The Information Retrieval Series

Peter Knees
Markus Schedl

Music Similarity and Retrieval

An Introduction to Audio- and Web-based Strategies

Springer

The Information Retrieval Series Volume 36

More information about this series at http://www.springer.com/series/6128

Peter Knees • Markus Schedl

Music Similarity and Retrieval

An Introduction to Audio- and Web-based Strategies

Springer

Peter Knees
Department of Computational Perception
Johannes Kepler University
Linz
Austria

Markus Schedl
Department of Computational Perception
Johannes Kepler University
Linz
Austria

ISSN 1387-5264
The Information Retrieval Series
ISBN 978-3-662-49720-3 ISBN 978-3-662-49722-7 (eBook)
DOI 10.1007/978-3-662-49722-7

Library of Congress Control Number: 2016940377

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH Berlin Heidelberg

To our families and everyone who shares our passion for music.

Foreword

When Peter Knees and Markus Schedl invited me to write the foreword to their upcoming book on music information retrieval, they attached an early sketch of the contents as a copy to the email. Upon reading through their notes, I was struck by five strongly emotive thoughts in quick succession: First, Peter and Markus have done a great job capturing and explicating the many different facets that come together to make up Music Information Retrieval (MIR) research and development. Second, this is exactly the kind of text I would assign to my students and recommend to colleagues who are newcomers to MIR. Third, MIR certainly has matured and evolved in some truly interesting ways since 2000 when the first International Symposium for Music Information Retrieval was convened in Plymouth, Massachusetts. Fourth, forget 2000, how about 1988, when a much younger me was beginning my personal relationship with MIR? If it were possible to describe the state of MIR in 2016 to the 1988 version of me or any of my fellow music school students, would we be able to comprehend at all just how magically advanced music access, processing, search, and retrieval would become in just under 30 years? Fifth, and finally, if it were possible to send this book back in time to 1988, would it convince us of the marvels that lay ahead and help us to understand how they came about? Since the answer to that last question was a resounding yes, I decided to accept the kind invitation of Peter and Markus. Thus, it was my pleasure and honor to write the foreword you are currently reading.

In 1988, I was studying for my undergraduate degree in music at the University of Western Ontario in London, Ontario, Canada. I was a decent second-rate theorist. I was, however, a famously bad flautist who spent more time drinking coffee, smoking, and playing cards than I did practicing. In the spring of 1988, I became desperate to find some easy-to-learn flute music to play at my fast-approaching final flute exam. I knew vaguely of a baroque sonata in my favorite key of C major. It took me several weeks of false starts, questions to librarians, browsing CD collections, and searching the library catalogue before we were finally able to discern that the piece in question was Bach's Flute Sonata in C major, BWV 1033. The exam itself was a bit of a flop (remember, I rarely practiced) so I am now understandably a professor of library and information science, and not music. Notwithstanding that my not-so-successful flute exam was the proximate cause to a change of careers for me, the search for the music did leave me with two complementary questions that have informed my academic work ever since: (1) Why is finding music so damn hard and annoying? and (2) What can we do to make it not hard and perhaps even enjoyable? The last 30 years have shown that I am far from alone in having these questions.

After my illustrious career as a second-rate music theory student, I found a welcoming home at the Graduate School of Library and Information Science (GSLIS), University of Western Ontario. I was encouraged to explore my two questions by the Dean of the faculty, Jean Tague-Sutcliffe. Professor Sutcliffe was a leading international figure in the formal evaluation of text retrieval systems, and I was quite privileged to have studied with her. My graduate student days at GSLIS shaped my thinking about what it would mean to do MIR research. First, my graduate work dealt with music only at a symbolic level in an attempt to retrofit music information into current state-of-the-art text retrieval engines. Second, library and information science has a strong tradition of user-centered service and research. It is important to note that I paid mostly lip-service to user-centered issues while a student as all the "serious" work was being done by systems students building and evaluating new retrieval techniques. In retrospect, I was very shortsighted to have fixated on classic IR techniques and a complete idiot to have downplayed the absolute importance of the user in any successful MIR system design and evaluation. 2016 me waggles his finger at 1995 me, accompanied by a firm "tsk, tsk."

Fast forward to October of 2000 and the first ISMIR meeting. I find myself sitting in the conference room of the John Carver Inn, Plymouth, Massachusetts. In front of me, Beth Logan is presenting her paper introducing Mel frequency cepstral coefficients (MFCC) as a possible feature for building machine learning models from music audio. This paper is historic for suggesting MFCCs as an audio feature to the MIR community that has become almost ubiquitous. The presentation is also semilegendary for the spirited debate among attendees concerning the appropriateness of using MFCCs, originally designed for speech, in music applications. I must admit, the pros and cons of the debate were completely lost on me. At this point in the meeting, my head was still swimming with one overwhelming thought: "Audio! We can do real MIR things with music audio! Who knew? Too cool!" I cannot overstate how thunderstruck I was by the fact that I was in the presence of very solid researchers making very real progress working with actual music audio. Digital signal processing (DSP) seemed like alchemy to me, except this alchemy was yielding gold. Remember, my MIR world vision until this time derived from my text-based IR graduate work and my undergrad studies in manipulating and analyzing the symbols of music notation. The first major conference publication of my MIR work was, for example, at the Association for Computing Machinery, Special Interest Group on Information Retrieval (ACM SIGIR), held in June of 2000 which discussed converting melodic intervals into text for use in standard search engines.

As you can see, ISMIR 2000 was a paradigm shifting event for me. This paradigm shift involved more than my naiveté surrounding the existence of useful music audio processing techniques. Until this point, I had not thought that MIR might not be bound by the classic query-index-response model (i.e., Cranfield model) that pervades traditional conceptions of IR. I had not yet considered that MIR might be expanded to include different kinds of tasks beyond simple search and retrieval. Papers about discovering and visualizing music structures, identifying themes, classifying genres, labeling instruments, describing content, summarizing audio, and visualizing patterns illuminated new areas of possibilities that I had overlooked previously. Associated closely with these newfound MIR areas was a bundle of techniques that I had heard of but had never seen in action. For me, machine learning (ML) and its older cousin, artificial intelligence (AI), had been the stuff of science fiction and spy novels. Prior to ISMIR 2000, I had no idea how ML/AI could possibly apply to MIR research. I recall being deeply skeptical of our choice of keynote, Marvin Minksy, who was world-renowned for his groundbreaking AI work. After seeing the excellent presentations that made use of such things as neural nets and Gaussian mixture models, I became convinced.

While flying home from ISMIR, I tried my best to make sense of what I learned. I knew that I had to broaden my conception of what it meant to do MIR research. I needed to come up with better way of summarizing succinctly what the MIR project was all about. To this end, I came up with the pseudo-equation $\mathrm{MUSIC}_{(Audio)} + \mathrm{DSP} + \mathrm{ML} = \mathrm{MIR}$. I am fully aware that this model leaves out a great many important aspects of MIR such as the symbolic side of music and, of course, users in their individual and social contexts; however, it did capture quite accurately what was to become (and continues to be) the most popular, and possibly the most productive (up to now), approach to MIR research and development. While there are many different and interesting facets to MIR that do not involve the three components of my pseudo-equation—music audio, digital signal processing, and machine learning—it is an inescapable fact that one must have at least an introductory understanding of these topics, and their interactions, if one is to comprehend the accomplishments and promise of MIR research. This book you have in your hand right now provides just such an introduction, and much more.

The Music Information Retrieval Evaluation eXchange (MIREX) convened for the first time at ISMIR 2005, London, UK. A quick review of the tasks evaluated in 2005 reinforces the notion that the model represented by my pseudo-equation above had become the dominant MIR research paradigm. Most of the MIREX 2005 tasks were based upon evaluation techniques popular in the ML and DSP communities rather than those of the information retrieval domain. That is, rather than using a set of predefined test queries against which systems are evaluated, most MIREX tasks have been designed to use pre-constructed ground truth sets upon which systems are trained and then tested in n-fold cross-validation experiments. MIREX has grown substantially since 2005; however, the ML evaluation worldview still predominates. The use of ground truth sets and train-test experiments have many benefits including ease of administration, simple reproducibility, known and accepted scoring methods, and generally understandable results. There are also several rather important shortcomings to the wholesale adoption of DSP and ML evaluation methods. First, the methodology has a tendency to shape the tasks evaluated in a tail-wagging-the-dog kind of way. That is, there is a quite strong tendency for tasks to be held because there exists ground truth (or ground truth would be easy to create) rather than the inherent importance of the task topic. Second, because ground truth is so expensive to create, it is usually donated by graduate students working on specific MIR subproblems, like beat tracking or chord estimation. This means that MIREX tasks are rather narrowly scoped and thus, even when taken all together, generate a fragmented vision of the MIR problem space. Because of this, MIREX has become very good at analyzing a wide variety of interrelated trees in some detail but has failed to describe the forest in which they live. Third, MIREX is almost completely mute about the most important component of any successful MIR system: the user.

As a leader of MIREX, I bear considerable responsibility for the shortcomings listed above. As MIREX was being developed prior to 2005, I must admit a strong desire to privilege the formal evaluation of content-based approaches because they fit the model implied by my MIR pseudo-equation. When MIR researchers like Julián Urbano assert (justifiably) that MIREX, in particular, and MIR research, in general, must now focus on user-centric tasks, I can offer no counterargument. They are absolutely correct in arguing that both MIREX and MIR research are now hitting an upper limit to the useful insights it can provide absent user-inspired tasks and user-centric experiments and studies. In retrospect, I should have put more effort into developing a MIREX community to define and create user-centered evaluation tasks. One thing that remains a bit of a mystery to me about this state of affairs is the fact that my own personal MIR research evolved away from system-centric to user-focused through my marvelous interactions with such notable "user needs and uses" scholars as Sally Jo Cunningham, Jin Ha Lee, Audrey Laplante, Charlie Inskip, and Xiao Hu. These folks, and others like them, need to be applauded for always keeping the user at the heart of their MIR research thinking.

Peter and Markus, when they were not writing this fine book, have also made significant contributions to the cause of putting the user at the heart of MIR research. Their ongoing work based upon an examination of the social contexts in which all kinds of users seek and interact with music and other music lovers is very well regarded. It is Peter's and Markus's music-in-context line of inquiry that informs the second and third part of the text. For me, these are the most important and potentially influential contributions of the book. Parts II and III put all the technical DSP and ML pieces discussed in Part I in their proper place, that is, at service of real users doing real things with real music. If I were able to somehow send one part of the book back to 1988, it would be these sections. Armed with their knowledge and ideas about users and how they live, use and enjoy music in their everyday lives, I would have spent less time fussing about making my MIR research look like classic IR research. I would have taught myself to more joyfully embrace the noisy, contradictory, and difficult-to-comprehend data generated by users in the wild. MIREX would definitely have a large set of user-centered tasks even if they are harder to manage and run. Most significantly, my pseudo-equation would have read from the beginning, MUSIC + DSP + ML + *USER* = MIR.

Champaign, IL, USA
February 2016

J. Stephen Downie

Preface

Music is an omnipresent topic in our society—it is everywhere and for everyone. Music is more than just the pure acoustic perception. It is a pop cultural phenomenon, maybe even the most traditional and most persistent in human history. It takes a central role in most people's lives, whether they act as producers or consumers, and has the power to amplify or change its listener's emotional state. Furthermore, for many people, their musical preferences serve as a display of their personality. Given its cultural importance, it seems no wonder music was the first type of media that underwent the so-called digital revolution. Based on the technological advancements in encoding and compression of audio signals (most notably the invention of the mp3 standard) together with the establishment of the Internet as the mainstream communication medium and distribution channel and, in rapid succession, the development of high capacity portable music players, in the late 1990s, digital music has not only stirred up the IT industry but also initiated a profound change in the way people "use" music. Today, a lot more people are listening to a lot more music in many more situations than ever before. Music has become a commodity that is naturally being traded electronically, exchanged, shared (legally or not), and even used as a means for social communication. Despite all these changes in the way music is *used*, the way music collections are *organized* on computers and music players and the way consumers *search* for music within these structures have basically remained the same for a long time.

Nowadays, we are witnessing a change in this behavior. Intelligent music listening applications are on the rise and become more and more important in high-end systems for music aficionados and everyday devices for casual listeners alike. For music retrieval systems, results are often required to serve a particular purpose, e.g., as background music that fits a specific activity such as dining or working out. Moreover, the purpose and usage of music are not limited to the auditory domain in the sense that the ways that personal music collections are presented often function as displays of personality and statements of distinction. In fact, a large portion of the aura that is surrounding collecting and structuring music stems from the rich and, technically speaking, multimodal context of music. This context spans from the aesthetics of the artwork to the type of packaging and included paraphernalia to liner notes to statements made by artists in accompanying media to gossip about the members of a band. In that sense, modern music information and retrieval systems must acknowledge the long tradition of collecting analog records to gain wide acceptance. This amplifies the requirement to provide highly context-aware and personalized retrieval and listening systems. However, not all of the aforementioned context can be digitized, and therefore preserved, using today's predominant means of media delivery. Then again, this loss of detail in context is exchanged for the sheer amount of instantly accessible content.

Applications such as *Shazam* (music identification), *Pandora* (automatic personalized radio stationing), *Spotify* (music streaming), or *Last.fm* (music recommendation, information system, and social network) today are considered essential services for many users. The high acceptance of and further demand for intelligent music applications also make music information retrieval as a research topic a particularly exciting field as findings from fundamental research can find their way into commercial applications immediately. In such a setting, where many innovative approaches and applications are entering the competition and more and more developers and researchers are attracted to this area, we believe that it is important to have a book that provides a comprehensive and understandable entry point into the topics of music search, retrieval, and recommendation from an academic perspective. This entry point should not only allow novices to quickly access the field of music information retrieval (MIR) from an information retrieval (IR) point of view but also raise awareness for the developments of the music domain within the greater IR community.

To this end, the book at hand gives a summary of the manifold audio- and web-based approaches and subfields of music information retrieval research for media consumption. In contrast to books that focus only on methods for acoustic signal analysis, this book is focused on music as a specific domain, addressing additional cultural aspects and giving a more holistic view. This includes methods operating on features extracted directly from the audio signal as well as methods operating on features extracted from contextual information, either the cultural context of music pieces as represented on the web or the user and usage context of music. With the latter, we account for the paradigm shift in music information retrieval that can be seen over the last decade, in which an increasing number of published approaches focus on the contextual feature categories, or at least combine "classical" signal-based techniques with data mined from web sources or the user's context.

We hope the reader will enjoy exploring our compilation and selection of topics and keep this compendium at hand for exciting projects that might even pave the way for "the next big music thing."

Vienna, Austria/Linz, Austria
January 2016

Peter Knees
Markus Schedl

Acknowledgments

> *No one tells the truth about writing a book. . . . The truth is, writing is this: hard and boring and occasionally great but usually not. —Amy Poehler*

Writing this book would not have been possible without the support of numerous people. We would like to express our gratitude to *Elias Pampalk*, *Tim Pohle*, *Klaus Seyerlehner*, and *Dominik Schnitzer* for sharing their audio feature extraction implementations and their support in preparing the examples in this book. For allowing us to include illustrations and screenshots, we are most grateful to *Emilia Gómez*, *Masataka Goto*, *Edith Law*, *Anita Shen Lillie*, *Irène Rotondi*, *Klaus Scherer*, *Mohamed Sordo*, and *Sebastian Stober*. We also thank *Dmitry Bogdanov* for his valuable input on hybrid recommender systems. Furthermore, we would like to thank *Gerhard Widmer* for his support and for granting us the time and freedom it takes to face the endeavor of writing a book.

We thank *Ralf Gerstner*, Executive Editor (Computer Science) for Springer, for his support, understanding, and patience. Also the constructive feedback provided by the anonymous reviewers was highly appreciated and helped a lot in sharpening the manuscript.

Finally, we appreciate the research funding organizations which provided gracious financial support, in particular, the Austrian Science Funds (FWF) and the European Commission.[1] By financing our work and projects, these organizations allowed us to gain the knowledge it takes to write such a book and provided a fruitful ground to contribute to and extend the state of the art in various tasks covered in the book.

[1] We received funding from the Austrian Science Funds (FWF) for the projects "Operational Models of Music Similarity for Music Information Retrieval" (L112-N04), "Music Retrieval Beyond Simple Audio Similarity" (L511-N15), "Personalized Music Retrieval via Music Content, Music Context, and User Context" (P22856), "Social Media Mining for Multimodal Music Retrieval" (P25655), and "Culture- and Location-aware Music Recommendation and Retrieval" (AJS3780). The European Commission supported our research within the projects "PHENICX—Performances as Highly Enriched aNd Interactive Concert eXperiences" (FP7-ICT-2011-9: 601166) and "GiantSteps—Seven League Boots for Music Creation and Performance" (FP7-ICT-2013-10: 610591).

Contents

Part III User-Centric MIR

Part IV Current and Future Applications of MIR

Chapter 1
Introduction to Music Similarity and Retrieval

Traditionally, electronically searching for music, whether in collections of thousands (private collections) or millions of tracks (digital music resellers), is basically a database lookup task based on meta-data. For indexing a collection, existing music retrieval systems make use of arbitrarily assigned and subjective meta-information like *genre* or *style* in combination with objective meta-data like *artist name*, *album name*, *track name*, *record label*, or *year of release*. On top of that, often, the hierarchical scheme *(genre–) artist–album–track* is then used to allow for browsing within the collection. While this may be sufficient for small private collections, in cases where most contained pieces are not known a priori, the unmanageable amount of pieces may easily overstrain the user and impede the discovery of desired music. Thus, a person searching for music, e.g., a potential customer, must already have a very precise conception of the expected result which makes retrieval of desired pieces from existing systems impractical and unintuitive.

Obviously, the intrinsic problem of these indexing approaches is the limitation to a rather small set of meta-data, whereas neither the musical content nor the cultural context of music pieces is captured. Archival and retrieval of music is historically a librarian's task, and structure and format of databases are optimized for access by experts. Today, the majority of users are not experts—neither in database search nor in terms of musical education. When searching for music, particularly when trying to discover new music, users rarely formulate their queries using bibliographic terms but rather describe properties like emotion or usage context [207]. Therefore, different search and retrieval scenarios become more important.

Retrieval systems that neglect musical, cultural, and personal aspects are far away from the manifold ways that people organize, deal with, and interact with music collections—or expressed in information retrieval (IR) terms, these system neglect their users' *music information needs* [90, 258]. For music, information needs can be quite distinct from standard text-related information needs. Music as a media is heavily intertwined with pop culture as well as with hedonistic and recreational activities. The need to find music might not be as much one that is targeted at

P. Knees, M. Schedl, *Music Similarity and Retrieval*, The Information Retrieval Series 36, DOI 10.1007/978-3-662-49722-7_1

information but merely one targeted at pure entertainment. Thus, one could argue that for most popular and mainstream music, the average user accessing a music information system has primarily an *entertainment need* (cf. [34, 325]).

1.1 Music Information Retrieval

As a response to the challenges, specifics, and needs of retrieval in the music domain, the area of research known as *music information retrieval* (MIR) has evolved in the 1990s and emancipated itself as a dedicated field at the beginning of the millennium with the organization of the ISMIR[1] conference series [54]. Among others, MIR is researching and developing intelligent methods that aim at extracting musically meaningful descriptors either directly from the audio signal or from contextual sources. These descriptors can then be used, e.g., to build improved interfaces to music collections. In this section, we give an overview of the field of MIR. We start by looking into definitions found in the literature and proceed by describing the predominant retrieval paradigms found in MIR, illustrated by exemplary tasks and applications. We round this overview up by pointing to research areas of MIR that go beyond traditional IR tasks.

In the literature, one can find several definitions of MIR—each focusing on specific aspects. We give a selection of these in order to sketch the bigger picture. In an early definition, Futrelle and Downie emphasize the multi- and interdisciplinarity of MIR and its origins in digital library research:

> MIR is a(n) ... interdisciplinary research area encompassing computer science and information retrieval, musicology and music theory, audio engineering and digital signal processing, cognitive science, library science, publishing, and law. Its agenda, roughly, is to develop ways of managing collections of musical material for preservation, access, research, and other uses. [138]

Later on, Downie highlights research on content analysis, i.e., the automatic extraction of music descriptors from the audio signal itself, interfaces, and infrastructure:

> MIR is a multidisciplinary research endeavor that strives to develop innovative content-based searching schemes, novel interfaces, and evolving networked delivery mechanisms in an effort to make the world's vast store of music accessible to all. [110]

Finally, in our own definition, we highlight the multimodality of the field:

> MIR is concerned with the extraction, analysis, and usage of information about any kind of music entity (e.g., a song or a music artist) on any representation level (for example, audio signal, symbolic MIDI representation of a piece of music, or name of a music artist). [401]

[1] Previously: *International Symposium for Music Information Retrieval* and *International Conference on Music Information Retrieval*; since 2009: *International Society for Music Information Retrieval Conference*.

The diversity of these different definitions demonstrates the width of this field. In this book, we cannot review every aspect of music information retrieval. We focus instead on the very central part of music search and retrieval and the notion of music similarity—the underlying concept that is essential to all of the presented methods. To this end, we continue the introduction of MIR with a discussion of music retrieval tasks from an information retrieval perspective.

1.2 MIR from an Information Retrieval Perspective

Similar to other domains, we can identify three main paradigms of music information access:

Retrieval The user has a specific music information need, e.g., finding a specific musical item, and actively expresses this need using a query. This query can be represented in different modalities such as text, a symbolic music representation, or as a piece of audio. The result can be (specific parts of) audio pieces, scores, or meta-data, potentially in a ranked list.

Browsing The user has an undirected information need and wants to explore the available music collection. Browsing is an interactive and iterative task that relies on intuitive and effective user interfaces for discovering items. Like search, browsing is initiated and actively pursued by the user.

Recommendation The system filters the collection for potentially relevant items based on the user's actions or preferences. These preferences can be given explicitly by the user or derived from implicit feedback, e.g., by observing actions during retrieval and/or browsing or by tapping listening data. The user is not required to actively search for music and is presented with a personalized view.

Figure 1.1 gives a schematic overview of these access paradigms and how they connect to data structures. As can already be seen, features and similarity take a central role in this layout. To facilitate music information retrieval, naturally, a lot of research is targeted at tasks that relate to music signal processing, feature extraction, and similarity measurement, e.g., to build systems for content-based querying and retrieval; cf. Part I. As we will see, particularly in existing browsing interfaces, timbre plays an important role as a descriptor for sound similarity. Chapter 3 will therefore prominently deal with timbre-related features (as well as touching upon aspects of rhythmicity and tonality). Other musical dimensions related to pitch (key, chords, harmonies, melodies), temporal facets (rhythm, tempo), and structure, as well as retrieval tasks specific to these properties, are outside the scope of this book. For computational methods to derive this kind of information and use it for retrieval, we refer the reader to other sources (see the *further reading* Sect. 1.6). We discuss the role of features and the multifaceted notion of music similarity in Sect. 1.3. Before this, we elaborate on the three music information access paradigms *retrieval*, *browsing*, and *recommendation* by describing typical tasks in MIR and pointing to

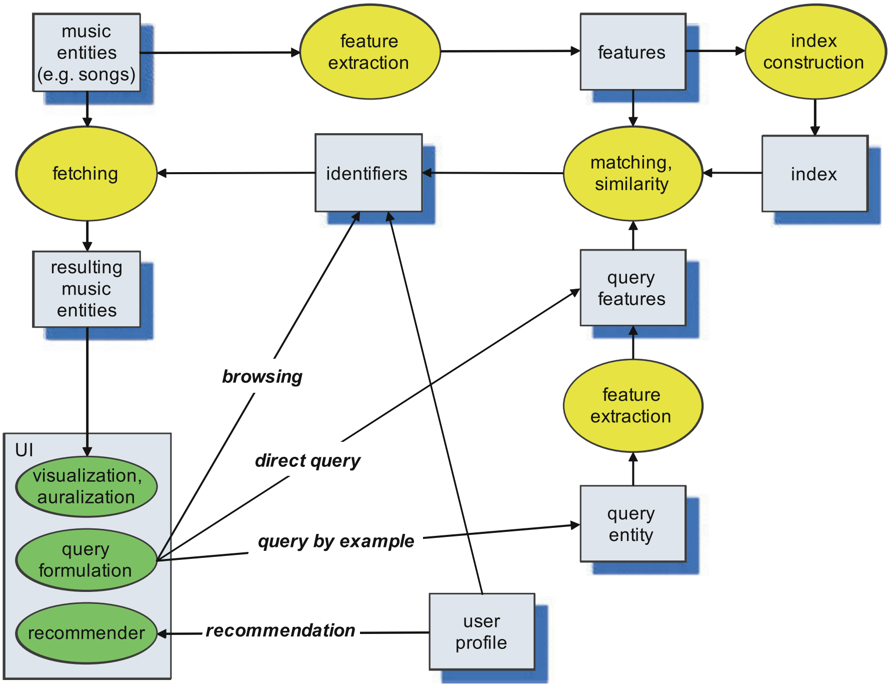

Fig. 1.1 Traditional information retrieval view on music retrieval with a focus on content-based document indexing

exemplary applications and research prototypes. Furthermore, we highlight areas of MIR beyond these three paradigms.

1.2.1 Retrieval Tasks and Applications in MIR

A retrieval approach that directly corresponds to text IR methods is searching for symbolic music using a query consisting of a symbolic representation of the same type, thus following the *query by example* retrieval scheme. The term “symbolic music” denotes digital data representations of music notation. An example of such a data representation would be the MIDI format.[2] For matching, the relevant part of the encoded information typically relates to pitch and timing information.

The *Themefinder* web search engine,[3] for instance, allows for querying a symbolic database by entering melodic sequences in proprietary text formats [240].

[2] Abbreviation for *musical instrument digital interface*.

[3] http://www.themefinder.org.

Other systems follow more intuitive approaches and are therefore also usable for less musically educated users. For example, in **query by singing/humming (QBSH)** systems, the user can hum or sing a part of the searched piece into a microphone. From that recording, musical parameters (mostly related to melody) are extracted, and the obtained sequence serves as a query to the database; cf. [143]. An example of a search engine offering exhaustive possibilities for querying is *Musipedia*.[4] *Musipedia* indexes a large number of music pieces by crawling the web for MIDI files that can then be used for identification of pieces. For indexing of pieces, the melodic contour, pitches and onset times, and a rhythm representation are extracted. To find a piece in the database, a theme (i.e., the query) can be either entered in Parsons code notation [353] or whistled into a microphone (to find matching melodies), played on a virtual piano keyboard (to find matching pitch and onset sequences), or tapped on the computer keyboard (to find matching rhythms). For a detailed explanation of the incorporated techniques, as well as a comprehensive comparison of symbolic music retrieval systems and MIR systems in general, we refer the reader to [485]. Please note that symbolic music retrieval is not within the scope of this book. Part I of this book addresses query by example systems that make use of audio similarity approaches to find the most similar recordings in a collection for a given track.

While the abovementioned systems aim at retrieving a ranked list of documents similar to the query, for the task of **audio identification** the goal is to find, i.e., to identify, the query within a large database of recordings. The query is typically a short snippet of a song, possibly recorded in low quality and in an environment with background noise, e.g., using a cellular phone. The expected result is the meta-data of the entry in the database, such as artist, album, and song name. The underlying technique is known as music fingerprinting. To this end, for each song in a music collection, a compact unique feature representation is created (the so-called fingerprint) which can be matched against the fingerprint of the query. A requirement is that fingerprints must be robust against all kinds of distortions, e.g., caused by factors such as cheap microphones or cellular phone connections; cf. [505]. Fingerprinting can be used to detect copyright infringements or in music identification services, with the most popular commercial example being *Shazam*.[5] Other examples of services that provide audio identification are *SoundHound*,[6] which also provides methods for QBSH, *Gracenote MusicID*,[7] *MusicBrainz Fingerprinting*,[8] and *Echoprint*.[9]

[4] http://www.musipedia.org.

[5] http://www.shazam.com.

[6] http://www.soundhound.com.

[7] http://www.gracenote.com/music/recognition.

[8] http://musicbrainz.org/doc/Fingerprinting.

[9] http://echoprint.me.

Conceptually related to both symbolic retrieval and audio identification is **cover song identification** or version identification. Here the goal is to find different renditions and stylistic interpretations of the query song. State-of-the-art algorithms for this task extract descriptors relating to melody, bass line, and harmonic progression to measure the similarity between two songs [391].

Another retrieval scenario is **text-based retrieval of audio and music** from the web. Some search engines that use specialized (focused) crawlers to find all types of sounds on the web exist. As with web image search, the traced audio files are then indexed using contextual information extracted from the text surrounding the links to the files. Examples of such search engines are *Aroooga* [229] and *FindSounds*.[10] Other approaches utilize text information from the web to index arbitrary music pieces. Hence, a textual context has to be constructed artificially by finding web pages that mention the meta-data of tracks. We discuss such approaches in Part II.

1.2.2 Browsing Interfaces in MIR

Next to hierarchical text-based information systems for browsing of music collections is an important access modality in MIR. Such interfaces should offer an intuitive way to sift through a music collection and to encounter serendipitous music experiences. We discuss intelligent music interfaces in detail in Sect. 9.2; however, here we want to point out some exemplary innovative interfaces that support the user in discovering music using MIR technology.

Figure 1.2 shows the *Intelligent iPod* interface that aims at providing "one-touch access" to music on mobile devices [429]. Just by using the scroll wheel of a classic *iPod*, the user can browse through the whole collection that is organized on a circular playlist according to acoustic similarity, i.e., neighboring songs are expected to sound similar and, overall, there should be smooth transitions between the different styles. Additionally, regions on the playlist are described using collaborative tags for easier navigation. After selecting a region, the same style of music continues playing. The combination of automatic, intelligent organization and the simple hardware interface resembles operating a radio dial that can be tuned to find desired music.

A combination of audio-based structuring and tag-based navigation support is also given in the *nepTune* interface [222]. Similar-sounding music is clustered and a virtual landscape is created to visualize the structure. This landscape can then

[10] http://www.findsounds.com.

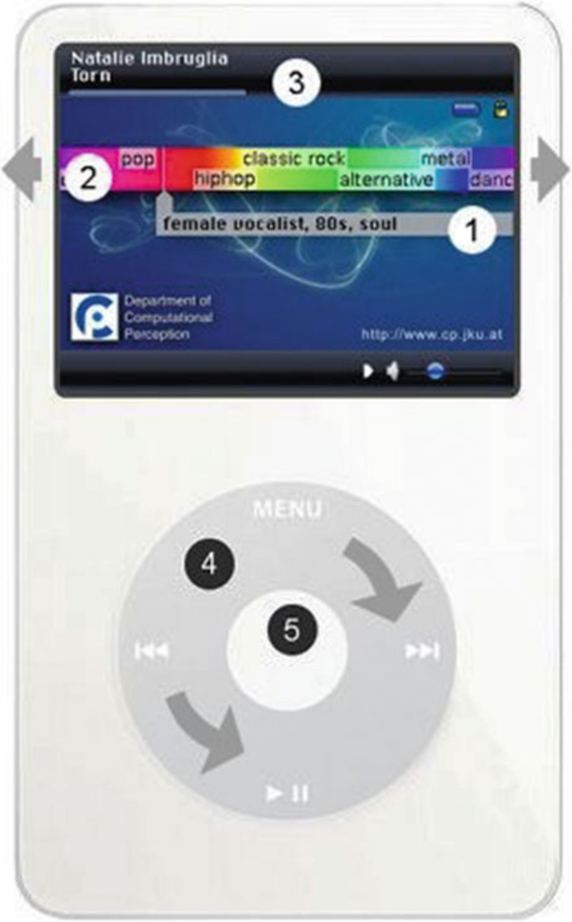

Fig. 1.2 The *Intelligent iPod* mobile browsing interface. (1) shows tags describing the music in the selected region. (2) represents the music collection as a stripe, where different styles are colored differently. The collection can be browsed by using the scroll wheel (4). The Scroll wheel can be used to browse the collection. (5), The central button to select the track. The currently playing track is shown in (3)

be navigated in the fashion of a computer game with the closest tracks auralized; cf. Fig. 1.3. Terms that describe the contents of the clusters can be displayed in order to facilitate orientation.

The final example is *Musicream*,[11] an interface that fosters unexpected, serendipitous music discoveries [153]. *Musicream* uses the metaphor of water taps that release flows of songs when opened; cf. Fig. 1.4. Different taps release music of different styles. The user can browse through the collection by grabbing songs and listen to them or create playlists by sticking songs together. To achieve consistent playlists, similar-sounding songs are easier to connect than dissimilar sounding.

[11] http://staff.aist.go.jp/m.goto/Musicream.

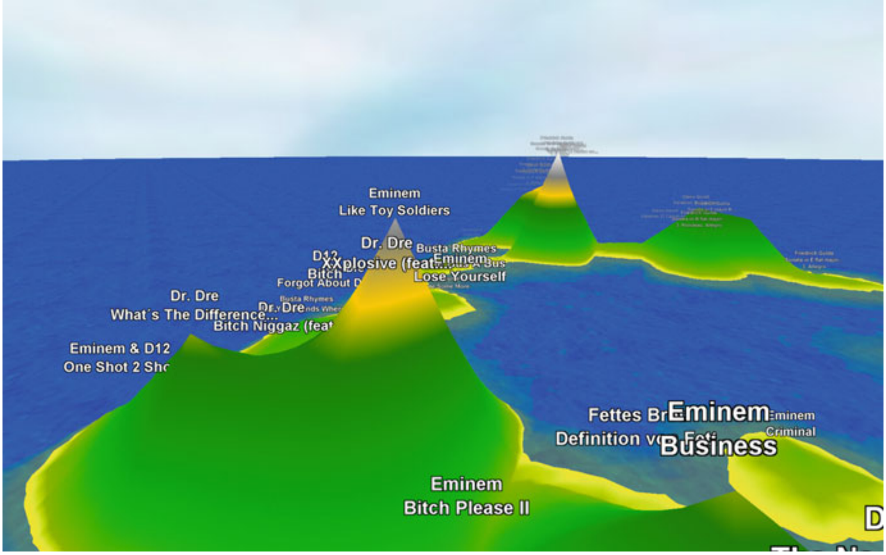

Fig. 1.3 The *nepTune* browsing interface

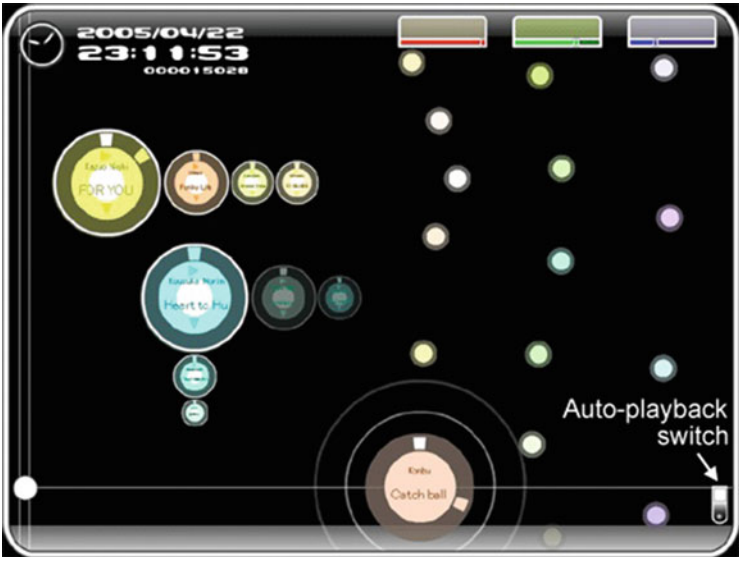

Fig. 1.4 The *Musicream* browsing interface [153] (reprinted with permission)

1.2.3 Recommendation Tasks and Applications in MIR

With virtually the complete catalog of music that has ever been commercially produced ubiquitously available and instantly streamable from the cloud, methods to provide the listener with the "right music at the right time" without requiring much (or any) interaction have become very relevant. One recommendation task that is very particular to the music domain is the **automatic generation of personalized music playlists**, i.e., recommending a sequence of songs "that is pleasing as a whole" [172]. This topic will be addressed in Sect. 9.3 of this book. Starting with a query (which can be a seed song or artist, a user profile, or the user's current context), the aim is to continuously play music that the user wants to listen to, sometimes also referred to as creating "personalized radio stations." Even though subsequent songs should sound similar, an important requirement is that the generated playlists are not boring. Hence, it is important to consider the trade-off between similarity and diversity [418, 536].

Automated playlist generation has received growing attention for about a decade and is now a standard feature of major online music retailers. Examples are *Pandora*,[12] *Last.fm Player*,[13] *Spotify Radio*,[14] *iTunes Radio*,[15] *Google Play Access All Areas*,[16] and *Xbox Music*.[17] Recommendations are typically made using (undisclosed) content-based retrieval techniques, collaborative filtering data, or a combination thereof. Moreover, most systems account for explicit user feedback on their playlists given as binary ratings to improve the personalization of recommendations.

An example of a truly content-based recommender is the *FM4 Soundpark* music player[18] that suggests other songs purely based on sound similarity [140]. The *FM4 Soundpark* is a moderated open platform for up-and-coming artists hosted by the Austrian public broadcasting station *FM4* and targets primarily alternative music. In this case, where artists are generally unknown, content-based similarity is the method of choice for recommendation. The system also offers to automatically create a "mix tape" based on a start and an end song [128]. The underlying technology makes use of content-based retrieval methods like those described in Part I.

[12] http://www.pandora.com.

[13] http://www.last.fm/listen.

[14] http://www.spotify.com.

[15] http://www.apple.com/itunes/itunes-radio/.

[16] http://play.google.com/about/music/.

[17] http://www.xbox.com/music.

[18] http://fm4.orf.at/soundpark.