Music Retrieval and Recommendation

Proposal for a full-day tutorial to be held at SIGIR 2015

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

1. FORMAT AND TARGET AUDIENCE

We propose a **full-day tutorial** (6 hours plus breaks) to give an **introductory course** to the field of and state of the art in music information retrieval (MIR). The turorial will particularly spotlight 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. In the tutorial, we will review approaches that extract features from all three data sources and combinations thereof. We will show how these features can be used for (large-scale) music indexing, music description, music similarity measurement, and recommendation. These methods are then 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.

The fact that music is omnipresent in today's society and heavily reflected on the web and in social media demands for a serious consideration of MIR also by experts from related disciplines. We hence expect the **target audience** to consist of two groups. First, both people from academia and industry engaging in music and **multimedia information retrieval**. Second, researchers from the fields of **web mining and text-based IR** who will be given the opportunity to enhance their expertise towards the innovative and uprising area of web- and community-based music information retrieval.

Since the tutorial will introduce foundations and basic concepts of music information retrieval and music-related web- and social media mining, it will also be accessible to individuals not familiar with the field. We hence will **not rely on any particular prior knowledge**.

SIGIR 2015 Santiago, Chile

2. 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 Adaptive Multimedia Retrieval 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.

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 90 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* [12].

3. EXTENDED ABSTRACT

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 three classes: *music content*, *music context*, and *user context*. 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. Extracting 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. The user context, in contrast to the other two, includes environmental aspects as well as physical and mental activities of the music listener and is very difficult to derive. Methods aiming at modeling this dimension thus typically exploit traces of user-music interaction, such as listening histories or geo-location data, in order to derive a representation of (possibly latent) user context.

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 "classical" signal-based techniques with data mined from web sources or the user's context. For this reason, we believe that MIR has matured to be considered interesting by the general IR community.

Regarding the structure of the tutorial, we will follow the structure of our book "Music Similarity and Retrieval: Audio- and Web-based Strategies" soon to be published in the Springer Information Retrieval series. We will **start the tutorial with an introduction** to the field of music information retrieval and give an overview of different types of music retrieval systems and typical MIR tasks. Thus we will present 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. To outline the following three sessions of the tutorial, we will summarize the ideas behind the three categories of computational features (content, music-context, and user-context) and discuss advantages and disadvantages of each before the first break.

The **second segment (after the first break)** will start with 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., [26]). We will subsequently review some standard approaches to audio feature extraction on frame and block level as well as state-of-the-art similarity measures. To this end, we present MFCCs [31], the block-level-framework [16], and pitch class profiles [35], as well as briefly addressing features for related MIR tasks such as beat detection [22], melody extraction [36], or score following [18]. Furthermore, aspects of large-scale indexing [37] and the problem of hubness in high-dimensional features spaces will be tackled [15]. Another aspect will be on evaluating MIR systems beyond the traditional IR-related measures and the difficulties entailed by the need for objective quantification, e.g., [23, 17].

In the **third part (after lunch break)**, we will focus on the contextual aspects of music which are accessible through web technology. To this end, we will give an introduction to the field of web-based MIR and a detailed description and comparison of contextual data sources on music (e.g., web pages [2], blogs [10], micro-blogs [9, 11], user tags [30, 28], and lyrics [32]) and discuss related methods to obtain this data (web mining, games with a purpose [29], etc.). Then we will present approaches to exploit these sources to

- *mine descriptive and relational metadata* (e.g., band members and instrumentation, country, album covers, genres, related artists, e.g., [7, 5]), to
- *construct similarity measures* for music artists and songs based on collaborative and cultural knowledge (e.g., [2, 39], and to
- *automatically index and retrieve* music [4, 3, 20].

The **fourth session** will deal with the user context, music recommendation, and personalization strategies. This includes a discussion on sources of music interaction traces (e.g., playlists [19, 33], ratings [24], postings and micro-blogs [11], peer-to-peer networks [27, 38], and social networks [25, 34]) and possibilities to mine the context directly from sensor data using smart devices [1]. Methods that use this data are then applied for tasks such as playlist generation, tag prediction, and (location-aware) music recommendation. We will further present 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 [14]. Additionally, user requirements such as need for novelty, diversity, or serendipity are addressed [21, 40]. We conclude the tutorial by giving an outlook to the next years of MIR and the biggest challenges the field is facing.

Throughout the tutorial, all presented concepts are illustrated and discussed using exemplary applications and case studies. After this tutorial, the participants will have a solid knowledge of current research in MIR with respect to content-based and context-based methods, its potential and limitations, and future directions.

As mentioned in the beginning, the tutorial is based on the work that has gone into an upcoming Springer IR series book dealing with the matter (earlier dissemination of some of this work can also be found in two survey articles [6, 12] and one book chapter [8] co-authored by the proponents. The instructors' teaching experience of several years in MIR and related areas further allow to base additional material on the corresponding lecture slides.

History of the Tutorial

A first edition of the tutorial was presented as a half-day tutorial at the 34^{th} **European Conference on Information Retrieval** (ECIR), Barcelona, Spain in April 2012. In June 2012, Markus Schedl (jointly with Masataka Goto) gave a half-day tutorial on music information retrieval at the **ACM International Conference on Multimedia Retrieval** (ICMR), Hong Kong that included material from the ECIR tutorial on similarity estimation. At **SIGIR 2013**, we presented an overhauled version of the ECIR half-day tutorial focusing more on user and evaluation aspects. Parts of this version were again used by Markus Schedl in a half-day tutorial on "Multimedia Information Retrieval: Music and Audio" held together with Emilia Gómez and Masataka Goto at **ACM Multimedia 2013**. The material was then extended to serve as classroom material for a 5-day summer school course at the **7th Russian Summer School in Information Retrieval** (RuSSIR 2013) in Kazan.

The presentation of the topic at ECIR 2012 has also sparked the interest of Springer to issue a book on music information retrieval which will soon be published. The in-depth research of the topics and collection of material and examples will now feed back into the tutorial, allowing us to present an extended full-day version.

Outline

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**

Course Objectives

The intended full-day tutorial aims at reporting on the state of the art in mining music-related information from the web and further giving the interested audience an introduction to content-based feature extraction. The main goal is to give a sound and comprehensive, nevertheless easy-to-understand, introduction to audio processing and feature extraction as well as the scientific use of web- and community-based media in the music domain. The presented approaches are highly valuable for tasks and applications such as automated music playlist generation, personalized web radio, music search engines, music recommender systems, and intelligent user interfaces to music, therefore attendees will leave the tutorial with a very concise conception of how to

- extract exemplary musical features from audio signals to calculate content-based piece-wise similarity,
- mine music-relevant data from web sources to calculate similarity between music artists,
- extract music-related metadata and automatically index music collections for (personalised) retrieval, and
- evaluate music information retrieval systems.

As a consequence, attendees will also gain a deeper understanding of the anatomy of the above mentioned intelligent music applications. Last but not least, the tutorial should also promote the field of MIR to the broader SIGIR community.

Relevance to the IR Community

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.

Unfortunately, the general IR and the MIR communities are not too closely tied. This tutorial should therefore also pave the way for increasing the mutual understanding of the two communities and to obtain a better awareness of each others goals and challenges.

Required Infrastructure

For the presentation of the tutorial, we need a `projector` and an `audio system` (compatible with a typical notebook's 3.5mm stereo jack), capable of playing music sufficiently loud to be perceivable by the entire audience (to allow playback of videos and sound material for demonstration of the discussed approaches).

Planned Course Material

For the proposed tutorial we plan to publish a comprehensive set of slides, including references to articles presenting

state-of-the-art methods to MIR. In addition, implementations of several of the presented techniques (music content- and music context-based feature extractors and similarity measures) are available via our *CoMIRVA*[1] framework [13]. We will also point to other open implementations of presented algorithms.

4. REFERENCES

References by the Proponents

[1] M. Gillhofer and M. Schedl. Iron Maiden while jogging, Debussy for dinner? - An analysis of music listening behavior in context. In *Proceedings of the 21st International Conference on MultiMedia Modeling (MMM 2015)*, Sydney, Australia, January 2015.

[2] P. Knees, E. Pampalk, and G. Widmer. Artist Classification with Web-based Data. In *Proceedings of the 5th International Symposium on Music Information Retrieval (ISMIR 2004)*, pages 517–524, Barcelona, Spain, October 10–14 2004.

[3] P. Knees, T. Pohle, M. Schedl, D. Schnitzer, and K. Seyerlehner. A Document-centered Approach to a Natural Language Music Search Engine. In *Proceedings of the 30th European Conference on Information Retrieval (ECIR'08)*, Glasgow, Scotland, UK, March 30-April 3 2008.

[4] 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 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007)*, Amsterdam, the Netherlands, July 23–27 2007.

[5] P. Knees and M. Schedl. Towards Semantic Music Information Extraction from the Web Using Rule Patterns and Supervised Learning. In *Proceedings of the 2nd Workshop on Music Recommendation and Discovery (WOMRAD)*, Chicago, IL, USA, October 2011.

[6] P. Knees and M. Schedl. A Survey of Music Similarity and Recommendation from Music Context Data. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 10(1), 2013.

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

[8] M. Schedl. *Music Data Mining*, chapter Web-Based and Community-Based Music Information Extraction. CRC Press/Chapman Hall, 2011.

[9] 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.

[10] M. Schedl. *Social Media Retrieval*, chapter Exploiting Social Media for Music Information Retrieval. Springer, December 2012.

[11] M. Schedl. Leveraging Microblogs for Spatiotemporal Music Information Retrieval. In *Proceedings of the 35th European Conference on Information Retrieval (ECIR 2013)*, Moscow, Russia, March 24–27 2013.

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

[13] M. Schedl, P. Knees, K. Seyerlehner, and T. Pohle. The CoMIRVA Toolkit for Visualizing Music-Related Data. In *Proceedings of the 9th Eurographics/IEEE VGTC Symposium on Visualization (EuroVis 2007)*, Norrköping, Sweden, May 23–25 2007.

[14] 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 fuer Informatik, Dagstuhl, Germany, 2012.

[15] 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.

[16] K. Seyerlehner, G. Widmer, M. Schedl, and P. Knees. Automatic Music Tag Classification based on Block-Level Features. In *Proceedings of the 7th Sound and Music Computing Conference (SMC 2010)*, Barcelona, Spain, July 2010.

[17] 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*, 2012.

Other References

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

[19] C. Baccigalupo, E. Plaza, and J. Donaldson. Uncovering Affinity of Artists to Multiple Genres from Social Behaviour Data. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR'08)*, Philadelphia, PA, USA, September 14–18 2008.

[20] L. Barrington, D. Turnbull, M. Yazdani, and G. Lanckriet. Combining audio content and social context for semantic music discovery. In *Proceedings of the 32nd ACM SIGIR*, Boston, MA, USA, 2009.

[21] O. Celma and P. Herrera. A new approach to evaluating novel recommendations. In *Proceedings of the 2008 ACM Conference on Recommender Systems*, RecSys '08, pages 179–186, New York, NY, USA, 2008. ACM.

[22] S. Dixon, F. Gouyon, and G. Widmer. Towards Characterisation of Music via Rhythmic Patterns. In *Proceedings of the 5th International Symposium on Music Information Retrieval (ISMIR 2004)*, pages 509–516, Barcelona, Spain, October 10–14 2004.

[23] 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 2004)*, Barcelona, Spain, October 10–14 2004.

[1] `http://www.cp.jku.at/comirva`

[24] G. Dror, N. Koenigstein, and Y. Koren. Yahoo! music recommendations: Modeling music ratings with temporal dynamics and item taxonomy. In *Proceedings of the 5th ACM Conference on Recommender Systems*, pages 165–172, Chicago, IL, USA, 2011. ACM.

[25] 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 2008 International Computer Music Conference (ICMC 2008)*, Belfast, UK, August 2008.

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

[27] N. Koenigstein, Y. Shavitt, and T. Tankel. Spotting Out Emerging Artists Using Geo-Aware Analysis of P2P Query Strings. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, Las Vegas, NV, USA, 2008.

[28] 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.

[29] E. Law, L. von Ahn, R. Dannenberg, and M. Crawford. Tagatune: A Game for Music and Sound Annotation. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, Austria, September 2007.

[30] M. Levy and M. Sandler. A semantic space for music derived from social tags. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, Austria, September 2007.

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

[32] R. Mayer, R. Neumayer, and A. Rauber. Rhyme and Style Features for Musical Genre Classification by Song Lyrics. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR'08)*, 2008.

[33] B. McFee and G. Lanckriet. Hypergraph Models of Playlist Dialects. In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, Porto, Portugal, October 2012.

[34] C. S. Mesnage, A. Rafiq, S. Dixon, and R. P. Brixtel. Music discovery with social networks. In *2nd Workshop on Music Recommendation and Discovery (WOMRAD 2011)*, pages 7–12, Oct. 2011.

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

[36] 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 2012.

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

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

[39] B. Whitman and S. Lawrence. Inferring Descriptions and Similarity for Music from Community Metadata. In *Proceedings of the 2002 International Computer Music Conference (ICMC 2002)*, pages 591–598, Göteborg, Sweden, September 16–21 2002.

[40] Yuan Cao Zhang, Diarmuid O Seaghdha, Daniele Quercia, Tamas Jambor. Auralist: Introducing Serendipity into Music Recommendation. In *Proceedings of the 5th ACM Int'l Conference on Web Search and Data Mining (WSDM)*, Seattle, WA, USA, February 8–12 2012.