharp, Rogers, & Preece 2007, p. xvii). This involves three main types of activity: 1) understanding users, 2) developing prototype designs, and 3) evaluation (Jones & Marsden 2006, p. 94). One important sub-task of prototype design is the user interface design. In the following sections, the different phases of interaction design are discussed in the context of MIR.

2.3.1 Understanding users

Understanding users means *"having a sense of people's capabilities and limitations; gaining a rich picture of what makes up the detail of their lives, the things they do and use"* (Jones & Marsden 2006, p. 94). To get better knowledge of the target users, techniques such as observation, interviews and focus groups, surveys, contextual inquiry, probing, and diary studies can be used (Jones & Marsden 2006, pp. 129-155).

So far, relatively few empirical MIR user studies have been reported (Weigl & Gustavino 2011). The studies have primarily focused on understanding the users, rather than evaluating the performance of the systems (Hu & Liu 2010). For example, Vignoli (2004) studied how music consumers organized and accessed their digital music collections, and Cunningham et al. (2003) conducted an ethnographic study of the searching and browsing techniques employed by local people in a library and in music stores.

Weigl and Gustavino (2011) have presented an overview of the user studies in the MIR literature. The authors searched through 719 articles from the ISMIR conference (www.ismir.net) for selected keywords such as user study, usability, and participants, and searched the ISI Web of Knowledge database (wokinfo.com) using the query string "'music information retrieval' AND user." The identified publications were divided into the following topics:

- User requirements and information needs (two publications)
- The information needs of specific groups (four publications) and in specific contexts (one publication)
- Insights into specific aspects of music perception and preference, such as the factors that cause listeners to dislike certain songs (one publication), the impact of social relations on

music acquisition and taste (one publication), and the effects of demographic factors and musical background on the semantic descriptions of music (two publications)

- Analyses of textual MIR queries—symbolic representation of the melody sought (one publication), and natural language expressions of music information needs (two publications)
- Employment of user studies to generate ground-truth data for use in training and evaluation corpora (three publications)
- The organisation of digital music information (two publications)
- Search strategies and relevance criteria used when actively seeking new music (two publications)
- Information behaviour in passive or serendipitous encounters with new music (one publication)

The majority of these user studies were qualitative, and the approaches ranged from ethnographic observation and interviews to diary studies and online surveys. While the list by Weigl and Gustavino is not complete (for example, many articles published in other conferences than ISMIR are missing), it gives a good overview of the emphasis of the research so far.

2.3.2 Prototype and user interface design

Once enough knowledge on the users has been gathered, the next step in interaction design is to design and implement a prototype. Prototyping is a way to "fail fast" (Jenson 2002, p. 124). It is very unlikely that the first version is perfect in every way, but re-designing is much cheaper in the early phase of the work than after the product has already been launched. Prototypes may vary from simple paper prototypes to complex beta releases. They can be categorized, e.g., according to their resemblance to the final product (low and high-fidelity ("hi-fi") prototypes) or the functionality that they provide. Horizontal prototypes show a lot of functionality with little detail, while vertical prototypes provide a lot of functionality for only a few functions (Sharp, Rogers, & Preece 2007, p. 537).

One important part of designing the prototype application is the user interface (UI) design. Most current interfaces belong to the category of graphical user interfaces (GUIs), which is sometimes referred to as WIMP (windows, icons, menus and a pointing device) systems. One special case of UI design is the design for small mobile devices, which create new type of challenges for the designers; for example, how to utilize the tiny keypad, and how to fit all the relevant information to the screen?

A wide variety of heuristics (rules of thumb), guidelines, and standards on designing prototypes and UIs exist. For example, Jones and Marsden (2006, pp. 101-102) give the following examples of design starting pointers: 1) design for truly direct manipulation, 2) design for ecological use, 3) design for maximum impact through minimum user effort, 4) design for personalization, 5) design for play and fun, and 6) design for one-handed use. Nielsen (1993, p. 20) and Shneiderman (1997) present some well-known basic heuristics for usable interfaces, including simple and natural dialogue ("less is more"), minimizing the user's memory load ("recognition over recall"), consistency, and shortcuts. Norman (2004) emphasizes the importance of creating aesthetically pleasing designs. In addition to the general heuristics, many developers have also published more detailed UI guidelines for their own platform(s).

Several examples of MIR prototypes and UI designs are given in Chapter 4 in the context of visualizing musical metadata.

2.3.3 User evaluations

The last step of interaction design is to evaluate the prototype with end-users, experts, or without users (automated testing). Sometimes the last two categories are referred to as usability inspections, and the term "usability testing" is reserved for user-based testing (Lazar, Feng, & Hocheiser 2010, p. 256). User-based testing refers to *"a group of representative users attempting a set of representative tasks"* (Lazar, Feng, & Hocheiser 2010, p. 260); such techniques include, e.g., observation, interviewing, and surveys/questionnaires.

While the ultimate goal of MIR systems is to help users seek music or music-related information, so far the evaluation of MIR algorithms and prototypes has been dominated by automated *"system-centered approaches which typically measure how well the systems (algorithms) classify music and how relevant their retrieved music was"* (Hu & Liu 2010). For example, in the annual Music Information Retrieval Evaluation eXchange (www.music-ir.org/mirex) competition, systems from different research groups perform selected MIR tasks; their performance is compared using system-centered measures such as accuracy for mood and genre classification, average precision for music similarity retrieval, and so forth (Hu & Liu 2010).

The problem has already been acknowledged (Weigl & Gustavino 2011). Focusing on users has been identified as one of the five most important MIR challenges for the next ten years (Downie, Byrd, & Crawford 2009), and the number of user-based tests is increasing steadily. In the following, a couple of examples of evaluating MIR applications with end-users are given.

Hoashi et al. (2009) conducted a comparative evaluation of a traditional list-based and a 2D visualization for a content-based MIR system. Based on the results, the authors concluded that visualizations can improve the usability and efficiency of the system and give a better impression of the accuracy of the MIR results. Leitich and Topf (2007) conducted initial user experiments with their Globe of Music application, and received promising results in terms of high user acceptance. For example, the authors learned that the visualizations provided the user with an *"intuitive interface which is easy to handle and fun to explore."*

Chen and Kluber (2010) studied the automatic generation of visual thumbnails for music. In the initial user study, *"all the participants commented that these visualizations were overall too complicated for them,"* and the participants did not appreciate the concepts much. However, the Music Icon concept was appreciated by seven DJs who also saw the concepts later. Lehtiniemi (2008) conducted a five-week user trial of the SuperMusic system with 42 participants. Based on the results, the author concluded that the SuperMusic concept *"satisfied 97% of the users and it was seen as a potential killer application in the music domain with some modifications."*

Other examples of user-testing music discovery related applications include, e.g., Haro et al. (2010), Cunningham & Zhang (2008), Pauws & Eggen (2002), Pauws & van de Wijdeven (2005), Vignoli & Pauws (2005), and Celma (2008).

3. Information visualization

Instead of using textual lists, a promising alternative for representing musical metadata is information visualization. Information visualization (Infovis) is a multidisciplinary research field bridging many areas including HCI, computer graphics, and cognitive psychology. Infovis tools and applications can help us to transform data into useful information, form mental models and gain insight, perform tasks more effectively and accurately, and thus also save time and money (Spence 2007).

The dictionary definition for visualization is *"forming a mental image or making something visible to the eye"* (McKean 2005). While the two concepts differ from each other, they can both help in the process of forming a mental model, i.e., *"knowledge of how to interact with a system and, to a lesser extent, how that system works"* (Sharp, Rogers, & Preece 2007, p. 116). One way to decrease the gap between data and the user's mental model of the data is to use information visualization (Yi et al. 2008).

Gershon and Eick (1995) characterize information visualization as *"a process of transforming data and information that are not inherently spatial, into a visual form allowing the user to observe and understand the information."* While this definition also covers the time before computers, Infovis tools are increasingly computer-based, and thus some other authors limit information visualization to computer-related activity. For example, Card et al. (1999, p. 6) define information visualization as *"the use of computer-supported, interactive, visual representation of abstract data to amplify cognition."* Amplifying cognition may be done in different ways, including shifting part of the workload from human's cognitive system to the perceptual system, reducing searching, and enhancing the recognition of patterns (Ware 2004).

At their core, Infovis systems have two main components: representation and interaction. The representation component "*concerns the mapping from data to representation and how that representation is rendered on the display,"* and it has its roots in computer graphics. The interaction component, on the other hand, has roots in HCI; it *"involves the dialog between the user and the system as the user explores the data set to uncover insights."* So far, the majority of Infovis research (including this thesis) has concentrated on the representation component. (Yi et al. 2007)

In the following, the basics of information visualization are discussed based on representation (Section 3.1) and interaction (Section 3.2). In addition, the concept of casual information visualization (Casual Infovis) is introduced in Section 3.3.

3.1 Representing data

Before the data can be presented to the user, it first has to be represented (encoded) visually. The most appropriate encoding mechanism depends on various factors, including the planned tasks and the context of use; if one selects the wrong type of representation, the user may have a hard time finding the required information from it.

In the first stage of the encoding process, raw input data is mapped into data tables which combine the data with descriptive metadata (Card, Mackinlay, & Shneiderman 1999, pp. 17-18). The data values can be either nominal (unordered, i.e., they are only = or ≠ to other values), ordinal (ordered, such as weekdays), or quantitative (numeric) (Card, Mackinlay, & Shneiderman 1999, p. 20). In information visualization, it is common to categorize data into 1D (a single number or a collection of numbers), 2D, 3D and multidimensional data, while in statistics the corresponding terms are univariate, bivariate, trivariate and multivariate (hypervariate) (Siirtola 2007, p. 16).

Next, the tables are transformed into visual structures (Card, Mackinlay, & Shneiderman 1999, p. 23) which can be processed more effectively by the human vision. The building blocks of visual structures are called visual variables (visual attributes). Originally, Bertin (1981) divided the variables into two groups: 1. "variables of the image" (location on the plane, size, and grayscale value) and 2. "differential variables" (texture, color, orientation, and shape). Each variable can be applied to a point, a line, or an area. Since 1981, there have been several updates on Bertin's taxonomy, including Mackinlay (1986), see Table 1.

Naturally, some visual variables are more efficient than others in representing different types of data and tasks. Human memory has a property called pre-attentive processing, which means that it can recognize certain visual attributes (e.g., orientation, size, spatial position, color and flicker) rapidly, automatically, and subconsciously without any cognitive effort from the human being (Ware 2004, pp. 149-154). Due to pre-attentive processing, these variables can be identified even after a very brief 30-300 ms exposure, i.e., they "pop out" from their surroundings (Spence 2007, pp. 47-48). As the variables can be processed in parallel and more efficiently than textual information, they should be prioritized in tasks involving searching.

Guidelines for encoding the data have been given in several sources including Bertin (1981) and Mackinlay (1986). Bertin (1981) discussed four low-level interpretation tasks and ordered a set of visual variables according to the number of tasks they can support. The tasks included association (can the variables be perceived as similar), selection (can they be perceived as different), order (can they be perceived as ordered), and quantity (can they be perceived as proportional to each other). The rank order of the variables was location, size, value, texture, color, orientation, and shape. For example, color was considered to perform well in association and selection tasks; while colors can be useful for selecting certain items from a group, they should not be used to visualize quantity (e.g., one cannot say that blue is two times red). In Mackinlay (1986) and Parry (2007), the graphical encodings were ranked according to their accuracy to convey quantitative, ordinal, and nominal data (Table 1). The most effective visual variable, position, was said to perform well for all data types.

According to Miller (1956), humans can hold only 7±2 chunks of information in short-time memory at one time. Thus, as a general rule humans can recognize 7±2 different values of a 1D variable. While the capacity can be improved by increasing the number of independent variables, this happens at a decreasing rate. In addition, as more variables are added the accuracy of any particular variable is decreased. Visualizations with more than three dimensions can be challenging for our cognitive system (Siirtola 2007, p. 2).

Table 1 Graphical encodings ranked according to their accuracy to convey quantitative, ordinal, and nominal data (reproduced from [Mackinlay 1986]). The most accurate encodings are on the top row and the least accurate on the bottom.

Quantitative	**Ordinal**	**Nominal**
Position	Position	Position
Length	Density	Color hue
Angle	Color saturation	Texture
Slope	Color hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density (value)	Containment	Color saturation
Color saturation	Length	Shape
Color hue	Angle	Length
Texture	Slope	Angle
Connection	Area	Slope
Containment	Volume	Area
Shape	Shape	Volume

Different visual variables also perform together in different ways. In the case of integral data dimensions (Figure 1), *"two or more attributes of a visual object are perceived holistically and not independently"* (Ware 2004, p. 177). For example, the rectangular shape is a combination of the rectangle's width and height. In the case of separable dimensions, each dimension (e.g., color and size of a ball) is judged separately. Other closely related terms include so-called "associative" and "dissociative" variables. While the associative variables *"can be ignored while one is inspecting the values of other variables"* (Siirtola 2007, p. 17), the dissociative variables interfere with the observation of other variables. For example, if a circle is too small, it is impossible to see its exact color.

Integral dimension pairs

red-green	yellow-blue
red-green	black-white
shape height	shape width
shape	size
color	size
direction of motion	shape
color	shape
color	direction of motion
x, *y* position	size, shape, or color

Separable dimension pairs

Figure 1 Integral and separable dimension pairs (reproduced from [Ware 2004, p. 180]).

Glyphs

One way to encode data with more than one dimension is to use glyphs, which can be understood as *"a graphical data object designed to convey multiple data values"* (Ware 2004, p. 145) or *"a graphical entity whose shape or appearance is modified by mapping data values to some of its graphical attributes"* (Wegenkittl, Löffelmann, & Gröller 1997). Different glyph types vary from stars and sticks to arrows and Chernoff faces.

When designing a glyph, a number of issues should be taken into account. As some variables are more accurate than others in conveying quantitative, ordinal and nominal data, guidelines such as those in Table 1 should be studied. To increase the recognition rate, one should rely on visual variables that are processed pre-attentively. Each variable should only have a small number of different values. One should preferably use the so-called "separable" dimensions (variables) which are judged separately from each other (Figure 1).

The most useful graphical attributes in glyph design have been summarized in Ware (2004, p. 183). The included table lists seven visual variables (spatial position, color, shape, orientation, texture, motion and blink) with their number of available dimensions.

3.2 Interacting with data

After the data has been encoded and rendered on the display, the user can start interacting with it. Instead of entering data into systems, the emphasis of Infovis interaction is typically on changing and adjusting the visual representation. This can greatly help the user to, e.g., perform the required task(s), to form a mental image of the data, and/or to acquire insight. Yi et al. (2007) view the different Infovis interaction techniques as *"the features that provide users with the ability to directly or indirectly manipulate and interpret representations."* According to this view, static images do not have associated interaction techniques, but changing the view from one to another is such a technique.

In Yi et al. (2007), the authors summarize several Infovis taxonomies relevant to interaction techniques. The taxonomies have different levels of granularity, varying from low-level interaction to users' tasks. Some of them are relatively system-centric, while the others focus on user goals. For example, Shneiderman (1996) lists seven information-seeking actions that users wish to perform: overview, zoom, filter, details-on-demand, relate (view relationships between items), history (keep of history of actions for undo etc.), and extract (make sub-collections). Shneiderman's publication also includes his famous Visual Information Seeking Mantra: *"Overview first, zoom and filter, then details-on-demand."*

Based on the taxonomy summary, Yi et al. divide Infovis interaction techniques into seven different categories around the user's intent: select (mark something as interesting), explore (show me something else), reconfigure (show me a different arrangement), encode (show me a different representation), abstract/elaborate (show me more or less detail), filter (show me something conditionally), and connect (show me related items). For example, "explore techniques" help the users in examining difference subsets of the data (e.g., different musical genres), and "encode techniques" enable the user to change the visual appearance (e.g., size or color) of each data element, which change can then affect pre-attentive cognition.

As the focus of this thesis is on the representation component of Infovis, the different interaction techniques are not discussed in more detail.

3.3 Casual information visualization

In Pousman et al. (2007), the authors present a set of systems that are positioned at the boundary between traditional Infovis and other domains. While the traditional Infovis systems are designed for work tasks, the given examples are more casual in nature. The authors focus on three borderline categories: 1) Social Infovis, 2) Artistic Infovis, and 3) Ambient Infovis.

Social Infovis refers to visualizations of social information such as social processes, networks, situations, and user-generated taxonomies such as tag clouds. In Viegas and Wattenberg (2008), tag clouds are also referred to as "vernacular visualizations", i.e., visualizations that come outside the visualization community and violate some of the golden rules of traditional visualization design.

Artistic Infovis refers to works of data-driven art. While the traditional Infovis systems are designed to be helpful in analytic tasks, Artistic Infovis systems *"have the explicit goal of challenging preconceptions of data and representation"* (Pousman, Stasko, & Mateas 2007).

Ambient information systems (Pousman & Stasko 2006) visualize (changes in) data in an aesthetically pleasing and subtle manner and sit in peripheral locations. While simple interactions are possible, *"exploring the data by changing representations is often beyond the capability of Ambient Infovis"* (Pousman, Stasko, & Mateas 2007). One good example of such a system is Informative Art.

Ljungblad et al. (2003, p. 215) define Informative Art as *"a playful combination of traditional wall decorations (such as posters and painting) and dynamic computer displays. A piece of informative art looks like a piece of abstract art, but instead of providing a static image its visual appearance is continuously updated to reflect some dynamically changing information. The resulting visualization is then shown on a wall-mounted display to give the impression of an ordinary painting."* In an example given by Ljungblad et al., the display presents a five-day weather forecast using colors, squares, and lines in such way that the visuals represent the work of Piet Mondrian (Figure 2). As stated by the authors, Informative Art is useful only if the user knows how to decode it. While the art can be visually pleasing, it is not the most efficient way to display detailed information, and should usually be used only for giving overviews or summaries of data.

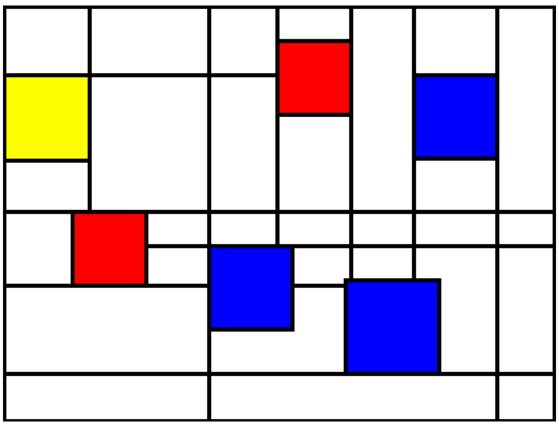

Figure 2 Example of Informative Art (Skog, Ljungblad, & Holmquist 2003). © 2003 IEEE.

While Social, Artistic, and Ambient Infovis systems do visualize information in some way (and thus, can be seen as Infovis systems), they are also quite different from the traditional work-centered Infovis systems. Pousman et al. (2007) list the following four main differences:

- *"User population: The user population is enlarged to include a wide spectrum of users from experts to novices. Users are not necessarily expert in analytic thinking, nor are they required to be experts are reading visualizations.*
- *Usage pattern: Usage expands past work, to focus on other parts of life. Systems are intended for usage that is momentary and repeatable (over weeks and months), or contemplative (a long moment at an art gallery).*
- *Data type: The data is typically personally important and relevant, as opposed to work-motivated. This means that a user's relationship to the data is often a more tightly coupled one.*
- *Insight: We propose that the kinds of insight that Casual Infovis may support are different from more traditional systems. We suggest that developers are interested in providing insight about data that is not analytical, but instead of a different sort."*

Based on the list, the authors introduce the term "Casual Infovis" and define it to be *"the use of computer mediated tools to depict personally meaningful information in visual ways that support everyday users in both everyday work and non-work situations."* Just like traditional Infovis systems, Casual Infovis systems can turn data into information and information into insight. However, such systems focus on awareness, social, and reflective insights, instead of being explicitly analytical and goal-oriented. While the data and the casual visualization may be meaningful to the user, strangers may find them merely curious (Pousman, Stasko, & Mateas 2007).

In Lau and Vande Moere (2007), the combination of Social Visualization, Ambient Visualization and Informative Art has also been referred to as "information aesthetics."

4. Visualizing musical metadata

In the context of MIR, information visualization can help in various tasks including searching, exploring, navigating, organizing, getting new music recommendations, analyzing, and Knowledge Discovery (Baumann 2005). Ways to visualize and interact with music collections vary from textual metadata (as used in most mainstream music player applications) to networks (e.g., www.liveplasma.com), geographical maps (e.g., Islands of Music [Pampalk 2001]), and more experimental designs such as rainbows (Pampalk & Goto 2006) and album covers arranged in the shape of a spiral (Lamere & Eck 2007). Most applications group similar types of content together using either content-based (audio-based) or contextual (human-labeled) data; many applications also somehow visualize the relationships between the different groups. For a summary on different ways to visualize and interact with music, see for example (Lillie 2008) or (Baumann 2005).

In this chapter, the focus is on applications that are based on metadata types discussed in this thesis, i.e., release year, tempo, genre, and emotion/mood metadata. Various encodings varying from mapping release year to colors and genres to icons are presented. While the references include visualizations related to both single tracks and large music collections, the temporal or structural representations of single tracks have not been included. Examples of such representations include techniques such as waveforms, common music notation (CMN), spectrograms and visualizations for musical structure, and they have been discussed in, e.g., (Isaacson 2005) and (Damm et al. 2008).

4.1 Release year

In the mainstream music player applications, release year metadata is typically represented using simple text strings. For example, with the Windows Media Center the user can arrange songs according to year, genre, album, artist, composer, song, and playlist. The year mode arranges album covers in a grid in such a way that albums from the same year are grouped together next to the corresponding textual header (e.g., "2007"). The release year decreases from left to right.

The UI of Apple's iTunes media player application includes a row of headings with metadata labels such as Artist, Albums and Time, and the user can easily personalize which labels are shown on the screen. If year information is not visible yet in the UI, the user can right-click any location of the row and activate it from the resulting list. The users can also create smart playlists that contain only songs that match certain conditions (e.g., add a song to the playlist if its release year is earlier than 1990). The parameters are set using pull-down menus and text boxes. However, the process can be too tedious for the average consumer, and it has been shown that smart playlists are not very popular in everyday use (Holm 2008).

Instead of showing the exact release year, the UI can also be based on decade. In (Bainbridge, Cunningham, & Downie 2003), the authors indicate the need to support "fuzzy" or imprecise metadata values for searches. In a study on posting questions to Google's service, the users rarely specified the exact date of a composition or a recording. Instead, the users gave the decade or used terms such as “recent” or “old”. One practical example of using this idea is the radio channel of Spotify (www.spotify.com). Spotify lets the user select the decade and genre of music that he/she wants to listen to, but the generated playlist then shows the exact release years below the corresponding album covers.

In addition to text strings, common ways to visualize release year include sliders and location on the *x*- or *y*-axis. For example, Musicovery (2011) is an interactive web radio with three different types of radio. In the “Mood radio” mode, the user can select music according to decade, genre, and mood (dark, energetic, calm and positive), and in the “Dance radio” mode mood has been replaced with tempo and “danceability.” In both modes, decade is presented as a horizontal range slider increasing from left to right. Another example of using sliders is the Musiclens (2007), which contains several vertical different sliders that can be used to get new recommendations. One of the sliders adjusts the release year of music, and the year decreases from top to bottom.

Torrens et al. (2004) present three different approaches (a disc, a rectangle and a tree-map) for visualizing music collections. Each UI presents four dimensions: genre, artist, release year, and a user-selectable quantitative criterion such as playcount, rating, or last-played date. In the disc visualization, the radius represents time in such a way that the oldest tracks are located in the center of the disc. In the rectangle visualization (Figure 3), the time axis goes along the vertical axis.

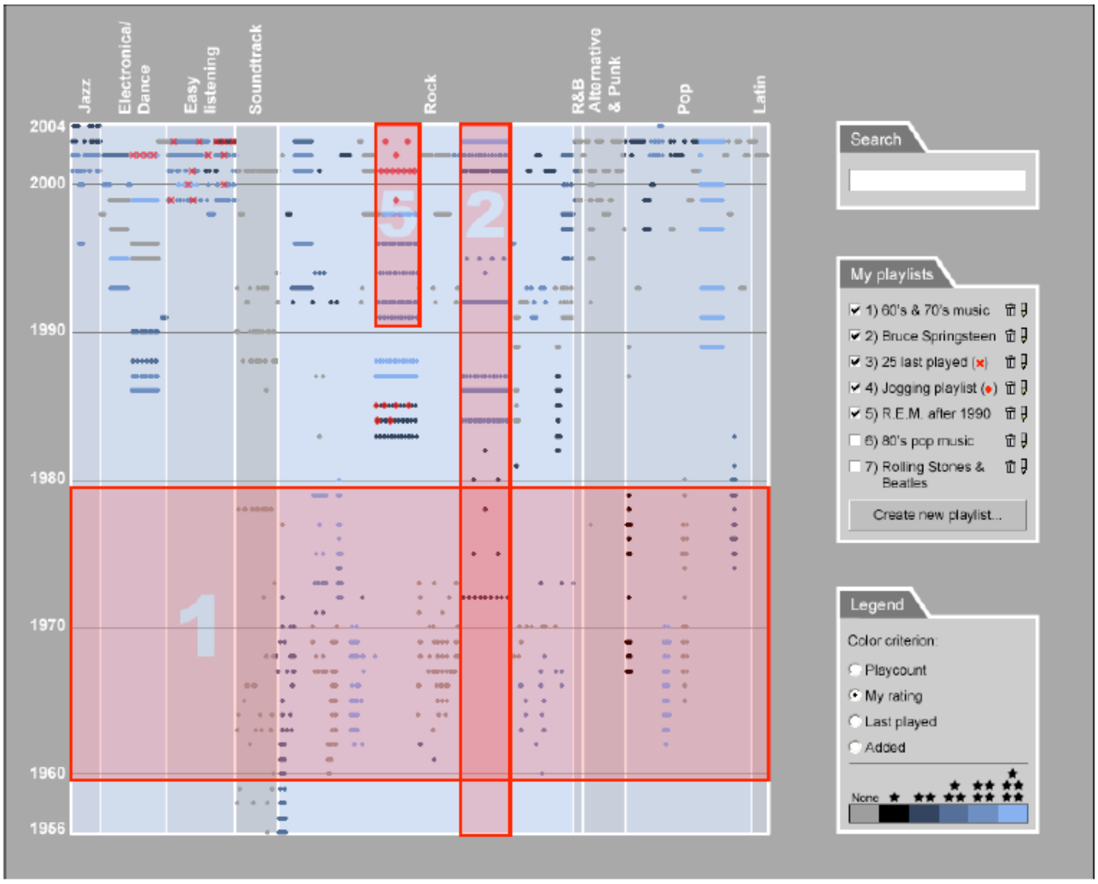

Figure 3 Example of visualizing release year metadata on the *y*-axis (Torrens, Herzog, & Arcos 2004).

The artist map application (Van Gulik & Vignoli 2005) visualizes artist similarities with colored circles located on an *x*-*y* space. The user can select his/her preferred visualization by selecting two metadata attributes (e.g., tempo, release year, mood or tempo) for the different axes, and either attribute can also be mapped to color. In one of the given examples, clustering is based on the release year (*x*-axis) and tempo (*y*-axis). The color coding is done on the year as follows: 60's = dark gray, 70's = gray, 80's = light gray, 90's = white, and 2000 = pink & red. In another given example, the coloring changes gradually from black (<1960) to white (2000+).

In (Nguyen, Tominski, & Schumann 2011), the authors present several interesting ideas to visualize year tags. For example, the paper proposes to map the year to the brightness of the tags (recent tags are darker than older tags) and tag size (recent tags are larger).

4.2 Tempo

The most distinct level of musical meter is the beat or tactus. The rate of the tactus pulse is typically represented in beats per minute (BPM), which is *"the rate at which most people tend to tap their foot on the floor while listening to music"* (Eronen 2009, p. 3). The BPM value can be measured using automatic music tempo detection algorithms such as those described in (Eronen 2009), or it can be retrieved from an existing database such as the Echo Nest (www.echonest.com). As the exact BPM value has little value to the average user, categorical labels such as "slow", "medium", and "fast" can also be used. For example, the artist map application (Van Gulik & Vignoli 2005) divides songs into five categories: very slow, slow, medium, fast, and very fast.

In modern music player applications, the desired musical tempo is typically selected using pull-down menus and text boxes, sliders, or by clicking on an *x*- or *y*-axis location. By specifying the desired value or value range, the user can select fast or slow music fitting his/her current mood or listening context. Fast tempi are generally associated with expressions such as happiness, activity, and pleasantness, while slow tempi are associated with sadness, tranquilness, and the like (Juslin & Sloboda 2001).

For example, in the case of Apple iTunes the user can specify tempo conditions for the smart playlists using pull-down menus and text boxes. The user can also activate the BPM value to be shown in the headings row located on top of the screen. In the Musiclens application (Musiclens 2007), the user can move sliders to affect the playlist of new music recommendations. One of the sliders adjusts tempo, which decreases from top to bottom.

Van Gulik's artist map user interface (Van Gulik & Vignoli 2005) visualizes artist similarities with colored circles located on an *x*-*y* space. In one of the given examples (Figure 4), clustering is based on release year (*x*-axis) and tempo (*y*-axis), and the coloring is done on tempo as follows: very slow = blue, slow = cyan, medium = green, fast = yellow, and very fast = red. In the "Dance radio" mode of the Musicovery application (2011), tempo has been mapped to the *y*-axis and it decreases from top to bottom. Finally, Zhu and Lu (2002) map music clips into a 2D timbre-rhythm space. Brightness decreases from top to bottom and tempo from right to left.

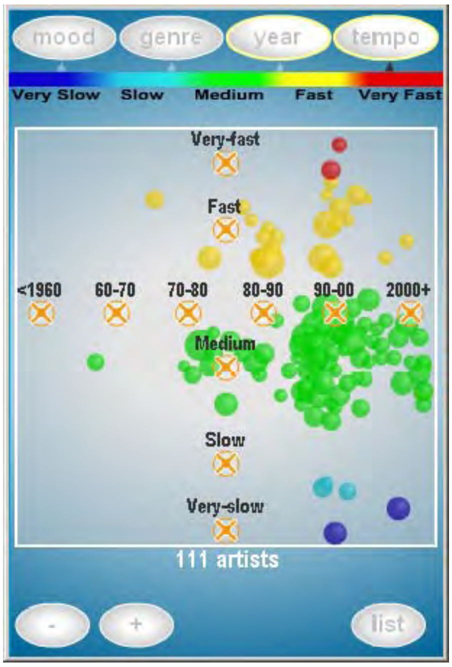

Figure 4 Screenshot from the artist map by Van Gulik and Vignoli (2005). In this example, tempo is represented with color and location on the *y*-axis, and release year is represented with location on the *x*-axis.

Musical tempo can also be represented with icons, pictures, or "visual thumbnails" (~glyphs, see Section 3.1). In the Lyricon system (Machida & Itoh 2011), the musical structure of a song is represented using multiple icons, which are automatically selected based on the lyrical and musical features (e.g., tempo) of the song. Oda and Itoh (2007) propose a technique for automatically selecting icons to represent MIDI (1996) files. Songs are matched with images based on user-given "sensitivity words" (e.g., bright or dark), keywords of the images (e.g., "flower" or "cloud"), and calculated musical features such as tempo, key, and the number of notes per time frame. Using the YCbCr color system, musical tempo has been mapped to the Y value of the icon image.

The MusCat music browser from Kusama and Itoh (2011) visualizes songs in the music collection using abstract pictures. The pictures are generated based on the musical features of the songs. In the given example design, tempo has been mapped to the number of circles in the picture in such a way that a larger number of small circles represents faster music and a smaller number of large circles represents slower music. The paper also discusses briefly some feedback received from the MusCat users. However, none of the comments was related to tempo.

Chen and Kluber (2010) studied the automatic generation of visual thumbnails for music. In one of the four concepts selected for the initial user study (the "Music Icon"), the musical tempo was mapped to the number of petals around a circle in such a way that higher numbers represented faster music (Figure 5, top). The shape of the petals was mapped to the aggressiveness of songs. The user study was conducted in the form of a questionnaire, and in total answers from 38 participants (9 female and 29 male with the average age of 27) were received. According to the paper (p. 566), *"all the participants commented that these visualizations were overall too complicated for them"*, and the participants did not appreciate the four concepts much. However, the Music Icon concept was appreciated by seven DJs who also saw the concepts later.

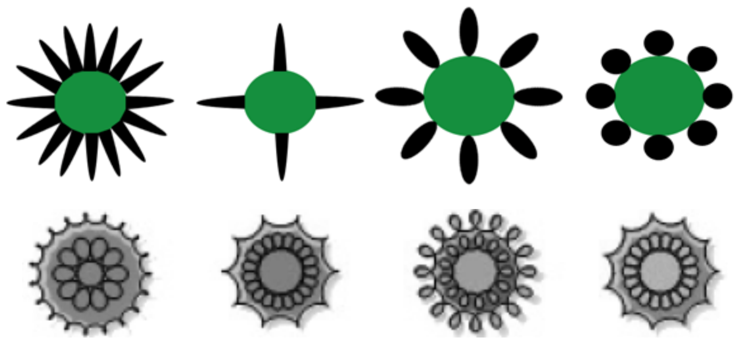

Figure 5 Examples of visual thumbnails for music. Top: Chen and Kluber (2010) and bottom: Kolhoff et al. (© 2006 IEEE).

In (Kolhoff, Preuss, & Loviscach 2006), the authors describe another music thumbnail concept with more promising results. While the described implementation does not analyze the rhythmic features of the songs, this has been addressed as one potential topic for future work. To start with the application, the user first selects a set of "training icons" to represent a set of prototypical songs, i.e., creates his/her own personalized mapping between music and icons. The icons have a bloom-like shape with a varying number of petals, shape of the petals, and color (Figure 5, bottom). The songs are analyzed using Mel-Frequency Cepstral Coefficients (MFCCs) to form representative audio thumbnails, after which the system is able to automatically create icons for the other tracks in the music collection. According to the authors, the users seemed to learn to "read" the icons in terms of music.

4.3 Genre

Musical genres or music genres are *"categorical labels created by humans to characterize pieces of music"* (Tzanetakis & Cook 2002). Songs belonging to the same genre share some intra- and extra-signal features that separate them from other types of music. Ways to categorize music include, e.g., art/popular music, vocal/instrumental, time period, female/male vocals, the instrumentation used, and the country of origin. However, in everyday use the most popular way to characterize music is to use genre labels such as rock, classical, reggae, and jazz.

Genre taxonomies are commonly used in physical and online record stores, radios, traditional music player applications such as Apple iTunes and Windows Media Player, and streaming music services such as Spotify (www.spotify.com). Consumers are accustomed to browsing by musical genres, and genres *"provide one with a vocabulary that can be used to discuss musical categories"* (McKay & Fujinaga 2006). Using genres, one can rapidly obtain "a manageable set of items" that can then be browsed based on some other criteria. In a study by Laplante (2010), genre was shown to be the most popular criterion of young adults to start a music search, and in (Lee & Downie 2004) (a study with 427 respondents) the probability of using music style/genre to search for music was as high as 62.7%. The most popular genres are also often easy to recognize.

On the downside, dividing songs into genres (and especially sub-genres) can sometimes be very difficult, imprecise, and artificial. The genre boundaries are often fuzzy; there can be overlap between the different genres (cf. rock and alternative rock); and the genre definitions may change over time. One way to deal with this is the multi-labeling of genres (Lukashevisch et al. 2009), i.e., categorizing a single song to more than one genre. Genres can also be culture specific (cf. world music and Christian music), and they may not be related to the actual musical content (cf. Christmas and children's music).

Building a consistent genre hierarchy can be extremely difficult even for experts, and the experts and the general public do not necessarily agree on all the definitions (e.g., rock, pop and electronica & dance [Sordo et al. 2008]). As there is no agreement on the general genre taxonomy, the number of used genres can vary greatly depending on the source. For example, in the year 2000 All Music Guide (www.allmusic.com) used five "meta-genres" with 531 genres, Amazon (www.amazon.com) used 18 and 719, and MP3.com used 16 and 430. Only 70 words were common in the three taxonomies (Pachet & Cazaly 2000). Furthermore, Wikipedia (Wikipedia List of Music Genres 2008) defines hundreds of different musical genres, and the MP3 ID3V2.3 specification (http://id3.org) lists 79 genres (or 147 if the WinAmp extensions are taken into account).

However, selecting the best genre label for a given song/album/artist is the task of a record company, a service provider, or an automatic genre classification algorithm (see, e.g., [Chang, Jang, & Iliopoulos 2010]), and thus the topic is outside the scope of this thesis. All visualization methods presented in the thesis can be mapped to the ID3v2 genre taxonomy or other commonly used genre metadata. In the following, different ways to visualize musical genres are described in more detail.

4.3.1 Album covers

In traditional music player applications, the most common way to visualize songs and albums (in addition to pure textual lists) is to rely on album cover art. For example, Apple iTunes includes a 3D "Cover Flow" view for browsing the music collection based on album front covers. The UI can be operated using, e.g., the on-screen scrollbar or a mouse wheel, and the covers flip from left or right into the center of the screen.

Album covers evoke certain feelings in the consumer, and in an optimal case the design will also convey the emotional message of the music or music style. To achieve this, album covers may utilize colors, fonts, and symbols that are typical for the genre of the album (for example, heavy metal bands often use sharp and edgy fonts in their designs). In Van Egmond (2004), the author tested the graphical and auditory similarity of 25 CDs, and came to the conclusion that the covers were quite successful in this respect.

Album covers have also been used to discover similar-sounding music. For example, Lamere's (Lamere & Eck 2007) Album Cloud visualization shows the cover art of the current album surrounded by covers from related albums. Other visualization from the same paper, the Album Grid, organizes the covers of a music collection in a grid in such a way that similar-sounding albums are close to each other. In addition to the basic planar grid, other shapes such as spirals, boxes, and ellipses can also be used. In the Musicmesh (2007) application, the user first types in the name of an album or an artist, after which the corresponding album cover and some related albums are shown on the screen. The covers are connected together with lines. When the user clicks on an album cover, a list of track names from that album is shown, and when he/she clicks on a track name, a video of that track is streamed from YouTube.

Instead of showing multiple static covers on the screen at the same time, another interesting possibility would be to use the Rapid Serial Visual Presentations (RSVP) method (Brinded 2011), where a series of images (in this case, album cover art) is shown in a rapid sequence on the screen. When the user notices an interesting and/or familiar album cover, he/she could return to that cover and start listening to the corresponding music.

4.3.2 Colors

In the context of music player applications and music recommendation systems, colors have often been used to visualize musical similarity. Colors are excellent for labeling and categorization (Ware 2004, p. 98), and thus they are also a natural choice for visualizing musical genres.

Several applications visualize artist or track similarities using colored circles or spheres on a planar *x-y* space. For example, the Finnish Hitlantis service (www.hitlantis.com) visualizes a collection of unknown artists as a circular cloud of colored circles. Musical genres are color-coded using 13 different colors. For example, R&B music has been mapped to the green color, electro to cyan, country to light green, blues to yellow, and red to punk. Frequently listened artists are located closer to the center of the screen.

In the UPF Music Surfer (2008), the similarity maps can be arranged according to various parameters such as timbre, rhythm, danceability, tonality, tempo, and genre. The UI shows seven color-coded genres at the same time. In one of the given examples, classical music has been represented with green color, New Age with yellow, rock 'n' roll with cyan, ethnic music with purple, and instrumental, blues and hard rock with different shades of blue.

In the artist map user interface (Van Gulik & Vignoli 2005), the user can select his/her preferred visualization by selecting two metadata attributes (e.g., genre, release year, mood or tempo) for the different axes. Either attribute can also be mapped to color. In one of the given examples (Figure 6), artists have been clustered according to genre and linked with colors as follows: rock music is presented with red, popular music with pink, Americana with blue, and dance/lounge with green.

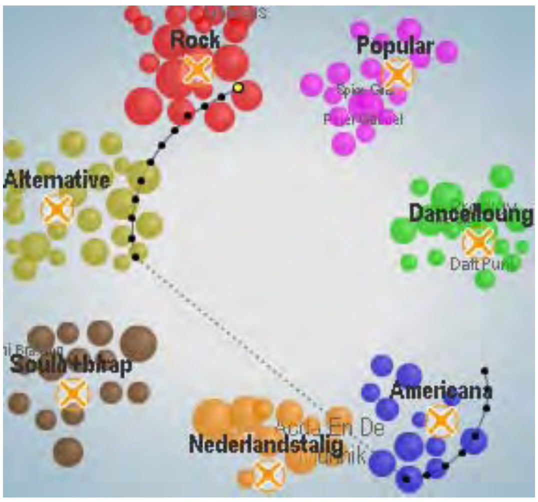

Figure 6 Screenshot from the artist map by Van Gulik and Vignoli (2005). In this example, artists have been clustered according to color-coded genres.

Other types of color-based UIs also exist. For example, the Music Rainbow (Pampalk & Goto 2006) arranges similar artists close to each other on a circular rainbow. Each color represents a different genre, and the rings are ordered in such a way that the most frequent term is located on the outside of the rainbow (and thus represented with red color) and the least frequent term on the inside (purple color). In the "Mood radio" mode of the Musicovery (2011) web radio, the user can select music according to decade, mood, and genre. The 18 genres are color-coded using various shades of rainbow colors. For example, rock and metal have been mapped to blue, reggae and soul to green, and rap to red. In Shiroi et al. (2011), genres and sub-genres are represented as color-coded nodes that are linked together using curved lines. Unfortunately, the paper does not describe the used genre-color mapping in more detail.

4.3.3 Icons

Icons are small visual symbols used in computers, mobile phones, and other electronic devices. By using icons instead of text strings, the designer can reduce the amount of text that has to be localized. As images are easily recognized after only little exposure, icons can serve as memory aids that help to recall the functionality of the system (Ware 2004, p. 230). Descriptive icons can also make the application more fun to use and more pleasing to look at. While photorealistic icons may look pleasing to the eye, studies have shown that simple icons can outperform complex ones (Byrne 1993), and that illustrations are learned faster than photographs (Gooch, Reinhard, & Gooch 2004).

The most important thing in icon design is the metaphor (Holm & Holm 2008). Metaphors can help in matching the design with user's mental model and link the visualization with real-world knowledge (Yi et al. 2008). Whenever possible, the designs should be based on existing metaphors and/or real-life physical objects (so-called resemblance icons [Nielsen 1993, p. 239]). In the case of international UIs, the used graphic language and metaphors should also be kept global.

In addition to their use in album cover art, icons and symbols have not often been used to visualize music collections. One exception to this is Apple's iTunes application, where the grid view of musical genres is based on icons. The most important genres, such as jazz and classical, are visualized with representative icons and the lesser-known genres with album covers from that genre. For example, the blues icon includes the name of the genre on a blue and wooden background, the reggae icon includes the Jamaican flag, world music is represented with a globe filled with small symbols, pop music with stars, and the hip-hop icon includes a loudspeaker and shiny "bling-bling" text. Genre names have been written with representative fonts on top of the icons, and pointing to a genre icon switches between the various album covers from that category.

The Globe of Music application (Leitich & Topf 2007) visualizes the music collection on the face of the globe. Each track is represented using an icon which is textured with album artwork. The user can access tracks by clicking on the corresponding icon; pointing an icon displays the related artist and song names on the screen. In GenreGram (Tzanetakis, Essl, & Cook 2001), each genre is represented with a cylinder that moves up and down in real time based on the confidence of genre classification. The cylinders are texture-mapped with representative images (Figure 7). For example, classical and country music have been visualized using pictures of an old composer and a cowboy, respectively.

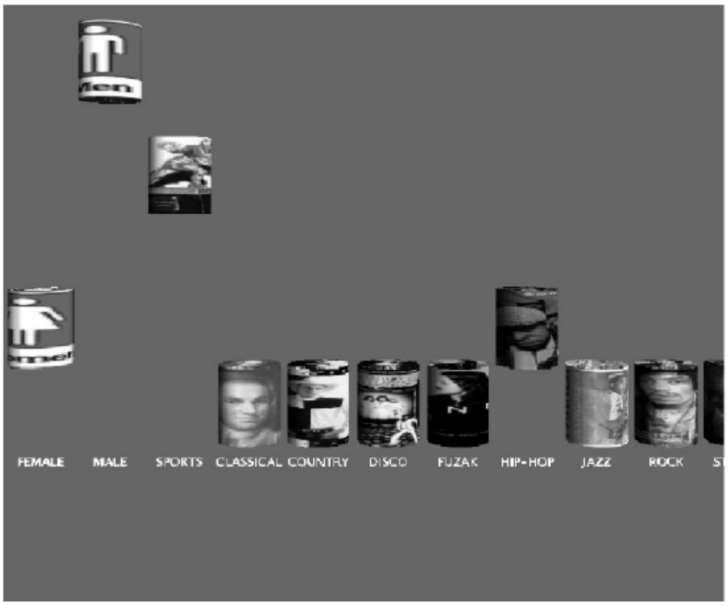

Figure 7 Representing genres with texture-mapped cylinders (Tzanetakis, Essl, & Cook 2001).

4.3.4 Fonts

Especially in the case of mobile phones and other small devices, the amount of screen estate that can be used for presenting information is very limited. In the context of music, one interesting method to add additional information to the screen is to use different fonts to represent musical genres, i.e., embed the genre information in text strings for song, artist, and/or album names without sacrificing the legibility of the text. For instance, heavy metal could be represented using a sharp font, classical music with a penmanship font, and so forth. This technique is commonly used in album cover art, where the artist names are often written with representative fonts.

To our knowledge, the idea has been rarely used in the MIR community. One exception to this is the grid view of Apple's iTunes application, where the genre names have been written with representative fonts on top of the genre icons. For example, the hip-hop icon includes a picture of a loudspeaker and text "hip-hop" written with a shiny "bling-bling" font. Pop music has been represented with a round font, R&B with a golden and shiny font, country with a Western font, and so forth.

4.3.5 Virtual worlds and avatars

A virtual world is a persistent and shared online environment where users can interact and socialize with each other in real-time (Book 2004). Virtual worlds are used for various purposes, including social networking, gaming, education and marketing; recently their potential has also been noticed in the music industry. For example, Second Life (www.secondlife.com) features clubs and radio stations playing different musical genres. There are also live concerts where the performance of an amateur or a famous real-life artist (e.g., U2 or Duran Duran) is streamed to the virtual world and performed by representative avatars.

An avatar can be defined as the *"representation of the self in a given physical environment"* (Castranova 2003, p. 4) or an embodiment of the user (Boberg, Piippo, & Ollila 2008). While the term avatar can also refer to icons or textual descriptions such as names (Boberg, Piippo, & Ollila 2008), many digital environments, such as virtual worlds and games, are based on human-like avatars. These avatars can be modified in various ways to reflect the identity and the current mood of the user, and they can be a convenient and entertaining way to interact with the environment.

Photorealistic or realistic human-like avatars may look pleasing to the eye, but they are often not a good selection for the avatar designs. Iconic cartoon style characters can be faster and simpler to implement, and they are also more universal and subjective (McCloud 1993, p. 46). Gooch et al. (2004) have shown that caricatures are faster to learn than photographs, and Fasel and Luettin (2003) state that exaggerated expressions can be easier to recognize than realistic ones.

Personality has impact on an individual's musical preferences, and music is often used as a means to send messages to other people about one's identity (Rentfrow & Gosling 2003). As a consequence, the user of a virtual world or other digital environment could also modify his/her avatar to reflect his/her musical taste. However, so far the idea has been rarely used in the MIR literature.

In Haro et al. (2010), the authors describe a method for generating a musical avatar representing the user's musical preferences. The system first calculates over 60 low-level audio features (e.g., tempo and spectral centroid) for each track in the user's music collection; these features are then used to generate a set of 77 "semantic descriptors." A subset of 17 descriptors (six genres, six moods and five others such as "party") is mapped to the various visual characteristics of a musical avatar (Figure 8). For example, musical genre (classical, jazz, metal, dance, rock or electronic) is mapped to the head, hair, hat, and instrument of the avatar. The authors conducted a user study with 11 participants, and concluded that *"the Musical Avatar provides a reliable, although coarse, visual representation of the user's music preferences"* (Haro et al. 2010).

Figure 8 Examples of musical avatars by Haro et al. (2010). Used with the permission of the authors.

4.4 Mood and emotion

Music can express, communicate, and evoke emotions in both listeners and performers (Juslin & Sloboda 2001). Active or passive music listening can have an effect on people's mood, facial expressions, and physiological reactions, and music is often used for relaxation, mood enhancement, motivating certain tasks such as sports, and moderating or boosting arousal levels. Therefore, it is also natural for humans to categorize music in terms of emotional reactions.

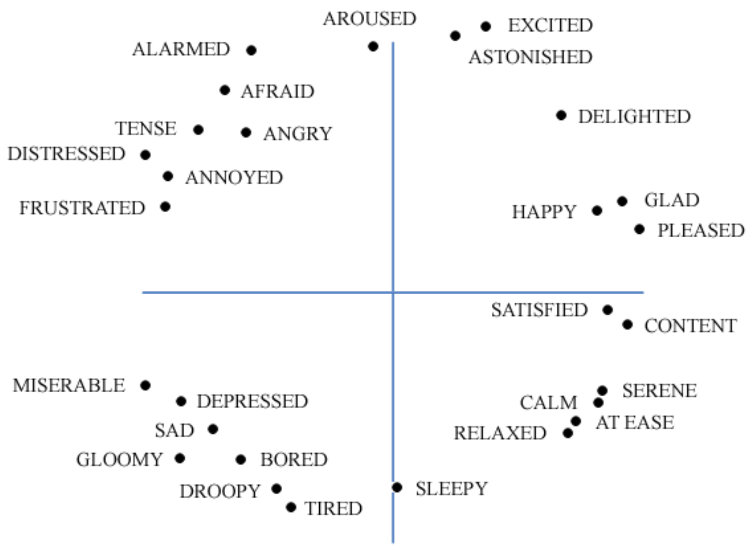

Figure 9 Russell's circumplex model of emotions (reproduced from [Russell 1980]).

In practical systems, the emotions or moods of the listener cannot be directly measured with sensors; they have to be asked from the user using, e.g., emoticons or textual labels. Given the mood information, the music player application can then select suitable music by comparing the information to the mood metadata stored within the music library.

Mood metadata for individual music tracks in the library can be determined using content-based audio analysis, human annotation, or hybrid approaches. Human-annotated mood information can be retrieved from various sources, including expert-annotated labels (e.g., AllMusic), social tags (e.g., Last.fm), annotation games (e.g., MoodSwings [Kim, Schmidt, & Emelle 2008]), web page content, and lyrics.

To classify and recognize emotions, MIR systems typically use categorical or dimensional approaches, which have their roots in psychological research. The categorical approach is based on the concept of basic emotions, meaning that *"there is a limited number of innate and universal emotion categories from which all other emotional states can be derived"* (Sloboda & Juslin 2001, p. 76). The number of basic emotions varies depending on the used reference. For example, Ekman (2004) has proposed six universally recognized emotions: anger, disgust, fear, joy or happiness, sadness, and surprise. In the case of automatic music classification, the number of emotion terms is typically higher than in psychological research. For example, in the year 2007 the All Music Guide service (www.allmusic.com) used 179 mood labels (Hu & Downie 2007), Hu et al. (2009) used 18 categories containing 135 mood tags, and the recent MIREX (www.music-ir.org/mirex) evaluations have used 29 labels divided into five clusters (Kim et al. 2010).

The dimensional approach focuses on *"indentifying emotions based on their placements on a small number of dimensions, such as valence, activity, and potency"* (Sloboda & Juslin 2001, p. 77). One widely used dimensional emotional scale is Russell's (1980) circumplex model (Figure 9), which maps y-axis to activation level and x-axis to valence. In the case of music, the most popular dimensional model is Thayer's two-dimensional valence-arousal space, which has been derived from Russell's more general model (Kim, Schmidt, & Emelle 2008). Thayer's model maps x-axis to the amount of stress and *y*-axis to the amount of energy (Trohidis et al. 2008), and thus categorizes music into four quadrants: high valence and arousal (joy, exuberance), high valence and low arousal (contentment), low valence and high arousal (anger), and low valence and arousal (depression). Both axes are continuous (Kim, Schmidt, & Emelle 2008). Regardless of the used approach, the collection of “ground truth” emotion labels remains a challenging problem (Kim et al. 2010).

One example of a music application based on the dimensional approach is Moody (2011), which is a mood-based playlist generator for iTunes. To use the application, the user has to first tag his/her iTunes music collection according to mood, or else download existing tags for the songs. This is done along two axes, where the vertical axis represents intensity and the horizontal axis denotes happiness. As a default, the axes are color coded in such way that red represents intensive but sad music, yellow intense and happy music, blue calm and sad music, and green happy but calm music. In the Mood Player from Fraunhofer Institute for Digital Media Technology (2009), both images and music are mapped to a 2D valence-arousal space. The system can be used for, e.g., automatically selecting suitable background music for holiday pictures. Yet another example of using the dimensional approach is the “Mood radio” mode of Musicovery (2011), where the user can select music according to decade, genre, and mood (dark, positive, calm or energetic).

Another potential visualization approach is to concentrate solely on colors. As the meaning of colors can vary greatly between different cultures and individual people, there have also been studies on which type of mappings perform well in a musical context. For example, Julià and Jordà (2009) arranged a survey on associating mood labels with colors. The label “aggressive” was mapped to red color with an agreement of 100%, “relaxed” to cyan with 43.5%, “acoustic” to brown with 52%, “happy” to yellow with 39%, “party” to magenta with 48%, and “sad” to blue with 56.5%. In the Colour Player from Voong (2007), the user has to assign tracks to colors manually based on the mood that they convey. After this, the user can select songs, filter tracks, and create playlists by interacting with the resulting color map. In a user study with nine participants, it was learned that the users liked the interface and recalled the used colors well.

Other ways to visualize moods vary from discs to vertical bars, icons (Kim et al. 2009), and emoticons. For example, the Musicream application (Goto & Goto 2005) visualizes songs using discs flowing from taps in the top of the screen. The colors of taps and discs reflect the mood of the songs in such a way that similar songs share the same color. With the Moodagent application (www.moodagent.com), the user can adjust the height of five “mood” bars (sensual, tender, joy, aggressive and tempo) or select a seed song to generate new music playlists. In the case of MoodTunes Lite (WhoopApps 2009), the user can select one of the five predefined moods (angry, bumpin, grooving, happy or pumped) or create new mood definitions with custom pictures. The predefined moods have been visualized with colored emoticons (e.g., angry mood with a red and angry-looking emoticon). Each mood can be mapped to any number of genres from the user’s music library; when the user picks a mood, songs from the associated genres start to play.

4.5 Summary and gaps in research

As shown in Sections 4.1-4.3, the encodings used in different academic and commercial applications vary greatly. In the case of release year metadata, common visualization methods include text strings, sliders, and location on the *x*- or *y*-axis. The desired musical tempo is typically selected using pull-down menus and text boxes, sliders or by clicking on an *x*- or *y*-axis location, and different icons, pictures and "visual thumbnails" have also been used to visualize tempo metadata. To our knowledge, the different methods have not been compared against each other (or other less popular or more "exotic" methods) in any earlier research.

Musical genres have been visualized in several ways including cover art, colors, icons, and to a lesser extent, using fonts and avatars. As the collection of used musical genres varies between the different applications, it is difficult to compare the performance of the different methods against each other, and these types of results have not been published in related work.

To our knowledge, systematic user studies on visualizing release year, tempo, genre, or other type of musical metadata have not been conducted. The evaluation of MIR applications has been dominated by automated system-centered approaches (Hu & Liu 2010), and the focus of music visualization-related research has been on other aspects of the systems than the performance of the used encodings. In most cases, the developers have mainly relied on their initial feelings when selecting the encodings between the different visual variables and musical metadata. Still, many developers could benefit greatly from using more intuitive mappings in their user interfaces. Especially in the case of colors and genres, some of the used encodings felt almost random.

This thesis aims to fill the above-mentioned gaps in the research. The purpose of the thesis is to explore the feasibility of visualizing music collections based on genre, tempo, and release year metadata. More specifically, the research goal is to study which visual variables and structures are best suitable for representing the metadata, and how the visualizations can be used in the design of novel user interfaces for music player applications such as music recommendation systems.

5. Research approach and methods

The work presented in this thesis was started in Nokia Research Center (NRC) as part of a larger project on music recommendation systems. In addition to developing the basic music recommendation technology, basic UI and infrastructure for the SuperMusic system (Section 2.2.2), the project also studied new ways to visualize and access music collections, subsets of the collections, and individual songs.

The original goal of the research was to find out quickly what type of visualizations could be feasible for use in future SuperMusic prototypes. The plan was to design novel visualizations and GUIs to minimize the need for explicit user input and to memorize things, while maximizing the satisfaction and joy of use when accessing music collections. All the different aspects of interaction design (understanding the users, prototype design and evaluation) were involved. Although some other members of the project did work on new MIR signal processing algorithms, the focus was on such visualization methods that could be implemented based on existing MP3 ID3v2 (http://id3.org) or other existing musical metadata.

This chapter gives an overview of the used research approach and methods. After describing the research approach in Section 5.1, the different phases of the research (surveys and evaluations of application prototypes) are discussed in Sections 5.2 and 5.3. The main research questions and contributions of individual publications are summarized in Tables 2 and 3.

5.1 Research approach

As part of the HCI approach, the project used a user-centered and constructive design-science research approach for visualizing music collections. In (Mao et al. 2005, p. 105), user-centered design (UCD) has been specified as *"a multidisciplinary design approach based on the active involvement of users to improve the understanding of user and task requirements, and the iteration of design and evaluation."* Gould & Lewis (1985) described three principles (early focus on users and tasks, empirical measurement and iterative design) for producing useful and easy-to-use computer systems, and these principles now form the basis of the user-centered approach (Sharp, Rogers & Preece 2007, p. 425).

In design-science research, knowledge and understanding of a problem domain and its solutions are achieved through building and evaluating innovative and purposeful artifacts (Hevner et al. 2004). The resulting knowledge can then be used in the design, realization (construction), and improvement of existing artifacts (Van Aken 2004). The artifacts can be constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), or instantiations (implemented and prototype systems) (Hevner et al. 2004). During NRC's music recommendation project, several innovative and purposeful artifacts (graphical representations and prototypes) for music retrieval and visualization were developed and evaluated, and part of the results of that work form the basis of this thesis.

Other closely-related terms include constructive research, design research, and constructive design research. The terms are often intermixed in academic literature. Constructive research refers to the *"building of an artifact (practical, theoretical or both) that solves a domain specific problem in order to create knowledge about how the problem can be solved (or understood, explained or modeled) in principle"* (Dodig-Crnkovic 2010), while design research *"involves the analysis of the use and performance of designed artifacts to understand, explain and very frequently to improve on the behavior of aspects of Information Systems"* (Vaishnavi & Kuechler 2004). Constructive design research can be understood as *"design research in which construction – be it product, system, space, or media – takes center place and becomes the key means in constructing knowledge"* (Koskinen et al. 2011, p. 5). Although all these definitions are closely related, this thesis leans on the definition of (Hevner et al. 2004) as described above.

5.2 Research methods

The conducted music visualization research followed closely Jenkins' (1985) research model, which consists of eight different steps: 1. idea, 2. library research, 3. research topic, 4. research strategy, 5. experimental design, 6. data capture, 7. data analysis, and 8. publish results. After inventing the original idea, a comprehensive literature study was conducted, and the findings were then used to define the research strategy in more detail. The rest of the research was divided into two different empirical phases.

The first phase of the research consisted of a series of surveys studying different graphical encodings (visual variables and structures) for tempo, release year, and genre metadata. For each survey, a collection of visual variables or structures was first developed based on the findings of the literature study and other relevant prior art, and their performance was then measured by asking the survey participants to map them to musical metadata. The quantitative results of the surveys were analyzed and published in international conferences (**P1**, **P2**, **P4**, **P5** and **P9**) and journals (**P3**). These publications contributed mainly to research questions RQ1 and RQ2.

The second phase of the research consisted of designing, implementing, and user-testing novel GUIs and simple visualizations for the SuperMusic system (Section 2.2.2). During the project, in total six interactive hi-fi prototypes for a Windows PC were designed. Three of the prototypes (potentiometers, avatars and a virtual world) were partially based on the findings of the visualization surveys and are thus included in this thesis. In addition, two simple visualizations (icons and emoticons) for the final (unpublished) Symbian Series 60 version of SuperMusic were implemented. The hi-fi prototypes were user-tested using observation, a semi-structured interview, and questionnaires. Both qualitative and quantitative questions were included, and both first impressions and longer-term user experience were studied. The results of the user studies were analyzed and published in international conferences (**P6**, **P7** and **P8**), and these publications contributed mainly to research questions RQ2 and RQ3.

5.3 Surveys

The first part of the research consisted of a series of surveys studying different graphical encodings (visual variables and structures) for musical metadata. Three different types of metadata (tempo, release year and musical genre) were selected, and in the case of genre, five different methods (colors, emoticons, icons, fonts and avatars) were studied.

Inspiration for the graphical designs was drawn from various sources including information visualization literature, modern and digital art, the history and culture of musical genres, comics, movies, and MTV. Several album cover art books such as (Aldis & Sherry 2006), (Emery 2004), (Kinsgsbury 2003), (Marsh 1996), (Marsh 2002), (Morrow 1999), (Pesch 1998), and (Yglesias 2004) were also studied to find the most common elements used in cover art design.

To find a small but representative set of genres for the surveys, we first studied the categories used in 14 European and American online music stores, recommendation services, music guides and radio stations of the time (see Appendix A for more details), and browsed through some MIR and other academic literature until certain patterns started to emerge. The most common genres were selected for further analysis, some popular genres were combined together, and the criteria for making the final decisions were specified. The criteria included, e.g., the target user group of the SuperMusic system and mobile music in general, popularity of musical genres in 2007, the genre list of MP3 ID3v2 meta-data (http://id3.org) as well as genre definitions used in Heittola (2003), Whitman (2005) and All Music Guide (currently known as AllMusic, www.allmusic.com).

After analyzing the genres and finding the commonalities with a good coverage, the following 17 main genres were selected:

1. alternative & indie
2. blues
3. classical
4. country
5. electronica & dance
6. folk
7. gospel
8. (heavy) metal
9. hip-hop & rap
10. jazz
11. Latin
12. New Age
13. pop
14. reggae
15. rock
16. soul, RnB & funk
17. world music

We also designed a mapping from the ID3V2.3 genres and the associated WinAmp extensions to these categories. Appendix B explains the selected genres and their history in more detail.

The surveys were conducted between 2007 and 2009. The survey data was collected using an online questionnaire tool, which was open to the participants for two to three weeks. The questionnaires included mostly closed questions, and the answers consisted mostly of nominal 1D data. For example, in the case of musical genres the participants were shown various images one at a time and asked questions such as "which musical genre comes to your mind when you are looking at picture *X*?" The answer had to be selected from the predefined and numbered list of genres.

As the concepts could not be revealed to the public at this point, the participants had to be recruited using convenience and snowball sampling from the various business and research units of the Nokia Corporation. Convenience sampling is a type of non-probabilistic sampling, where the sample is drawn from that part of population which can be reached easily and conveniently. Snowball sampling refers to situations where the participants may also recruit somebody else to take part in the study. In the case of HCI research, both probabilistic and non-probabilistic samples are considered to be valid data (Lazar, Feng, & Hocheiser 2010, p. 107).

Participation in the surveys was voluntary, and the participants did not get any rewards for their work. In theory, each survey could have had a different set of participants (between-group design [Lazar, Feng, & Hocheiser 2010, p. 46]), but in practice the groups shared some common members because some interested employees wanted to participate in several surveys. The majority of the participants (68-86%) were male, Finnish (59%-85%) and 25-40 years old, with engineering backgrounds. While the resulting sample of users may be slightly skewed, we did our best to find representative users who were interested in music and technology. Thus, the reliability of the surveys was as high as possible under the given constraints.

The main research questions and contributions of each survey have been summarized in Table 2. All these surveys contributed mainly to research questions RQ1 and RQ2 of this thesis.

Publications **P1** and **P2** studied the suitability and performance of different visual variables to represent tempo and release year metadata. In both cases, the research approach differed from the work of other authors (Chapter 4) mainly in the choice of visual parameters and the level of user involvement. Some of the visual features (e.g., location) were selected because they are prominent and logical, while the others were based on previous research and our sketches for visual music player user interfaces. **P1** focused on mapping musical tempo to various visual variables, including the number of objects, shape, size, orientation, color, and blur, and **P2** studied the feasibility of representing release year with variables such as lightness, location, color, and shape.

Publications **P3** and **P5** concentrated on representing musical genres with colors and avatars. To our knowledge, such detailed studies on these topics have not been conducted elsewhere. Parts of the designs in **P5** were based on the previous findings of **P3**, **P4**, and (Holm & Holm 2008). To improve the performance of the designs and make them (potentially) more entertaining, **P5** neglected the most common guidelines for cross-cultural design and internalization (Horton 1994; Aykin 2005) and included some exaggerated and stereotypical designs. The design choice was intentional, and it was done at the risk of not pleasing all the users.

Publication **P4** studied visualizing musical genres in the context of context-aware music recommendation systems. The aim of the study was to investigate which musical genres people listen to in different emotional states, and whether these genres and states can be represented with descriptive emoticons. Based on the results of an online questionnaire, the publication presents a list of genres that could be used as a starting point for making recommendations fitting the current mood of the user.

Finally, publication **P9** summarized the results of representing musical genres with colors (**P3**), icons (Holm & Holm 2008), fonts (Holm, Aaltonen, & Seppänen 2009), and avatars (**P5**). The performance of the different visualization methods was also compared against each other in terms of the Finnish participants.

Table 2 Summary of publications documenting the results of the surveys. *N* refers to the number of questionnaire participants.

Publication	Main research question(s)	Research method	Main contributions
P1	Which visual variables are best suited to represent tempo metadata? What is their performance from the user perspective?	Online questionnaire (*N*=75)	Performance of selected visual variables and guidelines for visualizing tempo metadata
P2	Which visual variables are best suited to represent release year metadata? What is their performance from the user perspective?	Online questionnaire (*N*=51)	Performance of selected visual variables and guidelines for visualizing release year metadata
P3	Can musical genres be represented with colors in a recognizable way? What is the performance of the visualizations from the user perspective?	Online questionnaire (*N*=104)	A compromise color-genre mapping for the Finnish users
P4	Can musical genres and moods be represented with emoticons? Which musical genres do people listen to in different emotional states?	Online questionnaire (*N*=87)	A list of genres that could be used as a starting point for making recommendations fitting the current mood of the user
P5	Can musical genres be represented with avatars in a recognizable and acceptable way? What is the performance of the visualizations from the user perspective?	Online questionnaire (*N*=71)	Performance of selected avatar designs and guidelines for representing genres with avatars
P9	Which is the best method for visualizing musical genres, and which genre is the easiest to visualize?	Online questionnaire (*N*=51-104)	Performance of colors, icons, fonts, and avatars for visualizing musical genres

5.4 Evaluations of application prototypes

In the second phase of the research, findings from the surveys were used to design novel GUIs and simple visualizations for the latest version of the SuperMusic system (Section 2.2.2). Instead of using scatterplots, lines, or other conventional and "engineer-like" representations used in many visual MIR applications (Chapter 4), the goal was to develop novel, entertaining, and easy-to-use means for accessing the SuperMusic service. Due to project constraints, the focus was only on musical genres.

5.4.1 Hi-fi prototypes for PC

During the project at Nokia Research Center, in total six interactive hi-fi prototypes for a Windows PC were designed. These included avatars (**P6**), a virtual world (**P7**), potentiometers (**P8**), mood pictures (Lehtiniemi & Holm 2012), a rotating cube (Lehtiniemi & Holm 2011a), and preview clips placed in an "album cover space" (Lehtiniemi & Holm 2011b). The prototypes were mostly intended at passive music discovery; they could be seen as a means to generate smart playlists or personalized radio channels, or as means to select seed songs for a music recommendation system without specifying explicit artist or song names. Three of the prototypes (avatars, potentiometers and the virtual world) were based on the findings of the visualization surveys; they were thus included in this thesis as well. In the light of Pousman et al. (2007), the designs could also be considered as Casual Infovis systems (Section 3.3): they visualize personal information (the preferred musical genre), and can be used by a wide spectrum of users in everyday music listening situations.

The prototypes were implemented as Flash applications to enable the future use of different platforms, including personal computers and mobile devices such as the Nokia N900. As suitable mobile devices or the Apple iPad were not available at the time of implementing the prototypes, the main test platform was an HP Pavilion tx2500 touch screen laptop. Five of the prototypes could be used with the device's touch screen, and the virtual world application only with the included keyboard.

In the avatar prototype (**P6**), music recommendations were generated based on the selected combination of avatar parts. Each avatar part represented a single musical genre, with the designs partially the same as in publication **P5**. In the potentiometer prototype (**P8**), the user was able to fine-tune the contents of the music recommendation playlist by turning tempo and energy level potentiometers. The look of the interface (skins and icons) was changed to reflect the currently selected genre. The used icon and skin designs were partially based on the findings of publications **P3**, (Holm & Holm 2008), and **P4**. Publication **P7** studied the concept of using virtual worlds as a music player interface. In the evaluated 3D prototype, the user moved around a city and interacted with various characters, the looks of which reflected the musical style that they were recommending. While we had very limited means to affect the looks of the characters in the development phase, they shared some of the ideas presented in **P3** and (Holm & Holm 2008). Figure 10 includes screenshots from all three prototypes.

Each prototype was evaluated by 40 or 41 participants during the year 2010. The participants were selected using convenience sampling from varying age groups and education levels. Participation was voluntary with a reward of two movie tickets per participant. Each prototype was evaluated by the same set of Finnish participants (within-group design [Lazar, Feng, & Hocheiser 2010, p. 46]), and this time both Nokia internal and external participants were used. In the case of publications **P6** and **P8**, 8% (3 participants) were 12-17 years old, 12% (5) 18-30 years old, 55% (22) 31-40 years old, and 25% (10) 41-55 years old. 63% of the participants were university graduates, and the rest were split between all levels of education from elementary school through college and university. In the case of publication **P7**, the corresponding numbers for the age groups were 7%, 12%, 56%, and 25%.

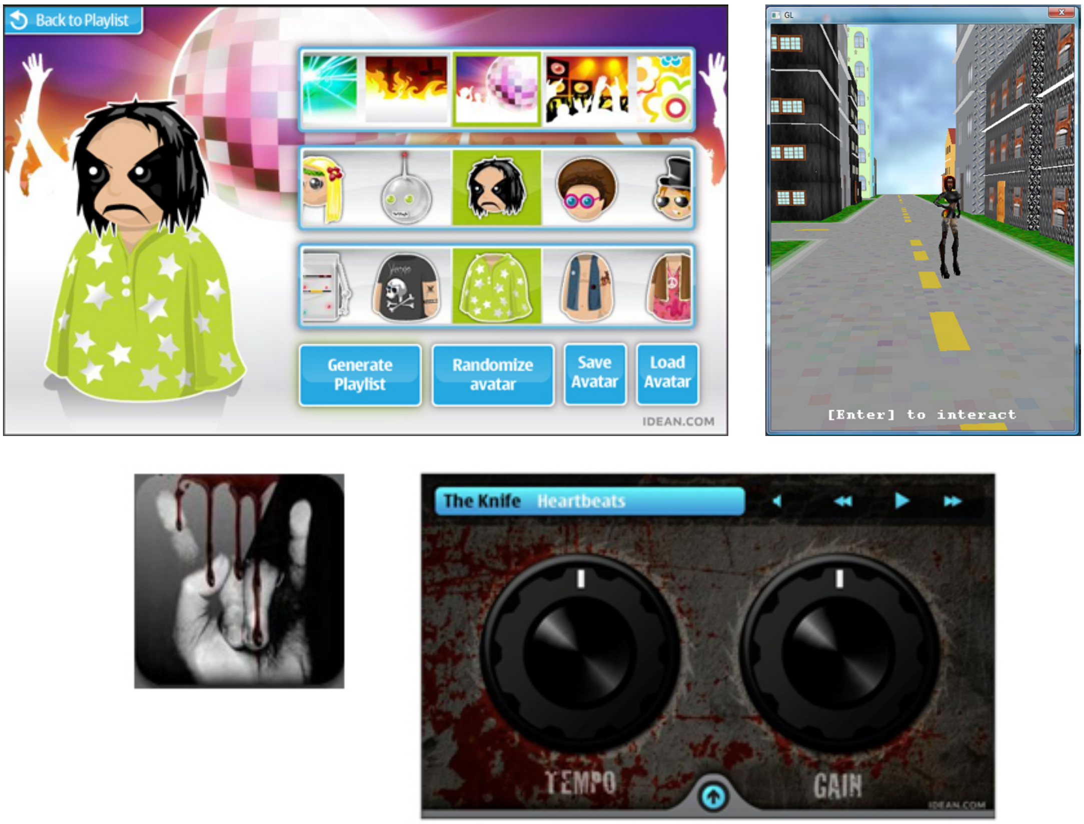

Figure 10 Screenshots from the evaluated prototypes. Top left: avatars (P6), top right: virtual world (P7), and bottom: potentiometers (P8).

All prototypes were tested during the same session in a random order. The used research method was a combination of observation, a semi-structured interview, and questionnaires. The user study included both qualitative and quantitative questions, with both first impressions and longer-term user experience studied. To increase the validity of the results, each prototype used a subset of the genres used in the genre surveys (Section 5.3). In the HCI literature, this use of multiple data sources to gather data (in this case, ratings for the graphical representations) is also known as data source triangulation (Lazar, Feng, & Hocheiser 2010, p. 295).

The interview sessions were arranged in various locations, including the participants' homes, the authors' homes, and the premises of Nokia Research Center and Tampere University of Technology. 37 interview sessions were arranged using the HP Pavilion tx2500 touch-screen laptop, with the rest using a basic Dell Inspiron laptop. In the beginning of the interview session, the users had to fill in a short background information questionnaire. Next, they used the software freely for 5-15 minutes while being observed. The participants were asked to describe their first impressions on the prototype, and various quality aspects of the prototype were studied by asking questions and filling in a questionnaire. Both hedonic and pragmatic aspects (Hassenzahl et al. 2000) were studied. In the questionnaire, the participants had to rate several aspects of the software on a seven-point Likert scale (1=totally disagree, 7=totally agree). In the case of the virtual world prototype, nine users also filled in the AttrakDiff questionnaire (http://www.attrakdiff.de).

After the interview session, the users were able to use online versions of the avatar and the potentiometer prototypes at home for approx. three weeks. When the evaluation period was over, they had to evaluate the prototypes by completing a questionnaire similar to the one used during the interview. Evaluations at home were performed using the participants' own computers, i.e., using a standard keyboard instead of the touch screen.

The results of the user studies were reported in **P6**, **P7**, and **P8**. Table 3 summarizes the main research questions and contributions of the individual publications. All the publications contributed mainly to research questions RQ1 and RQ3 of this thesis.

Table 3 Summary of publications documenting the results of the prototype user studies. *N* refers to the number of user study participants.

Publication	Main research questions	Research method	Main contributions
P6	How well does the avatar concept work in practice, and how can it be improved? How does the user experience change in longer-term use? Can musical genres be represented with avatars in a recognizable and acceptable way?	Observation, semi-structured interview, questionnaires (*N*=40)	Ratings for the short and longer-term quality aspects of the prototype, ratings for the performance of the avatars, and ideas for the future development of the concept
P7	How well does the virtual world concept work in practice, and how can it be improved?	Observation, semi-structured interview, questionnaires (*N*=41)	First impressions of the concept, ratings for the short-term quality aspects of the prototype, ratings for the performance of the avatars, and ideas for the future development of the concept
P8	How well does the potentiometer concept work in practice, and how can it be improved? How does the user experience change in longer-term use? Can musical genres be represented with potentiometer skins and icons in a recognizable and acceptable way?	Observation, semi-structured interview, questionnaires (*N*=40)	Ratings for the short and longer-term quality aspects of the prototype, ratings for performance of the potentiometer skins and icons, and ideas for the future development of the concept

5.4.2 Simple visualizations for Symbian

In addition to the hi-fi prototypes, NRC's music recommendation project developed a new (unpublished) version of the Symbian Series 60 client for the SuperMusic system. In addition to various other improvements, the application utilized a subset of the emoticons used in **P4** and the icons used in (Holm & Holm 2008). The users were able to select an emoticon describing their current mood, and the associated genres were then used as a starting point for recommending new music fitting that mood. Descriptive genre icons were used to visualize the genre of each item in lists of new music recommendations (Figure 11) and to visualize the genre of a "Now recommended" playlist located on the front page of the application. Unfortunately, the research was stopped before the visualizations were evaluated by any external users, and Nokia decided to focus on its (now closed) Comes with the Music service (Barrett 2011).

After evaluating all prototypes and visualizations and comparing them against each other, the next step of the project would have been to design the "ultimate" music player application based on the findings of the studies. Unfortunately, the project was stopped before this last step was taken.

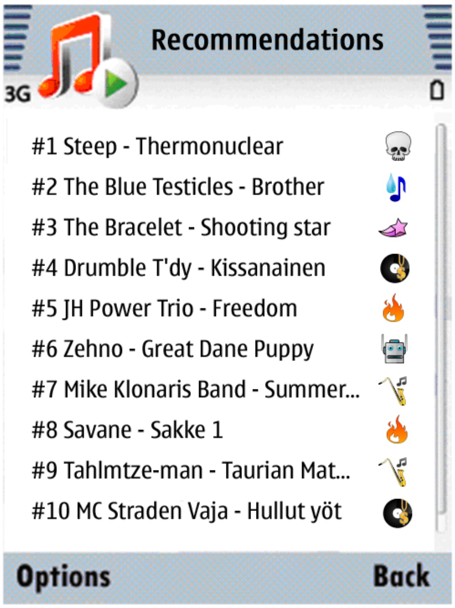

Figure 11 Illustrative example of supporting textual lists with genre icons.

6. Results

The purpose of the thesis was to explore the feasibility of visualizing music collections based on three types of musical metadata: musical genre, tempo, and release year. The three main research questions were specified in Section 1.2; in this chapter, the questions are discussed and answered one by one.

6.1 Visualizing musical metadata

The first research question (RQ1) was formulated as *"Can musical genres and other musical metadata be visualized in a recognizable and acceptable way?"* The question was not addressed specifically in a single publication; instead, the question was more general and related to the publications as a whole. In a sense, RQ1 was the starting point of the whole research project. Due to the scarcity of related MIR user studies, it was unclear how our ideas would be perceived and accepted by the users, and thus we wanted to study how far we could proceed with the ideas.

Based on the results of the publications, the short answer to RQ1 is "yes, but with certain limitations." Publications **P5**, **P6**, **P8** and **P9** demonstrate that it is possible to design recognizable, acceptable, entertaining, and easy-to-use visualizations for musical genres. However, there is a risk that some people may be offended by the stereotypical avatars (see Section 7.4), thus making that particular visualization method less acceptable in some cultural areas. In the case of tempo (**P1**) and release year (**P2**), there were some promising initial survey results, but we were not able to test the designs in practical applications. Other important metadata types from the users' perspective were also identified (Holm 2008), but they could not be studied yet due to the goals and time constraints of the project.

As shown in the following sections, the performance, suitability, and acceptability of (especially genre) visualizations depends on factors such the type of musical metadata, metadata vocabulary (e.g., the set of musical genres), used visual variables and/or structures, available screen estate (size of the visualization), user's preferred music discovery type, the target culture of the visualizations, and so forth. All the design decisions were based on Western views, and they were evaluated with mostly Finnish users. The results are valid for the cultural context in which they were studied, but further research is needed to validate them in other contexts as well.

6.2 Visual variables and structures

The second research question (RQ2) was expressed as *"Which visual variables and structures are best suited to represent the metadata, and what is their performance from the user perspective?"* Three different types of metadata (tempo, release year and musical genre) were studied, and in the case of genre, five different surveys (colors, emoticons, icons, fonts and avatars) were conducted.

To summarize, promising ways to represent tempo include the number of objects, shapes with a varying number of corners, and *y*-axis location combined with some other visual variable (e.g., the size of the objects) or clear labeling. The findings are also well in line with most previous research. Promising ways to represent release year include lightness and perceived location on the z-axis. These were not used in any of the studied prior art; instead, many applications mapped the year to the *x*-axis location. In the case of musical genres, the most successful method was avatars, followed by icons with a similar performance, with fonts coming third and the least successful method being colors. The best performance was achieved by combining different visualization methods, but this also consumed the most screen estate. With the exception of colors, the other studied methods have not been widely used in the MIR community.

In the following, the results for each metadata type are discussed in more detail.

6.2.1 Tempo

Publication **P1** focused on mapping musical tempo to selected visual variables including the number of objects, size, shape, blur, location, orientation, color, and lightness change. Based on the survey results, promising variables to represent tempo include the number of objects, shapes with a varying number of corners, and location on the *y*-axis combined with some other visual variable (e.g., the size of the objects) or clear labeling. In all of these cases, higher numbers mapped to faster tempos than lower numbers. For example, the star shape was associated with fast music by 84% of the participants, while a circle was generally associated with slow music (68%).

While the size of the objects alone was a relatively poor variable for representing tempo (e.g., 56% of participants associated small-sized circles with fast music and 28% with fast music), the performance can be improved by combining size with some other visual variable. For example, when participants were shown a picture where the small rectangles were located at the top and large ones at the bottom, 80% of the participants associated the top with fast music. The results also suggested that size dominates over location.

In the case of lightness, the results followed the logic that the direction of the lightness change defined the axis where the tempo should be mapped. Regardless of the direction of the change, top and right dominated over bottom and left. Thus, location also dominated over lightness, implying that lightness alone may not be a good way to present tempo. Other unsuccessful ways to visualize tempo metadata included colors (with the exception of red, which was associated with fast music by 60% of the participants), level of blur, depth order of objects, and orientation.

The findings are well in line with most previous research (Section 4.2). For example, Chen and Kuber (2010) map musical tempo to the number of petals around a circle in such a way that higher numbers represented faster music. In (Kusama & Itoh 2011), a larger number of small circles represent faster music and a smaller number of large circles represent slower music. The artist map (Van Gulik & Vignoli 2005) and Musicovery (2011) maps tempo to location on the *y*-axis in such a way that tempo decreases from top to bottom. In addition, the artist map represents tempo with color, i.e., uses two visual variables (redundant coding) for the same metadata. Zhu and Lu (2002) map music clips into a 2D timbre-rhythm space where tempo decreases from right to left. Based on the results of **P1**, using location on the *x*-axis is not a very intuitive solution, but Zhu and Lu have labeled the axis clearly.

6.2.2 Release year

Publication **P2** focused on mapping the release year of the music to selected visual variables including location, lightness change, shape, blur, and color. Based on the survey results, promising ways to represent the release year include lightness and perceived location on the *z*-axis. Light, sharp and foreground objects were most often associated with newer music, while dark, blurred and background objects were associated with older music. For example, 88% of the participants mapped foreground objects to new music. Unsuccessful ways to represent the release year included varying the number of points of a shape, location on the perimeter of a circle, and mapping colors to decades.

In the case of a rectangular shape with lightness change, the darker region should be located on the bottom or on the left side of the picture. In the case of circular shapes, darkness should increase towards the edges. Our other tests also supported the idea of mapping dark colors to old music and bright colors to new music.

The participants associated new songs with both high *x*- and *y*-coordinate values, and *y*-axis dominated slightly over *x*. Despite the quite low association percentages, location on the *x*-axis is a common way to represent time, and thus it may also be a good alternative for representing release year or decade. However, the axis should be labeled clearly or supported with some other visual variable (e.g., lightness change) to avoid confusion. Unlike in the case of tempo (where size dominated over location), location dominated over size. However, size had some effect on the results; low *x*- and *y*-values were more clearly associated with older songs when smaller objects were placed on the left or bottom and larger objects on the right or top.

The most successful variables (lightness and *z*-axis location) are not used in any publications from the other authors (Section 4.1). Instead, most applications map year to location on the *x*-axis. For example, the "Mood radio" mode of Musicovery (2011) presents decade as a horizontal range slider increasing from left to right. The artist map user interface (Van Gulik & Vignoli 2005) visualizes artist similarities with colored circles located in an *x*-*y* space. In one of the given examples, release year has been represented with both color and location on the *x*-axis, and the coloring changes gradually from black (<1960) to white (2000+). Examples of mapping time to the *y*-axis also exist. The Musiclens application (2007) maps release year information to a vertical slider, and the rectangle visualization of (Torrens, Herzog, & Arcos 2004) maps time to vertical location.

6.2.3 Genre

In the case of musical genres, one visual variable (color) and four different visual structures (emoticons, icons, fonts and avatars) were studied. In the following, the results are first presented separately for each method, and then compared against each other to find the most effective way to visualize genre metadata.

Colors

The performance of colors was studied in **P3**. Based on the results on an online survey with international participants, it was concluded that it is not possible to design a globally accepted color-genre mapping. For example, while a large majority (75%) of Western participants gave one of their 1-3 votes to blues, the corresponding number for Eastern participants was only 39%. Thus, the publication proposed that a music player application could be set-up with a default mapping that is most suitable for the given country or region. The following compromise mapping to be used in Finland was suggested: red – rock and alternative & indie; green – country and folk; yellow – reggae and Latin; blue – blues and jazz; black – metal; white – gospel and classical; pink – pop; cyan (or some other artificial color) – electronica & dance; gray – unclassified songs; orange – soul, RnB & funk and hip-hop & rap; brown – world music; and purple – New Age.

As the associations between colors and genres are highly subjective and the association percentages in general were relatively low, the results indicate that colors alone are not a good general solution for visualizing musical genres. However, should colors be used, the survey results are still the best starting point that there is. The color-genre mappings used in the academic literature vary greatly (Section 4.3.2), and none of them are reportedly based on the results of a scientific study. The authors have most likely relied on their initial feelings or used random colors; consequently most of the used encodings are not line with the findings of publication **P3**. For example, the UI of Hitlantis (www.hitlantis.com) visualizes blues music (which was mapped to blue color by 61% of the survey participants) with yellow color, and the "Mood radio" of Musicovery (2011) uses blue to represent rock and metal. While some successful color-genre associations also exist (e.g., the UPF Music Surfer (2008) maps blue to blues and Hitlantis maps electro to cyan), these are just individual examples.

Interestingly, the suggested compromise mapping has similarities with music associated with the different chakras, i.e., force centers of the human body. The concept of chakras is featured in tantric and yogic traditions of Hinduism and Buddhism; each chakra is associated with a certain color. In Luukkonen (2006), the associations are red – rock (especially punk and heavy rock); orange – soul; yellow – classical, jazz & Latin; green – folk & world music; blue – silence, water and wind sounds, etc.; and violet – new age. The compromise mapping is also relatively well in line with (Julià & Jordà 2009), where user study participants mapped the label "aggressive" to red color with an agreement of 100%, "relaxed" to cyan with 43.5%, "acoustic" to brown with 52%, "happy" to yellow with 39%, "party" to magenta with 48%, and "sad" to blue with 56.5%. Based on **P4** and AllMusic web pages (www.allmusic.com), sadness is commonly associated with blues music, aggressiveness with metal music, and so forth.

Emoticons

Publication **P4** studied visualizing musical genres in the context of context-aware music recommendation systems. In this type of recommenders, the user's current emotional state also plays an important role. The aim of the study was to investigate which musical genres people listen to in different emotional states, and whether these genres and states can be visualized with descriptive emoticons. Instead of using annotated mood information for the individual music tracks, the idea was to rely on associations between genres and moods and map each emoticon to a number of genres from the used music library.

To study the idea, seven emoticons ("sad", "angry", "happy", "neutral", "annoyed", "feeling great" and "sleepy") were designed. The used colors were well in line with colors used in some previous MIR research (e.g., Julià & Jordà [2009]) and color-coding conventions such as Morton (1997). While the size restriction (32x32 pixels) had some effect on the design decisions, all emoticon designs except the “neutral” one were recognized well and mapped to those emotion/mood words that we expected.

Based on the results of the conducted online survey, the publication presents collections of genres that could be used as a starting point for making recommendations fitting the current mood of the user. For example, 54% of blues music fans preferred to listen to blues when feeling sad (cf. 21% when feeling happy). From the music player’s point view, it would thus be a good starting point to recommend blues music to blues fans who have selected the “sad” emoticon to represent their current mood. When feeling happy, participants listened to lots of all types of music, and they were also more open to experiment and try new types of music. In the case of the “sleepy” emoticon, there was quite a lot scattering in the results, but in general the participants seemed to prefer either relaxing music or other music that was not too aggressive. Genres with strong beats (e.g., reggae, hip-hop & rap and Latin) were often avoided, but some participants also commented that their preferred musical style depends on whether they want to cheer up or fall asleep. The results of “annoyed” and “feeling great” were close to those of “angry” and “happy”, and thus they were not included in the latest Series 60 implementation of SuperMusic (Section 5.4.2).

Compared to previous work from the other authors (Section 4.4), a similar type of approach has only been taken in the MoodTunes Lite application (Whoop Apps 2009). In the case of MoodTunes Lite, the user can select one of five predefined emoticons (“angry”, “bumpin”, “grooving”, “happy” or “pumped”) or create new mood definitions with custom pictures. Each mood can be mapped to a number of genres from the user’s music library; when the user picks a mood, songs from the associated genres start to play.

Icons

Holm and Holm (2008) studied icons as a method for increasing the amount of useful musical information on the screen. Some selected symbols were against good icon design and ethic guidelines, but this was done on purpose, as we wanted to study how they competed with the less controversial symbols. Based on the results of an online questionnaire, it was shown that musical genres may be presented with icons and that the users can recognize the icons accurately. The most successful icons were country (96% of participants made the same association between icon and genre), world music (93%), classical (92%), gospel (89%), and jazz (79%). Interestingly, the icons used in iTunes’ grid view of musical genres (released after our publication) share many common elements with our designs. The winning icons were also implemented in the latest version of the SuperMusic Series 60 client, but that project unfortunately ended before the visualizations were evaluated by any external users. Still, the high association percentages (combined with the fact that the participants listened on the average to only six musical genres) imply that people would get accustomed to an icon based music player quickly and easily.

Some of the ideas presented in (Holm & Holm 2008), **P3**, and **P4** were later used in **P8**, where the look of the interface (skins and icons) was changed to reflect the currently selected genre. To simplify the prototype design, only four genres (metal, soul, rock, pop and electronica & dance) were included. The performance of the skins and the icons was studied both before listening to any music and after using the prototype at home. In the latter case, the users were asked to rate how well the look of the skins and the icons matched with the recommended music on a seven-point scale (1=did not match at all, 4= neutral, 7=perfect match).

The best ratings were given to the metal genre. During the initial interview, 93% of the participants associated the metal graphics with some type of metal music, and there was also a good match (median 6.0) between the recommended music and the designs. Soul came next with a median of 6.0. However, the initial impressions during the interview were not that good (48%), and thus the designs were not optimal. In the case of rock music, the corresponding numbers were 5.5 and 65%. Before listening to any music, the pop and electronica & dance designs were associated correctly by 93% and 83% of the participants. However, the medians of ratings given after using the prototype at home were only 5.0, implying that the seed songs for these genres were not optimal. In the case of pop music, another explanation may be the ambiguity of the label; the recommendations may not have matched all of the participants' own definitions of what pop music is.

Fonts

Holm, Aaltonen and Seppänen (2009) described the concept of embedding genre information in text strings for, e.g., song, artist, and album names. An online questionnaire to study the feasibility of representing genres with fonts was arranged, and the results suggested that the concept works reasonably well when a good representative font has been found. The most successful font collections included classical (91% of participants made the same association between font and genre), metal (87%), country (85%), electronica & dance (84%), and world music (58%). The findings are well in line with iTunes, the grid view of which (Section 4.3.3) shares many common elements with our designs. However, the results also suggest that is very difficult to find representative fonts for certain genres, and thus fonts should not be the first choice when selecting a method for visualizing musical style.

Avatars

The use of musical avatars was studied separately in three different publications (**P4**, **P6** and **P7**), and the results of **P4** were also included in the comparison of **P9**.

Publication **P4** studied the concept of mapping musical genres to stereotypical avatars consisting of head, body, and background elements. The avatar designs were partially based on the results of publications **P3**, **P4**, and (Holm & Holm 2008). To improve the performance of the designs and make them more entertaining, the avatars neglected the most common guidelines for cross-cultural design and internalization (Horton 1994; Aykin 2005) and included some exaggerated and stereotypical designs. The design choice was intentional, and it was done at the risk of not pleasing all the users.

Based on the results of the online survey, the basic idea works well when a good, representative avatar has been found. All the designs were associated with those genres that they were intended to represent, and in many cases the association percentages were close to 100%. The most successful avatars were classical (all parts of which were recognized correctly by at least 93% of the participants), electronica & dance (≥ 85%), gospel (≥ 89%), hip-hop & rap (≥ 83%), metal (≥ 79%), and reggae (≥ 79%). The least successful avatars were blues (individual avatar parts 13-42% and complete avatar 52%), folk (25-38% and 38%), and Latin (23-35% and 45%). These avatars would require a completely new approach and should thus be redesigned from scratch. Avatars requiring some minor modifications included alternative & indie, country, jazz, New Age, pop, rock, world music, and soul, RnB & funk.

With the exception of one comment on the alternative & indie torso, we did not receive complaints from the survey participants. In this respect, the use of exaggerated and stereotypical designs ended up being a good design decision. However, as discussed later in Section 7.4, there is a risk that some people from other nationalities may consider the stereotypes as racist, thus making the concept less suitable in certain countries or regions.

In **P6**, the avatars were used for discovering new music. After selecting the three avatar parts, the application generated a new (cross-genre) playlist by seeding the SuperMusic server with examples from the selected genres. The application included five complete avatar designs (titled “Metal”, “Rock”, “Electronica & Dance”, “Hippie” and “Soul & Disco”), which were partially the same as in the designs in **P5**. In this context, the Hippie avatar (the folk design of **P5**) was mapped to music from the 60’s/70’s (Grateful Dead, CCR, Free, The Beatles, Rory Gallagher, etc.) and newer artists playing in the same style.

During the conducted user study, the users were asked to rate how well the look of the avatars matched with the recommended music on a seven-point scale (1=did not match at all, 4= neutral, 7=perfect match). To summarize, the avatars performed well; most participants felt that the looks matched well with the related music recommendations (all the medians being either 5.0 or 6.0). The results were also well in line with the findings of **P5**.

In **P7**, we experimented with musical avatars in the context of virtual worlds. Due to technical limitations and resourcing problems, we had limited means to affect the looks of the characters in the development phase. This may have affected the fact that the results were less promising. Many characters were not representative enough, the female characters were considered to be too sexual, and many user study participants did not associate the included movie, cartoon, and game characters with soundtrack music.

As the only similar type of musical avatar application that we are familiar with is (Haro et al. 2010), it is very difficult to compare the results to prior MIR work. Haro et al. mapped musical genres to the head, hair, hat, and instrument of the avatar, but specific design details were not given. The authors also concluded that *“the Musical Avatar provides a reliable, although coarse, visual representation of the user’s music preferences.”*

A comparison of colors, icons, fonts, and avatars

Colors, icons, fonts, and avatars were also compared against each other in **P9**. To allow comparisons between the compromise color-genre mapping for the Finnish users and the other three visualization methods, the publication concentrated only on the Finnish participants of the surveys.

The comparison was done in terms of boxplots and a balloon plot, which has been replicated in Figure 12 for convenience.

In Figure 12, the association percentages have been rounded down to the nearest multiple of 20% and represented as circles of varying color and size. By looking at the columns, one can notice that the most successful method was using avatars to represent musical genres. Icons followed second with a similar performance, fonts came third, with the least successful method being colors. In the case of colors, there were two clear positive exceptions (metal and blues). By looking at the rows, one can see that metal, classical, country, and electronica & dance genres were the easiest to visualize, while the results for some genres, such as folk, were relatively poor. Based on the results, the best performance can be achieved by combining different visualization methods; while colors alone were not a good method for visualizing musical genres, the best performing method (avatars) was partially based on the findings of the color and icon surveys.

In the case of avatars, icons and fonts, removing the answers of other nationalities than Finnish had only marginal effect on the results. To illustrate this, Figure 13 shows the association percentages for the best-performing complete avatar designs.

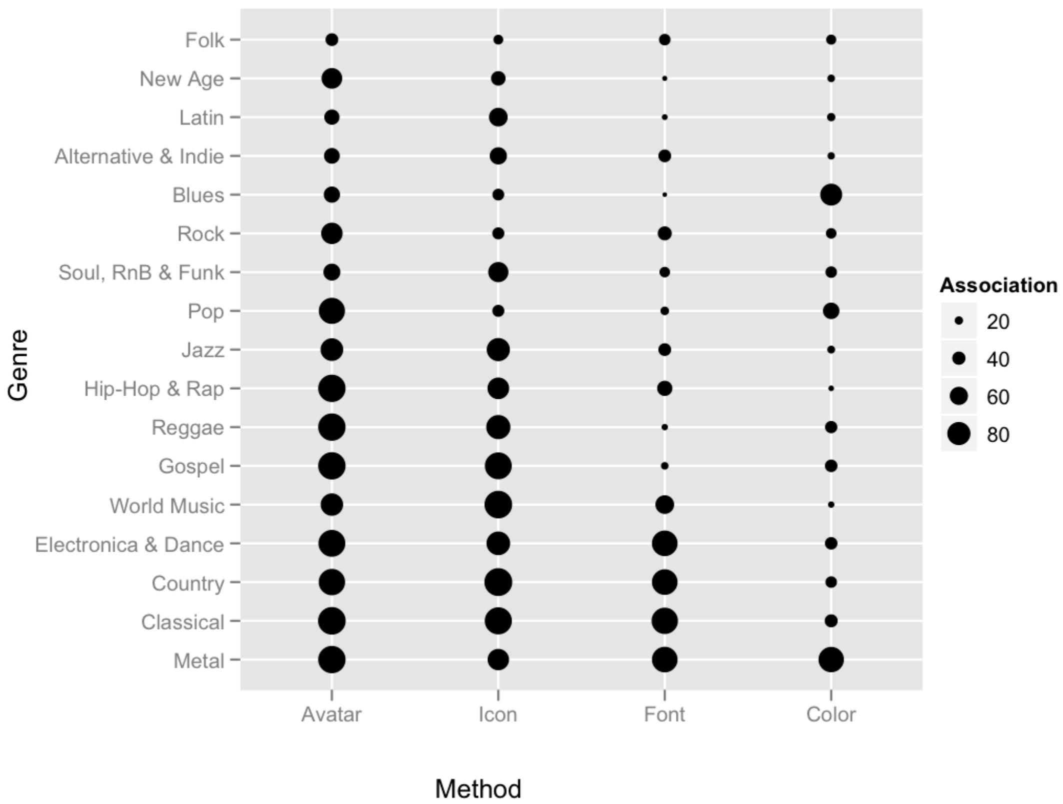

Figure 12 Comparison of different visualization methods and genres using a balloon plot (P9).

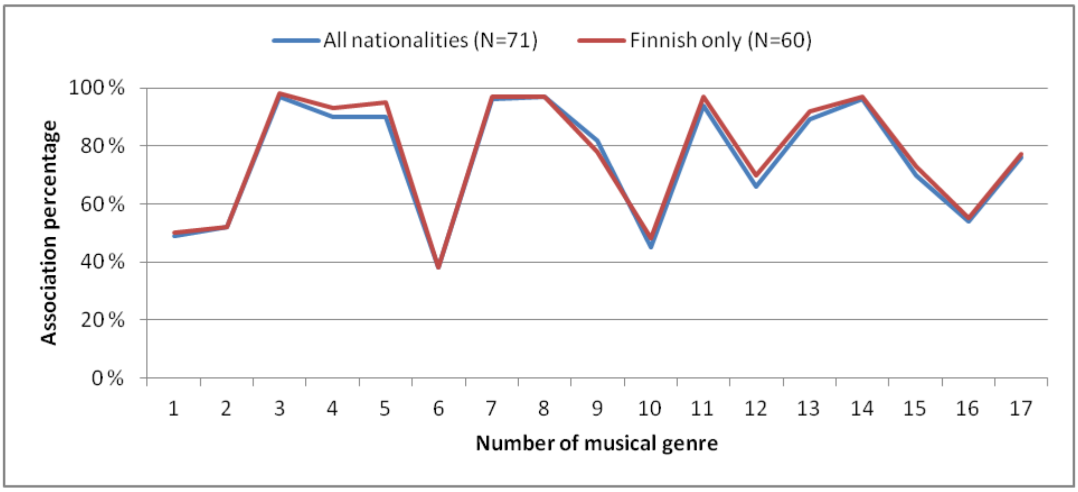

Figure 13 Association percentages for the best-performing complete avatar designs. Numbers on the *x*-axis refer to the numbers of used musical genres (Section 5.3).

In addition to association percentages, another way to look at the results is to consider the minimum amount of screen estate that the different visualization methods require to convey information. Naturally, the amount of available screen estate (type of device etc.) has an effect on how many cues to a certain musical genre (or other metadata type) can be included in the visualization; this was also reflected in the results.

Colors require the least space to represent musical genres. While a single colored object could fill a large part of the screen, the amount of conveyed information would still remain the same. Only if the size of a colored object becomes "very small" may it be difficult to distinguish the colors. The minimum size depends on various factors, including the set of used colors, used background color, etc. (Ware 2004, pp. 125), and for example Horton (1994, p. 57) instructs that the size for colored points should not be smaller than 5x5 pixels. On the other hand, the color method performed worst in the comparison, and **P3** suggested that it is not possible to design a globally accepted color-genre mapping.

Descriptive avatars require considerably more space than the other studied visualization methods, but they are also able to convey more information (cues about the genre) to the viewer. Whenever feasible, our avatar designs combined ideas (colors, facial expressions and symbols) from the previous publications. The high recognition rates do not necessarily mean that avatars are always the best method to visualize musical genres; especially in the case of mobile devices, the smaller-scale visualizations can be helpful in adding extra information to the screen.

The size of the designed icons (and emoticons) was as small as 32x32 pixels. While this had effect on many design decisions (we were not able to implement all our ideas using such a small number of pixels), the results show that it is possible to design small but recognizable genre icons. Representative fonts also consume very little (if any) extra space from the screen if selected properly. Still, there are some important differences between icons and fonts. When the amount of available screen space increases, one can easily embed more information (different colors, symbols, etc.) to the icon designs but the fonts remain (almost) the same – the characters will just be larger. Furthermore, using multiple different-looking fonts in the same UI view is generally considered to be a bad idea.

Multiple fonts can make the interface difficult and annoying to read, and the fonts may not work well together (Fox 2005, p. 95). The results are also likely to be aesthetically less pleasing. Nevertheless, there is at least one potential use case of using multiple fonts in music player applications. In the case of "geographical" maps such as Islands of Music (Pampalk 2001), the different genre regions could be labeled using descriptive fonts without cramming too many closely-located fonts onto the screen.

6.3 Use and acceptability of visualizations in music player applications

The third research question (RQ3) of the thesis was formulated as *"How can the genre visualizations be used in the design of novel graphical user interfaces for music player applications, and which visualization techniques are acceptable to the end users?"* Due to the sudden end of the project, GUIs for the other metadata types were not implemented.

Three interactive hi-fi prototypes (avatars, potentiometers, and a virtual world) for a touch screen PC and two simpler visualizations (icons and emoticons) for Symbian were developed. The avatar and potentiometer prototypes were mostly intended at passive music discovery, and they could be seen as means to generate smart playlists, personalized radio channels, or as a means to select seed songs for a music recommendation system without specifying explicit artist or song names. The virtual world prototype was more explorative in nature, but still the user was not able to specify explicitly what he/she wanted to hear. The icons were used to visualize the genres of individual songs in playlists (Section 5.4.2), and the emoticons were used as a means to fine-tune the contents of the recommendations. Unfortunately, the project was stopped before the icon and emoticon designs were evaluated by any users.

The hi-fi prototypes were studied separately in **P6**, **P7**, and **P8**. Based on the results of the publications, it is possible to design acceptable, entertaining, and easy-to-use genre visualizations that are also recognized and recalled well. However, all the prototypes were too simple for longer-term use; they would have benefited greatly from the addition of more features such as textual search functionality. Instead of replacing the traditional text-based music player applications, the prototypes were more suitable for complementing them. In the following, the results of each publication are discussed in more detail.

Avatars

Publication **P6** described an avatar-based application for discovering new music. To generate a new playlist, the users had to design a new avatar, randomize the avatar, or open a previously saved avatar. When designing a new avatar, the user had to select three avatar parts (head, body and background) reflecting his/her musical taste. There were five variations ("Metal", "Rock", "Electronica & Dance", "Hippie" and "Soul & Disco") for each part, and the designs were mostly the same as in **P5**. Based on the selected combination of parts, the application generated a new playlist of music recommendations. Each avatar part affected one third of the playlist.

Based on the results of the conducted qualitative and quantitative user study, it was concluded that the concept has great potential as a *complementary* UI for traditional text-based music applications. In the initial interview, the idea of listening to music through avatars was considered to be novel, fun, good, inspiring, innovative, and surprising. The application was considered to be easy to use and to provide a faster access to good music and cross-genre playlists than traditional music player applications. The concept was more appealing to the younger participants. However, in longer-term use (in this study, up to three weeks) the application lacked some important functionality, such as the possibility to do textual searches and play music from a certain artist. Other factors decreasing the longer-term appeal included the limited size of the music collection, the small number of genres, minor usability problems, and technical difficulties.

The results indicated that regardless of how entertaining or easy-to-use a music visualization might be, it cannot meet all the needs of a modern music consumer without the support for textual searches. Visualizations can be very suitable for casual listeners who prefer to listen to radio instead of buying albums, but they can also be too restricted for active music consumers who need more control over the music selection. In the case of the avatar prototype, the lack of search functionality was seen as a major drawback by 73% of the participants; only 55% felt that the prototype fit their current music consumption habits. While 65% were interested in using such as application to complement their other music player applications, only 10% thought that they could use the prototype as their main music player.

Most participants felt that the looks of the avatars matched well with the related music recommendations, and the results were well in line with the other avatar publication (**P5**). However, as discussed in Section 7.4, there is a risk that some people from other nationalities may consider the stereotypical designs as racist, thus making the concept less suitable for certain countries or regions.

Potentiometers

Publication **P8** presented the idea of using tempo and energy level potentiometers to fine-tune the contents of recommended playlists. The starting screen of the prototype was very minimalistic, basically consisting of only two potentiometers, the basic play controls, and an arrow button for changing the preferred musical genre. By turning the Tempo potentiometer clockwise, faster songs were selected to the playlist, and the Gain potentiometer had an effect on the energy level and aggressiveness of the songs. The look of the interface (the basic skin and icons used for selecting the genre) reflected the currently-selected genre; the designs were partially based on the findings of **P3**, **P4**, and (Holm & Holm 2008).

Based on the results of the conducted user study, it was concluded that the concept is suitable for discovering new music and for complementing traditional music player applications. The prototype was seen to be innovative, handy, easy to use and entertaining, and the graphical designs matched well with the musical genres. While there were some differences in the details (e.g., median ratings), the main findings were well in line with publication **P6**. In longer-term use, the prototype was too simple; it lacked some important functionality, such as textual search (a major drawback for 63% of the participants), and the music collection was too limited. There were also some technical problems and minor usability issues, which had an effect on the lasting appeal.

Virtual world

Publication **P7** studied the concept of using virtual worlds as an interface to access music collections. For example, different looking characters could recommend different types of music from the service, speeding a car could fast-forward a song, shooting a character could delete a song, and each bar or discotheque could represent a different musical genre. The world could also be dynamic; old buildings (representing rarely listened albums or genres) could be torn down, new bars opened, and so forth. To study the concept in practice, a simplified off-line version of such an application was developed and evaluated in a quantitative and qualitative user study. In the prototype, the user moves around a city and interacts with various characters, the looks of which reflect the musical style that they are recommending.

On the positive side, the concept was considered to be novel (median 6.0 on a seven-point Likert scale) and easy to use (median 5.0). 44% of the participants also considered the prototype to be more fun than the traditional music player applications. The prototype was more appealing to the younger participants, and they were also more interested in using it to complement other music player applications.

However, the results also suggested that it is not worth the time and money to start implementing game or virtual world-like features on top of an existing music application. Firstly, the concept did not fit well with the music consumption habits of the participants. Most people preferred to concentrate only on the music or do something else (e.g., read a book) while the music plays in the background. Finding new music should be as fast and easy as possible, and thus the prototype was too tedious to use. Secondly, providing only music recommendations in a virtual world environment does not provide enough entertainment and value for the users. The application should contain more activity (e.g., fighting) and interactive elements, and there should be a clear goal to strive for.

Rather than building the virtual world from scratch on top of a music player application, a better solution would be to enhance an existing social networking service (e.g., www.habbohotel.com), game (e.g., www.grandtheftauto.com) or virtual world (e.g., www.secondlife.com) with music recommendation and searching capabilities. Such a solution would automatically fulfill the users' need for more action and interactivity, and it would also be closer to the original concept that we had in mind. As such development resources are rarely available in academia, the preferable approach would be to collaborate with an external company, such as a developer of one of the above applications. Testing the concept on an existing service would also minimize the financial risks of developing a completely new type of service. In theory, such a concept would have some great potential for music recommendation. Music is already a part of applications such as Second Life, and for many consumers music listening is a secondary activity that is performed in the background for example while playing a computer game or browsing the web (Roberts, Foehr, & Rideout 2005).

7. Discussion

As shown in Chapter 6, it is possible to design recognizable, acceptable, entertaining, and easy-to-use music visualizations for selected musical metadata with certain limitations. In this chapter, the reliability and validity of the results is considered (Section 7.1), and selected factors affecting the performance and acceptability of the visualizations are discussed in more detail. These include the preferred music discovery and product usage modes (Section 7.2), the selection of musical genres (Section 7.3), the potential problem of avatars and racism (Section 7.4), as well as the possibilities and problems of increasing the number of metadata dimensions in the visualization (Section 7.5).

7.1 Reliability and validity

The reliability of a user test (in this case, a survey or a user evaluation) refers to whether one would get the same result if the test was to be repeated. As there can be huge differences between individual users, reliability can be improved by increasing the number of representative users. Validity, on the other hand, is *"the question of whether the usability test in fact measures something of relevance to usability of real products in real use outside the laboratory"* (Nielsen 1993, p. 169). Typical problems include using the wrong users, giving them the wrong tasks, or not including time constraints and social influences. While reliability can be measured with statistical tests, a high level of validity requires an understanding of the test method and some common sense. One way to increase validity is data source triangulation (Lazar, Feng, & Hocheiser 2010, p. 295), i.e., the use of multiple data sources such as interviews and questionnaires.

When NRC's music visualization research was started, music recommendation systems and online music services were still relatively unknown to the general public. While we were forced to use only Nokia internal participants in the surveys, these participants also represented well the potential users (early adopters or technologically-oriented users) of such systems. The hi-fi prototypes were later tested with people from different backgrounds and age groups. The results were well in line with the findings of the online surveys, thus increasing the validity of the results.

In the case of all surveys, the majority of the participants were Finnish. While this could descrease the validity of the results, significant cultural differences in the results were only found in the case of colors (**P3**). On the other hand, the selection of the studied musical genres was also targeted at Western users; if developing a commercial application for a certain market, both the genre set and the visualizations should be localized (Section 7.3) and verified with new user studies. This is true especially in the case of avatars, which were considered to be racist by some conference paper reviewers (Section 7.4), thus potentially limiting the use of the concept in commercial applications.

Other important factors affecting the validity of the results include the amount of prior knowledge and the number of metadata dimensions. In the case of all conducted surveys, the participants knew which musical attribute they should be paying attention to, and this has likely affected their way of looking at the visualizations. If the questionnaire participants were simply given a set of graphical objects and asked what type of music they represented, the participants could also have mapped them to other musical attributes than studied in the questionnaire. On the other hand, increasing the number of metadata dimensions (especially without prior knowledge, instructions or clear labels) is likely to make the visualizations more subjective and more difficult to encode (Section 3.1). As it is very hard to predict how well the different combinations of visual variables and musical metadata would perform in practice, the results of this thesis may not be valid in other cases than 1D metadata (one type of metadata visualized at a time).

In the original publications included in the thesis, the results were given mainly in the form of association percentages. Publications **P1**, **P2** and **P4** contained only the association percentages, and **P9** included association percentages, boxplots and a balloon plot. In **P3**, the statistical significance was calculated for all color-genre associations using Monte Carlo contingency table simulation and visualized using an association plot of Pearson residuals. In **P5**, the significance of avatar-genre associations was visualized using heatmaps and dendrograms. In **P6**, **P7**, and **P8**, the answers were given using a seven-point Likert scale, and the answers were visualized using boxplots. While the conditions for using the Chi-Square test were not met, measures such as Spearman's correlation and Wilcoxon's signed ranks test with related statistical significance were calculated for several statements. In the case of **P7**, nine users also filled in the AttrakDiff questionnaire (http://www.attrakdiff.de), and the results of that questionnaire were well in line with the results derived from the interview and our own questionnaire.

Based on the above discussion, the results of the thesis are valid for the cultural context in which they were studied; further research is needed to validate them in other contexts. The results may not be valid in other cases than 1D metadata. The statistical significancy (reliability) of the results was also determined in several cases.

7.2 Product usage and music discovery modes

While modern music consumers are now accustomed to (and even rely on) searching vast music collections for certain songs, albums or artists, this was not self-evident in 2007 when the visualization work was started. Thus, we decided to implement the hi-fi prototypes as vertical prototypes without the search functionality, wanting to investigate if the visualizations would be usable as they were. However, the lack of search functionality was seen as a major drawback in long-term use (in this case, two to three weeks), and thus the prototypes did not fit well with many participants' listening habits. All the prototypes were more suitable for complementing than replacing the traditional music player applications; none of the prototypes was acceptable as the only music player. Similar results were also found in publications (Lehtiniemi & Holm 2011a), (Lehtiniemi & Holm 2011b), and (Lehtiniemi & Holm 2012).

These results are easy to understand in the light of previous research on emotions and and user experience. Desmet (2004) explains how novel (sudden and unexpected) product features elicit surprise emotions. These emotions are often "one-time-only" emotions, which disappear when the user becomes familiar with the product. As all evaluated prototypes used a novel approach for recommending music, it is likely that many user study participants experienced this type of surprise emotions in the interview and thus felt more positive about the prototypes. When the novelty value of the prototypes disappeared at home, the participants became more concerned with the practical limitations of the prototypes. As stated in Karapanos et al. (2009), *"overall, while early experiences seemed to relate mostly to hedonic aspects of product use, prolonged experiences became increasingly more tied to aspects reflecting how the product becomes meaningful in one's life."*

Hassenzahl (2003) defines two different usage modes of a product. In the *goal mode*, reaching the current goal is important, and determines all the actions that the user performs. The user wants to be effective and efficient, and low arousal is preferred. In the *action mode*, the user is more playful and more spontaneous, and goals are determined "on the fly." If there is not enough stimulation, arousal starts to decrease and results in boredom. While traditional music player applications such as iTunes and Spotify are clearly more goal-oriented, NRC's virtual world prototype was targeted at the action mode. However, as shown in **P7**, it failed also in this respect. The potentiometer and avatar prototypes had characteristics of both goal and action modes. They supported well the goal of quickly playing a certain style of music, but did not support at all another important goal of the modern music consumer: specifying exactly which particular artist, album, or song to play.

In a way, the goal and action usage modes of a product can be seen as the two sides of a "perfect" music player UI. To be successful as the only music player application, the application should support both modes; should either mode fail, the user may need to use two different applications to satisfy his/her music consumption needs. Depending on the use context and his/her current mood, the user may want to, e.g., find a certain track/album/artist as quickly as possible (goal mode), start quickly the playback of music from a certain genre (goal mode), or spend some time exploring the music collection to find something new (action mode).

In the context of music, Hassenzahl's goal mode shares characteristics with Lillie's (2008) active (searching) and passive (music recommendation) modes of music discovery, while the action mode resembles more closely the exploratory (browsing) mode (Section 2.2). Especially in the case of large online music collections, the "perfect" application should support all three modes of discovery. One good example of such a system was the latest (unpublished) version of SuperMusic, which combined a state-of-the-art music recommendation system with icon and emoticon visualizations and different ways to search and access the music collection.

In addition to supporting certain goal mode tasks (e.g., the quick generation of smart playlists or personalized radio channels), visualizations can be an excellent choice for the explorative action mode. The users may spend a lot of time expanding their musical knowledge without a specific goal in mind (Weigl & Gustavino 2011), and it is thus natural to support this playful and spontaneous operation in the UI as well. A well-designed visualization can make the exploration of music collections a playful, entertaining and intuitive experience, and help the users greatly in finding certain type of songs from the collection. If designed properly, such visualizations can minimize the need to remember things and the need for explicit user input, increase the amount of useful information on the screen, and improve the usability and efficiency of the system (Hoashi et al. 2009). By relying on clear associations between visual and musical attributes, the user's cognitive load can be reduced, and users do not have to be musical experts to access the music collections.

7.3 Selecting the musical genres

As shown in Section 6.2.3, the performance of visualizations can vary significantly between the different genres. In this thesis, the used set of genres was selected based on Western ways to categorize music, and the genres were mainly targeted at the North American and European markets. Selecting the "perfect" set of genres was not in the scope of the thesis.

As a result of cultural globalization, many of the used genres have also gained international popularity. Still, all musical genres are not as popular in all parts of the world, and non-Western users may also feel that some important musical styles are missing from the list. For example, some of our Indian user study participants complained that the predefined genre list did not include Bollywood film music, which is a popular musical style in India. As another example, one of the most popular genres in Nairobi (Kenya) in 2008 was Afro-fusion (Impiö 2008). Based on genre definitions used in this thesis (see Appendix B), both Afro-fusion and Bollywood music would be placed under the "world music" umbrella term. Thus, in the case of a commercial visualization-based music player application, the set of genres should be localized, i.e., vary slightly depending on the target country, region, or continent.

In addition to localization, another way to modify the genre set is to reduce the total number of genres. For example, the least popular genres, such as gospel and New Age (at least among the Finnish participants, see **P3**), could be removed completely from the list or moved under some general category. Some current music services such as AllMusic also use only a few main genres and have more emphasis on the sub-genres.

As discussed in Section 4.3, building a consistent genre hierarchy can be extremely difficult, and the borders of genre definitions are not always clear. This also makes it more difficult to remove certain genres from the used genre set. For example, the relationship between country, folk, and world music is a complicated one. By one definition, they all could be placed under the "world music" umbrella term. By another definition, world music refers to non-Western music outside the mainstream (and thus, includes also non-Western folk), but the term folk music is often used to refer to only Western folk. Country music, on the other hand, is also original American folk music. Another problematic genre is pop, which is basically an umbrella category that may refer to any style that is popular at the moment.

The genres could also be combined. In current music services, common modifications include, e.g., combining rock with pop, rock with alternative & indie, and world music with folk. As many of these genres are also difficult to visualize (Section 6.2.3), removing them from the set could increase the average recognition rates of the presented visualization methods.

By removing and combining genres based on the above discussion, one promising set of genres for future visualization research would be: 1. blues 2. classical, 3. country, 4. electronica & dance, 5. hip-hop & rap, 6. jazz, 7. Latin, 8. metal, 9. reggae, 10. rock, 11. soul, RnB & funk, and 12. world music. Depending on the used visualization method, blues and jazz could also be combined. A comparison of this smaller set of genres and the genres used by AllMusic (www.allmusic.com) shows that the two lists are almost identical. The main genres used by AllMusic in 2011 were 1. blues, 2. classical, 3. country, 4. electronic, 5. rap, 6. jazz, 7. Latin, 8. reggae, 9. pop/rock, 10. R&B,

and 11. international. In addition to some naming conventions, the main difference between the two lists is the metal genre, which is missing from AllMusic. This similarity of the two genre sets is in fact a very important finding. AllMusic is a respected service that has been used as a reference in numerous academic publications, and the genres have been selected based on expert opinions. Now, the results of this thesis indicate that the genre set also has great potential in avatar and icon based music visualizations.

7.4 Avatars and racism

To improve the performance of the avatar designs (and make them potentially more entertaining), **P5** and **P6** neglected the most common guidelines for cross-cultural design and internalization (Horton 1994; Aykin 2005) and included some exaggerated and stereotypical designs. The design choice was intentional, and it was done at the risk of not pleasing all the users. However, the avatar designs performed well in both publications, and with the exception of one comment on the alternative & indie torso, we did not receive complaints from the survey participants. Thus, using exaggerated and stereotypical designs seemed to be a good design decision.

Still, several anonymous reviewers from three different conferences expressed their concern that such avatars may be considered to be racist. For example, one reviewer commented that "this will be offensive to some", "some of the avatars (esp. the Latin one) is downright racist", and "I would strongly recommend to try to stay away from designs that are offensive and too stereotypical." Another reviewer emphasized the sensitivity of the topic and commented that "I personally was not offended by any of the attribution, however I could see how certain populations might be." Two authors were also concerned about the ethical approval at the university the authors were in. Interestingly, while the use of skin colors was mentioned, there were not any negative comments on the use of religious symbols, flags, the peace mark, or the hints to drug abuse and violence.

According to the Oxford English Dictionary (2008), racism is *"the belief that all members of each race possess characteristics, abilities, or qualities specific to that race, especially so as to distinguish it as inferior or superior to another race or races. Hence: prejudice and antagonism towards people of other races, esp. those felt to be a threat to one's cultural or racial integrity or economic well-being; the expression of such prejudice in words or actions. Also occas. in extended use, with reference to people of other nationalities."* On the other hand, Rattansi (2007, p. 86) declines the existence of races entirely. As the concept of races does not have reliable scientific foundation, any doctrine that accepts the existence of races can itself be considered as racist. According to Rattansi (p. 131), racism *"involves specific beliefs about the existence of races and the possibility of their being classified hierarchically as superior and inferior on a number of physiological and cultural criteria, amongst other things."* The different forms of racism vary from biological (aka classical, old or conventional) to cultural (aka new), institutional, and religious racism.

While stereotypes are one potential form of cultural racism, it is often difficult to distinguish ethnicity and ethnocentrism as well as prejudices and stereotyping from race and racism (Rattansi 2007, pp. 162-163). As stated in Rattansi (p. 105), *"...generalizations, stereotypes, and other forms* *of cultural essentialism rest and draw upon a wider reservoir of concepts that are in circulation in popular and public culture. Thus, the racist elements of any particular proposition can only be judged by understanding the general context of public and private discourses in which ethnicity, national identifications, and race coexist in blurred and overlapping forms without clear demarcations."* One feature of human beings that is commonly associated with negative stereotypes, prejudice, and racism is skin color. To avoid the problem, HCI references such as (Horton 1994) propose using colorless (white) characters in the designs.

Regardless of which definition one relies on, the music avatars presented in this thesis do not fill the criteria for racism. Except for the skin color, the characters share similar facial characteristics, none of the characters is superior or inferior to the other characters, and thus the concept of races is not emphasized. Instead of negative stereotypes that are more often associated with racism, the designs rely mostly on neutral or historical stereotypes. The use of different skin colors is based on historical facts; as shown in Appendix B, the roots of many musical genres are in the Afro-American culture of the United States, and thus there was a clear reason to use especially brown skin colors in the avatar designs. The use of many other design elements (e.g., the saxophone, cannabis and the countryside view) can also be justified based on historical facts. The only (unfortunate and unintentional) important exception to this is the Latin avatar. As the design was not based on the origins of the musical genre, the character can be considered to be a poor and negative stereotype of South American men. As discussed in **P5**, a better alternative to represent Latin music would have been a female Samba dancer.

In an additional and unpublished part of the **P5** survey, we also investigated which skin colors the participants associated with different musical genres. Due to the sensitivity of the topic, the warning text "Please keep in mind that our intention is not to insult anybody - We just want to find the best, most stereotypical avatars to represent different musical genres" was added to the starting page of the online questionnaire. While two participants (both British males) refused to answer the question, the rest of the participants shared our views on the skin color selection. In most cases, the winning skin color matched our design, and in the case of no match the participants typically voted for a different shade of brown than used in the design.

Despite the above discussion, some people may still be offended by the stereotypical avatars, and thus there is a risk in using the avatar design in a commercial music player application. To reduce the problem, one could rely on colorless (white) and less stereotypical avatars as done in (Haro et al. 2010, see Section 4.3.5). However, this is in contradiction with the history of many musical genres (Appendix B), and it would potentially decrease the recognition rates of the avatars and make the application less fun. As another potential solution, the application could contain a large collection of avatars and automatically select a suitable (potentially less offending) character for each genre depending on the country that the user is living in (cf. Finland and the United States). The application could also let the users customize the avatars in the same style as in many modern video games and virtual world applications, thus leaving the choice of skin color and other characteristics to the users themselves.

7.5 Increasing the number of metadata dimensions

All the conducted surveys and prototypes studied representations for 1D musical metadata. In the case of avatar and virtual world prototypes, the only used metadata type was the musical genre. The potentiometer prototype relied on three types of metadata (tempo, energy level and genre), but only the genre metadata was visualized using skins and icons. Thus, a natural continuation to the visualization research would be to study representations and prototypes for 2D, 3D, and multidimensional musical metadata. However, as shown in Section 3.1, increasing the number of dimensions is not as trivial as it may seem to be.

While increasing the number of data dimensions increases the amount of information that can be embedded in the visualization, it can also make the visualization more difficult to encode. It is hard to predict how well the different combinations of visual variables and musical metadata would perform in practice; in fact this should be one of the next steps in future research. Depending on the selected visual variables and musical metadata, the associations may also be highly subjective. One potential way to deal with this is to use personalized mappings or let the users to modify the default mapping.

In the following, two examples of increasing the number of data dimensions in music visualizations are discussed. The first example (glyphs) is related to visualizing the contents of generated playlists, while the latter is related to generating new smart playlists with the help of visualizations.

7.5.1 Visualizing the contents of playlists

In the latest (unpublished) version of SuperMusic, small but descriptive genre icons were used to visualize the musical style of each item in the list of new music recommendations (Section 5.4.2). To increase the amount of data on the screen, each icon could also be replaced with a small glyph (Section 3.1)

In one interesting visualization concept, the genre, release decade and tempo region (e.g., slow, medium or fast) of a music track are mapped to the color, texture or degree of blur, and the number of edges (shape) of an abstract glyph. Based on Section 3.1, color and shape are highly separable dimensions, and they can also be processed pre-attentively. As size is a dissociative variable (Siirtola 2007, p. 17), the glyph should not be too "too small"; otherwise it is impossible to see its exact color.

Unfortunately, the idea may not be as straightforward and feasible as it seems to be. In (Chen and Kluber, 2010), the authors studied the automatic generation of thumbnails for music. One of the four concepts, the "Music Icon", mapped musical tempo and aggressiveness to the number and shape of petals around a circle. While the Music Icon was generally preferred as the best of the four concepts, test subjects found them all to be too complex and did not appreciate any of them much. One potential way to improve the understandability of glyphs is to rely on personalized mappings. In Kolhoff et al. (2006), the users selected their own set of "training icons" to represent prototypical songs, and then created their own personalized mappings between the music and the icons. According to the authors, the users seemed to learn to "read" the icons in terms of music.

7.5.2 Generating playlists using visualizations

Another potential use case for visualizations with more than one data dimension is the quick creation of smart playlists or personalized radio channels. In one potential UI concept, musical genres are represented with colored objects (e.g., circles, rectangles or patterns created using generative grammars [Redström, Skog, & Hallnäs 2000]) that are placed on the screen in a beautiful and artistic manner. To maximize the amount of useful data on the screen, only those genres that the user likes to listen to (typically around six [**P5**]) are shown on the screen. Optionally, the size of each object can be mapped to the popularity or number of songs from the corresponding genre (left side of Figure 14). When the user clicks on a part of an object, matching songs from the corresponding genre could start to play automatically. Double-clicking an object could open another view for fine-tuning the contents of the playlist or a traditional text-based view with artists from the selected genre.

To increase the number of data dimensions, the tempo and release year could be represented with other visual variables, such as location on the *x*- or the *y*-axis, degree of blur, texture, and/or lightness. To improve the performance of the visualizations, one could also use redundant coding. For example, tempo could be mapped to the size of the objects in such a way that faster songs are represented with smaller objects located on the top (Figure 14, second from the left), and release year could be mapped to lightness in such way that newer songs are located on the right and lighter side of each object (Figure 14, second from the right). Different combinations such as the one shown on the right side of Figure 14 would also be possible.

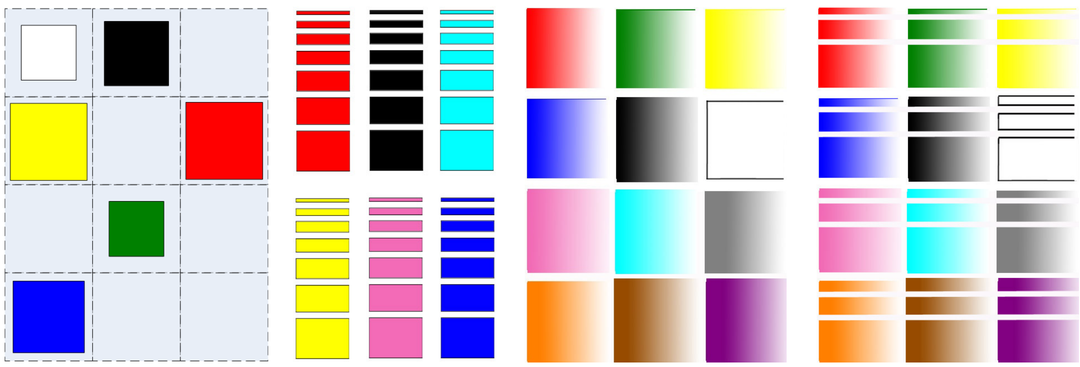

Figure 14 Illustrative example of visualizing music collections with colored objects (see text for explanation).

To combine songs from multiple genres, tempi, or decades, the user could do multiple clicks in a row or "draw" a path from one object to another. In the latter case, a rising line could create a playlist with increasing tempo, a left-to-right horizontal line could create a playlist with increasing release year, and so forth. One good example of this type of approach can be found from the artist map application (Van Gulik & Vignoli 2005) shown in Figure 15. However, as the artist map visualizes individual songs instead of subsets of the collection, the UI is better suitable for large screens (e.g., desktop or laptop computers instead of mobile devices) and relatively small music collections.

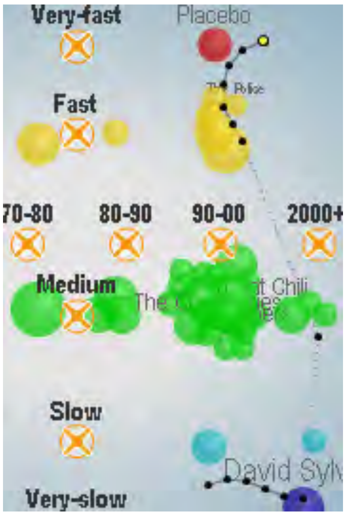

Figure 15 Example of generating playlists by drawing paths (Van Gulik & Vignoli 2005). In this example, tempo is represented with color and location on the *y*-axis and release year with location on the *x*-axis. The dotted line shows a playlist that includes some very fast songs from the 90s, medium tempo songs from 2000+, and slow songs from 2000+ and 90-00.

One should note that the examples given above are just illustrative ideas that are used here to demonstrate the different possibilities of increasing the number of metadata dimensions. As the optimal combination of musical and visual variables depends on various factors, including target device, context of use and so forth, the real application ideas should first be user-tested using questionnaires, interviews, lo-fi/hi-fi prototypes, or other research methods.

8. Conclusions and future work

The purpose of the thesis was to explore the feasibility of visualizing music collections based on three types of musical metadata: musical genre, tempo, and release year. More specifically, the research goal was to study which visual variables and structures are best suited to represent the metadata, and how the visualizations can be used in the design of novel user interfaces for music player applications such as music recommendation systems. The thesis took a user-centered and constructive design-science approach to the research of visualizing music collections, and it involved all the different aspects of interaction design: understanding the users, prototype design, and evaluation. The topic was approached through three research questions, which were addressed in a series of user studies.

8.1 Conclusions for the academic community

The first contribution of the thesis comprises a set of suggested visual variables and structures for visualizing tempo, release year, and genre metadata. The encodings were studied in a series of surveys with 51-104 (mostly Finnish) participants, and selected genre designs were further user-tested with three hi-fi prototypes. All the proposed visualization methods can be implemented based on MP3 ID3v2 or other musical metadata, and thus they do not require the development of new MIR signal processing algorithms. To the best of our knowledge, such systematic user studies on visualizing musical metadata have not yet been conducted elsewhere.

To summarize, promising ways to represent tempo include the number of objects, shapes with a varying number of corners, and *y*-axis location combined with some other visual variable (e.g., the size of the objects) or clear labeling. These findings are well in line with most previous work from the other authors, but according to our knowledge scientific studies on the encodings have not been reported elsewhere.

Promising ways to represent the release year include lightness and perceived location on the *z*-axis. These variables were not used in the studied previous research; instead, most applications mapped the year to the *x*-axis location, which is a common way to represent time.

In the case of musical genres, the most successful visualization method was the avatars, icons came second with a similar performance, with fonts coming third and the least successful method being colors. The best performance was achieved by combining several different visual elements (cf. avatars), and but this approach also consumed the most screen estate. With the exception of colors, the other methods have not been widely used in the MIR community. The easiest genre to visualize was metal; it performed best in both surveys and the user tests of the prototypes.

The use of emoticons to represent musical genres was investigated in a separate study, which resulted in collections of genres that could be used as a starting point for making recommendations fitting the current mood of the user. Compared to previous work from the other authors, a similar type of approach has only been taken in the MoodTunes Lite application (Whoop Apps 2009).

The academic community (and music service developers as well) can take these results as a basis for future research on music visualization. Especially the results of the tempo and release year surveys could be used as a starting point for developing visualizations with more than one data dimension, and the results of the genre surveys could form the basis for determining the “perfect” set of genres for visualizations for different cultures. The studied encodings between visual variables/structures and selected metadata could also be used in many academic (and commercial) applications by other authors. At the moment, the encoding used in the academic literature vary greatly, and they are usually not based on the results of scientific studies.

8.2 Conclusions for music service developers

The second contribution of the thesis includes a set of design implications for digital music service development. In the practical part of the research, findings from the surveys were used in the design of interactive hi-fi prototypes and simple visualizations for the latest version of Nokia Research Center’s SuperMusic system. The prototypes were user-tested with 40-41 participants. Based on the results, it is possible to design recognizable, acceptable, entertaining, and easy-to-use music visualizations for genre metadata with certain limitations. Regardless of how entertaining or easy-to-use the music visualization might be, it cannot meet all the needs of a modern music consumer without support for textual searches. Due to the lack of textual search and other limitations, all the evaluated prototypes were too simple for longer-term use (in this study, up to three weeks); they were suitable only for complementing rather than replacing the traditional text-based music player applications. Should a single “perfect” music player application be used, it should support several types of music discovery, including textual searches and (for example, visual) exploration.

Furthermore, the performance, suitability, and acceptability of visualizations depends on factors such the type of available musical metadata, the metadata vocabulary (e.g., the set of musical genres), used visual variables and/or structures, available screen estate (size of the visualization and type of used device), the user’s preferred music discovery type, the target culture of the visualizations, and so forth. For example, while the avatar concept was received well in Finland (especially by the younger consumers), our experience indicates that the concept of using stereotypes may be less acceptable in some other cultures.

Based on the results of the thesis, musical genre is a relatively safe choice for product implementation. Despite the on-going debate on the feasibility of using genres to categorize music, genre metadata is used in practically in every music collection and application, and it is also commonly used in everyday language. Our survey results for visualizing genre metadata (especially in the case of avatars and icons) were promising, and the average performance of the different methods is likely to improve after redesigning the worst designs. The encodings used in current commercial applications vary greatly, and companies could benefit greatly from adding more intuitive mappings to their UI. In this thesis, the selection of the studied musical genres was targeted at Western users; if developing a commercial application for other markets, both the genre set and the visualizations should be localized and verified with new user studies.

In the case of visualizing tempo and release year metadata, there were some promising survey results, but the ideas have not yet been tested in any practical applications. Examples of visualizing tempo and release year metadata can also be found from other authors, but their selection of visual variables differs somewhat from that of ours. Despite the interest towards using the tempo parameter (Holm 2008), many digital music collections still do not contain pre-calculated BPM information. Should it be missing, the BPM value can be measured, for example using automatic music tempo detection algorithms such as those represented in (Eronen 2009), or it can be retrieved from an existing database such as the Echo Nest. Unfortunately, the release year information is also sometimes missing from some songs in the music collections, making that metadata type potentially less feasible for real-world applications.

8.3 Future work

While the surveys presented in this thesis studied many other visual variables listed by authors such as Mackinlay (1986) and Marchese (2011), none of the surveys focused explicitly on textures. Still, Mackinlay (1986) ranks texture as the third best variable to convey nominal data and the fifth best to convey ordinal data. Thus, a logical continuum to the series of online surveys would be to study associating textures with selected musical metadata, such as genres. In the context of music visualization, different textures could be used in various UI elements, including buttons, icons, skins, potentiometers, and naturally the various elements of virtual worlds. Inspired by (Zuo et al. 2004), smooth surfaces or textures could potentially be mapped to "happy" genres, such as pop, and rough surfaces to more "aggressive" genres such as heavy metal or rock. Another potential approach would be to use natural materials such as metal and grass to visualize the different genres.

Based on the results of (Holm 2008), other interesting attributes for future surveys include friends, listening history, the popularities (play/skip count and ratings) of individual songs, albums, (genres) and artists, and collections of most played, not yet played and recently added songs, albums and artists. To emphasize some important items, such as new recommendations or most popular items, one could for example place them in front by varying the location on the *z*-axis or the degree of blur, or use halos or some irregular or distinctive shapes that call attention to themselves. As the amount of user-given and automatically generated metadata is on the increase, new potential metadata types for visualizing music collections are also likely to appear.

Combinations of different variables and/or structures should be studied, e.g., in the context of glyphs. Increasing the number of metadata dimensions (especially without prior knowledge, instructions or clear labels) is likely to make the visualizations more subjective and more difficult to encode. As it is very hard to predict how well the different combinations of visual variables and musical metadata would perform in practice, the current survey results may not be valid in other cases than 1D metadata.

During the different phases of the research, several new ideas on how to improve the different prototypes and individual visual structures (especially the avatar designs) were received. For example, the Latin avatar should be replaced with a female samba dancer, the blues avatar could be replaced with a Blues Brothers or a John Lee Hooker type of character holding an electric guitar and a bottleneck slide, and the dancers of the soul, RnB & funk background should be removed to avoid mixing the design with electronica & dance music. While some of the designs could be improved with minor changes, those with the worst performance should be redesigned from scratch. The improved designs should preferably be verified with new user tests.

Yet another interesting topic for future work would be to conduct international studies on stereotypical musical avatars and racism. While the current avatar designs could be acceptable for the Finnish market, this might not be the case in some other countries. The performance and acceptability of colorless avatars (as suggested by textbooks) and avatars with different skin colors could be studied with two separate but similar groups of test subjects, and the results should be compared against each other. To improve the performance of the genre visualizations, the set of musical genres should be localized for the target country or region and verified with new user tests.

In the future, it would be fruitful to test an updated version of one of the existing prototypes or some new visualization concept as a complementary view to a traditional list-based music player application such as Spotify. Testing the existing icon and emoticon designs in the context of a mobile music player application would also be beneficial. In the case of the virtual world concept, one could study the possibilities of collaborating with a commercial company such as Sulake (creators of Habbo Hotel) or Linden Lab (creators of Second Life) to test the concept in a real-world application.

During the research, several new ideas for future prototypes were also received. These include, e.g., coloring calendar views based on the most-listened genres, visualizing individual songs with glyphs, widgets for social networks to give an overview of one's music library, UIs for generating smart playlists or personalized radio channels, and FilmFinder (Ahlberg & Shneiderman 1994) type of scatterplot visualizations. Depending on the amount of available screen space and other such factors, the complexity of the prototypes can vary from simple Casual Infovis systems to more complex Infovis systems. One interesting idea is to map the media collection to a universe, where the different media items are represented as planets and stars. Continuing the idea of (Kirn 2011b), one could modify the surface (color and texture) of the planets based on the musical genre, the rotational speed could be mapped to the tempo metadata of songs, and some other visual variable such as location on the z-axis or lightness change (shades on top of the planets) could represent release year.

One fascinating way to display the visualizations is Informative Art. An abstract, dynamic, and interactive piece of art could be shown on a large, high-definition touch screen hanging on the wall of a home or a public space, and touching the screen would automatically start the playback of music mapped to that part of the screen. Another approach would be to let the user modify the visualization to start the playback of music. For example, the user could select a certain type of music by moving the related visual element(s) to the center of the screen, and a new playlist of recommendations would then be automatically generated. In the standby/screensaver mode, the screen could also be used to visualize, e.g., the Last.fm data of the user, his/her friends, or a selected group of users. Informative Art could also be context-aware; for example, the selection of visible genres could depend on the day (cf. reading on a quiet weekday and partying on Saturday night).

References

Ahlberg, C., & Shneiderman, B. (1994). Visual information seeking using the FilmFinder. In *CHI'94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 433-434). ACM Press.

Aldis N., & Sherry, J. (2006). *Heavy metal thunder: Album covers that rocked the world*. London: Mitchell Beazley.

Anderson C. (2006). *The long tail: Why the future of business is selling less of more*. Hyperion.

Aykin, N. (2005). *Usability and Internalization of Information Technology*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bainbridge, D., Cunningham, S. J., & Downie, J. S. (2003). How people describe their music information needs: A grounded theory analysis of music queries. In *ISMIR'03: Proceedings of the 4th International Conference on Music Information Retrieval* (pp. 221–222). ISMIR.

Barrow, S. (2001). *The rough guide to reggae*. London: Rough Guides.

Baumann, S. (2005). Visualization for music IR, Tutorial part 2. Tutorial in *ISMIR'05: The 6th International Conference on Music Information Retrieval*. ISMIR.

Bertin, J. (1981). *Graphics and graphic information-processing*. (W. J. Berg & P. Scott, Trans.). Berlin: Walter de Gruyter.

Birosik, P. J. (1989). *The New Age music guide*. Collier Books.

Boberg, M., Piippo, P., & Ollila, E. (2008). Designing avatars. In *DIMEA'08: Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts* (pp. 232-239). ACM Press.

Bogdanov, D., & Herrera, P. (2011). How much metadata do we need in music recommendation? A subjective evaluation using preference sets. In *ISMIR'11: Proceedings of the 12th International Society for Music Information Retrieval Conference* (pp. 97-102). ISMIR.

Bohlman, P. V. (2002). *World music: A very short introduction*. New York, NY: Oxford University Press Inc.

Book. B. (2004). Moving beyond the game: social virtual worlds. In *Proceedings of State of Play 2 Conference*.

Brinded, T., Mardell, J., Witkowski, M., & Spence, R. (2011). The effects of image speed and overlap on image recognition. In *IV'11: Proceedings of 15th International Conference on Information Visualisation* (pp. 3-11). IEEE Computer Society.

Broughton, S., Ellingham, M., Muddyman, D., & Trillo, R. (Eds.). (1994). *World music: The rough guide*. London: Rough Guides.

Bukszpan, D. (2010). *Heavy metal: raskaan musiikin pioneerit, jättiläiset ja kapinalliset*. Helsinki: Kustannusosakeyhtiö Nemo.

Burrows, J. (2005). *Klassinen musiikki*. Helsinki: Werner Söderström Oy.

Byrne, M. D. (1993) Using icons to find documents: Simplicity is critical. In *CHI'93: Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems* (pp. 446-453). ACM Press.

Cano, P., Koppenberg, M., & Wack, N. (2005). Content-based music audio recommendation. In *MM'05: Proceedings of the 13th Annual ACM International Conference on Multimedia* (pp. 211-212). ACM Press.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. San Francisco, CA: Morgan Kaufmann Publishers.

Cascone, K. (2002). The aesthetics of failure: "post-digital" tendencies in contemporary computer music. *Computer Music Journal 24(4)*, 12-18.

Casey, M., Veltcamp, R., Goto, M. Leman, M., Rhodes, C., & Slaney, M. (2008). Content-based music information retrieval: Current directions and future challenges. In Proceedings of the IEEE, 96(4), 668-696.

Celma, O. (2008). *Music recommendation and discovery in the long tail*. Doctoral dissertation, Universitat Pompeu Fabra, Barcelona.

Celma, O., & Lamere, P. (2007). Music recommendation tutorial. Tutorial *in ISMIR'07: The 8th International Conference on Music Information Retrieval*. ISMIR.

Chang, J. (2008). *Can't stop, won't stop: Hiphopsukupolven historia*. Helsinki: Like.

Chang, K. K., Jang, J-S R., & Iliopoulos, C. S. (2010). Music genre classification via compressive sampling. In *ISMIR'10: Proceedings of the 11th International Conference on Music Information Retrieval* (pp. 387–392). ISMIR.

Charlton, K. (1998). *Rock music styles: A history*. New York, NY: McGraw-Hill.

Charlton, K. (2008). *Rock music styles: A history* (5th ed.). New York, NY: McGraw-Hill.

Chen, Y-X., & Kluber, R. (2010). ThumbnailDJ: Visual thumbnails of music content. In ISMIR'10: *Proceedings of the 11th International Conference on Music Information Retrieval* (pp. 565–570). ISMIR.

Cohen, R. D. (2006). *Folk music: The basics*. Taylor & Francis.

Corthaut, N., Govaerts, S., Verbert, K., & Duval, E. (2008). Connecting the dots: Music metadata generation, schemas and applications. In *ISMIR'08: Proceedings of the 9th International Conference on Music Information Retrieval* (pp. 249–254). ISMIR.

Cunningham, S. J., & Zhang, Y. (2008). Development of a music organizer for children. In *ISMIR'08: Proceedings of the 9th International Conference on Music Information Retrieval* (pp. 185–190). ISMIR.

Cunningham, S. J., Reeves, N., & Britland, M. (2003). An ethnographic study of music information seeking: implications for the design of a music library. In *JCDL'03: Proceedings of the 2003 Joint Conference on Digital Libraries* (pp. 5-16). IEEE.

Damm, D., Fremerey, C., Kurth, F., Muller, M., & Clausen, M. (2008). Multimodal presentation and browsing of music. In *ICMI'08: Proceedings of the 10th International Conference on Multimodal Interfaces* (pp. 205-208). ACM Press.

Desmet, P. M. A. (2004). From disgust to desire: How products elicit emotions. In D. McDonagh, P. Hekkert, J. Van Erp & D. Gyi (Eds.), *Design and Emotion: The Experience of Everyday Things*. London: Taylor & Francis.

Dodig-Crnkovic, G. (2010). Constructive research and info-computational knowledge generation. In L. Magnani, W. Carnielli, & C. Pizzi (Eds.), *Model-Based Reasoning in Science and Technology*. Berlin: Springer.

Downie, J. S., Byrd, D., & Crawford, T. (2009). Ten years of ISMIR: reflections on challenges and opportunities. In *ISMIR'09: Proceedings of the 10th International Conference on Music Information Retrieval* (pp. 13–18). ISMIR.

Ekman, P. (2004). *Emotions revealed – Understanding faces and feelings*. London: Orion Books.

Emery, A. (2004). *The book of hip hop cover art*. London: Mitchell Beazley.

Eronen, A. (2009). *Signal processing methods for audio classification and music content analysis*. Doctoral dissertation. Tampere University of Technology, Tampere.

Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis: A survey. *Pattern Recognition, 36*, 259-275.

Fox, B. (2005). *Game interface design*. Boston, MA: Thomson Course Technology.

Frith, S. (2007). *Taking popular music seriously: Selected Essays*. Aldershot: Ashgate.

Gershon, N., & Eick. S. G. (1995). Foreword. In *InfoVis'95: Proceedings of the IEEE Symposium on Information Visualization* (pp. vii-viii). IEEE Computer Society Press.

Gooch, B. Reinhard, E., & Gooch, A. (2004). Human facial illustrations: Creation and psychophysical evaluation. *ACM Transactions on Graphics*, 23(1), 27-44.

Goto, M., & Goto, T. (2005). Musicream: New music playback interface for streaming, sticking, sorting, and recalling musical pieces. In *ISMIR'05: Proceedings of the 6th International Conference on Music Information Retrieval* (pp. 404-411). ISMIR.

Gould, J. D., & Lewis, C. H. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM* 28(3), 300-311.

Goussevskaia O., Kuhn M., & Lorenzi, M. (2008). From web to map: Exploring the world of music. In *WI-IAT'08: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (pp. 242-248). IEEE Computer Society.

Haro, M., Xambo, A., Fuhrmann, F., Bogdanov, D., Gomez, E., & Herrera, P. (2010). The musical avatar – a visualization of musical preferences by means of audio content description, In *AM'10: Proceedings of the Audio Mostly Conference*. ACM Press.

Hassenzahl, M. (2003). The thing and I: Understanding the relationship between user and product. In M. A. Blythe, K. Overbeeke, A. F. Monk & P. C. Wright (Eds.). *Funology: From Usability to Enjoyment*. Dordrecht: Kluwer Academic Publishers.

Hassenzahl, M., Platz, A., Burmester, M., & Lehner, K. (2000). Hedonic and ergonomic quality aspects determine a software's appeal. In *CHI'00: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 201-208). ACM Press.

Heittola, T. (2003). *Automatic classification of music signals*. MSc thesis, Tampere University of Technology.
Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28(1)*, 75-105.
Hoashi, K., Hamawaki, S., Ishizaki, H., Takishima, Y., & Katto, J. (2009). Usability evaluation of visualization interfaces for content-based music retrieval systems. In *ISMIR'09: Proceedings of the 10th International Society for Music Information Retrieval Conference* (pp. 207-212). ISMIR.
Holm, J. (2008). Popularity of music player parameters and features. In *ICMC'08: Proceedings of the International Computer Music Conference*. ICMA.
Holm, J., & Holm, H. (2008). Associating icons with musical genres. In *IMSA'08: Proceedings of the 12th IASTED International Conference on Internet and Multimedia Systems and Applications*. ACTA Press.
Holm, J., Aaltonen, A., & Seppänen, J. (2009). Associating fonts with musical genres. In *Afrigraph'09: Proceedings of the 6th International Conference on Virtual Reality, Computer Graphics, Visualisation and Interaction in Africa* (pp. 33-36). ACM Press.
Holmes, T. B. (1985). *Electronic and experimental music*. Charles Schribner's Sons.
Horton, W. (1994). *The icon book: Visual symbols for computer systems and documentation*. New York, NY: John Wiley & Sons.
Hu, X., & Downie, S. (2007). Exploring mood metadata: Relationships with genre, artist, and usage metadata. In *ISMIR'07: Proceedings of the 8th International Conference on Music Information Retrieval* (pp. 67-72). ISMIR.
Hu, X., & Liu, J. (2010). Evaluation of music information retrieval: Towards a user-centered approach. In *HCIR'10: Proceedings of the 4th Workshop on Human-Computer Interaction and Information Retrieval*. ACM Press.
Hu, X., Downie, J. S., & Ehmann, A. F. (2009). Lyric text mining in music mood classification. In *ISMIR'09: Proceedings of the 10th International Conference on Music Information Retrieval* (pp. 411–416). ISMIR.
Impiö, J. (2008). *A study of the informal music industry in Africa*. Nairobi: Nokia Research Center.
ISO (International Organization for Standardization) (2009). *ISO FDIS 9241-210:2009. Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems*. International Organization for Standardization.
Isaacson, E. (2005). What you see is what you get: on visualizing music. In *ISMIR'05: Proceedings of the 6th International Conference on Music Information Retrieval* (pp. 389-395). ISMIR.
Jenkins, A. M. (1985). Research methodologies and MIS research. In E. Mumford, R. Hirscheim, G. Fitzgerald & A. T. Wood-Harper (Eds.). *Research Methods in Information Systems*. Amsterdam: North-Holland.
Jenson, S. (2002). *The simplicity shift: Innovative design tactics in corporate world*. Cambridge: Cambridge University Press.
Jones, M., & Marsden, G. (2006). *Mobile interaction design*. Chichester: John Wiley & Sons.
Julià, C. F., & Jordà, S. (2009). Songexplorer: A tabletop application for exploring large collections of songs. ISMIR 2009. In *ISMIR'09: Proceedings of the 10th International Society for Music Information Retrieval Conference* (pp. 675–680). ISMIR.
Juslin, P. N., & Sloboda, J. A. (2001). *Music and emotion: Theory and research*. New York, NY: Oxford University Press.
Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J-B. (2009). User experience over time: An initial framework. In *CHI '09: Proceedings of the 27th International Conference on Human Factors in Computing Systems* (pp. 729-738). ACM Press.
Kemp, C. (2004). *Towards a holistic interpretation of musical genre classification*. Doctoral dissertation. University of Jyväskylä, Jyväskylä.
Kim, H-J., Kwon, J-Y., Yoo, M-J., & Lee, I-K. (2009). Generating affective music icons in the emotion plane. In *CHI '09: Proceedings of the 27th International Conference on Human Factors in Computing Systems* (pp. 3389-3394). ACM Press.
Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., Speck, J. A., & Turnbull, D. (2010). Music emotion recognition: A state of the art review. In *ISMIR'10: Proceedings of the 11th International Society for Music Information Retrieval Conference* (pp. 255–266. ISMIR.
Kim, Y., Schmidt, E., & Emelle, L. (2008). Moodswings: A collaborative game for music mood label collection. In *ISMIR'08: Proceedings of the 9th International Conference on Music Information Retrieval* (pp. 549–554). ISMIR.

Kingsbury, P. (2003). *Vinyl hayride: Country music album covers 1947-1989*. San Francisco, CA: Chronicle Books.

Kolhoff, P., Preuss, J., & Loviscach, J. (2006). Music icons: Procedural glyphs for audio files. In *SIBGRAPI'06: Proceedings of the XIX Brazilian Symposium on Computer Graphics and Image Processing* (pp. 289-296). IEEE Computer Society.

Koskinen, I., Zimmerman, J., Binder, T., Redström, J., & Wensveen, S. (2011). *Design research through practice from the lab, field, and showroom*. Morgan Kaufmann.

Kosmos, K. (2002). *Natty Congo Bongo: Juurijuttuja Jamaikalta*. Helsinki: Like.

Kot, G. (2009). *Ripped: How the wired generation revolutionized music*. New York, NY: Scribner.

Kusama, K. & Itoh, T. (2011). MusCat: A music browser featuring abstract pictures and zooming user interface, In *SAC'11: Proceedings of the 2011 ACM Symposium on Applied Computing* (pp. 1222-1228). ACM Press.

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

Lamere, P., & Eck, D. (2007). Using 3D visualizations to explore and discover music. In *ISMIR'07: Proceedings of the 8th International Conference on Music Information Retrieval* (pp. 173-174). ISMIR.

Laplante, A. (2010). User's relevance criteria in music retrieval in everyday life: An exploratory study. In *ISMIR'10: Proceedings of the 11th International Conference on Music Information Retrieval* (pp. 601–606). ISMIR.

Lau A., & Vande Moere, A. (2007). Towards a model of information aesthetic visualization. In *IV'07: Proceedings of the 11th International Conference on Information Visualisation* (pp. 87-92). IEEE Computer Society.

Law, E. L. M., von Ahn, L., Dannenberg, R. B., & Crawford, M. (2007). TagATune: A game for music and sound annotation. In *ISMIR'07: Proceedings of the 8th International Conference on Music Information Retrieval* (pp. 361-364). ISMIR.

Lazar, J., Feng, J. H., & Hocheiser, H. (2010). *Research methods in Human-Computer Interaction*. Chichester: John Wiley & Sons.

Lee, J. H., & Downie, J. S. (2004). Survey of music information needs, uses, and seeking behaviours: Preliminary findings. In *ISMIR'04: Proceedings of the 5th International Conference on Music Information Retrieval* (pp. 441-446). ISMIR.

Lehikoinen, J., Aaltonen, A., Huuskonen, P., & Salminen, I. (2007). *Personal content experience: Managing digital life in the mobile age*. Chichester: John Wiley & Sons.

Lehtiniemi, A. (2008). Evaluating SuperMusic: Streaming context-aware mobile music service. In *ACE'08: Proceedings of the International Conference on Advances in Computer Entertainment Technology* (pp. 314-321). ACM Press.

Lehtiniemi, A., & Holm, J. (2011a). Evaluating a cube-based user interface for exploring music collections. In *IV'11: Proceedings of the 15th International Conference on Information Visualisation* (pp. 40-46). IEEE Computer Society.

Lehtiniemi, A., & Holm, J. (2011b). Easy access to recommendation playlists: Selecting music by exploring preview clips in album cover space. In *MUM'11: Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia* (pp. 94-99). ACM Press.

Lehtiniemi, A., & Holm, J. (2012). Using animated mood pictures in music recommendation. In IV'12: *Proceedings of the 16th International Conference on Information Visualisation* (pp. 143-150). IEEE Computer Society.

Leitich, S., & Topf, M. (2007). Globe of Music – music library visualization using GeoSOM. In *ISMIR'07: Proceedings of the 8th International Conference on Music Information Retrieval*. ISMIR.

Liikkanen, L. (2010). *Are digital sales killing the Finnish recording industry? Analysis of Finnish music consumption trends in 2009*. Electronic publication by Helsinki Institute for Information Technology HIIT, Aalto University, Finland.

Lillie, A. S. (2008). *MusicBox: Navigating the space of your music*. BSc thesis, Stanford University.

Ljungblad, S., Skog, T., & Holmquist, L. E. (2003). From usable to enjoyable information displays. In M. A. Blythe, K. Overbeeke, A. F. Monk & P. C. Wright (Eds.). *Funology: From Usability to Enjoyment*. Dordrecht: Kluwer Academic Publishers.

Lukashevisch, H., Abesser, J., Dittmar, C., & Grossmann, H. (2009). From multi-labeling to multi-domain-labeling: a novel two-dimensional approach to music genre classification. In *ISMIR'09: Proceedings of the 10th International Society for Music Information Retrieval Conference* (pp. 459–464). ISMIR.

Luukkonen, H. (2006). *Parantava sateenkaari: väriterapiaopas*. Delfiini kirjat.

Machida, W. & Itoh, T. (2011). Lyricon: A visual music selection interface featuring multiple icons. In *IV'11: Proceedings of the 15th International Conference on Information Visualisation* (pp. 145-150). IEEE Computer Society.
Mackinlay, J. D. (9186). Automating the design of graphical presentations of relational information. *ACM Transaction on Graphics, 5(2)*, 110-141.
Magno, T., & Sable, C. (2008). A comparison of signal-based music recommendation to genre labels, collaborative filtering, musicological analysis, human recommendation, and random baseline. In *ISMIR'08: Proceedings of the 9th International Conference on Music Information Retrieval* (pp. 161-166). ISMIR.
Malone, B. (2002). *Country music, U.S.A.* (2nd ed.). University of Texas Press.
Mao, J-Y., Vredenburg, K., Smith, P. W., & Carey, T. (2005). The state of user-centered design practice. *Communications of the ACM, 48(3)*, 105-109.
Marchese, M. T. (2011). Art 4 visualizers: Tutorial. Tutorial in *IV'11: The 15th International Conference on Information Visualisation*. IEEE Computer Society.
Marsh, G. (1996). *The blues: Album cover art*. San Francisco, CA: Chronicle Books.
Marsh, G. (2002). *Blue Note: The ultimate collection*. San Francisco, CA: Chronicle Books.
McCloud, S. (1993). *Understanding comics: The invisible art*. Kitchen Sink Press.
McKay, C., & Fujinaga, I. (2006). Musical genre classification: Is it worth pursuing and how can it be improved? In *ISMIR'06: Proceedings of the 7th International Conference on Music Information Retrieval* (pp. 101-106). ISMIR.
McKean, E. (Ed.). (2005). *The new Oxford American dictionary* (2nd ed.). New York, NY: Oxford University Press.
MIDI Manufacturers Association (1996). MIDI 1.0 detailed specification, document version 4.2. Revised February 1996. In *The complete MIDI 1.0 detailed specification, document version 96.1*. Los Angeles, CA: MMA Inc.
Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review, 63*, 81-97.
Morrow, C. (1999). *Stir it up! Reggae album cover art*. London: Thames & Hudson.
Morton, J. (1997). *A guide to color symbolism*. PDF document ISBN 0-9679080-0-0. COLORCOM.
Moynihan, M., & Soderlind, D. (2003). *Lords of chaos: The bloody rise of the satanic metal underground* (2nd ed.). Los Angeles, CA: Feral House.
Neumayer, R., Lidy, T., & Rauber, A. (2005). Content-based organization of digital audio collections. In *Proceedings of the 5th Open Workshop of MUSICNETWORK*.
Nguyen, D. Q., Tominski, C., & Schumann, H. (2011). Visualizing tags with spatiotemporal references. In *IV'11: Proceedings of the 15th International Conference on Information Visualisation* (pp. 32-39). IEEE Computer Society.
Nielsen, J. (1993). *Usability engineering*. London: Academic Press.
Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things*. New York, NY: Basic Books.
Oda, M., & Itoh, T. (2007). MIST: A music selection icon selection technique using neural network. In *Proceedings of NICOGRAPH International*.
Oliver, P. (1969). *The story of the blues*. Barrie & Rockliff / The Crescent Press.
Pachet, F., & Cazaly, D. (2000). A taxonomy of musical genres. In *RIAO'00: Proceedings of the Content-Based Multimedia Information Access Conference* (pp. 1238-1246).
Pampalk, E. (2001). *Islands of Music: Analysis, organization, and visualization of music archives*. MSc thesis, Vienna University of Technology.
Pampalk, E., & Goto, M. (2006). MusicRainbow: A new user interface to discover artists using audio-based similarity and web-based labeling. In *ISMIR'06: Proceedings of the 7th International Conference on Music Information Retrieval* (pp. 367-370). ISMIR.
Pauws, S., & Eggen, B. (2002). PATS: Realization and evaluation of an automatic playlist generator. In *ISMIR'02: Proceedings of the 3rd International Conference on Music Information Retrieval* (pp. 222-230). ISMIR.
Pauws, S., & van de Wijdeven, S. (2005). User evaluation of a new interactive playlist generation concept. In *ISMIR'05: Proceedings of the 6th International Conference on Music Information Retrieval* (pp. 638-643). ISMIR.
Pesch, M. (1998). *Techno style: Album cover art*. Zurich: Edition Olms.
Pousman, Z., & Stasko, J. (2006). A taxonomy of ambient information systems: Four patterns of design. In *AVI'06: Proceedings of the Working Conference on Advanced Visual Interfaces* (pp. 67-74). ACM Press.

Pousman, Z., Stasko, J. T., & Mateas, M. (2007). Casual information visualization: depictions of data in everyday life. *IEEE Transactions on Visualization and Computer Graphics archive, 13(6)*, 1145-1152.

Pressing, J. (1995). *Synthesizer performance and real-time techniques*. A-R Editions.

Rattansi, A. (2007). *Racism: A very short introduction*. New York, NY: Oxford University Press.

Redström, J., Skog, T., & Hallnäs, L. (2000). Informative art: using amplified artworks as information displays. In *DARE'00: Proceedings of DARE 2000 on Designing Augmented Reality* (pp. 104-114). ACM Press.

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84(6)*, 1236–1256.

Rich, E. (1979). User modeling via stereotypes. *Cognitive Science: A Multidisciplinary Journal, 3(4)*, 329–354.

Roberts, D.F., Foehr, U.G., & Rideout, V. (2005) *Generation M: Media in the lives of 8 to 18 year-olds*. Menlo Park, CA: Kaiser Family Foundation.

Roto, V. (2007). User experience from product creation perspective. In *Proceedings of Towards a UX Manifesto workshop, in conjunction with HCI 2007* (pp. 31-34).

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161-78.

Seyerlehner, K., Knees, P., Schnitzer, D., & Widmer, G. (2009). Browsing music recommendation networks. In *ISMIR'09: Proceedings of the 10th International Society for Music Information Retrieval Conference* (pp. 129–134). ISMIR.

Sharp H., Rogers, Y., & Preece, J. (2007). *Interaction design: Beyond Human-Computer Interaction*. Chichester: John Wiley & Sons.

Shiroi, S., Misue, K., & Tanaka, J. (2011). A tool to support finding favorite music by visualizing listeners' preferences. In *IV'11: Proceedings of the 15th International Conference on Information Visualisation* (pp. 176-181). IEEE Computer Society.

Shneiderman, B. (1996) The eyes have it: a task by data type taxonomy for information visualizations. In *VL'96: Proceedings of the 1996 IEEE symposium on visual languages* (pp. 336-343). IEEE Computer Society.

Shneiderman, B. (1997). *Designing the user interface*. Addison Wesley.

Siirtola, H. (2007). *Interactive visualization of multidimensional data*. Doctoral dissertation, University of Tampere, Tampere.

Skog, T., Ljungblad, S., & Holmquist, L. E. (2003).Between aesthetics and utility: designing ambient information visualizations. In *Infovis'03: Proceedings of the IEEE Symposium on Information Visualization* (pp. 233-240). IEEE Computer Society.

Sloboda, J. A., & Juslin, P. N. (2001). Psychological perspectives on music and emotion. In P.N. Juslin & J. A. Sloboda (Eds.). *Music and emotion: Theory and research*. New York, NY: Oxford University Press.

Sordo, M., Celma, O., Blech, M., & Guaus, E. (2008). The quest for musical genres: Do the experts and the wisdom of crowds agree? In *ISMIR'08: Proceedings of the 9th International Conference on Music Information Retrieval* (pp. 255–260). ISMIR.

Spectrum Tietokeskus (1976). *Spectrum tietokeskus # 6 (Kop-Lip)*. Latinalainen Amerikka: Musiikki. WSOY.

Spence, R. (2007). *Information visualization: Design for interaction* (2nd ed.). Harlow: Pearson Education Limited.

Torrens, M., Hertzog, P., &. Arcos, J-L. (2004). Visualizing and exploring personal music libraries. In *ISMIR'04: Proceedings of the 5th International Conference on Music Information Retrieval* (pp. 421-424). ISMIR.

Trohidis, K., Tsoumakas, G., Kalliris, G., & Vlahavas, I. (2008). Multi-label classification of music into emotions. In *ISMIR'08: Proceedings of the 9th International Conference on Music Information Retrieval* (pp. 325–330). ISMIR.

Tzanetakis, G., Essl, G., & Cook, P. (2001). Automatic musical genre classification of audio signals, In *ISMIR'01: Proceedings of the 2nd Annual International Symposium on Music Information Retrieval.*

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

Van Aken, J. E. (2004). Management research based on the paradigm of the design sciences: the quest for field-tested and ground technological rules. *Journal of management Studies, 41(2)*, 219-246.

Van Egmond, R. (2004). Compact disc cover design: transfer of musical style and emotional contents into graphical design. In D. McDonagh, P. Hekkert, J. Van Erp & D. Gyi (Eds.). *Design and emotion: The experience of everyday things*. London: Taylor & Francis.
Van Gulik, R., & Vignoli, F. (2005). Visual playlist generation on the artist map. In *ISMIR'05: Proceedings of the 6th International Conference on Music Information Retrieval* (pp. 520–523). ISMIR.
Viegas, F. B., & Wattenberg, M. (2008). Tag clouds and the case for vernacular visualization, *Interactions, 15(4)*, 49-52.
Vignoli, F. (2004). Digital music interaction concepts: a user study. In *ISMIR'04: Proceedings of the 5th International Conference on Music Information Retrieval*. ISMIR.
Vignoli, F., & Pauws, S. (2005). A music retrieval system based on user driven similarity and its evaluation. In *ISMIR'05: Proceedings of the 6th International Conference on Music Information Retrieval* (pp. 272–279). ISMIR.
Voong, M. (2007). Music organisation using colour synaesthesia. In *CHI EA '07: CHI '07 Extended Abstracts on Human Factors in Computing Systems* (pp. 1869-1874). ACM Press.
Ward, G. C., & Burns, K. (2000). *Jazz: A history of America's music*. New York, NY: Alfred A. Knopf.
Ware, C. (2004). *Information visualization: Perception for design*. San Francisco, CA: Morgan Kaufmann Publishers.
Wegenkittl, R., Löffelmann, H., & Gröller, E. (1997). Visualizing the behavior of higher dimensional dynamical systems. In *VIS'97: Proceedings of the 8th Conference on Visualization* (pp. 119-125). IEEE Computer Society.
Weigl, D. M., & Gustavino, C. (2011). User studies in the music information retrieval literature. In *ISMIR'11: Proceedings of the 12th International Society for Music Information Retrieval Conference* (pp. 335–340). ISMIR.
West, K., Cox, S., & Lamere, P. (2006). Incorporating machine-learning into music similarity estimation. In *AMCMM '06: Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia* (pp. 89–96). ACM Press.
Whitman, B. (2005). *Learning the meaning of music*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge.
Yglesias, P. E. (2004). *Cocinando! Fifty years of Latin album cover art*. New York, NY: Princeton Architectural Press.
Yi, J. S., Kang, Y. A., Stasko, J. T., & Jacko, J. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics, 13(6)*, 1224-1231.
Yi, J. S., Kang, Y-A., Stasko, J. T., & Jacko, J. (2008). Understanding and characterizing insights: How do people gain insights using information visualization? In *BELIV'08: Proceedings of the 2008 Conference on beyond Time and Errors: Novel Evaluation Methods for Information Visualization*. ACM Press.
Yoshii, K., Goto, M., Komatani, K., Ogata, T., & Okuno, H. G. (2008). An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model. *IEEE Transactions on Audio, Speech, and Language Processing, 16(2)*, 435-447.
Zhu, J., & Lu, L. (2002). Perceptual visualization of a music collection. *IEEE Transactions on Speech and Audio Processing, 10(5)*, 1058-1061.
Zuo, H., Hope, T., Jones, M., & Castle, P. (2004). Sensory interaction with materials. In D. McDonagh, P. Hekkert, J. Van Erp & D. Gyi (Eds.). *Design and emotion: The experience of everyday things*. London: Taylor & Francis.

Web references

Andrews, R. (2011). Spotifies may steal customers from Pandoras, with Facebook's help. Retrieved January 10, 2012 from http://paidcontent.co.uk/article/419-spotifies-may-steal-customers-from-pandoras-with-facebooks-help

Barrett, B. (2011, January 17). Nokia kills off Ovi Music service you didn't know existed. Retrieved June 19, 2011 from http://gizmodo.com/#!5735707/nokia-kills-off-ovi-music-service-you-didnt-know-existed

Castranova, E. (2003). Theory of the avatar. Retrieved January 17, 2012 from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=385103

Chevannes, B. (1994). Rastafari, roots, and ideology. Retrieved November 3, 2010 from http://www.oneworldmagazine.org/focus/etiopia/rasta.html

Dahlen, C. (2006, May 22). Better than we know ourselves. Retrieved September 20, 2011 from http://pitchfork.com/features/articles/6340-better-than-we-know-ourselves

DeGusta, M. (2011, February 18). The REAL death of the music industry. Retrieved June 19, 2011 from http://www.businessinsider.com/these-charts-explain-the-real-death-of-the-music-industry-2011-2

Fraunhofer Institute for Digital Media Technology (2009). Mood Player – cross-medial, audio-visual mood analysis. Retrieved September 22, 2011 from http://www.idmt.fraunhofer.de/eng/research_topics/moodplayer.html

fRoots (1987). Minutes of meeting between the various 'World Music' record companies and interested parties, Monday 29th June 1987. Retrieved November 4, 2010 from http://www.frootsmag.com/content/features/world_music_history/minutes/page03.html

Goldmann, D. (2010, February 2). Music's lost decade: Sales cut in half. Retrieved June 19, 2011 from http://money.cnn.com/2010/02/02/news/companies/napster_music_industry

Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G., & Verplank. B. (1992). ACM SIGCHI Curricula for Human-Computer Interaction, Section 2.1: Definition of HCI. Retrieved November 1, 2011 from http://old.sigchi.org/cdg/cdg2.html#2_1

IFPI (2011). IFPI digital music report: Music at the touch of a button. Retrieved June 19, 2011 from http://www.ifpi.org/content/section_statistics/index.html

Kirn, P. (2011a). TuneCore: Apple iCloud will transform industry, make streaming the norm: Wait, really? Retrieved June 28, 2011 from http://createdigitalmusic.com/2011/06/tunecore-apple-icloud-will-transform-industry-make-streaming-the-norm

Kirn, P. (2011b). Sound the final frontier: audio collections as planets in space intelligently related. Retrieved September 22, 1011 from http://createdigitalmusic.com/2011/05/sound-the-final-frontier-audio-collections-as-planets-in-space-intelligently-related/#more-18951

Music Industry Report (2010, August 18). Case study: Streaming leads to increased music consumption. Retrieved June 19, 2011 from http://musicindustryreport.org/?p=24752

Moody (2011). Moody web pages. Retrieved August 8, 2011 from http://www.crayonroom.com/moody.php

Musiclens (2007). Musiclens web pages. Retrieved November 3, 2007 from http://www.musicline.de/de/$255dae7db14eabefc87799d69cb952e7/recoengine04/start/ml/musiclens

Musicovery (2011). Musicovery web pages. Retrieved August 16, 2010 from http://www.musicovery.com

Musicmesh (2007). Musicmesh web pages. Retrieved December 21, 2007 from www.musicmesh.net

Oxford English Dictionary (2008). *Oxford English Dictionary* (3rd ed., June 2008). New York, NY: Oxford University Press. Online version retrieved September 21, 2011 from http://www.oed.com/view/Entry/157097

Parry, J. (2007, August 27). Visual variables. Retrieved September 25, 2011 from http://www.joeparry.com/blog/?p=41

Pulley, B. (2011, November 11). Record sales increase as Lady Gaga, Adele find a future with Spotify, Rdio. Retrieved January 16, 2012 from http://www.bloomberg.com/news/2011-11-14/record-sales-rise-as-lady-gaga-adele-find-a-future-with-spotify.html

Qloud (2011). Qloud web pages. Retrieved June 3, 2011 from www.qloud.com

Richards, T. (2002, January 29). Rhythm and blues. St. James encyclopedia of pop culture. Retrieved November 17, 2010 from http://findarticles.com/p/articles/mi_g1epc/is_tov/ai_2419101026/?tag=content;col1

UPF Music Surfer (2008). UPF Music Surfer web pages. Retrieved March 10, 2008 from http://musicsurfer.iua.upf.edu

Vaishnavi, V., & Kuechler, W. (2004, January 20). Design research in information systems. Retrieved October 10, 2011 from http://ais.affiniscape.com/displaycommon.cfm?an=1&subarticlenbr=279

Van Buskirk, E. (2011, October 7). 10 awesome music apps powered by Spotify. Retrieved January 15, 2012 from http://evolver.fm/2011/10/07/10-awesome-music-apps-powered-by-spotify

WhoopApps (2009). iTunes preview: MoodTunes Lite by WhoopApps. Retrieved September 21, 2011 from http://itunes.apple.com/us/app/moodtunes-lite/id344998031?mt=8

Wikipedia Contemporary R&B (2010). Retrieved November 17, 2010 from Wikipedia: http://en.wikipedia.org/wiki/Contemporary_R%26B

Wikipedia Electronica (2010). Retrieved November 22, 2010 from Wikipedia: http://en.wikipedia.org/wiki/Electronica

Wikipedia Gospel Music (2010). Retrieved November 16, 2010 from Wikipedia: http://en.wikipedia.org/wiki/Gospel_music

Wikipedia List of Music Genres (2008). Retrieved March 10, 2008 from Wikipedia: http://en.wikipedia.org/wiki/List_of_music_genres

Wikipedia New Age Music (2011). Retrieved August 3, 2011 from Wikipedia: http://en.wikipedia.org/wiki/New_Age_music

Wikipedia World Music (2010). Retrieved November 4, 2010 from Wikipedia: http://en.wikipedia.org/wiki/World_music

Appendices

Appendix A: Summary of main genres used by selected music services in 2007

Table 4 summarizes the main genres used by selected music services in 2007. The table was created during the first literature review on music visualization, and it was used as a basis for selecting the genre set for the conducted surveys. The web sites were last accessed in April 2007, and thus some of them no longer exist. The used genre taxonomies may also have changed.

In the table, letters A-M refer to music services as follows:

A= Play (www.play.com)
B=CDON (www.cdon.com)
C=Amazon (www.amazon.com) and CD Now (www.cdnow.com)
D=CD Universe (www.cduniverse.com)
E=MP3 (www.mp3.com)
F=Nokia (www.musicrecommenders.com)
G=All Music Guide (allmusic.com)
H= Music Match (www.mmguide.musicmatch.com)
I=iTunes, J=MSN Music (http://music.msn.com)
K=RealMusic (http://europe.music.euro.real.com)
L=HMV (www.hmv.co.uk)
M=Virgin (www.virgindigital.co.uk)

In the case of each service, the supported genres are marked with '1' in the corresponding column. The "Sum" column shows the number of music services using each genre, and the "Selected label" column lists the final genre labels selected for the surveys.

Table 4 Main genres used by selected music services in 2007 and labels selected for the surveys.

	A	B	C	D	E	F	G	H	I	J	K	L	M		Sum		Selected label
Genre																	
Alternative, Indie and/or Alternative Rock	1		1	1	1	1			1	1	1	1	1		10		Alternative and Indie
Asian pop						1									1		
Avant-Garde							1								1		
Best of the Best		1													1		
Bluegrass								1							1		
Blues	1	1	1	1			1	1	1	1					8		Blues
Box sets			1	1											2		
Britrock										1					1		
Cajun							1								1		
Celtic							1								1		
Childrens	1		1					1	1	1					5		
Chillout											1				1		
Christmas	1														1		
Classical	1	1	1	1		1			1	1	1	1	1		10		Classical
Comedy							1	1		1					3		
Compilation	1	1													2		
Country	1	1	1	1	1	1	1	1	1	1		1			11		Country
Crooners	1														1		
Easy listening							1	1		1		1			4		
Electronica, Dance and/or DJ	1		1	1	1	1	1	1	1	1		1	1		11		Electronica and Dance
Finnish		1													1		
Folk	1		1				1	1	1	1		1			7		Folk
Folk/country/easy											1				1		
Gospel, Christian or Religious	1		1				1	1		1					5		Gospel
Heavy Metal and/or Hard Rock	1	1	1	1	1					1					6		Metal
Hip-Hop and/or Rap	1	1	1	1	1	1	1		1	1	1	1			11		Hip-Hop and Rap
Holiday								1		1					2		
Imports			1												1		
Instrumental			1												1		
Jazz	1	1	1	1	1		1	1	1	1					9		Jazz
Jazz & Blues						1						1	1		3		
Latin or Brazilian			1	1	1	1	1	1	1	1					8		Latin
Live performance				1											1		
Miscellaneous or Other			1							1					2		Other
New Age	1		1				1	1	1	1					6		New Age
Oldies				1						1					2		
Pop			1						1	1	1	1			5		Pop
Pop legends											1				1		
Rap / R&B								1							1		
R & B			1	1			1				1				4		
Reggae	1	1					1	1	1	1		1			7		Reggae
Rock, Classic Rock or Rock (inc. metal)			1				1		1	1	1	1			6		Rock
Rock & Pop	1	1		1	1			1					1		6		
Singles		1											1		2		
Soul & R'n'B	1	1				1			1	1		1			6		Soul, RnB, and Funk
Soul or Soul & Urban					1						1		1		3		
Soundtracks	1	1	1	1			1	1	1	1		1	1		10		
Specialist & roots													1		1		
Spoken or Spoken Word & Childrens										1		1	1		3		
Techno & Trance		1													1		
Vocal or Broadway & Vocalists			1				1	1		1					4		
Western pop						1									1		
World or International	1	1	1	1		1	1	1	1	1		1			10		World Music
World/Jazz/Blues											1				1		
50s/60s											1				1		
70s											1				1		
80s											1				1		

Appendix B: History and definitions of musical genres used in the thesis

In this appendix, the musical genres used in this thesis are presented and defined in more detail. The genres are presented in roughly chronological order.

Folk

Music has been an important part of human life throughout history. Before the invention of the phonograph in 1877, most songs were passed orally from performer to performer, as musical folklore, and changed slightly on the way as the result of the personality of each performer. So-called "folk" music has traditionally been defined as having the following five attributes (Cohen 2006, pp. 1-2): *"1) its origins can perhaps be located in a particular culture or region; 2) authorship has historically been unknown, although authors did emerge over the past two centuries; 3) it has traditionally been performed by nonprofessionals, perhaps playing acoustic instruments; 4) its composition has been fairly simple, with perhaps little complexity so that it can be performed and shard communally; and 5) the songs have historically been passed down through oral transmission."* Folk songs have both spatial and temporal attributes (Bohlmann 2006, p. 71). They may tell stories about individuals, communities, or specific historical events. While being primarily connected to a previous era, they can also be shaped by contemporary events such as the Vietnam War in America.

Just like in the case of world music (see below), it is not possible to list the main instruments for folk. The selection of instruments varies greatly from the harp and bagpipes of Celtic music to the acoustic guitar of American folk. Since folk and world music can share many common elements, it can often be very difficult to decide to which genre an artist belongs. In many music services, the label "folk" is used to refer to all Western folk music, while non-Western folk music styles have been placed under the "world music" umbrella term.

For many people, the first image of folk is either some local folk music style or an American folk artist from the 1960's and 1970's hippie era. Examples of the latter include Bob Dylan and Joan Baez, who sang topical protest songs and accompanied themselves with acoustic guitars, and popular singer-songwriters such as Joni Mitchell or Van Morrison, who came originally from a folk background or had such elements in their music. The outlook of folk performers varied from the "down-home and casual" image of the sixties folk-rock groups (Charlton 1998) to the wild and colorful clothing of the hippies. To make a clear difference to world music, all folk music visualizations presented in this thesis relied on the hippie theme.

World Music

It is impossible to find a single, comprehensive definition for the label "world music." From the ethnomusicologist point of view, it encompasses all the music of the world, be it folk, art, or popular music (Bohlmann 2002, pp. i-v). However, using such a broad definition has little practical value for the music industry. According to another definition, world music is *"a general categorical term for global music, such as the traditional music or folk music of a culture that is created and played by indigenous musicians and is closely related to the music of the regions of their origin"* (Wikipedia

World Music 2010). In fact, folk music and folk songs can be considered as the original world music (Bohlmann 2002, p. 69).

As a commercial phenomenon, world music is roughly 25 years old. In 1987, the members of several Western record companies and other interested parties decided to start using the umbrella term “world music” to label music outside the Anglo-American and European mainstreams (fRoots 1987; Bohlman 2002, pp. iv-v). After the meeting, the term has been widely used to categorize artists and albums in marketing, music catalogues, and record stores. In this thesis, I have also adopted the Western, commercial view of world music.

Examples of world music forms include Eastern European folk music, taarab, mbalax, soukous, throat singing, and flamenco. Popular “non-Western” styles such as Jamaican reggae and Latin music are usually classified as their own genres. World music may be based on non-Western scales and modes, and it often features traditional ethnic instruments such as the kora, the sitar or the didgeridoo (Wikipedia World Music 2010).

Nowadays, world music is *“not simply the music of an exotic ‘other’”* (Bohlmann 2002, p. vi). As a result of cultural globalization, people nowadays more frequently encounter music from other cultures. World music is used in commercials, movie soundtracks, and played as background music in coffee shops. Today, many mainstream artists have adopted features from world music, and even heavy metal bands such as Nile have included non-Western elements as part of their music.

Classical

Historically speaking, the term “classical music” refers to only music from the so-called Classical period (1750-1820). Famous composers from the period include, e.g., Wolfgang Amadeus Mozart and Ludwig van Beethoven. In a broader sense, the genre compasses several Western art music styles from year 1000 to the present: early music (medieval and renaissance, 1000-1600), Baroque (1600-1750), Classical (1750-1820), Romantic (1810-1920), and Modern music (1900-present). (Burrows 2005)

Classical music can be performed by any suitable group of musicians varying from solo artists to chamber groups and symphony orchestras. Examples of commonly used instruments include brass instruments (e.g., trumpet and tuba), reed instruments (e.g., clarinet and oboe), string instruments (e.g., violin and contra bass), percussion, and piano. (Burrows 2005)

Latin

Societally and geographically speaking, term Latin America refers to those American territories which once were inhabited by people who spoke languages derived from Latin (e.g., Spanish, Portuguese, and French) (Spectrum Tietokeskus 1976, pp. 397-414). However, in everyday language the term usually compasses all of the Americas south of the United States (including Mexico, Central and South America, and the Caribbean). Naturally, the “Latin” musical genre encompasses the various musical styles originating from those regions.

Latin music is a mix of European, African, and Indian music styles. Examples of subgenres include rumba, samba, bossa nova, tango, calypso and salsa. The songs are typically melodic and danceable, and therefore popular in clubs and dancehalls all over the world. Brazil is also well-known for its samba schools and carnivals such as the annual Rio de Janeiro carnival. (Spectrum Tietokeskus 1976; Broughton et al. 1994)

Many Caribbean music styles express the cultures of the islands, and therefore have developed to their own musical directions. One such example is Jamaican reggae, where the African-based percussion styles are less apparent due to the repression of the British colonizers. Reggae is often more political than the other Latin subgenres, and usually classified as its own main genre. (Broughton et al. 1994)

Country

From the 17th through the 19th century, large numbers of British immigrants moved to the Southern parts of United States, bringing the music and instruments of the Old World with them. The resulting rural folk music was a blend of British folk songs, ballads, dances, instrumental pieces, and musical contributions of other ethnic groups (particularly African-Americans) of the region. Country music, originally known as hillbilly music, grew out of these folk songs. (Malone 2002)

In the twenties, the music of the South received the attention of the rest of the country via popular barn dance radio programs. Hillbilly music developed into different subgenres including western swing, bluegrass, and honky tonk. The cowboy image of country music was developed in the southwestern states Louisiana, Oklahoma, Texas, and California, and Hollywood films popularized cowboy songs during the 1930s and 1940s. While the importance of cowboy themes in country has decreased since, the cowboy image has stuck with the genre. (Charlton 1998; Malone 2002)

Since its birth, country music has spread all over the world. The largest fan base can still be found in the US, and country artists have had number one hits in the local charts. The typical musical elements of country include steady beat, triadic harmonies, two-beat bass (alternating between the root and fifth of the chord on beats one and three), and sliding notes. Common instruments include fiddle, five-string banjo, guitar, dulcimer, piano, ukulele, steel guitar, and mandolin. (Charlton 1998; Malone 2002)

Blues

Blues music developed its style in the rural areas of the southern United States in the end of the 19th and the beginning of the 20th century. The roots of the genre are in African-American spirituals, work songs, field hollers, and songsters' ballads. After the American Civil War (1861-1865) and the official end of slavery, the African-American people were still not accepted by the white society and many continued working in agricultural production (e.g., the cotton fields). For these people, the blues was an integral part of life, and blues singers accompanied themselves on acoustic guitars and sang rough but expressive texts. Because of the rural origins, the earliest known blues style was called country blues. (Oliver 1969; Charlton 1998)

The first blues recordings were made in the twenties. As time passed by, the genre started to fragment and other styles such as urban blues, rhythm and blues, and Chicago blues emerged. The typical elements of blues include the use of bent "blue" notes, a twelve-bar structure, three-chord progressions, call-and-response between for example the vocalist and the guitarist, shuffle beat, and rough but expressive vocals. Common instruments include acoustic and electric guitars, bass, drums, piano, and horns. (Charlton 1998)

Gospel

While the term "gospel music" can refer to any type of Christian music, most people tend to associate it with African-American gospel. The roots of this "black gospel" can be traced to the hymns and spirituals of the 18th century. Before the American Civil War, the African-American slaves attended churches with their white owners to sing religious hymns with white European traditions (aka "white gospel"). After the war, the African-Americans built their own churches and developed their own style of religious music. While the term "spiritual" referred originally to both black and white religious music, it became more commonly associated with songs composed by the African-Americans. In the 1930s, the term "gospel" was taken into use to refer to modern religious music. (Charlton 1998)

The typical musical elements of African-American gospel music include religious lyrics, call-and-response, vocal harmonies, and improvised vocals (Charlton 1998). Common instruments include piano, Hammond organ, guitar, horns, drums, and bass (Wikipedia Gospel Music 2010).

Jazz

Jazz music was born in the African-American communities of New Orleans in the beginning of the 20th century. It was originally dance music that was influenced by ragtime piano music, the sacred music of the Baptist church, and the blues arriving from the Mississippi Delta. The original term, "jass", was invented in 1916, and the present term "jazz" in 1917. According to the legend, jazz was invented in the red-light district of Basin Street (aka Storyville), but in reality Storyville was just one of the places where the musical style was played. (Ward & Burns 2000)

After its birth, jazz spread first to the bigger American cities such as New York and Chicago. By the mid-1920s, different versions of it were played all over the country, and gradually various subgenres, including bebop, cool, fusion, free, soul, and nu jazz were developed. The typical musical elements of jazz include group interplay and improvisation, call-and-response, swing, and the use of polyrhythms, although these can vary depending on the subgenre. Common instruments include the saxophone and other wind instruments, piano, guitar, acoustic bass, and drums.

Soul, RnB & Funk

Rhythm and Blues (aka R&B or RnB) was born in the African-American ghettos of the United States. Based on blues and jazz, it was the popular black music of the 1940s and 1950s. Before Billboard magazine used the term "Rhythm and Blues" in 1949, the style was known as "race music." Unlike the blues, rhythm and blues was urban dance music that focused on the positive sides of life. (Richards 2002; Charlton 1998)

Soul music was born in the cities of the Northern United States as a combination of fifties gospel and rhythm and blues. Many soul singers were originally gospel singers. Instead of God, the message of soul was love and relationships, and typical musical elements included the use of melismas, call-and response, horn solos, and backbeat (accenting the 2nd and 4th beats of a bar). The term "soul" came into common use in the sixties, and different cities such as Chicago, New York, Detroit and Memphis soon developed their own characteristic styles such as Memphis Soul and (Detroit based) Motown. (Charlton 1998)

In the '60s and '70s, soul music became even more scattered. Disco music was developed based on the soul of Detroit and Philadelphia, and James Brown recorded the first funk songs. Funk *was "a very distinctive style that used polyrhythms, syncopated bass lines, and short vocal phrases with a considerable amount of repetition of those rhythm patterns and phrases"* (Charlton 2008, p. 270). Chord changes were usually kept to a minimum.

To add to the confusion, yet another style called contemporary R&B (often referred to as R&B) was born in the 1980s. Contemporary R&B combines elements of rhythm and blues, hip-hop, and soul, but has little to do with the original rhythm and blues. While soul, funk, and rhythm and blues rely on the use of traditional instruments such as horns, keyboards, electric guitar, bass, drums, contemporary R&B often relies on the use of electronic instruments such as synthesizers and drum machines. (Wikipedia Contemporary R&B 2010)

Rock

Finding the exact point at which rock music was born may be impossible, but the genre has its roots in the 1940s blues, rhythm and blues, and 1950s rock and roll (Charlton 2008, p. 46-62). While several authors (e.g., Frith 2007) state that rock music was born in the United Kingdom and the United States in the 1960s, many rock historians also consider some of the 1940s recordings to be rock music (Charlton 2008, p. 46).

The main instruments of rock include (usually distorted) electric guitar, bass, and drums with occasional keyboards such as a piano, synthesizer, or the Hammond organ. The songs are usually in 4/4 with emphasis on the backbeat. According to some authors such as (Frith 2007), rock has more emphasis on musicianship and live performance than pop music. Since the times of Chuck Berry and the Beatles, rock music has transformed into various subgenres including psychedelic rock, country rock, hard rock, and heavy metal. Like pop, rock has nowadays become a wide umbrella category and thus lost part of its original importance.

Pop

In a way, popular or popular music is the opposite of elite, art, or "serious" music. It is often used to refer to commercially successful contemporary music or "music for the masses." According to Kemp (2004, p. 36), pop music is *"light entertainment intended for dancing, leisure, relaxation and enjoyment"* and *"primarily youth cultural."* One important form of pop music is teenage or teen idol pop music, which originated in the 1950s with artists such as Pat Boone and Richie Valens (Charlton 2008, pp. 71-76).

The main problem with the pop music label is that it is an umbrella category that may refer to basically any style that is popular at the moment. For example, the pop/rock page of AllMusic (www.allmusic.com) includes genres varying from heavy metal to J-Pop. Some consumers (at least in Finland) also associate the term with guitar-oriented music that is slightly lighter than rock (e.g., Britpop).

Reggae

Reggae music was developed in Jamaica in the late 1960s (Barrow 2001). It is a successor of Jamaican ska and rocksteady, and normally slower in tempo. As stated in AllMusic (www.allmusic.com), *"During one very hot summer, it was too hot to either play or dance to ska, so the beat was slowed down and reggae was born."*

The rhythmic style of reggae emphasizes the second and fourth beats of each bar. Typical instruments include drums and other percussion, guitar, bass, keyboards, and horns. The topics of lyrics vary from radical messages (Rastafarianism, Pan-Africanism, political corruption, nuclear war, etc.) to lighter topics such as tributes to American Westerns, dancing, and smoking cannabis (Morrow 1999). The most common visual sign of a reggae artist is the dreadlocks, which are *"the long, uncombed locks of hair that many Rastafarians wear as a sign of their faith"* (Morrow 1999, p. 25). Since the 1970s, reggae has gained word-wide success as "rasta music", and in peoples' minds it is strongly connected with the Rastafari movement.

The Rastafari movement and ideology was born in the Christian culture of Jamaica in the 1930s. The Rastafarians (Rastas) worship Haile Selassie I, the former Emperor of Ethiopia (1930–1936 and 1941–1974), as the second coming of Jesus Christ onto the Earth. Before his coronation, Haile Selassie was also known as Ras Tafari. The prominent themes of the Rastafari movement include the spiritual use of cannabis, the rejection of Babylon (i.e., Western society), proclaiming Zion (i.e., Africa and especially Ethiopia) as the promised land, and embracing Pan-Africanism (such as the political views of the prophet Marcus Garvey). While it is true that many reggae musicians practice Rastafarianism, most "true" Rastafarians have no interest in or respect for reggae music; they consider it to be a form of useless and earthly entertainment. (Morrow 1999; Chevannes 1994; Kosmos 2002)

Metal

Heavy metal (or simply metal) developed as an aggressive subgenre of rock in the late 1960's and early 1970's. The most influential early bands were Black Sabbath, Deep Purple, and Led Zeppelin, all of which were from the United Kingdom. For many people, heavy metal was born when the first Black Sabbath record was released in February 1970. Since its birth, heavy metal has become a multibillion dollar industry consisting of several subgenres such as death, doom, NWBHM (New Wave of British Heavy Metal), hair, and black metal.

The typical elements of metal music include distorted electric guitars, power chords, riffs as the basic structural elements of the music, guitar solos, powerful drums, and aggressive, evil or macho lyrics. Most heavy metal bands rely on the combination of electric guitars, bass, drums, and vocals, with occasional keyboards added to the mix. (Charlton 1998)

The classic heavy metal "uniform" consists of long hair, jeans, and a black band t-shirt. However, the style can vary greatly from the pink lipstick and glitter of 80's hair metal bands (Aldis & Sherry 2006) to the black and white corpsepaint and pentagrams of Norwegian black metal bands (Moynihan & Soderlind 2003). In addition to clothing, another distinctive visual element of heavy metal is the devil horns hand gesture, which was popularized by the late singer Ronnie James Dio (Bukszpan 2003, p. 65).

Hip-Hop & Rap

Hip-hop culture was born in the 1970's in the South Bronx district of New York City. As a result of building the Cross-Bronx Expressway in the 1950's, many jobs had been lost and the area had turned into a slum. Several African-American and Latino gangs had also grown in power. As a response to these poor living conditions, hip-hop culture emerged with its four cornerstones DJing, MCing (aka emceeing or rapping), b-boying (aka breakdancing), and graffiti. (Chang 2008)

The core components of hip-hop music are beats and rhymes produced by DJs (Disc Jockeys) and MCs (Masters of Ceremony). The genre originated from the NYC block parties where DJs played popular musical styles such as soul and funk to the dancing crowds. DJs such Jamaican-born DJ Kool Herc began repeating selected percussion breaks (breakbeats) from the songs using two similar vinyl records and an audio mixer. Other turntable techniques such as scratching and beatmatching were later added to the repertoire. Nowadays, beats are also generated using other equipment such as samplers, synthesizers, computers, and real instruments. (Chang 2008)

Today, hip-hop has become a world-wide, multibillion dollar industry. Hip-hop music consists of several subgenres, including gangsta, Christian, old-school, West Coast, and East Coast rap, and hip-hop clothing has become a part of mainstream street culture and popular fashion. While the style has changed from year to year, hip-hop is often associated with items such as baggy jeans, bling-bling jewelry, baseball caps, and the like.

Electronica & Dance

Electronically produced music has existed as long as electronic technology itself. The first electronic music instrument, the Musical Telegraph from 1874, amplified the sound of telegraph transmitters through an electromagnet-equipped washbasin (Holmes 1985). The first mass-market analogue synthesizers appeared in the 1960s and polyphonic synthesizers in the early 1970s (Pressing 1995). However, synthesizers did not reach mainstream popularity until the 1980s when low-cost microprocessor and integrated circuits started to appear. Replacing the analogue components with digital ones resulted in cheaper, more reliable, and more user-friendly synthesizers, and in 1983 the control signals between synthesizers from different companies, computers, and other devices were standardized (MIDI 1996).

After the invention of MIDI, the creation of electronic music became significantly easier for the average musician, and new subgenres started to emerge. The term "electronica" soon became an umbrella term containing several electronic music styles such as ambient, drum'n'bass, house, and techno (Cascone 2002). Unlike electronic dance music, which is targeted for dancing at nightclubs and raves, electronica also contains genres intended as foreground or background listening. (Wikipedia Electronica 2010)

New Age

In (Birosik 1989), New Age has been defined as music *that "encourages personal empowerment, earth connectedness, space consciousness, and interpersonal awareness."* New Age music has been used by for example yoga and meditation practitioners to create a peaceful atmosphere and to manage stress. As the definition of New Age music focuses on feelings rather than the type of instruments, the term may include and overlap with a wide variety of styles including classical, folk, and world music.

The roots of New Age music are in European experimental music from the late 1960s and early 1970s. The songs are often long, repetitive, and hypnotic. Commonly-used instruments include synthesizers, strings, ethnic instruments, acoustic guitars, and flutes. The songs may also utilize recordings of nature sounds. The themes of New Age music vary from nature and space to wellness, harmony, and dreams. (Wikipedia New Age Music 2011)

Alternative & Indie

"Alternative music" is a term commonly used to describe music outside the mainstream. While the term does not imply any certain musical style, it is often used to refer to rock-type music such as Britpop, grunge, or indie rock. Many alternative rock groups try to stay close to their fans and not "sell out" to large record companies (Charlton 2008, p. 314).

The term "indie" or "independent music" refers to independent artists, i.e., artists that have not signed a recording deal with a major commercial record company. Independent artists may release their albums at their own cost (the "do-it-yourself" approach) or through a small record label that works independently from the majors.

Some music services such as Amazon (www.amazon.co.uk) combine alternative and indie music under the same category.

Original publications

This thesis is based on the following publications, which are reproduced here with permission. The publications are referred to in the text by **P1**, **P2**, and so forth.

P1 Holm, J., & Aaltonen, A. (2007). Associating graphical objects with musical tempo. *Proceedings of the 2nd Audio Mostly Conference* (pp. 174-179). Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany, 27-28 September. (online proceedings http://www.audiomostly.com/amc2007)

P2 Holm, J., & Aaltonen, A. (2008). Associating graphical objects with release year of music. *Proceedings of the 3rd IASTED International Conference on Human-Computer Interaction* (pp. 218-223). ACTA Press.

P3 Holm, J., Aaltonen, A., & Siirtola, H. (2009). Associating colours with musical genres. *Journal of New Music Research, 38(1)*, 87-100. Routledge.

P4 Holm, J., Holm, H., & Seppänen, J. (2010). Associating emoticons with musical genres. *Proceedings of the International Conference on New Interfaces for Musical Expression*. University of Technology Sydney, Australia, 7-11 June, 2010. (online proceedings http://www.educ.dab.uts.edu.au/nime/PROCEEDINGS/)

P5 Holm, J., Siirtola, H., & Laaksonen, L. (2010). Associating avatars with musical genres. *Proceedings of the 14th International Conference on Information Visualisation* (pp. 186-193). IEEE Computer Society.

P6 Holm, J., Lehtiniemi, A., & Eronen, A. (2010). Evaluating an avatar-based user interface for discovering new music. *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*. ACM Press.

P7 Holm, J., & Lehtiniemi, A. (2011). A virtual world prototype for interacting with a music collection. *Proceedings of the 4th International Conference on Online Communities and Social Computing, held as Part of HCI International 2011* (pp. 326-335). Springer-Verlag.

P8 Lehtiniemi, A., & Holm, J. (2011). Evaluating a potentiometer-based graphical user interface for discovering new music. *Proceedings of the 15th International Conference on Information Visualisation* (pp. 110-118). IEEE Computer Society.

P9 Holm, J., & Siirtola, H. (2012). A comparison of methods for visualizing musical genres. *Proceedings of the 16th International Conference on Information Visualisation* (pp. 636-645). IEEE Computer Society.

The publications are not included in the electronic version of the dissertation.

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-2897-2
ISSN 1459-2045