User-Aware Music Retrieval

Markus Schedl[1], Sebastian Stober[2], Emilia Gómez[3], Nicola Orio[4], and Cynthia C. S. Liem[5]

1 **Department of Computational Perception**
Johannes Kepler University, Linz, Austria
markus.schedl@jku.at
2 **Data & Knowledge Engineering Group**
Otto-von-Guericke-Universität, Magdeburg, Germany
stober@ovgu.de
3 **Music Technology Group**
Universitat Pompeu Fabra, Barcelona, Spain
emilia.gomez@upf.edu
4 **Department of Information Engineering**
University of Padova, Italy
orio@dei.unipd.it
5 **Multimedia Information Retrieval Lab**
Delft University of Technology, the Netherlands
c.c.s.liem@tudelft.nl

Abstract

Personalized and user-aware systems for retrieving multimedia items are becoming increasingly important as the amount of available multimedia data has been spiraling. A personalized system is one that incorporates information about the user into its data processing part (e.g., a particular user taste for a movie genre). A context-aware system, in contrast, takes into account dynamic aspects of the user context when processing the data (e.g., location and time where/when a user issues a query). Today's user-adaptive systems often incorporate both aspects.

Particularly focusing on the music domain, this article gives an overview of different aspects we deem important to build personalized music retrieval systems. In this vein, we first give an overview of factors that influence the human perception of music. We then propose and discuss various requirements for a personalized, user-aware music retrieval system. Eventually, the state-of-the-art in building such systems is reviewed, taking in particular aspects of *similarity* and *serendipity* into account.

1998 ACM Subject Classification H.5.5 Sound and Music Computing, J.5 Arts and Humanities–Music, H.5.1 Multimedia Information Systems, I.5 Pattern Recognition

Keywords and phrases user-aware music retrieval, personalization, recommendation, user context, adaptive systems, similarity measurement, serendipity

Digital Object Identifier 10.4230/DFU.Vol3.11041.135

1 Introduction

Multimodal music processing and retrieval can be regarded as subfields of music information research (MIR), a discipline that has substantially gained importance during the last decade. Multimodality can be recognized at several levels in MIR, for example, different modalities to access music collections (query-by-example, direct querying, browsing, metadata-based search, visual user interfaces) or different representations of music items themselves – score sheet, symbolic MIDI, digital audio waveform, or textual lyrics, just to name a few.

Multimodal Music Processing. *Dagstuhl Follow-Ups, Vol. 3.* ISBN 978-3-939897-37-8.
Editors: Meinard Müller, Masataka Goto, and Markus Schedl; pp. 135–156
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

In this article, multimodality relates to the integration of *various knowledge sources* in music processing systems. A key source of knowledge is given by aspects linked to the user and his or her usage of the system, which is the focus of the present study. The article at hand hence gives an overview of the state-of-the-art in modeling and determining properties of music and listeners using features of different kinds. These features all relate to how music is perceived by humans. First, a broad categorization of such features is presented in Section 2. Also references to related work on extracting and processing the respective features is given for each feature category. Subsequently, various research endeavors and directions deemed to be important by the authors for the future of personalized, multimodal music retrieval are presented. More precisely, we present a set of requirements important for user-aware music retrieval systems in Section 3. Two vital prerequisites to build user-aware music retrieval applications, such as personalized music recommender systems or user-adaptive browsing interfaces, are first *elaborating similarity measures* that are capable of revealing similarity relations as perceived by humans and second provide a *serendipitous experience* to the user. In order to develop the mentioned, sophisticated similarity measures, we need methods that capture musical similarity at different levels using different modalities, for example, timbre, rhythm, harmony, lyrics, or co-listening information. A review of the state-of-the-art in building such adaptive similarity measures is presented in Section 4. The latter requirement, ensuring a certain degree of serendipity in retrieval results, necessitates to take into account various user-dependent factors. For example, it is important for a serendipitous system to have information about the user's music taste and preference, where taste refers to a long-term inclination and preference describes a rather short-term, situation-dependent affection. Both are likely to change over time, although taste usually changes only gradually and at a slower rate than preference. More details on serendipity aspects in personalized music retrieval are given in Section 5. Finally, in Section 6, we draw conclusions and indicate some directions for future research.

2 Computational Aspects of Music Perception and Similarity

Developing computational features that encode knowledge on how we humans perceive music is one of the grand challenges in MIR. It is a particular endeavor for various reasons. Among others, music perception is very subjective and influenced for example by the listener's music preferences, but also highly dependent on his or her musical training as well as social and sociographic background. Moreover, perceptually relevant features may be extracted from very different media and representations of music, which describe a wide variety of aspects. Media encoding music or music-related data range from score sheets to digital audio files and from textual lyrics to images of cover artwork. Which of these multimodal aspects influence human perception of music, in which way and to which extent is still an open research question.
Computational music features can be broadly categorized into three classes, according to the authors: *music content*, *music context*, and *user context*, cf. Figure 1.

2.1 Music Content

In traditional MIR, features extracted by applying signal processing techniques to audio signals were dominant. Such features are commonly denoted as *signal-based*, *audio-based*, or *content-based*. In addition to audio signals, the music content may be described by various other modalities, such as handwritten or digitized score, or video clips.

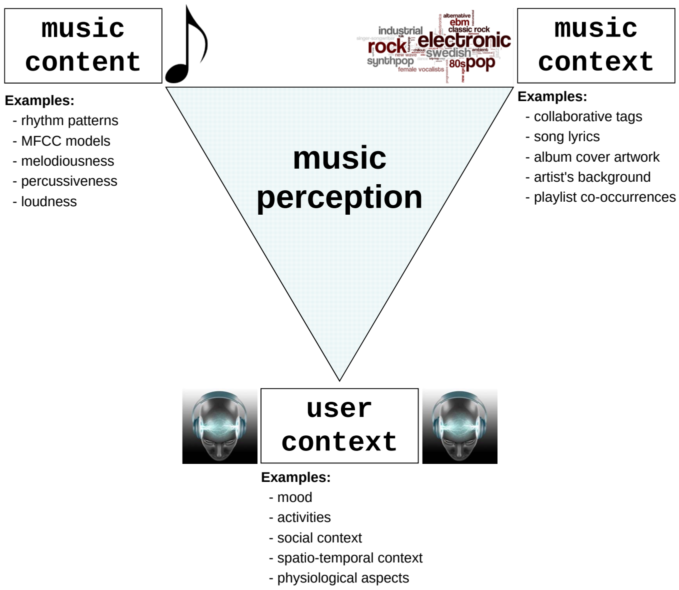


Figure 1 Feature categories to describe music.

Thorough overviews of common extraction techniques are presented in [17, 28, 65]. Music content-based features may be low-level representations that stem directly from the audio signal, for example Mel Frequency Cepstral Coefficients (MFCCs) [54], zero-crossing rate [29], amplitude envelope [15], bandwidth and band energy ratio [52], spectral centroid [82], fundamental frequency or chroma features [11]. As mentioned in [28], most low-level features do not make sense to the majority of the listeners, although they are easily exploited by computing systems.

Alternatively, content-based features may be derived or aggregated from low-level properties, and therefore represent aspects on a higher level of music understanding. Such features are often named mid-level features. Machine learning, statistical modeling and models of the human auditory system make mid-level descriptors possible, usually by gathering large sets of observations. Mid-level features usually aim at capturing either *timbral aspects* of music, which were traditionally modeled via *MFCCs* [2], *rhythmic aspects*, for example described via beat histograms [92] or fluctuation patterns [78, 69], and *tonal aspects* such as predominant melody[73], key or chord progression [27], often derived from chroma features.

Recent work aims at inferring more specific high-level concepts, meaningful to users, such as melodiousness, complexity, danceability, aggressiveness [70, 68, 90], mood [44], or genre [31]. The transition from low- or mid-level descriptors to high-level descriptors requires bridging the semantic gap. According to [28], high-level or semantic feature extractors require to include an induction procedure that has to be carried out by means of a user model, and not only a data model as in the case of mid-level descriptors.

2.2 Music Context

The *music context* can be described as all information relevant to the music item under consideration, albeit not directly extractable from the music manifestation itself. For example, the meaning of a song's lyrics [40, 36], the political background of the musician, or the geographic origin of an artist [30, 81, 80] are likely to have a strong impact on how music is perceived and interpreted, 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 [76]. The majority of the approaches covering the music context are strongly related to *Web content mining* [53] as the Web provides contextual information on music artists in abundance. For example, in [34] the authors construct term profiles created from *artist-related Web pages* to derive music similarity information. *RSS feeds* are extracted and analyzed in [18]. Alternative sources to mine music context-related data include *playlists* (e.g., radio stations and mix tapes, i.e., user-generated playlists) [3, 16, 67] and *Peer-to-Peer networks* [83, 55, 24, 96]. In these cases, *co-occurrence analysis* is commonly employed to derive similarity information on the artist- or track-level. Co-occurrences of artist names on Web pages are also used to infer artist similarity information [77] and for artist-to-genre classification [79]. *Song lyrics* as a source of music context-related information are analyzed, for example, in [56] to derive similarity information, in [45] for mood classification, and in [60] for genre classification. Another source for the music context is *collaborative tags*, mined for example from *last.fm* [43] in [25, 51] or gathered via *tagging games* [59, 91, 46].

2.3 User Context

Scientific work on MIR that takes into account aspects of the user context is still relatively sparse and covers diverse topics. It can be broadly divided into user music-seeking behavior studies, user preferences elicitation, multifaceted user and similarity models, and personalized, user-aware recommender systems.

User Music-Seeking Behavior Studies

Several MIR researchers, largely with backgrounds in library and information sciences, have devoted studies to music-seeking behavior and information requirements of users. While these studies typically are conducted on a much smaller-scaled population than usually found in engineering settings, they are detailed and give qualitative insight into real-life and every-day music behavior. Many of them strikingly point out how the reception of music is not just guided by the characteristics of the music audio signal, but is strongly influenced by multimodal influences that do not necessarily have to do with the music.

Cunningham et al. [22] conducted an ethnographic study of music searching and browsing techniques. Important findings regarding this chapter were that music shopping often was a collaborative activity, with a social function going beyond music listening, and a 'surprisingly visual' activity too, with shoppers identifying music genres that they liked through the appearance of album covers. The influential role of visual means in musical settings also appears in other user studies, e.g. Bainbridge et al. [6], in which a user-centered personal digital library is designed with the spatial hypermedia paradigm, and recently Barthet and Dixon [10], describing an ethnographic study of musicologists at the British Library. In the latter study, visualization of audio signals aided the musicologists with exploring and studying music recordings, but also could steer the users' attention towards specific details. A visual spectrogram display pointed out signal features (e.g. vibrato) that the user was not aware of, but also deemphasized sound aspects that could not be seen: "I completely forgot about the bassoon, it feels like it is unimportant now, but I was once struck by it".

Social context is a strong influence on music taste. Laplante [42] found that young adults had a strong penchant for informal channels (e.g. friends), but a low trust of experts (e.g. music store staff). Furthermore, it was noted that music discoveries often were the result of passive rather than active search behavior – this points towards serendipitous finds, which will be discussed in the upcoming sections of this chapter.

The reasons why we remember, like or hate music also are strongly determined by context. In a study of reasons why people dislike songs [21], the factors of influence were lyrics, the earworm effect (getting a song stuck in your head without wanting this), quality of the singing voice, dislike of music videos, over-exposure of a song, pretentiousness of the performing artist, clashing taste cultures (disliking the social community associated with a certain style) and unfortunate personal associations. An extensive study by Lee [49] of natural language music queries also illustrates frequent associative notions: dormant searches get rekindled because similar thematic context settings are encountered (e.g. searching for information on a 'spooky tune' that has been used in cartoons to signify that someone has died, after hearing it being played on Halloween), and songs get a special affective meaning because they had been heard in special affective settings ("My grandfather, who was born in 1899, used to sing me to sleep with this song and I can't remember the words").

Findings from user studies as described in this paragraph have not widely been adopted in the design of MIR systems yet, but still will be very relevant when studying user context.

User Preferences Elicitation

An obvious way to obtain information about the taste, preferences and behavior of a user is context logging. However, this can pose privacy issues. In a study on users' acceptance of context logging in the context of music applications by Nürnberger and Stober [89], 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 results might be biased as 70% of the 305 participants were from Germany.

An alternative to context logging is to explicitly ask the users to provide means to characterize their musical preferences. One example of this methodology is presented in [32]. This study proposes a method to automatically generate, given a provided set of preferred music tracks, an iconic representation of a user's musical preferences – the Musical Avatar. Starting from the raw audio signals, they compute a set of semantic descriptors which are mapped to the visual domain by creating a humanoid cartoony character that represents the user's musical preferences. Examples of possible avatars are provided in Figure 2. This representation of a users's musical preferences is then used to provide personalized recommendations in [13].

User and Similarity Models

One of the earliest works in user modeling for MIR is [19], where Chai and Barry present some general considerations on modeling the user in a music retrieval system. They also suggest an XML-based user modeling language for this purpose.

Zhang et al. present *CompositeMap* [100, 101], a model that takes into account similarity aspects derived from music content as well as social factors. The authors propose a multimodal

Figure 2 Examples of Musical Avatars representing the user's musical preferences [58].

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 do neither take the user context into consideration, nor do they try to learn a user's preferences.

In [63] a multimodal music similarity model on the artist-level is proposed. To this end, McFee and Lanckriet 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.

In [72] 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) [48, 98]. 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 [100] a very similar kind of personalization strategy via user-adjusted weights.

Knees and Widmer present in [37] an approach that incorporates *relevance feedback* [74] into a text-based music search engine [35] to adapt the retrieval process to user preferences. The search engine proposed by Knees et al. builds a model from music content features (MFCCs) and music context features (term vector representations of artist-related Web pages). To this end, a weight is computed for each (term, music item)-pair, based on the term vectors. These weights are then smoothed, taking into account the closest neighbors according to the content-based similarity measure (Kullback-Leibler divergence on Gaussian Mixture Models of the MFCCs). To retrieve music via natural language queries, each textual query issued to the system is expanded via a *Google* search, resulting again in a term weight vector. This query vector is subsequently compared to the smoothed weight vectors describing the music pieces, and those with smallest distance to the query vector are returned.

Nürnberger and Detyniecki present in [66] a variant of the *Self-Organizing Map* (SOM) [38] 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.

In [99] Xue et al. present 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). Although not explicitly targeted at music retrieval, the idea of integrating data about the user, his peer group, and global data to build a social retrieval model might be worth considering for MIR purposes.

User-Aware Music Recommendation

Baltrunas et al. present a user-aware music recommender system for usage in cars [7]. They aim at learning relations between user aspects and music genres. As contextual aspects, Baltrunas et al. look into driving style, road type, landscape, sleepiness, traffic conditions, mood, weather, and time of day. Using a Web-based tool, the authors first assess in a user study which of these contextual aspects influence the preference for music of a particular genre, either in a positive of negative way. According to the study, driving style strongly influences the choice for music from the genres Blues, Classical, and Metal, whereas sleepiness seems to foster the decision for Pop, Country, and Reggae music. Furthermore, Baltrunas et al. investigate the impact of user context on user ratings and found that in most cases the awareness of a particular contextual situation had a negative effect on the ratings. The most significant (negative) influence on user ratings had the conditions "sleepy" and "traffic jam". The authors of [7] then propose a music recommendation approach that employs an extended Matrix Factorization [39] algorithm to predict item ratings. Their model includes contextual condition and genre vectors.

Bogdanov et al. [13] present a system which automatically generates recommendations from a user's musical preferences, given her/his accounts on popular online music services. Using these services, the system retrieves a set of tracks preferred by a user, and further tries to infer a semantic description of musical preferences from raw audio information. Thereafter, the system generates music recommendations, using a semantic music similarity measure.

Even though no detailed information on their approach is publicly available, *last.fm* [43] 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. Another commercial example employing a *collaborative filtering* (CF) [14] approach can be found in *amazon.com*'s music Web store [1]. Again, no details of the exact approach are publicly available.

2.4 Further Remarks

Having presented the three basic feature categories (music content, music context, and user context), we would like to note that there is an overlap between some of these. Indeed, particular features cannot only be assigned to one group, but combine aspects of several categories. For example, song lyrics are in principal music content. However, even state-of-the-art techniques do not allow for converting sung lyrics into textual representations from the audio signal, or even to derive some kind of higher level meaning. On the other hand, several lyrics portals on the Web (music context sources) offer such textual representations. Another example is similarity measures based on collaborative filtering. They are music context-related in the sense that the process is collaborative, however CF is used for personalizing a music recommendation to a user or a group of users, hence it takes into account the user context.

3 Important Aspects for Personalized Music Retrieval

Traditionally, evaluating music retrieval approaches focused on the concept of musical similarity, meaning that the performance of a retrieval system is judged the better the more similar the returned pieces are to a given seed. Although this is a very intuitive manner of assessment, it does not take into account that the information need of the user might be different. Indeed, for many common and popular MIR tasks, such as automated playlist generation and music recommendation, the listener does not necessarily want to be offered a list of closest matches in terms of acoustic similarity, as usually given by today's content-based music recommenders. User studies focusing on the perceived quality of automated, content-based playlist generation [71, 50] showed that playlists with items that were acoustically very similar were often deemed too perfect or homogeneous, and thus boring. In addition, users were shown to judge playlist items differently based on the amount of (metadata) information accompanying the playlist item [9, 50].
We therefore believe that a new generation of user-aware music retrieval systems should not only focus on traditional similarity scores derived via applying audio signal processing techniques, but also take other factors, including information from different modalities, into account. More precisely, such factors include the following:

Similarity

Similarity relations in various dimensions should be taken into account. One set of dimensions might be based on music properties such as rhythm, harmony, or timbre, inferred from the audio signal; another might take into account the resemblance according to other data sources, such as collaborative tags, playlist co-occurrences, or even images of album covers. A third set of dimensions might be learned from a user's listening preferences, for example, by relating certain properties of the user context to particular categories of music. To give an example, similarity could be defined as pieces that are usually listened together while a user is jogging or while being together with friends.

Moreover, the user's preferred music material should also influence the features and their relevance for similarity computation. For instance, a retrieval system focusing in classical music would need musically meaningful descriptors and similarity measures, while in a retrieval scenario of mainstream popular music timbre can be informative enough for distinguishing different types of music.

Diversity

Although the items in the results set of a music retrieval request should be similar, they should also reveal a certain degree of diversity. For example, there is the well-known "album" effect [95], i.e., due to same recording settings, tracks on one and the same album usually show a higher level of similarity than other tracks (even by the same artist). To alleviate this issue, some retrieval systems filter results from the same album or even by the same artist as the seed. Producing a well diversified result set for a given query is thus a common requirement for IR systems.

Familiarity/Popularity vs. Hotness/Trendiness

These four terms or aspects are related to each other. Familiarity or popularity describes how well-known an artist or song is, whereas hotness or trendiness relates to the amount of buzz or attention an artist is currently getting [41]. Popularity has a more positive connotation than the neutral expression of familiarity. However, we will use the terms interchangeably in the remainder of the paper, likewise the terms hotness and trendiness. In terms of temporal aspects, popularity can be seen as a longer lasting property, whereas hotness usually relates to recent appreciation of typically shorter duration, although hot artists might also be very familiar/popular to many people. To give an example, "The Beatles" are certainly popular, whereas "Lady Gaga" currently tends to rank higher on the hotness dimension.

Recentness

This aspect distinguishes recently released pieces from pieces that are older and therefore have a longer (playing) history. In contrast to the aspect of hotness, novelty does not require an artist to be recently popular, just a temporal closeness to the present.

Novelty

This aspect describes whether a music item is novel to the user of the system. If a music recommender keeps on suggesting tracks/artists well-known to the user, he or she will not be satisfied, even if the recommended items are perfectly suited otherwise. Hence, presenting novel recommendations is a vital requirement for a personalized recommender system.

Serendipity

Serendipity is a requirement often mentioned in the context of recommender systems. It means that a user is surprised in a positive way since he discovered an item he did not expect. In the context of music retrieval, we believe that the listener's music preference and taste as well as aspects of artist and song popularity have to be taken into account when we aim at providing serendipitous results. For instance, a fan of medieval folk metal might be rather disappointed and bored if the system recommends the band "Saltatio Mortis", which is very well known for this style of music. In contrast, for a user occasionally enjoying "Metallica" but also "Bob Dylan", the former mentioned band may be a serendipitous recommendation.

Apart from the listener's music preference and taste, a user profile for a serendipitous recommendation algorithm should take into account different categories of users as well as their different cultural backgrounds. For instance, music perception of musicians is likely to be quite dissimilar to that of music experts and editors, which is again different from untrained, passive listeners.

Transparency

For the acceptance of user-aware music retrieval systems it is crucial how the results are presented and explained. The presentation and explanation should be adapted to the users' musical training and preferences. For instance, the system should provide clues about why certain songs have been retrieved: "These two songs are similar because they share the same harmonic progression, the same tempo, are from the same artists, were recorded by the same producer" or "This song was suggested because you are currently in an aggressive mood while driving your car", or even "This was your favorite song during the Summer you met your future spouse".

4 Adaptive Music Similarity Measures

Users of MIR systems may have a varying (musical) background and experience music in different ways. Consequently, when comparing musical pieces with each other, opinions may diverge. Moreover, different retrieval tasks may also require different views on music similarity. In order to support individual user perspectives and multiple retrieval tasks, an adaptable model of music similarity is required. Often, (dis-)similarity is modeled by a distance measure. Either way, parameters need to be introduced that allow to adapt the measure.

Direct Manipulation (Adaptability)

Depending on how complex the resulting model is, users may be able to manually adjust and tweak the parameters according to their needs. For instance, Baumann et al. [12] describe a joystick interface to control the weights of three similarity facets in a linear combination. From a study with 10 users, it was concluded that users tend to use nearly similar joystick settings throughout different environments for finding a set of similar songs given an anchor song. Though the joystick interface was considered very intuitive by the users, it is unclear whether it may be applied to more than three similarity facets. Similarly, the *E-Mu Jukebox* described by Vignoli et al. [93] allows changing the similarity function that is applied to create a playlist from a seed song. Here, five similarity facets (sound, tempo, mood, genre and year) are visually represented by adapters that can be dragged on a bull's eye. The closer a facet is to the center, the higher is its weight in the similarity computation. Again, a linear weighting scheme is used here. This interface is to some extent scalable with respect to the number of facets but less intuitive. Indeed, a user study with 22 participants showed that the interface is harder to use, but more useful compared to two control systems.

With an increasing number of facets, direct manual manipulation is likely to become more difficult – even for a simple similarity model such as weighted linear combination. Moreover, specific similarity preferences often exist only subconsciously and thus are hard to specify explicitly. Instead of asking the user to explicitly state how he compares music, adaptive MIR systems aim to learn suitable parameter settings from ground truth data (such as expert annotations) or in an interactive way from user feedback.

Query and Relevance Feedback

The content-based MIR system for symbolic music described by Rolland [75] adjusts its similarity model based on feedback received during successive interactions with the user (search sessions). To model the similarity between a transcribed query and a melody, the concept of *pairings* is introduced: A pairing is a part of an alignment (between query and melody) that may comprise several notes and rests. Pairings can be classified into types and for each type, a weight is defined that specifies the importance of the pairing type in the similarity computation. In a ranked list of search results, the user can point out the correct match and optionally some reasonable secondary matches. Given this feedback, the weight for each pairing type is reinforced by a constant update factor if it contributes more to the similarity in the correct match than in the higher ranked false matches or otherwise decreased respectively. This way, the system can adapt to the user's way of comparing melodies.

The *MUSIPER* system developed by Sotiropoulos et al. [85] constructs music similarity perception models of its users. To this end, users are asked to specify the degree of similarity for retrieved music pieces. The system uses this relevance feedback to train several Radial Basis Function Networks (RBFN) – a special form of neural network – in parallel. Each RBFN represents a different similarity measure based on a different (content-based) feature subset. The model parameters that are adapted during learning are the internal weights of the networks. Finally, the network (and the respective feature subset) which best approximates the similarity ratings specified by the user is selected. The authors report significant improvement of perceived similarity in subsequent music retrievals during an evaluation with 100 participants and argue that the relation between subsets of features and personalized music similarity could be verified.

Collection Clustering

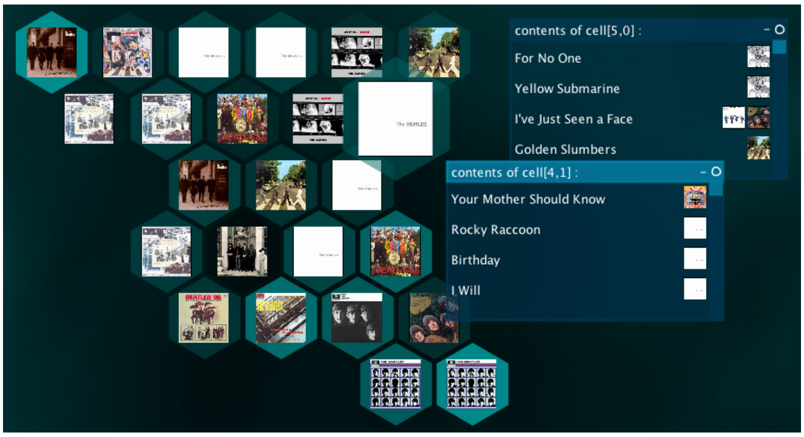
Slaney et al. [84] apply several algorithms based on second-order statistics (whitening, Linear Discriminant Analysis (LDA) [23], Relevant Component Analysis (RCA) [8]) and optimization techniques (Neighborhood Component Analysis (NCA) [26], Large-Margin Nearest Neighbor (LMNN) [94]) to learn Mahalanobis distance metrics for clustering songs by artist, album or blog they appear on. For the optimization, an objective function that mimics the k-nearest neighbor leave-one-out classification error is chosen. Songs are represented as vectors containing various acoustic features. From their experiments, the authors conclude that all algorithms lead to a significant improvement over the baseline. In particular, NCA and RCA showed higher robustness with (artificially generated) noisy features.

The *BeatlesExplorer* [87] (Figure 3, top) is a prototype system for organization and exploration of music collections that adapts to the user's perceived similarity in that it learns weights for different aspects of music similarity. Initially, a growing Self-Organizing Map (SOM) is induced that clusters the music collection. The user has then the possibility to change the location of songs on the map by simple drag-and-drop actions. Each movement of a song causes a weight change in the underlying similarity measure based on a quadratic programming scheme. As a result, the location of other songs may be modified as well. Experiments simulating user interaction with the system show, that during this stepwise adaptation the similarity measure indeed converges to one that captures how the user compares songs.

The *SoniXplorer* [57] shown in Figure 3 (bottom) is another SOM-based system that also adapts a weighted linear combination of basic similarities. Here, the SOM is displayed as video-game-like virtual 3-D landscape accompanied by spatialized playback of songs. Apart from moving songs on the map, the user can raise or lower the terrain to increase or decrease barriers between regions. For the adaptation, a target distance matrix is derived from the arrangement. Then a linear regression learner adapts the weighting accordingly.

Metric Learning with Relative Distance Constraints

In many publications, adapting music similarity is considered as a metric learning problem subject to so-called *relative distance constraints*. A relative distance constraint (s, a, b) demands that the object a is closer to the seed object s than object b, i.e., $d(s, a) < d(s, b)$. Such constraints can be seen as atomic bits of information fed to the adaptation algorithm. They can be derived from a variety of higher-level application-dependent constraints. For



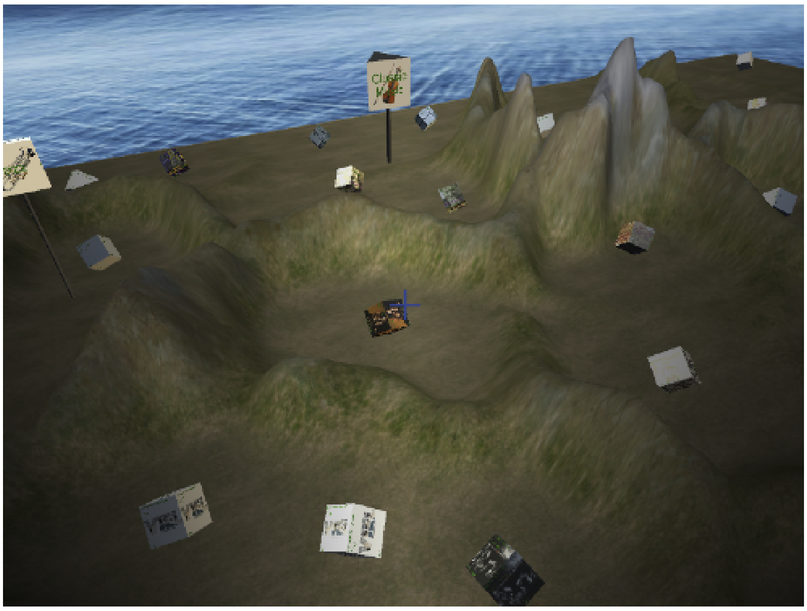


Figure 3 Prototype interfaces for music collection structuring w.r.t. user-adaptive similarity. Top: BeatlesExplorer [87]. Bottom: SoniXplorer [57].

instance, if the user moves a song s from one cluster to a different one in the *BeatlesExplorer* described above, this can be interpreted by the following set of relative distance constraints:

$$d(s, c_t) < d(s, c) \qquad \forall c \in C \setminus \{c_t\}$$

where C is the set of cluster cells of the SOM (each represented by a prototype) and c_t is the target cluster of the user's drag-and-drop action. Bade et al. describe how relative distance constraints can be derived from expert classifications of folk songs [4] or from an existing personal hierarchy of folders with music files [5]. Alternatively, it is also possible to ask the users directly to state the opinion for a triplet of songs as in the bonus round of the *TagATune* game [47]. McFee et al. [64] use artist similarity triples collected in the web survey described by Ellis et al. [24]. They further describe a graph-based technique to detect and remove inconsistencies within sets of constraints such as direct contradictions.

Using relative distance constraints, the task of learning a suitable adaptation of a similarity measure can be formulated as constraint optimization problem. Approaches are manifold and very much depend on the underlying adaptable model of similarity and its parameters. McFee et al. [64] apply a partial order embedding technique that maps artists into multiple non-linear spaces (using different kernel matrices), learns a separate transformation for each kernel, and concatenates the resulting vectors. The Euclidean distance in the resulting embedding space corresponds to the perceived similarity. In further work [62], they use the metric learning to rank (MLR) technique [61] – an extension of the Structural SVM approach [33] – to adapt a Mahalanobis distance according to a ranking loss measure. This approach is also applied by Wolff et al. [97] whose similarity adaptation experiments are based on the *MagnaTagATune* dataset derived from the *TagATune* game [47].

Instead of adapting a Mahalanobis distance, the work of Stober et al. focuses on simpler linear combination models. In [86], they describe various applications and respective adaptation algorithms which they evaluate and compare in [88] also using the *MagnaTagATune* dataset. Their distance model, which is a weighted sum of m facet distances $\delta_{f_1}, \dots, \delta_{f_m}$, is less expressive because of fewer parameters than the Mahalanobis distance but it can easily be understood and directly manipulated by the user. This design choice specifically addresses the users' desire to remain in control and not to be patronized by an intelligent system that "knows better". Furthermore, this similarity model allows to reformulate the metric learning task as a binary classification problem as described by Cheng et al. [20], which creates the possibility to apply a wide range of sophisticated classification techniques such as SVM. As Figure 4 illustrates, the idea is to rewrite each relative distance constraint $d(s, a) < d(s, b)$ as

$$\sum_{i=1}^{m} w_i(\delta_{f_i}(s, b) - \delta_{f_i}(s, a)) = \sum_{i=1}^{m} w_i x_i = \mathbf{w}^T \mathbf{x} > 0$$

where x_i is the *distance difference* w.r.t. facet f_i. The positive training example $(\mathbf{x}, +1)$ then represents the satisfied constraint whereas the negative example $(-\mathbf{x}, -1)$ represents its violation (i.e., inverting the relation sign). For these training examples, the normal vector of the hyperplane that separates the positive and negative instances contains the adapted facet weights.

5 Novelty and Serendipity in Music Recommendation

The ability to recommend "interesting new music" is considered an important social factor inside communities, especially among groups of young users (and groups of musicians). In

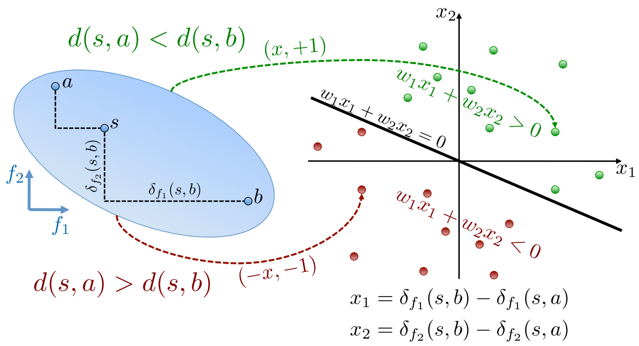


Figure 4 Transformation of a relative distance constraint for linear combination models into two training instances of the corresponding binary classification problem as described in [20].

this context, we use the generic term "music" to address different kinds of recommendations, from individual songs, albums, bands, or sub-genres. A good human recommender takes into account two main components to highlight his role as a music connoisseur:

- He is the first who is aware of music that the others do not know yet, although it is part of their music genre of interest and thus it is likely that sooner or later this music would have been found also without the recommendation.
- He discovers music that might be enjoyed by others, disregarding some aspects of the music content and context that would have suggested the opposite.

In the former case, the emphasis is on the *novelty* of the recommendations, where the role of the human recommender is related to his/her ability to mine music collections and to be at the same time up-to-date with the music market. In the latter case, the emphasis is on *serendipity* because the human recommender can prove his ability to find unexpected relations between music content, pointing towards music that will not be known without his/her recommendation.

Obviously, automatic recommender systems do not have to establish their role inside a community, yet these considerations about what motivates a human recommendation can be a starting point in the development of recommender systems that take into account both novelty and, more important, serendipity. This approach can take advantage of the fact that the user who receives the recommendations can evaluate them also considering how his role in the community will be affected by receiving given recommendations.

From this point of view, the concept of novelty may be extended to include also the process of finding new music. For instance, a user who has in his profile an interest for the recent work of a particular rock band can give a low value to the recommendation of a novel song taken from the band's first recorded album, which can be easily found in any catalogue and a high value to the recommendation of a novel song by another band where some of the musicians he likes appear as guest stars. According to these considerations, the novelty of an item can be measured depending also on the difficulties that a user would encounter to retrieve that particular item in a search session.

Also the concept of serendipity can be partially reconsidered depending on how human recommendations are provided. A central role is played by the fact that the user would not expect to like the recommended music item, because its average characteristics place it far from his listening profile. In order to enjoy the recommended item, the user is required to concentrate on a reduced set – maybe a single aspect – of the music dimensions that characterize it. For instance, a serendipitous experience for a user with a special interest for classical music for flute is to discover that many background music in movies of the 1970s is played on the flute. Or a serendipitous experience for a user interested in rock music with strong rhythm is to discover Scottish music for drums only.

According to these considerations, serendipity can be related to the ability of selectively suppress some dimensions of music content and context while recommending a list of music items. As a side note, perhaps one of the reasons why pure text-based search systems are still very popular among users of music recommender systems is that they suppress the information which is not explicitly represented in tags and metadata, thus promoting this aspect of serendipity.

6 Conclusions

The contribution of this article is threefold. First, we presented a broad categorization of aspects that influence human music perception, namely computational features related to music content, to music context, and to user context. We briefly reviewed the state-of-the-art in extraction and use of features in each category. Second, we proposed several aspects to take into account when elaborating user-aware music retrieval systems, more precisely, similarity, diversity, familiarity, hotness, recentness, novelty, serendipity, and transparency. Eventually, we thoroughly reported on recent developments in research on adaptive music similarity measures and music recommendation focusing on novelty and serendipity aspects.

We believe that a lot of research is still needed to understand the mechanisms involved in the perception of music similarity according to the three broad categories of aspects. Investigating the relations between computational features and human music perception will eventually pave the way to personalized, user-aware music retrieval systems and therefore is a research endeavor worth pursuing.

References

1 `http://www.amazon.com/music` (access: January 2010).

2 Jean-Julien Aucouturier and François Pachet. Improving Timbre Similarity: How High is the Sky? *Journal of Negative Results in Speech and Audio Sciences*, 1(1), 2004.

3 Claudio Baccigalupo, Enric Plaza, and Justin 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 Korinna Bade, Jörg Garbers, Sebastian Stober, Frans Wiering, and Andreas Nürnberger. Supporting folk-song research by automatic metric learning and ranking. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 741–746, Utrecht, the Netherlands, August 2010.

5 Korinna Bade, Andreas Nürnberger, and Sebastian Stober. Everything in its right place? learning a user's view of a music collection. In *Proceedings of NAG/DAGA 2009, International Conference on Acoustics, Rotterdam*, pages 344–347, 2009.

6 David Bainbridge, Brook J. Novak, and Sally Jo Cunningham. A user-centered design of a personal digital library for music exploration. In *Proceedings of the 2010 Joint Conference on Digital Libraries (JCDL '10)*, pages 149–158, 2010.

7 Linas Baltrunas, Marius Kaminskas, Bernd Ludwig, Omar Moling, Francesco Ricci, Karl-Heinz Lüke, and Roland Schwaiger. InCarMusic: Context-Aware Music Recommendations in a Car. In *International Conference on Electronic Commerce and Web Technologies (EC-Web)*, Toulouse, France, Aug–Sep 2011.

8 Aharon Bar-Hillel, Tomer Hertz, Noam Shental, and Daphna Weinshall. Learning a mahalanobis metric from equivalence constraints. *Journal of Machine Learning Research*, 6(1):937, 2006.

9 Luke Barrington, Reid Oda, and Gert Lanckriet. Smarter than Genius? Human Evaluation of Music Recommender Systems. In *Proc. ISMIR*, pages 357–362, October 2009.

10 Mathieu Barthet and Simon Dixon. Ethnographic observations of musicologists at the British Library: implications for Music Information Retrieval. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pages 353–358, Miami, USA, October 2011.

11 Mark A. Bartsch and pages=15–18 Gregory H. Wakefield, year=2001. To Catch a Chorus: Using Chroma-based Representations for Audio Thumbnailing. In *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics 2001*, October.

12 S. Baumann and J. Halloran. An ecological approach to multimodal subjective music similarity perception. In *Proceedings of 1st Conference on Interdisciplinary Musicology (CIM'04)*, 2004.

13 Dmitry Bogdanov, Martín Haro, Ferdinand Fuhrmann, Anna Xambó, Emilia Gómez, and Perfecto Herrera. A Content-based System for Music Recommendation and Visualization of User Preferences Working on Semantic Notions. In *9th International Workshop on Content-based Multimedia Indexing (CBMI 2011)*, Madrid, Spain, 2011.

14 John S. Breese, David Heckerman, and Carl 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.

15 Juan José Burred and Alexander 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.

16 Pedro Cano and Markus 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.

17 Michael A. Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. Content-Based Music Information Retrieval: Current Directions and Future Challenges. *Proceedings of the IEEE*, 96:668–696, April 2008.

18 Òscar Celma, Miguel Ramírez, and Perfecto 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.

19 Wei Chai and Barry 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.

20 Weiwei Cheng and Eyke Hüllermeier. Learning similarity functions from qualitative feedback. In *Proceedings of the 9th European Conference on Advances in Case-Based Reasoning (ECCBR'08)*, pages 120–134, 2008.

21 Sally Jo Cunningham, J. Stephen Downie, and David Bainbridge. "The Pain, The Pain": Modelling Music Information Behavior And The Songs We Hate. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pages 474–477, London, UK, September 11–15 2005.

22 Sally Jo Cunningham, Nina Reeves, and Matthew Britland. An Ethnographic Study of Music Information Seeking: Implications for the Design of a Music Digital Library. In

Proceedings of the 2003 Joint Conference on Digital Libraries (JCDL '03), pages 5–16, 2003.

23 J. Duchene and S. Leclercq. An optimal transformation for discriminant and principal component analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(6):978–983, 1988.

24 Daniel P.W. Ellis, Brian Whitman, Adam Berenzweig, and Steve 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.

25 Gijs Geleijnse, Markus Schedl, and Peter 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.

26 J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov. Neighbourhood components analysis. In *Advances in Neural Information Processing Systems (NIPS)*, 2004.

27 Emilia Gómez. *Tonal Description of Music Audio Signals*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2006.

28 Fabien Gouyon, Perfecto Herrera, Emilia Gomez, Pedro Cano, Jordi Bonada, Alex Loscos, Xavier Amatriain, and Xavier Serra. Content processing of music audio signals. In Pietro Polotti and Davide Roccheso, editors, *Sound to Sense, Sense to Sound: A State-of-the-art in Sound and Music Computing*, pages 83–160. Logos Verlag, Berlin GmbH, 2008.

29 Fabien Gouyon, François Pachet, and Olivier 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.

30 Sten Govaerts and Erik 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.

31 E. Guaus. *Audio Content Processing for Automatic Music Genre Classification: Descriptors, Databases, and Classifiers*. PhD thesis, Universitat Pompeu Fabra, 2009.

32 Martín Haro, A. Xambó, F. Fuhrmann, D. Bogdanov, E. Gómez, and P. Herrera. The Musical Avatar - A Visualization of Musical Preferences by Means of Audio Content Description. In *5th Audio Mostly Conference: A Conference on Interaction with Sound*, Piteå, Sweden, September 2010.

33 T. Joachims. A support vector method for multivariate performance measures. In *International Conference on Machine Learning (ICML)*, pages 377–384, 2005.

34 Peter Knees, Elias Pampalk, and Gerhard 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.

35 Peter Knees, Tim Pohle, Markus Schedl, and Gerhard 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.

36 Peter Knees, Markus Schedl, and Gerhard 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.

37 Peter Knees and Gerhard 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.

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

39 Yehuda Koren. Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 426–434, Las Vegas, NV, USA, August 2008.

40 Jan Korst and Gijs Geleijnse. Efficient lyrics retrieval and alignment. In Wim Verhaegh, Emile Aarts, Warner ten Kate, Jan Korst, and Steffen Pauws, editors, *Proceedings of the 3rd Philips Symposium on Intelligent Algorithms (SOIA 2006)*, pages 205–218, Eindhoven, the Netherlands, December 6–7 2006.

41 Paul Lamere. Artist similarity, familiarity and hotness. `http://musicmachinery.com/2009/05/25/artist-similarity-familiarity-and-hotness` (access: September 2011).

42 Audrey Laplante. *Everyday life music information-seeking behaviour of young adults: an exploratory study*. PhD thesis, McGill University, Montréal, Canada, 2008.

43 `http://last.fm` (access: October 2011).

44 C. Laurier and P. Herrera. *Automatic Detection of Emotion in Music: Interaction with Emotionally Sensitive Machines*, chapter 2, pages 9–32. IGI Global, 2009.

45 Cyril Laurier, Jens Grivolla, and Perfecto 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.

46 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.

47 Edith Law and Luis von Ahn. Input-agreement: a new mechanism for collecting data using human computation games. In *Proceedings CHI '09*, pages 1197–1206, 2009.

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

49 Jin Ha Lee. Analysis of user needs and information features in natural language queries seeking user information. *Journal of the American Society for Information Science and Technology (JASIST)*, 61:1025–1045, 2010.

50 Jin Ha Lee. How Similar Is Too Similar?: Exploring Users' Perceptions of Similarity in Playlist Evaluation. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pages 109–114, Miami, USA, October 2011.

51 Mark Levy and Mark 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.

52 Dongge Li, Ishwar K. Sethi, Nevenka Dimitrova, and Tom McGee. Classification of General Audio Data for Content-based Retrieval. *Pattern Recognition Letters*, 22(5):533–544, 2001.

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

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

55 Beth Logan, Daniel P.W. Ellis, and Adam 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.

56 Beth Logan, Andrew Kositsky, and Pedro 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.

57 Dominik Lübbers and Matthias Jarke. Adaptive multimodal exploration of music collections. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pages 195–200, Kobe, Japan, October 2009.

58 `http://mtg.upf.edu/project/musicalavatar` (access: October 2011).

59 Michael I. Mandel and Daniel P.W. 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.

60 Rudolf Mayer, Robert Neumayer, and Andreas 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.

61 B. McFee and G. R. G. Lanckriet. Metric learning to rank. In *Proceedings of the 27th International Conference on Machine Learning (ICML'10)*, 2010.

62 Brian McFee, Luke Barrington, and G.R.G. Lanckriet. Learning similarity from collaborative filters. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 345–350, Utrecht, the Netherlands, August 2010.

63 Brian McFee and Gert 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.

64 Brian McFee and Gert Lanckriet. Heterogeneous embedding for subjective artist similarity. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pages 513–518, Kobe, Japan, October 2009.

65 M. Müller, D.P.W. Ellis, A. Klapuri, and G. Richard. Signal processing for music analysis. *IEEE Journal on Selected Topics in Signal Processing*, 5(6):1088–1110, October 2011.

66 Andreas Nürnberger and Marcin Detyniecki. Weighted Self-Organizing Maps: Incorporating User Feedback. In Okyay Kaynak and Erkki 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.

67 François Pachet, Gert Westerman, and Damien 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.

68 Elias Pampalk. *Computational Models of Music Similarity and their Application to Music Information Retrieval*. PhD thesis, Vienna University of Technology, March 2006.

69 Elias Pampalk, Andreas Rauber, and Dieter 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.

70 Tim Pohle. *Automatic Characterization of Music for Intuitive Retrieval*. PhD thesis, Johannes Kepler University Linz, Linz, Austria, 2009.

71 Tim Pohle, Peter Knees, Markus Schedl, Elias Pampalk, and Gerhard Widmer. "Reinventing the Wheel": A Novel Approach to Music Player Interfaces. *IEEE Transactions on Multimedia*, 9:567–575, 2007.

72 Tim Pohle, Peter Knees, Markus Schedl, and Gerhard 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.

73 G. Poliner, D. Ellis, A. Ehmann, E. Gómez, S. Streich, and B. Ong. Melody transcription from music audio approaches and evaluation. *IEEE Transactions on Audio, Speech and Language Processing*, 15:1247–1256, 2007.

74 Joseph J. Rocchio. Relevance Feedback in Information Retrieval. In Gerard Salton, editor, *The SMART Retrieval System - Experiments in Automatic Document Processing*, pages 313–323. Englewood Cliffs, NJ: Prentice-Hall, 1971.

75 P.Y. Rolland. Adaptive user modeling in a content-based music retrieval system. In *Proceedings of the 2nd International Conference on Music Information Retrieval (ISMIR 2001)*, Bloomington, Indiana, USA, October 2001.

76 Markus Schedl and Peter 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.

77 Markus Schedl, Peter Knees, and Gerhard 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.

78 Markus Schedl, Elias Pampalk, and Gerhard Widmer. Intelligent Structuring and Exploration of Digital Music Collections. *e&i - Elektrotechnik und Informationstechnik*, 122(7–8):232–237, July–August 2005.

79 Markus Schedl, Tim Pohle, Peter Knees, and Gerhard 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.

80 Markus Schedl, Cornelia Schiketanz, and Klaus 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.

81 Markus Schedl, Klaus Seyerlehner, Dominik Schnitzer, Gerhard Widmer, and Cornelia 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.

82 Eric Scheirer and Malcolm 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.

83 Yuval Shavitt and Udi 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.

84 Malcolm Slaney, Kilian Q. Weinberger, and William White. Learning a metric for music similarity. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pages 313–318, Philadelphia, PA, USA, September 2008.

85 Dionysios N. Sotiropoulos, Aristomenis S. Lampropoulos, and George A. Tsihrintzis. Musiper: a system for modeling music similarity perception based on objective feature subset selection. *User Modeling and User-Adapted Interaction*, 18(4):315–348, 2008.

86 Sebastian Stober. Adaptive distance measures for exploration and structuring of music collections. In *Proceedings of AES 42nd Conference on Semantic Audio*, 2011.

87 Sebastian Stober and Andreas Nürnberger. Towards user-adaptive structuring and organization of music collections. In *Proceedings of the 6th international workshop on Adaptive Multimedia Retrieval (AMR'08)*, 2008.

88 Sebastian Stober and Andreas Nürnberger. An experimental comparison of similarity adaptation approaches. In *Proceedings of 9th International Workshop on Adaptive Multimedia Retrieval (AMR'11)*, 2011.

89 Sebastian Stober, Matthias Steinbrecher, and Andreas Nüurnberger. A Survey on the Acceptance of Listening Context Logging for MIR Applications. In *Proceedings of 3rd Work-*

shop on Learning the Semantics of Audio Signals (LSAS 2009), Graz, Austria, December 2009.

90 Sebastian Streich. *Music Complexity: A Multi-faceted Description of Audio Content.* PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2007.

91 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.

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

93 Fabio Vignoli and Steffen Pauws. A music retrieval system based on user driven similarity and its evaluation. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pages 272–279, London, UK, September 2005.

94 K.Q. Weinberger, J. Blitzer, and L.K. Saul. Distance metric learning for large margin nearest neighbor classification. In *Advances in Neural Information Processing Systems (NIPS)*, 2006.

95 Brian Whitman, Gary Flake, and Steve Lawrence. Artist Detection in Mmusic with Minnowmatch. In *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing*, pages 559–568, Falmouth, MA, USA, September 10-12 2001.

96 Brian Whitman and Steve 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.

97 Daniel Wolff and Tillman Weyde. Combining sources of description for approximating music similarity ratings. In *Proceedings of the 9th International Workshop on Adaptive Multimedia Retrieval (AMR'11)*, 2011.

98 Wei Xu, Xin Liu, and Yihong 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.

99 Gui-Rong Xue, Jie Han, Yong Yu, and Qiang Yang. User Language Model for Collaborative Personalized Search. *ACM Transactions on Information Systems*, 27(2), February 2009.

100 Bingjun Zhang, Jialie Shen, Qiaoliang Xiang, and Ye 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.

101 Bingjun Zhang, Qiaoliang Xiang, Ye Wang, and Jialie 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.