Multimedia Information Retrieval: Music and Audio

Markus Schedl
Department of Computational Perception
Johannes Kepler University
Linz, Austria
markus.schedl@jku.at

Emilia Gómez
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
emilia.gomez@upf.edu

Masataka Goto
National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Ibaraki, Japan
m.goto@aist.go.jp

Keywords

multimodal music information retrieval, audio and music description, similarity, classification, personalization

Categories and Subject Descriptors

Information systems [**Information search and retrieval**]: Music information retrieval

Motivation

Music is an omnipresent topic in our daily lives, as almost everyone enjoys listening to his or her favorite tunes. 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 categorized into: *music content*, *music context*, and *user context*. The music content refers to features extracted from the audio signal, while information about musical entities not encoded in the signal (e.g., image of an artist or political background of a song) are referred to as music context. The user context, in contrast, includes environmental aspects as well as physical and mental activities of the music listener. MIR research has been seeing a paradigm shift over the last couple of years, as an increasing number of recent approaches and commercial technologies combine content-based techniques (focusing on the audio signal) with multimedia context data mined, e.g. from web sources and with user context information.

Goals and Structure

In this half-day tutorial we (i) explain standard and state-of-the-art techniques for music content-based feature extraction, (ii) report on the basics and the state-of-the-art in mining music-related information from the web and social media to infer context-based features, and (iii) demonstrate attractive applications based on MIR technologies (using both content- and context-based methods).

The main goal is to give a sound and comprehensive, nevertheless easy-to-understand, introduction to the scientific use of multimedia data sources in the music domain. The presented approaches are highly valuable for tasks such as automatic music video analysis and generation, music-synchronized computer graphics, automated music playlist generation, personalized web radio music recommendation systems, and intelligent user interfaces to music.

To reach this goal, we first summarize the ideas behind various computational music features and discuss advantages and disadvantages of each. We review state-of-the-art techniques for the automatic description of music signals in terms of timbre (instrumentation), pitch-content information (melody, harmony and tonality), and rhythm (tempo or meter). We then present how these descriptors are used to measure similarity between musical pieces (e.g. to retrieve covers or versions of the same song) and classify music according to semantic concepts such as artist, genre, or mood.

Hereafter, we focus on the contextual aspects of music which are accessible through web technology. To this end, we give an introduction to the field of web-based MIR and an overview of popular data sources (e.g., web pages, microblogs, social networks, user tags, lyrics). Then we present approaches to exploit these sources to construct similarity measures in order to automatically index and retrieve music.

Relevance for the Multimedia Community

The Multimedia and the MIR communities have not been too closely tied, so far. As the 2013 edition of ACM Multimedia will see for the first time a *Music and Audio* track, this might change soon, however.

Given that both communities frequently use similar concepts and techniques, we believe that a tutorial on MIR in the context of ACM Multimedia will be of benefit for both communities, increase mutual understanding, raise awareness of each others goals and challenges, and help each other to reach these goals. For instance, some aspects of MIR are closely related to image and video processing and analysis, in particular when it comes to tasks such as deriving semantic information from music video clips (either official ones or user-generated music videos) or analyzing the importance of album cover artwork for the perception and organization of music, both aspects that have been looked into in MIR research. In addition, the challenge of creating visualizations and intelligent browsing interfaces to music collections is obviously highly related to the Multimedia community.

ACM Multimedia 2013 Barcelona, Spain

On the other hand, many tasks in the multimedia domain would benefit from incorporating audio and music analysis techniques more thoroughly, e.g., semantic analysis of video, video classification, synchronization of visual changes with musical beats, music video generation, or automatic slideshow generation. We are therefore sure that researchers in image and video processing could contribute to MIR and vice versa to realize mutual benefits.

In addition, the current boom of social media and user-generated multimedia content, which is also reflected in the frequent consideration of related scientific topics (e.g., Social Media Mining or Adaptive Multimedia Retrieval) in IR and multimedia conferences and journals, demands for a broader point of view on multimedia. The trend towards multimodal processing and retrieval of multimedia content requests audio- and music-related features of various kinds (considering, for instance, acoustic properties, textual data such as lyrics, image data such as album covers, and music video clips), in order to create holistic models.

Tutorial Outline

The tutorial lasts three hours; the presenters are researchers covering different subareas of MIR: content- and context-based music processing and retrieval, web- and social media-based MIR, similarity measurement, user interfaces including active music listening interfaces, and singing information processing.

The tutorial is accompanied by a comprehensive set of slides, including references to state-of-the-art methods. In addition, we provide links to existing software packages and toolboxes for music content- and music context-based feature extraction, similarity computation, knowledge discovery, and music visualization/exploration. As an example, we list here state-of-the-art technologies co-authored by the tutorial presenters:

- *CoMIRVA*[1], a framework providing a "Collection of Music Information Retrieval and Visualization Applications" [14, 13].
- *MELODIA*[2], a plugin for melody extraction from polyphonic music signals [11].
- *HPCP*[3], a plugin for chroma feature extraction from polyphonic music signals [2].
- *The Musical Avatar*[4], a system for generating an iconic representation of one's musical preferences [1].
- *The Mood Cloud*[9], a real-time music mood visualization tool.
- *Freesound*[5], a collaborative database of Creative Commons Licensed sounds with functionalities for user tagging and search by acoustic similarity.
- *NepTune*[6], a user interface to explore music by flying through artificial music landscapes [8, 12].
- *AGMIS*[7], a music information system automatically populated by mining the web for music-related information [15].
- *Music Tweet Map*[8], a visualization and exploration tool for music listening behavior inferred from microblogs [7].
- *SmartMusicKIOSK*[9], a music listening station with chorus-search function [3].
- *Songle*[10], a web service for active music listening improved by user contributions [6].

[1] http://www.cp.jku.at/comirva
[2] http://mtg.upf.edu/technologies/melodia
[3] http://mtg.upf.edu/technologies/hpcp
[4] http://mtg.upf.edu/project/musicalavatar
[5] http://www.freesound.org
[6] http://www.cp.jku.at/projects/nepTune
[7] http://www.cp.jku.at/projects/agmis
[8] http://www.cp.jku.at/projects/MusicTweetMap
[9] http://staff.aist.go.jp/m.goto/SmartMusicKIOSK
[10] http://songle.jp

- *Musicream*[11], an integrated music-listening interface for active, flexible, and unexpected encounters with songs [4].
- *DanceReProducer*[12], an automatic mashup music video generation system by reusing dance video clips on the web [10].
- *VocaListener*[13] and *VocaWatcher*[14] for imitating a human singer by using signal processing [5].

[11] http://staff.aist.go.jp/m.goto/Musicream
[12] http://staff.aist.go.jp/t.nakano/DanceReProducer
[13] http://staff.aist.go.jp/t.nakano/VocaListener
[14] http://staff.aist.go.jp/t.nakano/VocaWatcher

Acknowledgments

This research is supported by the Austrian Science Fund (FWF): P22856, P25655, and the EU FP7: 601166.

1. REFERENCES

[1] D. Bogdanov, M. Haro, F. Fuhrmann, A. Xambó, E. Gómez, and P. Herrera. Semantic Audio Content-based Music Recommendation and Visualization Based on User Preference Examples. *Information Processing & Management*, 49:13–33, January 2013.

[2] E. Gómez. *Tonal Description of Music Audio Signals*. PhD thesis, Universitat Pompeu Fabra, 2006.

[3] M. Goto. A chorus-section detection method for musical audio signals and its application to a music listening station. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5):1783–1794, 2006.

[4] M. Goto and T. Goto. Musicream: Integrated music-listening interface for active, flexible, and unexpected encounters with musical pieces. *IPSJ (Information Processing Society of Japan) Journal*, 50(12):2923–2936, 2009.

[5] M. Goto, T. Nakano, S. Kajita, Y. Matsusaka, S. Nakaoka, and K. Yokoi. VocaListener and VocaWatcher: Imitating a human singer by using signal processing. In *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 5393–5396, 2012.

[6] M. Goto, K. Yoshii, H. Fujihara, M. Mauch, and T. Nakano. Songle: A web service for active music listening improved by user contributions. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, pages 311–316, 2011.

[7] D. Hauger and M. Schedl. Exploring Geospatial Music Listening Patterns in Microblog Data. In *Proceedings of the 10th International Workshop on Adaptive Multimedia Retrieval (AMR)*, Copenhagen, Denmark, October 2012.

[8] P. Knees, M. Schedl, T. Pohle, and G. Widmer. Exploring Music Collections in Virtual Landscapes. *IEEE MultiMedia*, 14(3):46–54, July–September 2007.

[9] C. Laurier and P. Herrera. Mood cloud : A real-time music mood visualization tool. In *CMMR, Computer Music Modeling and Retrieval*, Copenhagen, 2008.

[10] T. Nakano, S. Murofushi, M. Goto, and S. Morishima. DanceReProducer: An automatic mashup music video generation system by reusing dance video clips on the web. In *Proceedings of the 8th Sound and Music Computing Conference (SMC)*, pages 183–189, 2011.

[11] 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.

[12] M. Schedl, C. Höglinger, and P. Knees. Large-Scale Music Exploration in Hierarchically Organized Landscapes Using Prototypicality Information. In *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR)*, Trento, Italy, April 2011.

[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)*, Norrköping, Sweden, May 2007.

[14] M. Schedl, P. Knees, and G. Widmer. Using CoMIRVA for Visualizing Similarities Between Music Artists. In *Proceedings of the 16th IEEE Visualization Conference (IEEE Vis)*, Minneapolis, MN, USA, October 2005.

[15] M. Schedl, G. Widmer, P. Knees, and T. Pohle. A Music Information System Automatically Generated via Web Content Mining Techniques. *Information Processing & Management*, 47, 2011.