Personalization in Multimodal Music Retrieval

Markus Schedl and Peter Knees

Department of Computational Perception
Johannes Kepler University
Linz, Austria
`http://www.cp.jku.at`

Abstract. This position paper provides an overview of current research endeavors and existing solutions in multimodal music retrieval, where the term "multimodal" relates to two aspects. The first one is taking into account the *music context* of a piece of music or an artist, while the second aspect tackled is that of the *user context*. The music context is introduced as all information important to the music, albeit not directly extractable from the audio signal (such as editorial or collaboratively assembled meta-data, lyrics in textual form, cultural background of an artists, or images of album covers). The user context, in contrast, is defined by various external factors that influence how a listener perceives music. It is therefore strongly related to user modeling and personalization, both facets of music information research that have not gained large attention by the MIR community so far. However, we are confident that adding personalization aspects to existing music retrieval systems (such as playlist generators, recommender systems, or visual browsers) is key to the future of MIR. In this vein, this contribution aims at defining the foundation for future research directions and applications related to multimodal music information systems.

1   Introduction and Motivation

Multimodal music processing and retrieval can be regarded as subfields of music information research, a discipline that has substantially gained importance during the last decade. The article at hand focuses on certain aspects of this field in that it will give an overview of the state-of-the-art in modeling and determining properties of music and listeners using features of different nature. In this introductory part, first a broad classification of such features is presented. Second, the three principle ways of *music retrieval* are introduced, each together with references to existing systems. Third, existing work on including personalization aspects in typical MIR tasks is reviewed.

The subsequent section points out various research endeavors and directions deemed to be important by the authors for the future of music information research, in particular, how to bring personalization and user adaptation to MIR. To this end, various data sources to describe the user context are introduced and discussed. Then, we present six steps towards the creation of a personalized multimodal music retrieval system.

1.1   Categories of Features

Estimating perceived musical similarity is commonly achieved by describing aspects of the music entity (e.g., a song, a performer, or an album) or the listener via computational features, and employing a similarity measure. These features can be broadly categorized into three classes, according to the authors: *music content*, *music context*, and *user context*, cf. Figure 1.
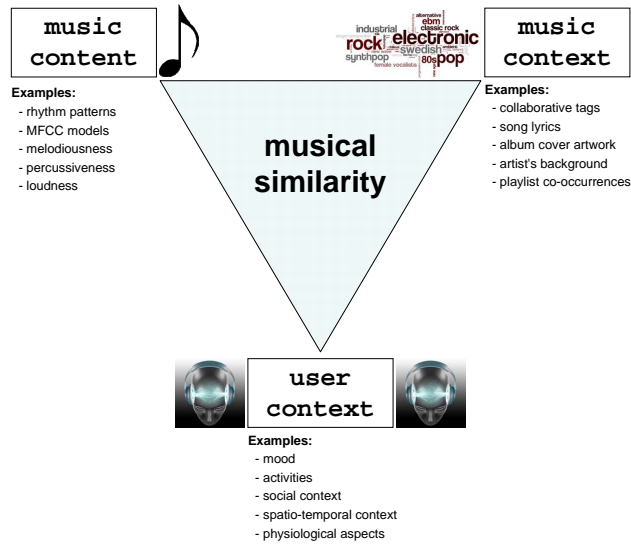

Fig. 1. Feature categories to describe music and listeners.

Music Content In content-based MIR, features extracted by applying signal processing techniques to audio signals are dominant. Such features are commonly denoted as *signal-based*, *audio-based*, or *content-based*. A good overview of common extraction techniques is presented in [7]. Music content-based features may be low-level representations that stem directly from the audio signal, for example zero-crossing rate [18], amplitude envelope [5], bandwidth and band energy ratio [37], or spectral centroid [67]. Alternatively, audio-based features may be derived or aggregated from low-level properties, and therefore represent aspects on a higher level of music understanding. Models of the human auditory system are frequently included in such derived features. High-level features usually aim at capturing either *timbral aspects* of music, which are commonly modeled via *MFCCs* [2], or *rhythmic aspects*, for example described via beat histograms [75]

or fluctuation patterns [63, 56]. Recent work addresses more specific high-level concepts, such as melodiousness and aggressiveness [57, 52].

Music Context The *music context* can be described as all information relevant to the music, albeit not directly extractable from the audio signal. For example, the meaning of a song's lyrics [29, 26], the political background of the musician, or the geographic origin of an artist [19, 66, 65] are likely to have a large impact on the music, but are not manifested in the signal.

An overview of the state-of-the-art in *music context*-based feature extraction (and similarity estimation) can be found in [61]. The majority of the approaches covering the music context are strongly related to *Web content mining* [38] as the Web provides contextual information on music artists in abundance. For example, in [21] the authors construct term profiles created from *artist-related Web pages* to derive music similarity information. *RSS feeds* are extracted and analyzed in [8]. Alternative sources to mine music context-related data include *playlists* (e.g., radio stations and mix tapes, i.e., user-generated playlists) [3, 6, 48] and *Peer-to-Peer networks* [70, 39, 11, 77]. In these cases, *co-occurrence analysis* is commonly employed to derive similarity on the artist- or track-level. Co-occurrences of artist names on Web pages are also used to infer artist similarity information [62] and for artist-to-genre classification [64]. *Song lyrics* as a source of music context-related information are analyzed, for example, in [40] to derive similarity information, in [33] for mood classification, and in [42] for genre classification. Another source for the music context is *collaborative tags*, mined for example from *last.fm* [32] in [12, 36] or gathered via *tagging games* [41, 74, 34].

User Context Existing work on incorporating user context aspects into MIR systems is relatively sparse. A preliminary study on users' acceptance of context logging in the context of music applications was conducted by Nürnberger and Stober [72]. The authors found significant differences in the participants' willingness to reveal different kinds of personal data on various scopes. Most participants indicated to eagerly share music metadata, information about ambient light and noise, mouse and keyboard logs, and their status in instant messaging applications. When it comes to used applications, facial expressions, bio signals, and GPS positions, however, a majority of users are reluctant to share their data. As for country-dependent differences, US-Americans were found to have on overall much lesser reservations to share personal data than Germans and Austrians. One has to note, however, that the study is biased towards Germans (accounting for 70% of the 305 participants).

In [59] Pohle et al. present preliminary steps towards a simple personalized music retrieval system. Based on a clustering of community-based tags extracted from *last.fm*, a small number of musical concepts are derived using *Non-Negative Matrix Factorization* (NMF) [35, 78]. Each music artist or band is then described by a "concept vector". A user interface allows for adjusting the weights of the individual concepts, based on which artists that match the resulting distribution

of the concepts best are recommended to the user. Zhang et al. propose in [80] a very similar kind of personalization strategy via user-adjusted weights.

Knees and Widmer present in [27] an approach that incorporates *relevance feedback* [60] into a text-based music search engine [23] to adapt the retrieval process to user preferences.

Even though no detailed information on their approach is publicly available, *last.fm* [32] builds user models based on its users' listening habits, which are mined via the "AudioScrobbler" interface. Based on this data, *last.fm* offers personalized music recommendations and playlist generation, however, without letting the user control (or even know) which factors are taken into account.

1.2   Categorizing Music Retrieval Systems

According to [76], music information retrieval systems to access music collections can be broadly categorized with respect to the employed *query formulation* method into *direct querying*, *query by example*, and *browsing* systems.

Direct querying systems take as input an excerpt of the feature representation to search for a piece of music. To give an example, *Themefinder* [73] and *musipedia* [44] support queries for sequences of exact pitch and of intervals, as well as for gross contour, using only up/down/repeat to describe the sequence of pitch changes.

A popular instance of a *query by example* system is *Shazam* [71], where the user records part of a music piece via his or her mobile phone, which is then analyzed on a server, identified, and meta-data such as artist or track name is sent back to the user's mobile phone. Another category of query by example retrieval applications is *query by humming* systems [13, 30, 49], where the search input consists of a user's recorded voice.

User interfaces that address the modality of *browsing* music collections exist in a considerable quantity. A fairly popular visualization and browsing technique employs the "Islands of Music" [53, 50] metaphor, which uses *Self-Organizing Maps* (SOM) [28], i.e., a non-linear, topology-preserving transform of a high-dimensional feature space to usually two dimensions. There exist also various extensions to the basic "Island of Music" approach. For example, [51] present "Aligned SOMs" that allow a smooth shift between SOMs generated on features representing diametric aspects of (music) properties. A mobile version of the "Islands of Music" is presented in [46]. This version also features a simple playlist generation method. In [25, 24] a three-dimensional extension is proposed to explore music collections in a fun way by further incorporating additional material mined from the Web. In addition, this three-dimensional version features intuitive playlist generation. It further makes use of an approach that is called "Music Description Map" [22] to calculate a mapping from music-related terms gathered from the Web to a SOM grid.

A browsing approach that offers an input wheel to sift through a cyclic playlist generated based on audio similarity is presented in [58]. A variant enriched with Web data and implemented on an "iPod touch" can be found in [68].

"The World of Music" [14] represents an appealing music artist visualizer and

browser, which calculates an embedding of high-dimensional data into the visualization space by employing *Semidefinite Programming* (SDP) [15]. *Multidimensional Scaling* (MDS) [31, 10] to visualize similar artist relations and browse in music collections is employed in [69]. Seyerlehner uses k-nearest neighbor graphs to reduce the computational complexity involved when dealing with medium- to large-sized collections and calculating a projection from the high-dimensional feature space to the two-dimensional visualization plane. Other interesting user interfaces for music collections include "MusicSun" [55], "MusicRainbow" [54], and "Musicream" [17].

1.3 Personalization Approaches

Aspects of the user context (cf. Section 1.1) are seldom taken into account when it comes to accessing music collections. One of the few commercial examples where the user context is considered in music search is the *collaborative filtering* [4] approach employed in *amazon.com*'s music Web store [1]. However, no details of the exact approach are publicly available.

In [9] Chai and Barry present some general considerations on modeling the user in a music retrieval systems and suggest an XML-based user modeling language for this purpose.

[47] presents a variant of the *Self-Organizing Map* (SOM) [28] that is based on a model that adapts to *user feedback*. To this end, the user can move data items on the SOM. This information is fed back into the SOM's codebook, and the mapping is adapted accordingly. [79] presents a *collaborative personalized search model* that alleviates the problems of *data sparseness* and *cold-start for new users* by combining information on different levels (individuals, interest groups, and global).

[80, 81] present *CompositeMap*, a model that takes into account similarity aspects derived from music content as well as from social factors. The authors propose a multimodal music similarity measure and show its applicability to the task of music retrieval. They also allow a simple kind of personalization of this model by letting the user weight the individual music dimensions on which similarity is estimated. However, they neither take the user context into consideration, nor do they try to learn a user's preferences.

In [43] a multimodal music similarity model at the artist-level is proposed. To this end, the authors calculate a *partial order embedding* using *kernel functions*. Music context- and content-based features are combined by this means. However, this model does not incorporate any personalization strategies.

2 User Modeling and Personalization in Music Retrieval

User profiling is without doubt key to enable personalized music services of the future. In the past, typical MIR applications, such as automated playlist generators or music browsers, employed approaches based on similarity measures computed on features derived from some representation of the music or artist, for

example, acoustic properties extracted from the audio signal [57] or term profiles calculated from music-related texts [21]. However, such approaches are known to be limited in their performance by some upper bound [2]. Furthermore, such approaches fall short of addressing the subjective component of music perception. What is it, for example, that makes you like a particular song when you are relaxing on Sunday morning? Do you prefer listening to happy or melancholic music when you are in a depressive mood? Which song do you relate to the first date with your beloved? The answer to these questions is most likely to be highly dependent on subjective factors. The sole use of music content- and music context-features described above is therefore insufficient to answer them.

That is where user modeling, personalization, and preference learning come into play. Models that combine different representation levels (e.g., low-level acoustic features and semantically meaningful tags) on different levels of data aggregation (e.g., segments within a piece of music, track-, artist-, or genre-level) and relate them to user profiles are crucial to describe user's preferences. The user model itself can also incorporate data on different levels of user representation. For example, [79] proposes a user model that comprises an *individual model*, a *interest group model*, and a *global user model*. We suggest adding a forth model, namely a *cultural user model*, that reflects the cultural area of the user. This cultural context can be given by an agglomeration, a whole country, or by a region that form a more or less homogeneous cultural entity.

2.1   Data Sources

There exists a wide variety of data sources for user context data, ranging from general location data (obtained by GPS or WiFi access points, for example) to highly personal aspects, such as blood pressure or intimate messages revealed by a user of a chatting software. Therefore, privacy issues play an important role for the acceptance of personalization techniques.
[16] provide a possible categorization of user context data. According to the authors, such data can be classified into the following five groups:

1. Environment Context
2. Personal Context
3. Task Context
4. Social Context
5. Spatio-temporal Context

The environmental context is defined as the entities that surround the user, for example, people, things, temperature, noise, humidity, and light. The personal context is further split into two sub-groups, namely, the *physiological context* and the *mental context*, where the former refers to attributes such as weight, blood pressure, pulse, or eye color, whereas the latter describes the user's psychic aspects, for example, stress level, mood, or expertise. The current activities pursued by the user is described by the *task context*. This context thus comprises actions, activities, and events the user is taking part of. Taking into account today's

mobile phones with multi-tasking capabilities, we suggest to extend this definition to include aspects of direct user input, running applications, and information on which application currently has the focus. Further taking the different messenger and microblogging services of the Web 2.0 era into consideration, we propose including them into the category of task context. These services, however, may also be a valuable source for a user's *social context*, which gives information about relatives, friends, enemies, or collaborators. Finally, the *spatio-temporal context* reveals information about a user's location, place, direction, speed, and time.

The recent emergence of "always on" devices equipped not only with a permanent Web connection, but also with various built-in sensors, has remarkably facilitated the logging of user context data from a technical perspective. Integrated GPS modules, accelerometers, light and noise sensors as well as interfaces to almost every Web 2.0 service makes user context logging easier than ever before, by providing data for all context categories described above.

2.2   Towards Personalized Music Services

We believe the following steps to be crucial to establish a foundation for personalized music retrieval.

1. Investigate the suitability and acceptance of different data sources to create user profiles.
2. Develop methods to mine the most promising data sources.
3. Create a model that reflects different aspects of the user context.
4. Investigate directions to integrate different similarity measures (content-based, music context-based, and user context-based).
5. Develop and thoroughly evaluate integrated models for the three kinds of music similarity.
6. Build a user-adaptive system for music retrieval, taking into account user-related factors.

The first step is to investigate the user's readiness to disclose various kinds of user-specific information, which will contribute to creating a user profile. Such a model is indispensable for personalized music recommendation that reflects various aspects of the music, the listener, and his or her environment. For example, aspects such as current position and direction of movement (e.g., is the user at home, doing sports, driving a car, in a train), weather conditions, times of the day, activities he or she is pursuing while listening to music, his or her current mood and emotion, demographic and socio-economic information about the user, information on the used music playback device (e.g., size, power, storage capacity, battery status), and information on music files (e.g., audio features, cultural meta-data extracted from the Web, editorial meta-data published by record companies, personal meta-data like playcounts or user tags) contribute to how a user judges the similarity between two artists or songs. We assume that the user's willingness to disclose partly private and sensible information, such as geographic location, listening habits, Web browser histories and bookmarks, or

content of shared folders in Peer-to-Peer networks, is strongly influenced by the benefits he or she can gain thereby (as one can easily see when looking at the overwhelming success of social networks). However, this willingness needs to be thoroughly evaluated, for example, by means of questionnaires and Web surveys.

Based on the results of the first step, it is possible to identify the most promising data sources, which a wide range of users are untroubled to share. Hence, the objective of the next step is to develop various data extractor components to gather user information, ranging from simple ones like date, time, and weekday monitoring, or recording user location and mouse clicking rates to complex ones such as bio-feedback measurements or user postings on social networks. For most data sources, employing post-processing to the gathered data will be required. To give an example from the Web mining domain, a study conducted in [20] revealed that about 50% of all user comments on *MySpace* [45] pages of popular music artists consists solely of spam, and 75% of the non-spam content failed linguistic parsing, meaning that 75% consists of broken sentences.

Step three subsequently aims at investigating which kinds of user context features relate to a listener's music taste, and at designing a user model that reflects and aggregates these user-specific factors. To this end, it is necessary to apply and refine machine learning techniques to learn user preferences, i.e., a mapping between individual, user-specific factors and the user's appeal to certain music categories, styles, or individual artists or tracks. In this step, various models of different scope and complexity need to be evaluated: for example, one model that takes only directly user-related data into account, another one similar to [79] that represents an integrated model comprising of an *individual user model*, a *group model* (cultural / peer group), and a *global model*.

Existing multifaceted models for music similarity measurement, such as [43, 80], seem to lack real personalization functionality beyond simple user-adjustable weights for certain feature dimensions. Therefore, looking into different ways of building an aggregate model of music similarity based on the three broad categories of sources (music content, music context, and user context) is the key part of step four. Besides the problem of dealing with the inhomogeneous nature of the data sources, another important issue to address is the dimensionality problem since some data sources (term profiles in the case of user tags or Web page content, for example) are very high-dimensional, and therefore require the application and evaluation of dimensionality reduction techniques.

Following different strategies to develop such a comprehensive, multifaceted model will result in various model prototypes. In the next step, these prototypes have to undergo a comprehensive evaluation, including user studies and Web surveys. The best performing models are then determined for various usage scenarios, e.g., recommender systems, playlist generation, or retrieval systems supporting very specific, cross-data source queries such as "give me music for listening to on my mobile device when I am driving my car (user context), that further has a strong harmonic component (music content) and sad lyrics (music context)".

The final step comprises creating, evaluating, and refining various prototypical music retrieval systems that adapt to the user's current listening preferences, which are derived from the user context. The systems will make use of the aggregate models of music similarity elaborated in step five. They may include automatic personalized playlist generation systems, personalized recommender systems, or adaptive user interfaces to music collections. In this step, evaluating the ergonomic as well as the qualitative aspects of the retrieval systems is necessary.

3   User-Awareness and Personalization are the Future of MIR

From the analysis and considerations presented so far, the authors' perspective of future research directions and music services can be summarized as follows.

Personalization aspects have to be taken into account when elaborating music retrieval systems. In this context, it is important to note the highly subjective, cognitive component in the understanding of music and judging its personal appeal. Therefore, designing user-aware music applications requires intelligent machine learning techniques, in particular, preference learning approaches that relate the user context to concise, situation-dependent music preferences.

User models that encompass different social scopes are needed. They may aggregate an individual model, an interest group model, a cultural model, and a global model.

Multifaceted similarity measures that combine different feature categories (music content, music context, and user context) are required. The corresponding representation models should then not only allow to derive similarity between music via content-related aspects, such as beat strength or instruments playing, or via music context-related properties, such as the geographic origin of the performer or a song's lyrics, but also to describe users and user groups in order to compute a listener-based similarity score. Such user-centric features enable the application of collaborative filtering techniques and eventually the elaboration of personalized music recommender systems.

Evaluation of user-adaptive systems is of vital importance. As such systems are by definition tailored to individual users, this is certainly not an easy task and far beyond the genre-classification-experiments commonly employed when assessing music similarity measures.

Nevertheless, we are sure that future research directions in MIR should be centered around intelligently combining various complementary music and user representations as this will pave the way for exciting novel music applications that keep on playing music according to the user's taste without requiring any explicit user interaction.

Acknowledgments

References

1. http://www.amazon.com/music (access: January 2010).
2. J.-J. Aucouturier and F. Pachet. Improving Timbre Similarity: How High is the Sky? *Journal of Negative Results in Speech and Audio Sciences*, 1(1), 2004.
3. 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.
4. J. S. Breese, D. Heckerman, and C. Kadie. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98)*, pages 43–52, San Francisco, CA, USA, 1998. Morgan Kaufmann.
5. J. J. Burred and A. Lerch. A Hierarchical Approach to Automatic Musical Genre Classification. In *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx-03)*, London, UK, September 8–11 2003.
6. P. Cano and M. Koppenberger. The Emergence of Complex Network Patterns in Music Artist Networks. In *Proceedings of the 5th International Symposium on Music Information Retrieval (ISMIR 2004)*, pages 466–469, Barcelona, Spain, October 10–14 2004.
7. M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-Based Music Information Retrieval: Current Directions and Future Challenges. *Proceedings of the IEEE*, 96:668–696, April 2008.
8. O. Celma, M. Ramírez, and P. Herrera. Foafing the Music: A Music Recommendation System Based on RSS Feeds and User Preferences. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, London, UK, September 11–15 2005.
9. W. Chai and B. Vercoe. Using user models in music information retrieval systems. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR 2000)*, Plymouth, MA, USA, 2000.
10. T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman & Hall, 1994.
11. D. P. Ellis, B. Whitman, A. Berenzweig, and S. Lawrence. The Quest For Ground Truth in Musical Artist Similarity. In *Proceedings of 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, Paris, France, October 13–17 2002.
12. G. Geleijnse, M. Schedl, and P. Knees. The Quest for Ground Truth in Musical Artist Tagging in the Social Web Era. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, Austria, September 23–27 2007.
13. A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith. Query by Humming: Musical Information Retrieval in an Audio Database. In *Proceedings of the 3rd Association for Computing Machinery (ACM) International Conference on Multimedia*, pages 231–236, San Fancisco, CA, USA, 1995.
14. D. Gleich, M. Rasmussen, K. Lang, and L. Zhukov. The World of Music: SDP Layout of High Dimensional Data. In *Proceedings of the IEEE Symposium on Information Visualization 2005*, 2005.
15. M. X. Goemans and D. P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfyability Problems Using Semidefinite Programming. *Journal of the Association for Computing Machinery*, 42(6):1115–1145, November 1995.

16. A. Göker and H. I. Myrhaug. User Context and Personalisation. In *Proceedings of the 6th European Conference on Case Based Reasoning (ECCBR 2002): Workshop on Case Based Reasoning and Personalization*, Aberdeen, Scotland, September 2002.

17. M. Goto and T. Goto. Musicream: New Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, London, UK, September 11–15 2005.

18. F. Gouyon, F. Pachet, and O. Delerue. On the Use of Zero-Crossing Rate for an Application of Classification of Percussive Sounds. In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx-00)*, Verona, Italy, December 7–9 2000.

19. S. Govaerts and E. Duval. A Web-based Approach to Determine the Origin of an Artist. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, Kobe, Japan, October 2009.

20. J. Grace, D. Gruhl, K. Haas, M. Nagarajan, C. Robson, and N. Sahoo. Artist Ranking Through Analysis of On-line Community Comments. In *Proceedings of the 17th ACM International World Wide Web Conference (WWW 2008)*, Bejing, China, April 21–25 2008.

21. 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.

22. P. Knees, T. Pohle, M. Schedl, and G. Widmer. Automatically Describing Music on a Map. In *Proceedings of 1st Workshop on Learning the Semantics of Audio Signals (LSAS 2006)*, Athens, Greece, December 6–8 2006.

23. 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.

24. P. Knees, M. Schedl, T. Pohle, and G. Widmer. An Innovative Three-Dimensional User Interface for Exploring Music Collections Enriched with Meta-Information from the Web. In *Proceedings of the 14th ACM International Conference on Multimedia (MM 2006)*, Santa Barbara, CA, USA, October 23–27 2006.

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

26. P. Knees, M. Schedl, and G. Widmer. Multiple Lyrics Alignment: Automatic Retrieval of Song Lyrics. In *Proceedings of 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pages 564–569, London, UK, September 11–15 2005.

27. P. Knees and G. Widmer. Searching for Music Using Natural Language Queries and Relevance Feedback. In *Proceedings of the 5th International Workshop on Adaptive Multimedia Retrieval (AMR'07)*, Paris, France, July 2007.

28. T. Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, Germany, 3rd edition, 2001.

29. J. Korst and G. Geleijnse. Efficient lyrics retrieval and alignment. In W. Verhaegh, E. Aarts, W. ten Kate, J. Korst, and S. Pauws, editors, *Proceedings of the 3rd Philips Symposium on Intelligent Algorithms (SOIA 2006)*, pages 205–218, Eindhoven, the Netherlands, December 6–7 2006.

30. N. Kosugi, Y. Nishihara, T. Sakata, M. Yamamuro, and K. Kushima. A Practical Query-by-Humming System for a Large Music Database. In *Proceedings of the 8th*

ACM International Conference on Multimedia, pages 333–342, Los Angeles, CA, USA, 2000.

31. J. B. Kruskal and M. Wish. *Multidimensional Scaling*. Paper Series on Quantitative Applications in the Social Sciences. Sage Publications, Newbury Park, CA, USA, 1978.

32. http://last.fm (access: January 2010), 2010.

33. C. Laurier, J. Grivolla, and P. Herrera. Multimodal Music Mood Classification using Audio and Lyrics. In *Proceedings of the International Conference on Machine Learning and Applications*, San Diego, CA, USA, 2008.

34. 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.

35. D. D. Lee and H. S. Seung. Learning the Parts of Objects by Non-negative Matrix Factorization. *Nature*, 401(6755):788–791, 1999.

36. 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.

37. D. Li, I. K. Sethi, N. Dimitrova, and T. McGee. Classification of General Audio Data for Content-based Retrieval. *Pattern Recognition Letters*, 22(5):533–544, 2001.

38. B. Liu. *Web Data Mining – Exploring Hyperlinks, Contents and Usage Data*. Springer, Berlin, Heidelberg, Germany, 2007.

39. B. Logan, D. P. Ellis, and A. Berenzweig. Toward Evaluation Techniques for Music Similarity. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003): Workshop on the Evaluation of Music Information Retrieval Systems*, Toronto, Canada, July–August 2003. ACM Press.

40. B. Logan, A. Kositsky, and P. Moreno. Semantic Analysis of Song Lyrics. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2004)*, Taipei, Taiwan, June 27–30 2004.

41. M. I. Mandel and D. P. Ellis. A Web-based Game for Collecting Music Metadata. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, Austria, September 2007.

42. 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.

43. B. McFee and G. Lanckriet. Heterogeneous Embedding for Subjective Artist Similarity. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, Kobe, Japan, October 2009.

44. http://www.musipedia.org (access: February 2010).

45. http://www.myspace.com (access: November 2009), 2009.

46. R. Neumayer, M. Dittenbach, and A. Rauber. PlaySOM and PocketSOMPlayer, Alternative Interfaces to Large Music Collections. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, London, UK, September 11–15 2005.

47. A. Nürnberger and M. Detyniecki. Weighted Self-Organizing Maps: Incorporating User Feedback. In O. Kaynak and E. Oja, editors, *Proceedings of the Joined 13th International Conference on Artificial Neural Networks and Neural Information Processing (ICANN/ICONIP 2003)*, pages 883–890. Springer-Verlag, 2003.

48. F. Pachet, G. Westerman, and D. Laigre. Musical Data Mining for Electronic Music Distribution. In *Proceedings of the 1st International Conference on Web Delivering of Music (WEDELMUSIC 2001)*, Florence, Italy, November 23–24 2001.

49. B. Padro. Finding Structure in Audio for Music Information Retrieval. *IEEE Signal Processing Magazine*, 23(3):126–132, May 2006.
50. E. Pampalk. Islands of Music: Analysis, Organization, and Visualization of Music Archives. Master's thesis, Vienna University of Technology, Vienna, Austria, 2001. http://www.oefai.at/∼elias/music/thesis.html.
51. E. Pampalk. Aligned Self-Organizing Maps. In *Proceedings of the Workshop on Self-Organizing Maps (WSOM 2003)*, pages 185–190, Kitakyushu, Japan, September 11–14 2003. Kyushu Institute of Technology.
52. E. Pampalk. *Computational Models of Music Similarity and their Application to Music Information Retrieval*. PhD thesis, Vienna University of Technology, March 2006.
53. E. Pampalk, S. Dixon, and G. Widmer. Exploring Music Collections by Browsing Different Views. *Computer Music Journal*, 28(3), 2004.
54. E. Pampalk and M. Goto. MusicRainbow: A New User Interface to Discover Artists Using Audio-based Similarity and Web-based Labeling. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*, Victoria, Canada, October 8–12 2006.
55. E. Pampalk and M. Goto. MusicSun: A New Approach to Artist Recommendation. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, Austria, September 23–27 2007.
56. E. Pampalk, A. Rauber, and D. Merkl. Content-based Organization and Visualization of Music Archives. In *Proceedings of the 10th ACM International Conference on Multimedia (MM 2002)*, pages 570–579, Juan les Pins, France, December 1–6 2002.
57. T. Pohle. *Automatic Characterization of Music for Intuitive Retrieval*. PhD thesis, Johannes Kepler University Linz, Linz, Austria, 2009.
58. T. Pohle, P. Knees, M. Schedl, E. Pampalk, and G. Widmer. "Reinventing the Wheel": A Novel Approach to Music Player Interfaces. *IEEE Transactions on Multimedia*, 9:567–575, 2007.
59. T. Pohle, P. Knees, M. Schedl, and G. Widmer. Building an Interactive Next-Generation Artist Recommender Based on Automatically Derived High-Level Concepts. In *Proceedings of the 5th International Workshop on Content-Based Multimedia Indexing (CBMI'07)*, Bordeaux, France, June 2007.
60. J. J. Rocchio. Relevance Feedback in Information Retrieval. In G. Salton, editor, *The SMART Retrieval System - Experiments in Automatic Document Processing*, pages 313–323. Englewood Cliffs, NJ: Prentice-Hall, 1971.
61. M. Schedl and P. Knees. Context-based Music Similarity Estimation. In *Proceedings of the 3rd International Workshop on Learning the Semantics of Audio Signals (LSAS 2009)*, Graz, Austria, December 2009.
62. M. Schedl, P. Knees, and G. Widmer. A Web-Based Approach to Assessing Artist Similarity using Co-Occurrences. In *Proceedings of the 4th International Workshop on Content-Based Multimedia Indexing (CBMI 2005)*, Riga, Latvia, June 21–23 2005.
63. M. Schedl, E. Pampalk, and G. Widmer. Intelligent Structuring and Exploration of Digital Music Collections. *e&i - Elektrotechnik und Informationstechnik*, 122(7–8):232–237, July–August 2005.
64. M. Schedl, T. Pohle, P. Knees, and G. Widmer. Assigning and Visualizing Music Genres by Web-based Co-Occurrence Analysis. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*, Victoria, Canada, October 8–12 2006.

65. M. Schedl, C. Schiketanz, and K. Seyerlehner. Country of Origin Determination via Web Mining Techniques. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2010): 2nd International Workshop on Advances in Music Information Research (AdMIRe 2010)*, Singapore, July 19–23 2010.

66. M. Schedl, K. Seyerlehner, D. Schnitzer, G. Widmer, and C. Schiketanz. Three Web-based Heuristics to Determine a Person's or Institution's Country of Origin. In *Proceedings of the 33th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2010)*, Geneva, Switzerland, July 19–23 2010.

67. E. Scheirer and M. Slaney. Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1997)*, pages 1331–1334, Munich, Germany, April 21–24 1997.

68. D. Schnitzer, T. Pohle, P. Knees, and G. Widmer. One-Touch Access to Music on Mobile Devices. In *Proceedings of the 6th International Conference on Mobile and Ubiquitous Multimedia (MUM 2007)*, Oulu, Finland, December 12–14 2007.

69. K. Seyerlehner. Inhaltsbasierte Ähnlichkeitsmetriken zur Navigation in Musiksammlungen. Master's thesis, Johannes Kepler Universität Linz, Linz, Austria, June 2006.

70. Y. Shavitt and U. Weinsberg. Songs Clustering Using Peer-to-Peer Co-occurrences. In *Proceedings of the IEEE International Symposium on Multimedia (ISM2009): International Workshop on Advances in Music Information Research (AdMIRe 2009)*, San Diego, CA, USA, December 16 2009.

71. http://www.shazam.com (access: February 2010).

72. S. Stober, M. Steinbrecher, and A. Nürnberger. A Survey on the Acceptance of Listening Context Logging for MIR Applications. In *Proceedings of 3rd Workshop on Learning the Semantics of Audio Signals (LSAS 2009)*, Graz, Austria, December 2009.

73. http://themefinder.org (access: February 2010).

74. D. Turnbull, R. Liu, L. Barrington, and G. Lanckriet. A Game-based Approach for Collecting Semantic Annotations of Music. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, Austria, September 2007.

75. G. Tzanetakis and P. Cook. Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

76. R. C. Veltkamp. Multimedia Retrieval Algorithmics. In *Proceedings of the 33rd International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 2007)*, Harrachov, Czech Republic, January 20–26 2007. Springer.

77. 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.

78. W. Xu, X. Liu, and Y. Gong. Document Clustering Based on Non-negative Matrix Factorization. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003)*, pages 267–273, Toronto, Canada, July 28–August 1 2003. ACM Press.

79. G.-R. Xue, J. Han, Y. Yu, and Q. Yang. User Language Model for Collaborative Personalized Search. *ACM Transactions on Information Systems*, 27(2), February 2009.

80. B. Zhang, J. Shen, Q. Xiang, and Y. Wang. CompositeMap: A Novel Framework for Music Similarity Measure. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2009)*, pages 403–410, New York, NY, USA, 2009. ACM.
81. B. Zhang, Q. Xiang, Y. Wang, and J. Shen. CompositeMap: A Novel Music Similarity Measure for Personalized Multimodal Music Search. In *MM '09: Proceedings of the seventeen ACM international conference on Multimedia*, pages 973–974, New York, NY, USA, 2009. ACM.