Music Retrieval and Recommendation

A Tutorial Overview

Peter Knees and Markus Schedl
Department of Computational Perception
Johannes Kepler University Linz
Altenberger Str. 69, 4040 Linz, Austria
peter.knees@jku.at, markus.schedl@jku.at

ABSTRACT

In this tutorial, we give an introduction to the field of and state of the art in music information retrieval (MIR). The tutorial particularly spotlights the question of music similarity, which is an essential aspect in music retrieval and recommendation. Three factors play a central role in MIR research: (1) the music content, i.e., the audio signal itself, (2) the music context, i.e., metadata in the widest sense, and (3) the listeners and their contexts, manifested in user-music interaction traces. We review approaches that extract features from all three data sources and combinations thereof and show how these features can be used for (large-scale) music indexing, music description, music similarity measurement, and recommendation. These methods are further showcased in a number of popular music applications, such as automatic playlist generation and personalized radio stationing, location-aware music recommendation, music search engines, and intelligent browsing interfaces. Additionally, related topics such as music identification, automatic music accompaniment and score following, and search and retrieval in the music production domain are discussed.

Categories and Subject Descriptors

H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing—*Methodologies and techniques*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

Keywords

Music Information Retrieval; Music Recommendation; Content; Context; Listener

1. INTRODUCTION

As the amount of music available via streaming services, online stores, platforms like YouTube, and other web sources has skyrocketed over the last couple of years. Retrieving relevant music that matches the user's taste is a challenging, albeit important task to make accessible the ever-growing digital music repositories in an intelligent manner. Given the current rise of social media and user-generated contents, retrieving information about music as well as retrieving music itself heavily relies on text-based IR techniques, as text is still the widest used means of communication on the web. On the other hand, multimodal retrieval schemes for multimedia content demand for acoustic features and make hybrid (signal- and text-based) approaches attractive.

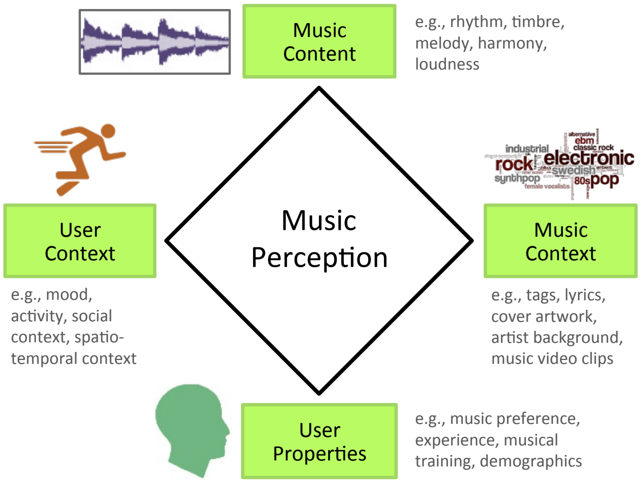
Music Content
e.g., rhythm, timbre, melody, harmony, loudness
User Context
e.g., mood, activity, social context, spatio-temporal context
Music Perception
industrial
ebm
classic rock
rock
electronic
swedish
synthpop
80s pop
Music Context
e.g., tags, lyrics, cover artwork, artist background, music video clips
User Properties
e.g., music preference, experience, musical training, demographics

Figure 1: Four different categories of factors that influence music perception.

Music information retrieval (MIR) is a research field that aims – among other things – at automatically extracting semantically meaningful information from various representations of music entities, such as a digital audio file, a band's web page, a song's lyrics, or a tweet about a microblogger's current listening activity.

A key approach in MIR is to describe music via computational features, which can be broadly categorized into *music content*, *music context*, *user properties*, and *user context*, cf. Figure 1. While music content-based features are derived directly from the audio signal of the music file, music context refers to pieces of information that are not encoded in the actual audio file, nevertheless play an important role in human perception of music. Such aspects include the meaning of song lyrics, the background of an artist, the cover of an album, the sequence of songs selected by a DJ to constitute a playlist, or collaborative tags describing a release. Extract-

SIGIR'15, August 09-13, 2015, Santiago, Chile.
ACM 978-1-4503-3621-5/15/08.
DOI: http://dx.doi.org/10.1145/2766462.2767880.

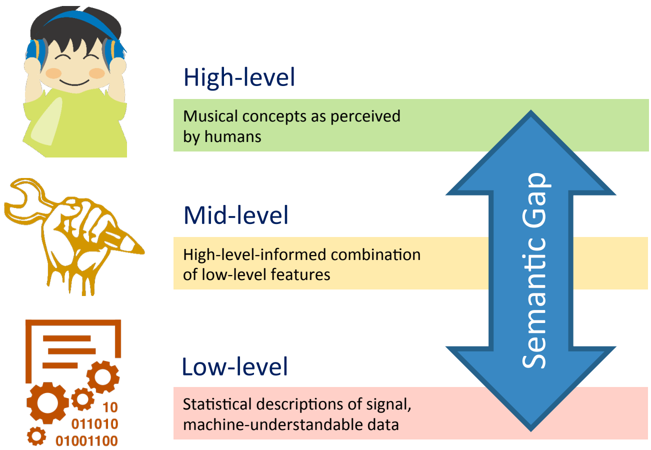
High-level
Musical concepts as perceived by humans
Mid-level
High-level-informed combination of low-level features
Low-level
Statistical descriptions of signal, machine-understandable data
Semantic Gap
10
011010
01001100

Figure 2: The different levels of feature abstraction and the "semantic gap" between them.

ing music content features requires access to the actual audio file; in contrast, contextual feature extractors require as input only editorial metadata (e.g., name of artist and song) to harvest music-related information from the web and consequently approximate similarities between music items. On the other hand, music content features are in general more objective than music context features as the underlying data source, i.e., the audio itself, does not change dynamically, in contrast to user-generated content or other kinds of contextual data sources. Both types of features (content and context), however, share a susceptibility to noise of different kinds. User properties refer to the listener's demographic information, musical education and experience, preferences and taste, as well as personality traits. The user context includes environmental aspects as well as physical and mental activities of the music listener. Particularly aspects of the latter two categories are often difficult to derive. Methods aiming at modeling these dimensions thus typically exploit traces of user-music interaction, such as listening histories or geo-location data, in order to learn a representation of (possibly latent) user state and context.

Depending on the chosen source, features might be closer to a concept as humans understand it (*high-level*) or closer to a strictly machine-interpretable representation (*low-level*). For instance, features derived from the music content or the sensor data captured with a user's personal device will be mostly statistical descriptions of signals that are difficult to interpret, i.e., low-level features, whereas information extracted from web sources, such as occurrences of words in web page texts, is typically easier to interpret. For all types of features in order to be processable, we need a numeric abstraction. Between these low-level numeric representations and the factors of musical perception that we want to model there is a discrepancy – commonly known as "semantic gap," cf. Figure 2. Particularly, the hybrid and user-centric techniques discussed strive to narrow this gap.

2. CONTENTS

MIR research has been seeing a paradigm shift over the last couple of years, as an increasing number of recently published approaches focus on the contextual feature categories, or at least combine audio-based techniques with data mined from web sources or the user's signals. This is reflected in the structure of the tutorial.

First, we introduce selected existing applications that rely on MIR technology to motivate the presented contents and relate them to real-world scenarios and applications, such as automated music playlist generation, personalized web radio, music recommendation systems, and intelligent user interfaces to music. Then, we summarize the ideas behind and discuss advantages and disadvantages of computational features extracted from music content and music context, as well as user-centric information. Each of these discussions is substantiated in a dedicated segment.

Regarding **music content** analysis, we give a brief introduction to signal processing methods (PCM, A/D Conversion, FFT, DCT, etc.) to lay the foundation for elaborated methods of music processing (e.g., [9]). We review some standard approaches to audio feature extraction on frame and block level as well as state-of-the-art similarity measures using features such as MFCCs [17], block-level features[29], and pitch class profiles [20]. We also briefly address features for related MIR tasks such as beat detection [4], melody extraction [21], or score following [1]. In addition, aspects of large-scale indexing [27] and the problem of hubness for retrieval in high-dimensional feature spaces are addressed [28]. Further attention is given to evaluating MIR systems beyond the traditional IR-related measures and the difficulties entailed by the need for objective quantification, e.g., [5, 31].

As for aspects of the **music context**, we focus on data accessible through web technology. To this end, we introduce the field of web-based MIR and give a detailed description and comparison of contextual data sources on music (e.g., web pages and blogs [10], micro-blogs [23], user tags [16, 14], and lyrics [18]) and discuss related methods to obtain this data (web mining, games with a purpose [15], etc.). These sources can be exploited in order to

- *mine descriptive and relational metadata* (e.g., band members and instrumentation, country, album covers, genres, related artists, e.g., [22, 12]), to

- *construct similarity measures* for music artists and songs based on collaborative and cultural knowledge (e.g., [10, 32], and to

- *automatically index and retrieve* music [11, 2].

Regarding the **user-centric aspects** (user properties and context) and their applications in music recommendation and other personalized systems, we discuss sources of music interaction traces (e.g., playlists [19], ratings [6], postings and micro-blogs [24], peer-to-peer networks [13, 30], and social networks [7]) and possibilities to mine the context directly from sensor data using smart devices [8]. Methods that use this data can then be applied for tasks such as playlist generation, tag prediction, and location-aware music recommendation. We further address methods that include information from both context data and content information, either by learning hybrid similarity measures or by optimizing audio-based or hybrid similarity functions in order to reflect preference of users [26]. Additionally, user requirements such as need for novelty, diversity, or serendipity are addressed [3, 33]. This last segment concludes with an outlook to the next years of MIR and the biggest challenges the field is facing.

The following outlines the structure of the tutorial:

1. **Introduction to Music Similarity and Retrieval**
   (a) The Information Retrieval Perspective
   (b) Factors of Music Similarity
   (c) Applications: Playlist Generation, User Interfaces, etc.
2. **Content-Based MIR**
   (a) Basic Methods of Audio Signal Processing
   (b) Audio Feature Extraction for Similarity Measurement
   (c) Music Understanding and Semantic Description
   (d) Evaluation of Music Similarity Algorithms
3. **Contextual Music Similarity, Indexing, and Retrieval**
   (a) Contextual Music Meta-Data: Comparison and Sources
   (b) Text-Based Features and Similarity Measures
   (c) Text-Based Indexing and Retrieval
4. **Collaborative Music Similarity and Recommendation**
   (a) Listener-centered Data Sources: Traces of Music Interaction
   (b) Collaborative Music Similarity and Recommendation
   (c) User-Awareness
   (d) Multi-Modal Combination
5. **Grand Challenges and Outlook**

3. BIOGRAPHIES OF THE PRESENTERS

Dr. Peter Knees is an assistant professor at the *Department of Computational Perception* at the *Johannes Kepler University Linz*. He holds a Master's degree in Computer Science from the *Vienna University of Technology* and a Ph.D. degree from the *Johannes Kepler University Linz*.

Since 2004, he co-authored over 60 peer-reviewed conference and journal publications, served as program committee member for several conferences relevant to the fields of music, multimedia, and text IR, including ISMIR, ACM Multimedia, ECIR Tutorials, and the Adaptive Multimedia Retrieval workshop and was an organizer of the *International Workshop on Advances in Music Information Research* series and the SIGIR 2014 *Workshop on Social Media Retrieval and Analysis*. He is teaching grad-level courses on *Multimedia Search and Retrieval*, *Learning from User-generated Data*, *Multimedia Data Mining*, and *Intelligent Information Systems* and has given tutorials and lectures on music IR at ECIR, SIGIR, and RuSSIR. In addition to music and web information retrieval, his research interests include multimedia systems, user interfaces, and recommender systems.

Dr. Markus Schedl is an associate professor at the *Johannes Kepler University Linz / Department of Computational Perception*. He graduated in Computer Science from the *Vienna University of Technology* and earned his Ph.D. in Computer Science from the *Johannes Kepler University Linz*. Markus further holds a Master's degree in International Business Administration from the *Vienna University of Economics and Business*. He (co-)authored more than 100 refereed conference/workshop papers and journal articles (published, among others, in SIGIR, ECIR, ACM Multimedia; Journal of Machine Learning Research, ACM Transactions on Information Systems, Springer Information Retrieval, IEEE Multimedia). He is an associate editor of the Springer International Journal of Multimedia Information Retrieval, serves on various program committees, and as reviewer (among others, for ACM Multimedia, ECIR, IJCAI, ICASSP, IEEE Visualization; Transactions of Multimedia, Transactions on Intelligent Systems and Technology, Information Sciences, Pattern Recognition Letters). He is co-founder of the *International Workshop on Advances in Music Information Research* (AdMIRe) and the *International Workshop on Social Media Retrieval and Analysis* (SoMeRA), the latter held in conjunction with SIGIR 2014. He recently co-authored an article titled "Music Information Retrieval: Recent Developments and Applications" in *Foundations and Trends in Information Retrieval* [25].

4. ACKNOWLEDGMENTS

This tutorial is supported by the European Union Seventh Framework Programme FP7/2007-2013 through grant agreements no. 601166 (PHENICX) and no. 610591 (GiantSteps), as well as by the Austrian Science Funds (FWF): P25655.

5. REFERENCES

[1] A. Arzt and G. Widmer. Towards effective 'any-time' music tracking. In *Proceedings of the Starting AI Researchers' Symposium (STAIRS)*, Lisbon, Portugal, 2010.

[2] L. Barrington, D. Turnbull, M. Yazdani, and G. Lanckriet. Combining audio content and social context for semantic music discovery. In *Proceedings of the ACM SIGIR*, Boston, MA, USA, 2009.

[3] O. Celma and P. Herrera. A new approach to evaluating novel recommendations. In *Proceedings of the ACM Conference on Recommender Systems (RecSys)*, New York, NY, USA, 2008.

[4] S. Dixon, F. Gouyon, and G. Widmer. Towards Characterisation of Music via Rhythmic Patterns. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, Barcelona, Spain, 2004.

[5] J. S. Downie, J. Futrelle, and D. Tcheng. The international music information retrieval systems evaluation laboratory: Governance, access and security. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, Barcelona, Spain, 2004.

[6] G. Dror, N. Koenigstein, and Y. Koren. Yahoo! music recommendations: Modeling music ratings with temporal dynamics and item taxonomy. In *Proceedings*

of the ACM Conference on Recommender Systems (RecSys), Chicago, IL, USA, 2011.

[7] B. Fields, M. Casey, K. Jacobson, and M. Sandler. Do You Sound Like Your Friends? Exploring Artist Similarity via Artist Social Network Relationships and Audio Signal Processing. In *Proceedings of the International Computer Music Conference (ICMC)*, Belfast, UK, 2008.

[8] M. Gillhofer and M. Schedl. Iron Maiden while jogging, Debussy for dinner? - An analysis of music listening behavior in context. In *Proceedings of the International Conference on MultiMedia Modeling (MMM)*, Sydney, Australia, 2015.

[9] B. Gold and N. Morgan. *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*. Wiley, 1999.

[10] P. Knees, E. Pampalk, and G. Widmer. Artist Classification with Web-based Data. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, Barcelona, Spain, 2004.

[11] P. Knees, T. Pohle, M. Schedl, and G. Widmer. A Music Search Engine Built upon Audio-based and Web-based Similarity Measures. In *Proceedings of the ACM SIGIR*, Amsterdam, Netherlands, 2007.

[12] P. Knees and M. Schedl. Towards Semantic Music Information Extraction from the Web Using Rule Patterns and Supervised Learning. In *Proceedings of the Workshop on Music Recommendation and Discovery (WOMRAD)*, Chicago, IL, USA, 2011.

[13] N. Koenigstein, Y. Shavitt, and T. Tankel. Spotting Out Emerging Artists Using Geo-Aware Analysis of P2P Query Strings. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, Las Vegas, NV, USA, 2008.

[14] P. Lamere. Social Tagging and Music Information Retrieval. *Journal of New Music Research: Special Issue: From Genres to Tags – Music Information Retrieval in the Age of Social Tagging*, 37(2):101–114, 2008.

[15] E. Law, L. von Ahn, R. Dannenberg, and M. Crawford. Tagatune: A Game for Music and Sound Annotation. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, 2007.

[16] M. Levy and M. Sandler. A semantic space for music derived from social tags. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, 2007.

[17] B. Logan. Mel Frequency Cepstral Coefficients for Music Modeling. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, Plymouth, MA, USA, 2000.

[18] R. Mayer, R. Neumayer, and A. Rauber. Rhyme and Style Features for Musical Genre Classification by Song Lyrics. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Philadelphia, PA, USA, 2008.

[19] B. McFee and G. Lanckriet. Hypergraph Models of Playlist Dialects. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Porto, Portugal, 2012.

[20] H. Purwins. *Profiles of Pitch Classes: Circularity of Relative Pitch and Key—Experiments, Models, Computational Music Analysis, and Perspectives*. PhD thesis, Technische Universität Berlin, Germany, 2005.

[21] J. Salamon and E. Gómez. Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech and Language Processing*, 20:1759–1770, 08/2012.

[22] M. Schedl. *Automatically Extracting, Analyzing, and Visualizing Information on Music Artists from the World Wide Web*. PhD thesis, Johannes Kepler University Linz, Austria, 2008.

[23] M. Schedl. #nowplaying Madonna: A Large-Scale Evaluation on Estimating Similarities Between Music Artists and Between Movies from Microblogs. *Information Retrieval*, 15:183–217, June 2012.

[24] M. Schedl. Leveraging Microblogs for Spatiotemporal Music Information Retrieval. In *Proceedings of the European Conference on Information Retrieval (ECIR)*, Moscow, Russia, 2013.

[25] M. Schedl, E. Gómez, and J. Urbano. Evaluation in music information retrieval. *Foundations and Trends in Information Retrieval*, 8(2–3):127–261, 2014.

[26] M. Schedl, S. Stober, E. Gómez, N. Orio, and C. C. Liem. User-Aware Music Retrieval. In M. Müller, M. Goto, and M. Schedl, editors, *Multimodal Music Processing*, volume 3 of *Dagstuhl Follow-Ups*, pages 135–156. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2012.

[27] D. Schnitzer. *Indexing Content-Based Music Similarity Models for Fast Retrieval in Massive Databases*. PhD thesis, Johannes Kepler University Linz, Austria, 2011.

[28] D. Schnitzer, A. Flexer, M. Schedl, and G. Widmer. Local and global scaling reduce hubs in space. *Journal of Machine Learning Research*, 13:2871–2902, October 2012.

[29] K. Seyerlehner, G. Widmer, M. Schedl, and P. Knees. Automatic Music Tag Classification based on Block-Level Features. In *Proceedings of the Sound and Music Computing Conference (SMC)*, Barcelona, Spain, 2010.

[30] Y. Shavitt, E. Weinsberg, and U. Weinsberg. Mining Music from Large-Scale, Peer-to-Peer Networks. *IEEE Multimedia*, 18(1):14–23, January 2011.

[31] J. Urbano and M. Schedl. Minimal test collections for low-cost evaluation of audio music similarity and retrieval systems. *International Journal of Multimedia Information Retrieval: Special Issue on Hybrid Music Information Retrieval*, 2(1):59–70, January 2013.

[32] B. Whitman and S. Lawrence. Inferring Descriptions and Similarity for Music from Community Metadata. In *Proceedings of the International Computer Music Conference (ICMC)*, Göteborg, Sweden, 2002.

[33] Yuan Cao Zhang, Diarmuid O Seaghdha, Daniele Quercia, Tamas Jambor. Auralist: Introducing Serendipity into Music Recommendation. In *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM)*, Seattle, WA, USA, 2012.