

Chapter 14

Emotion-Based Matching of Music to Places

Marius Kaminskas and Francesco Ricci

Abstract Music and places can both trigger emotional responses in people. This chapter presents a technical approach that exploits the congruence of emotions raised by music and places to identify music tracks that match a place of interest (POI). Such technique can be used in location-aware music recommendation services. For instance, a mobile city guide may play music related to the place visited by a tourist, or an in-car navigation system may adapt music to places the car is passing by. We address the problem of matching music to places by employing a controlled vocabulary of emotion labels. We hypothesize that the commonality of these emotions could provide, among other approaches, the base for establishing a degree of match between a place and a music track, i.e., finding music that “feels right” for the place. Through a series of user studies we show the correctness of our hypothesis. We compare the proposed emotion-based matching approach with a personalized approach where the music track is matched to the music preferences of the user, and to a knowledge-based approach which matches music to places based on metadata (e.g., matching music that was composed during the same period that the place of interest was built in). We show that when evaluating the goodness of fit between places and music, personalization is not sufficient and that the users perceive the emotion-based music suggestions as better fitting the places. The results also suggest that emotion-based and knowledge-based techniques can be combined to complement each other.

M. Kaminskas (✉)

Insight Centre for Data Analytics, University College Cork, Cork, Ireland
e-mail: marius.kaminskas@insight-centre.org

F. Ricci

Faculty of Computer Science, Free University of Bozen-Bolzano, Bozen-Bolzano, Italy
e-mail: francesco.ricci@unibz.it

14.1 Introduction

Music is generally considered an emotion-oriented form of content—it creates an emotional response in the listener, and therefore can be described by the evoked emotions. For instance, most people will agree with a description of Rossini’s “William Tell Overture” as *happy* and *energetic*, and Stravinsky’s “Firebird” as *calm* and *anxious* [18]. This music typically suggests, triggers those emotions in listeners.

Similarly to music, places can also trigger emotional responses in visitors [4] and therefore are perceived as associated to their generated emotions. Moreover, certain types of places have been shown to have a positive effect on people’s emotional health [14]. This phenomenon is particularly relevant for tourism: places (destinations) have become objects of consumption (much like books, movies, or music in the entertainment domain). Tourists “gaze upon particular objects, such as piers, towers, old buildings, or countryside ... with a much greater sensitivity to visual elements of landscape and townscape than is normally found in everyday life” [39]. Therefore, emotional responses to places are particularly strong in tourists.

In our work, we envision a scenario where music can augment the experience of places by letting places “sound with music”, the music that is perceived by people as the “right” one for the place. This scenario is suggested by the observations made above, namely that both music and places are naturally associated with emotions, and therefore, given a place, one can identify music that is associated with the same emotions that are associated to the place.

Besides, music is strongly connected to places for cultural and social reasons. Music is a cultural dimension and a human activity that contributes to give meaning to a place. For instance, consider how important flamenco music is for a city like Seville in Spain, or opera compositions for Vienna in Austria. We all deem music as profoundly related to and contributing to the image of such destinations. In fact, many music compositions have been motivated or inspired by specific places. Consider for instance the impressionistic compositions by Claude Debussy, such as “La Mer” or “Preludes”. They are all dedicated to places, e.g., “La Puerta de Vino” (The Wine Gate) and “La Terrasse Des Audiences Du Clair De Lune” (The Terrace of Moonlit Audiences).

Based on the observations made above, it is meaningful to explore the relationships between music and places and in particular to find music that “sounds well for a place” (see Sect. 14.2.3). However, automatically finding music artists and compositions related to a given place is not a simple task for a computer program. It requires knowledge of both domains, and a methodology for establishing relations between items in the two domains, which is clearly a difficult problem to be solved automatically by an intelligent computer-based system [16, 25]. In this chapter, we describe a technical approach that exploits the congruence of emotion-based descriptions of music and places of interest (POIs) to establish a degree of match between a place and a music track.

The proposed technique can be applied in location-aware music recommendation services. For instance, a mobile city guide may provide an enhanced presentation of the POI visited by a tourist, and play music that is related, i.e., emotionally associated to the place (e.g., Mozart in Salzburg, or a Bach's fugue in a Gothic Cathedral). Other examples include a car entertainment and navigation system that adapts music to the place the car is passing by, or a tourism website where the information on travel destinations is enhanced through a matching music accompaniment. Such information services can be used to enhance the user's travel experience, to provide rich and engaging cultural information services, and to increase the sales of holiday destinations or music content.

This chapter describes the emotion-based matching of music to places which has been developed and evaluated through a series of user studies [5, 22]. As a first step of our research, we have shown that a set of emotions can be employed by users to describe both music tracks and POIs [5]. We have adopted tags to represent the emotional characteristics of music and POIs and used this common representation language to match and retrieve music that fits a place. We decided to use tags because of the popularity of social tagging and user-generated tagging data are still growing on the Web [35, 38]. Several web mining and machine learning techniques have been developed to handle available tagging data. These include tag-based content retrieval and recommendation solutions, which are relevant for implementing our target functionality of retrieving music matching a given place.

Subsequently, we have evaluated the proposed music-to-POI matching technique in a mobile guide prototype that suggests and plays music tracks while users are visiting POIs in the city of Bolzano (Italy) [5]. The evaluation results show that users agree with the music-to-POI matches produced using our technique.

Having validated the usefulness of emotion tags for matching music to places, we have implemented a machine learning technique to scale up the tagging process. In fact, in the first experiments the music was tagged manually by users and this requires a consistent user effort. So, in order to reduce this cost we have shown that one can employ state-of-the-art music auto-taggers, and evaluated the performance of our matching approach applied to automatically tagged music against alternative music matching approaches: a simple *personalized* approach and a *knowledge-based* matching technique which matches music to places based on metadata (e.g., matching music that was composed during the same period that the place of interest was built in) [22].

In the following section we provide background information on research that addresses emotions in the context of searching for music and places. Subsequently, we describe the proposed emotion-based matching approach and experiments conducted to validate its effectiveness. Finally, we discuss the position of our technique within music adaptation research and present some open issues in the area of emotion-aware and location-aware music services.

14.2 Background

14.2.1 Emotions and Music

Emotional qualities of music are studied in the area of music psychology [30]. Music psychologists have been studying how people perceive and are affected by emotions in music [17, 41], or how music preferences correlate with the user's state-of-mind [29, 31]. Likewise, this topic has attracted the interest of computer science researchers. The research area of music information retrieval (MIR)—a computer science branch dedicated to the analysis of music content—devoted considerable attention to automatic detection of emotions conveyed by music [40] (see also Chap. 12). This subject attracts researchers due to the large possibilities it opens for the development of music delivery services. For instance, emotion-aware music services may be used for searching music collections using emotion keywords as a query, or for recommending music content that fits the user's mood.

Automatic recognition of emotion-related descriptors of music tracks is a challenging research problem, which requires large collections of training data—music labeled with appropriate emotion tags—and therefore needs the definition of a vocabulary of possible emotions conveyed by music. However, the definite set of emotions raised by music is not easy to determine. Despite numerous works in cognitive psychology, up to date no universal taxonomy of emotions has been agreed on. Human emotions are complex and multilayered, therefore, focusing on different aspects of emotions leads to different lists of emotions. This means that there may not exist a universal emotion model to discover, but emotions are to be chosen based on the task and domain of the research [9]. Two main groups of emotion models have been identified [11]: *dimensional models*, where emotional states are represented as a combination of a small number of independent (usually two) dimensions, and *category-based models*, where sets of emotion terms are arranged into categories.

In *dimensional models*, the general idea is modeling emotions as a combination of activeness and positiveness of each emotion. Thus, the first dimension represents the *Activation* level (also called *Activity*, *Arousal* or *Energy*), which contains values between *quiet* and *energetic*; and the second dimension represents the *Valence* level (also called *Stress* or *Pleasure*), which contains values between *negative* and *positive*. The most popular dimensional emotion models are Russell's circumplex model [32] and Thayer's model [37].

Category-based models offer a simpler alternative to modeling emotions where emotion labels are grouped into categories. For instance, Hevner [17] conducted a study where users were asked to tag classical music compositions with emotion adjectives and came up with 8 emotion clusters: *dignified*, *sad*, *dreamy*, *serene*, *graceful*, *happy*, *exciting*, and *vigorous*. A more recent and elaborated research on emotional response to music was carried out by Zentner et al. [41] who conducted a series of large-scale user studies to determine which emotions are perceived and felt by music listeners. The study participants rated candidate adjectives by measuring how often they feel and perceive the corresponding emotions when listening to music.

The studies resulted in a set of 33 emotion terms in 9 clusters which was called the Geneva Emotional Music Scale (GEMS) model. The representative emotions of the clusters are: *wonder*, *transcendence*, *tenderness*, *nostalgia*, *peacefulness*, *power*, *joyful activation*, *tension*, and *sadness*. We have adopted the GEMS model as a starting point in our research.

Having a vocabulary of emotions for labeling music content allows researchers to build automatic music emotion recognition algorithms. This task is more generally referred to as *music auto-tagging* and is typically performed employing a supervised learning approach: based on a training set of music content features and a vocabulary of labels, a classifier is trained and subsequently used to predict labels for new music pieces. In our work, we employed a variant of the auto-tagger presented by Seyerlehner et al. [34] which showed superior performance in the “Audio Tag Classification” and the “Audio Tag Affinity Estimation” tasks, run at the 2012 Music Information Retrieval Evaluation eXchange (MIREX).¹

14.2.2 *Emotions and Places*

The concept of *place* and its meaning has attracted attention among both philosophers and psychologists. Castello [7] defined place as a part of space that stands out with certain qualities that are perceived by humans. The author claimed that places stand out in our surroundings not because of certain features that they possess, but because of the meanings we attribute to these features. As such, places can have a strong impact on people’s memories, sentiments, and emotional well-being.

Debord [10] introduced *psychogeography*—the study of the effects of geographical environment on the emotions and behavior of individuals. Bachelard [4], in his work on the *Poetics of Space* discussed the effect that a house and its outdoor context may have on human. The author described places that evoke impressions (i.e., emotions) in humans as *poetic* places.

Place is most commonly analyzed in the context of an urban environment [4, 7, 10, 14]. This is not surprising since most people live in cities and therefore interact with urban places on a daily basis. In our work, we focus on places of interest (POIs)—places that attract the interest of city dwellers and especially tourists. In the tourism domain, a POI is arguably the main consumption object [39] as tourists seek out churches, castles, monuments, old streets, squares in search of authenticity, culture, and the sense of place.

Although urban places and their impact on human emotional well-being have been studied by psychologists [7, 13, 14], to our knowledge no taxonomy of emotions that a place may evoke in people has been proposed.

¹<http://www.music-ir.org/mirex>.

14.2.3 *Places and Music*

Although it has been argued that sound is as important as the visual stimuli in contributing to the sense of a place [19], few works have addressed the use of music to enhance people's perception of a place.

In his work on the *sonification* of place, Iosafat [19] discussed the philosophical and psychological aspects of the relation between places, their sounds, and human perception of places. The author presented an *urban portrait* project, where field recordings of prominent places in a city were mixed with musicians' interpretations of the places to create a sonic landscape of the city.

Ankolekar and Sandholm [2] presented a mobile audio application *Foxtrot* which allows its users to explicitly assign audio content to a particular location. The authors stressed the importance of the emotional link between music and location. According to the authors, the primary goal of their system is to "enhance the sense of being in a place" by creating its emotional atmosphere. *Foxtrot* relies on crowdsourcing—every user is allowed to assign audio pieces (either a music track or a sound clip) to specific locations (represented by the geographical coordinates of the user's current location), and also specify the visibility range of the audio track—a circular area within which the track is relevant. The system is then able to provide a stream of location-aware audio content to its users.

While sounds can enhance the sense of place, also places may contribute to the listener's perception of music or even act as stimuli for creating music [1]. The US music duo Bluebrain is the first band to record a location-aware album.² In 2011, the band released two such albums—one dedicated to Washington's park National Mall, and the second dedicated to New York's Central Park. Both albums were released as iPhone apps, with music tracks prerecorded for specific zones in the parks. As the listener moves through the landscape, the tracks change through smooth transitions, providing a soundtrack to the walk. Only by listening to the albums in their intended surroundings can the users fully experience music as conceived by the artist.

In music recommendation research, where the goal is providing users with personalized music delivery services, the place of a user at the time of receiving recommendations has been recognized as an important factor which influences the user's music preferences [15]. The location of the user (along with other types of information such as weather or time) represents an additional source of knowledge which helps generating highly adaptive and engaging recommendations, known as situational or contextual recommendations [21].

Research works that exploit the user's location for music recommendation typically represent it as categorical data [8] or as GPS coordinates [33]. For instance, Cheng and Shen [8] created a dataset of (*user*, *track*, *location*) tuples where the *location* value was set as one of {*office*, *gym*, *library*, *canteen*, *public transport*}. The authors proposed a generative probabilistic model to recommend music tracks by taking the location information into account.

²<http://bluebrainmusic.blogspot.com/>.

Schedl et al. [33] used a dataset of geo-tagged tweets related to music listening for location-aware recommendation. For each user in the dataset, the set of her music listening tweets was aggregated into a single location profile. The authors proposed two ways to build the user's location profile—aggregating the GPS coordinates of the user's listening events into a single Gaussian Mixture Model representation, or converting the coordinates into categorical data at the level of continent, country, or state. The location profiles of users were then exploited for user-to-user similarity computation in a collaborative filtering algorithm.

The above approaches to location-aware music recommendation differ from our work in that they consider places either at a very high level (e.g., the country of a user), or as generic types of location (e.g., an office environment). Conversely, in our work we consider specific places, their emotional associations, and model explicit relations between a place and a music track by leveraging these emotional content of both types of items. To the best of our knowledge, no other research works model the relation between music and places at this level.

14.3 Matching Music to Places of Interest

In this section, we describe the proposed approach for matching music to places of interest. As said in Sect. 14.1, we use emotions as the link between music and POIs, and we represent the emotions as tags attached to both music and places. The tag-based representation allows matching music tracks and POIs by comparing the tag profiles of the items. This approach requires both music tracks and POIs to be annotated with a common tag vocabulary. For this purpose, we rely on both user-generated annotations and a music auto-tagging technique.

First we describe the procedure of establishing vocabulary of emotions fit for annotating both places and music. Subsequently, we show how an appropriate similarity metric for music-to-POI similarity computation was selected and then describe the live user study where our approach was evaluated [5].

Finally, we describe the extension of the approach using a revised emotion vocabulary, and present the user study where our approach was compared against alternative matching techniques [22].

For fully detailed descriptions of our work, we refer the readers to the original publications describing the technique [5, 22].

14.3.1 *Establishing the Emotion Vocabulary*

As discussed in Sect. 14.2, there has been a number of works on emotion vocabularies for music, but none addressing the vocabulary of emotions evoked by places. Therefore, as a starting point in our research, we chose to use a well-established vocabulary of emotions evoked by music—the Geneva Emotional Music Scale (GEMS)

Table 14.1 The tag vocabulary used for matching music to POIs

GEMS tags	Category
Allured, Amazed, Moved, Admiring	Wonder
Fascinated, Overwhelmed, Thrills, Transcendence	Transcendence
Mellowed, Tender, Affectionate, In love	Tenderness
Sentimental, Dreamy, Melancholic, Nostalgic	Nostalgia
Calm, Serene, Soothed, Meditative	Peacefulness
Triumphant, Energetic, Strong, Fiery	Power
Joyful, Animated, Bouncy, Amused	Joyful Activation
Tense, Agitated, Irritated	Tension
Sad, Tearful	Sadness
Additional tags	
Ancient, Modern	Age
Colorful, Bright, Dark Dull	Light and Color
Open, Closed	Space
Light, Heavy	Weight
Cold, Mild and Warm	Temperature

model [41]. The GEMS model consists of nine categories of emotions, each category containing up to four emotion tags (Table 14.1).

However, since our approach deals with both music and POIs, we could not rely solely on tags derived from a music psychology research. Therefore, in addition to the tags from GEMS model, we have selected five additional categories of tags (Table 14.1). These were selected from the “List of Adjectives in American English” [28] through a preliminary user study on music and POI tagging [20]. We note that although these additional adjectives describe physical properties of items, they relate to the users’ emotional response to music (e.g., a user may perceive a music composition being *cold* or *colorful*).

Subsequently, we conducted a user study to evaluate the fitness of the proposed vocabulary for uniformly tagging both music tracks and POIs and to bootstrap a dataset of POIs and music tracks with emotion annotations. Figure 14.1 shows the interface of the web application used in the experiment.

We created a dataset consisting of 75 music tracks—famous classical compositions and movie soundtracks—and 50 POIs in the city of Bolzano (Italy)—castles, churches, monuments, etc. The tagging was performed by 32 volunteer users recruited via email—students and researchers from the Free University of Bolzano and other European universities. Roughly half of the study participants had no prior knowledge of the POIs. The users were asked to view or listen to one item at a time (the same interface was used for tagging POIs and music tracks) and to assign the tags that in their opinion fit to describe the displayed item.

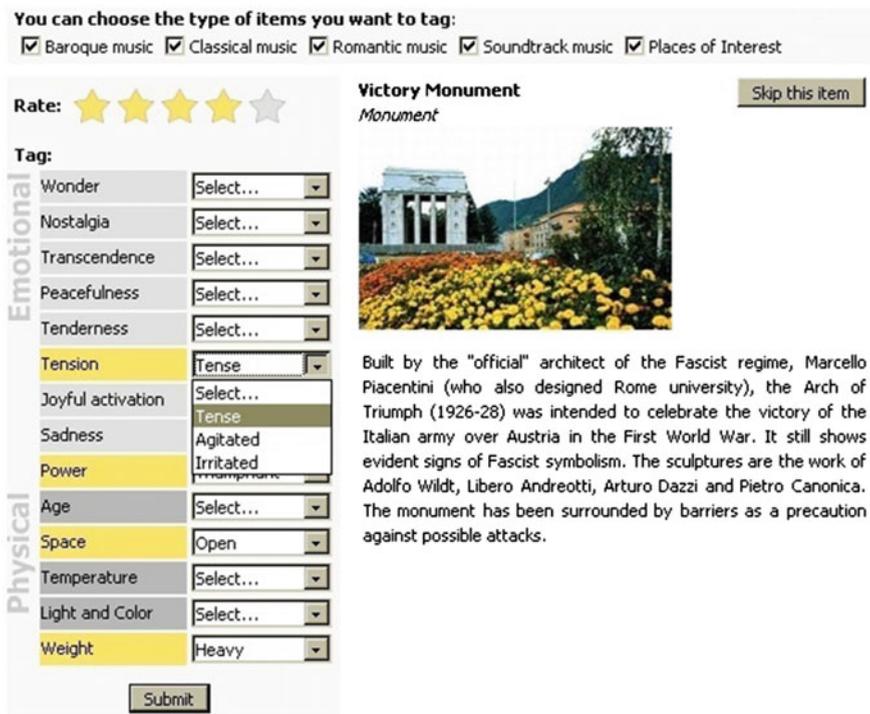


Fig. 14.1 Screenshot of the web application used for tagging POIs and music tracks

In total, 817 tags were collected for the POIs (16.34 tags per POI on average), and 1025 tags for the music tracks (13.67 tags per track on average). Tags assigned to an item by different users were aggregated into a single list, which we call the item’s *tag profile*. Note that by aggregating the tags of different users we could not avoid conflicting tags in the items’ profiles. This is quite normal when dealing with user-generated content. However, this does not invalidate the findings of this work. Conversely, we show that our approach is robust and can deal with such complication.

Figure 14.2 illustrates the distribution of tags collected for POIs (left) and music tracks (right). Larger font represents more frequent tags. We can see that the tag distributions for POIs and music tracks are different—music tends to be tagged with tags from the GEMS model more frequently (frequent tags include *affectionate, agitated, sad, sentimental, tender, triumphant*), while POIs tend to be labeled with adjectives describing the physical properties (e.g., *closed, cold*). This is not surprising, since the GEMS model was developed for the music domain. However, certain tags from the proposed vocabulary are uniformly applied to both types of items (e.g., *animated, bright, colorful, open, serene*). This result shows that a common vocabulary may indeed be used to link music and places.



Fig. 14.2 The tag clouds for POI (left) and music (right) annotations

14.3.2 Music-to-POI Similarity Computation

Having a dataset of POIs and music tracks tagged with a common emotion vocabulary, we need a method to establish a degree of similarity (relatedness) between the tagged items. To do so, we have considered a set of a well-established set of similarity metrics that are applicable to tagged resources [27] and performed a web-based user study to evaluate the alternative metrics. We have designed a web interface (Fig. 14.3) for

Victory Monument



Built by the "official" architect of the Fascist regime, Marcello Piacentini (who also designed Rome university), the Arch of Triumph (1926-28) was intended to celebrate the victory of the Italian army over Austria in the First World War. It still shows evident signs of Fascist symbolism. The sculptures are the work of Adolfo Wildt, Libero Andreotti, Arturo Dazzi and Pietro Canonica. The monument has been surrounded by barriers as a precaution against possible attacks.

Select the tracks that in your opinion fit the described location:

- R. Strauss - Also sprach Zarathustra - Fanf**
00:00 [progress bar] 00:00 [play/pause] [volume] [FMP3]
- Gluck - Dance of the Blessed Spirtis (from**
00:00 [progress bar] 00:00 [play/pause] [volume] [FMP3]
- Mendelssohn - Wedding March, from A Mid**
00:00 [progress bar] 00:00 [play/pause] [volume] [FMP3]
- Schumann - Traumerei**
00:00 [progress bar] 00:00 [play/pause] [volume] [FMP3]
- Tchaikovsky - Piano Concerto No. 1 in B fl**
00:00 [progress bar] 00:00 [play/pause] [volume] [FMP3]

Submit

Fig. 14.3 Screenshot of the web application used for evaluating music-to-POI matching

collecting the users' subjective evaluations, i.e., assessments if a music track suits a POI. The users were asked to consider a POI, and while looking at it, to listen to music tracks selected using the different similarity metrics. The user was asked to check all the tracks that in her opinion suit that POI.

The goal of this experiment was to see whether the users actually agree with the music-to-POI matching computed using our approach and to select the similarity metric that produces the best results, i.e., music suggestions that the users agree with. The obtained results showed the weighted Jaccard similarity metric to produce the best quality results (see [5] for a detailed description of the metrics' comparison). The weighted Jaccard metric defines the similarity score between a POI u and a music track v as:

$$\text{similarity}(u, v) = \frac{\sum_{t \in X_u \cap X_v} \log f(t)}{\sum_{t \in X_u \cup X_v} \log f(t)} \quad (14.1)$$

where X_u and X_v are the items' tag profiles and $f(t)$ is the fraction of items in our dataset (both POIs and music tracks) annotated with the tag t .

We note that in this evaluation study the users were asked to evaluate the matching of music to a POI while they were just reading a description of the POI. In order to measure the effect of the music-to-POI suggestions while the user is actually visiting the POI, we have implemented a mobile guide for the city of Bolzano and evaluated it in a live user study.

14.3.3 User Study: Mobile Travel Guide

This section describes the design and evaluation of an Android-based travel guide that illustrates the POI the user is close to and plays music suited for that POI. After the user has launched this application, she may choose a travel itinerary that is displayed on a map indicating the user's current GPS position, and the locations of the POIs in the itinerary (Fig. 14.4, left). Then, every time the user is nearby to a POI (either belonging to the selected itinerary, or not), she receives a notification alert conveying information about the POI. While the user is reading this information, the system plays a music track that suits the POI (Fig. 14.4, center). For example, the user might hear Bach's "Air" while visiting the Cathedral of Bolzano, or Rimsky-Korsakov's "Dance of the Bumble Bee" during a visit to the busy Walther Square.

The guide functions using a database of POIs and music tracks tagged as described in Sect. 14.3.1, as well as the precomputed similarity scores (Eq. 14.1) for each POI-music pair in the dataset. To suggest a music track for a given POI, the application sorts the music tracks by decreasing similarity score, and then randomly selects one of the top 3 music tracks. The motivation for not always choosing the top-scoring music track for each POI is to avoid, or at least reduce, the probability that the same music tracks are played for POIs that have been annotated with similar sets of tags, and therefore to ultimately suggest more diverse music tracks while the user is

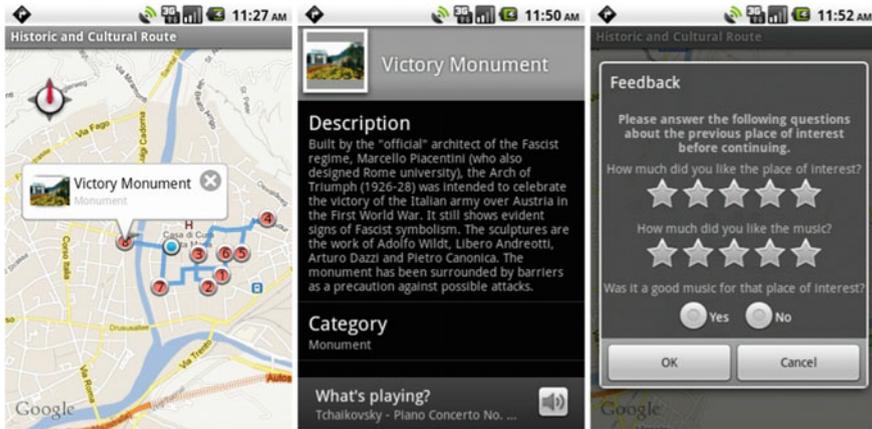


Fig. 14.4 Screenshots of the mobile guide application, showing the map view, the details of a POI, and a feedback dialog

visiting an itinerary. In the future, more sophisticated diversification techniques may be investigated.

In order to evaluate the proposed music-to-POI matching approach, we compared the performance of the guide with an alternative system variant having the same user interface, but not matching music with the POIs. Instead, for each POI, it suggests a music track that, according to our similarity metric, has a low similarity with the POI. We call the original system variant *match*, and the second variant *music*.

For the evaluation study we adopted a between-groups design, involving 26 subjects (researchers and students at the Free University of Bolzano). Subjects were assigned to the *match* and *music* variants in a random way (13 each). We note that the outcome of this comparison was not obvious, as without a careful analysis, even the low-matching tracks could be deemed suited for a POI, since all tracks belong to the same music type—popular orchestral music.

Each study participant was given a phone with earphones, and was asked to complete a 45 min walking route in the center of Bolzano. Whenever a subject was approaching a POI, a notification invited the user to inspect the POI's details and to listen to the suggested music track. If the suggested music track was perceived as unsuited, subjects could pick an alternative music track from a shuffled list of four possible alternatives: two randomly generated, and two with high music-to-POI similarity scores.

The users were then asked to provide feedback regarding the quality of music suggestions and their satisfaction with the system (Fig. 14.4, right). By analyzing the feedback, we were able to evaluate the performance of our technique against the baseline approach. A total of 308 responses regarding the various visited POIs and their suggested music tracks were obtained: 157 (51 %) from subjects in the *match* group, and 151 (49 %) from subjects in the *music* group.

The obtained results showed 77 % of the users in the *match* group to be satisfied with music suggestions, compared to 66 % in the *music* group. The difference in these proportions is statistically significant ($p < 0.001$ in a *chi-square* test). We can thus conclude that users evaluate the music tracks suggested by our proposed method to better suit the POIs than the music tracks suggested in the control setting.

Moreover, to additionally confirm this result, we have analyzed users' behavior when manually selecting alternative tracks for the POIs. If unsatisfied with the music suggestion, a user was shown a list of four tracks (presented in a random order)—two tracks retrieved by our approach, and two tracks randomly selected from the remaining tracks in our dataset. Even in this case, the users strongly preferred the music tracks matched with the POIs—out of 77 manual music selections, 58 (75 %) were chosen from the tracks matching to the POI and 19 (25 %) from the randomly selected tracks, i.e., the probability that a user selects a matched music track is about three times higher than that of selecting a random music track. This preference for matched music tracks is also statistically significant ($p < 0.001$ in a *chi-square* test), which proves our hypothesis that users prefer tracks for POIs that are generated by our music-to-POI matching approach.

14.3.4 Vocabulary Revision and Automatic Tag Prediction

A major limitation of the initial design of our approach was the high cost of acquiring emotion annotations of music tracks and POIs. Therefore, following the positive results obtained in the initial evaluation, our goal was to make the music matching technique more scalable and to evaluate it on a larger and more diverse dataset of POIs and music tracks.

In order to scale up our technique, emotion tags had to be predicted automatically for both POIs and music tracks. Automatic POI tagging would allow avoiding the costly user-generated tag acquisition and would make our technique easily applicable to existing sets of POIs (e.g., touristic itineraries or guidebooks). However, in the scope of this work, we do not address the problem of POI tagging and leave it for future work (see Sect. 14.4). Instead, we focus on the automatic tagging of music tracks. This is an equally, if not more, important scalability issue, as music auto-tagging would make our technique applicable to any music collection (e.g., the user's music library).

As discussed in Sect. 14.2, *auto-tagging* techniques can be used to tag music with a defined set of labels. However, these techniques are computationally expensive and require training a model for each label. Therefore, it was important to reduce the size of our vocabulary prior to applying an auto-tagging technique.

We observed that the original GEMS emotion vocabulary (Table 14.1) contained many synonyms (e.g., *triumphant-strong*, *calm-meditative*), and that certain tags were rarely employed by users during the tagging survey (Sect. 14.3.1). Therefore, we have revised the vocabulary by merging synonym adjectives, discarding the rarely used tags, and substituting some labels for clarity (*transcendence* was replaced with

Table 14.2 A revised version of the emotion vocabulary

Tags	Type
Affectionate, Agitated, Animated, Bouncy, Calm, Energetic, Melancholic, Sad, Sentimental Serene, Spiritual, Strong, Tender, Thrilling	GEMS tags [41]
Ancient, Modern, Bright, Dark, Colorful, Heavy, Lightweight, Open, Cold, and Warm	Additional tags

Table 14.3 Cities and POIs of the evaluation dataset

City	POIs
Amsterdam	Canals of Amsterdam
Barcelona	Sagrada Familia, Casa Batlló
Berlin	Brandenburg Gate, Charlottenburg Palace
Brussels	Royal Palace of Brussels
Copenhagen	Christiansborg Palace
Dublin	Dublin Castle
Florence	Florence Cathedral
Hamburg	Hamburg Rathaus
London	Big Ben, Buckingham Palace
Madrid	Almudena Cathedral, Teatro Real, Las Ventas
Milan	Milan Cathedral, La Scala
Munich	Munich Frauenkirche
Paris	Eiffel Tower, Notre Dame de Paris
Prague	National Theater
Rome	St. Peter's Basilica, Colosseum
Seville	Seville Cathedral
Vienna	Vienna State Opera

spiritual, and *light* with *lightweight*). The revision allowed us to reduce the size of the vocabulary from 46 to 24 tags (Table 14.2).

To apply the revised vocabulary to a more diverse dataset, we first collected a set of 25 well-known POIs from 17 major city tourism destinations in Europe (Table 14.3). The POI information was extracted using DBpedia knowledge base.³

Subsequently, to acquire a collection of music tracks, we used a technique developed in a previous work [23], which queries the DBpedia knowledge base and retrieves music composers or artists semantically related to a given POI (see Sect. 14.3.5.2). For each of the 25 POIs in Table 14.3, we queried DBpedia for top-5 related musicians and aggregated them into a single set. This resulted in a collection

³<http://dbpedia.org/>.



Fig. 14.5 Interface of the POI and music tagging application

of 123 musicians (there were two repetitions in the initial list of 125 musicians). Finally, we retrieved three music tracks for each musician by taking the top-ranked results returned by the YouTube search interface⁴ (the musician's name was used as a search query). Doing so ensured the collected tracks to be representative of the musicians in our dataset. We obtained a set of 369 music tracks belonging to nine music genres: Classical, Medieval, Opera, Folk, Electronic, Hip Hop, Jazz, Pop, and Rock.

To annotate the new dataset with the emotion vocabulary, we used a web application similar to the first tag acquisition application (Fig. 14.1). However, contrary to the earlier experiments, at this stage we only relied on user-assigned annotations to tag the 25 POIs and to bootstrap the automatic tagging of music tracks, since automatic tag prediction requires training a multi-label classifier on a set of tagged music tracks. We chose a subset of music tracks (123 tracks—one random track per musician) to be annotated as training data for the auto-tagging algorithm.

Figure 14.5 shows the interface used for tagging both POIs and music tracks. The users were asked to annotate POIs and music tracks using the revised emotion vocabulary of 24 adjectives. Note that unlike in the previous tagging application (Fig. 14.1), here we did not present the tags grouped by category, but they were simply displayed as a flat list. This was done to make the tagging process more straightforward. Moreover, the vocabulary revision discarded some redundant tag categories.

The tagging procedure was performed by 10 volunteers recruited via email—students and researchers from the Free University of Bolzano and other European universities. We note that the tagging of music and POIs is a subjective task and

⁴<http://www.youtube.com/>.

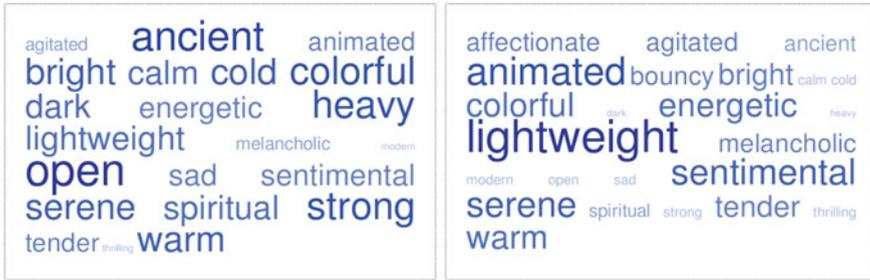


Fig. 14.6 The tag clouds for POI (*left*) and music (*right*) annotations from the revised vocabulary

users may disagree whether certain tags apply to an item [38]. To ensure the quality of the acquired annotations, we considered the agreement between users, which is a standard measure of quality for user-generated tags [24]. We cleaned the data by keeping for each item only the tags on which at least two taggers agreed. As a result, we obtained an average of 5.1 distinct tags for the POIs and 5.8 for the music tracks.

Figure 14.6 shows the distribution of collected tags for POIs (left) and music tracks (right) with larger font representing more frequent tags. Similarly to tag clouds produced for the original vocabulary (Fig. 14.2), the tag distributions differ for POIs and music tracks. We observe that certain tags (e.g., for music annotations—*cold*, *open*, *dark*, *thrilling*) diminished in frequency after cleaning the tags based on user agreement. This means that certain tags may be more subjective than others, and therefore more difficult to agree on. Nevertheless, tags like *colorful*, *energetic*, *lightweight*, *sentimental*, *serene*, *warm* are consistently applied to both music and POIs, which serves our goal of matching the two types of items.

Finally, we used a state-of-the-art music auto-tagger developed by researchers at the Johannes Kepler University (JKU) of Linz [34], training it on the 123 user-tagged tracks and predicting tags for the full set of 369 music tracks (see [22] for details). The auto-tagger outputs a probability for each tag in the vocabulary to be relevant for a track. However, the metric we used for music-to-POI similarity computation (Eq. 14.1) requires computing the intersection and union of the items’ tag profiles. Therefore, we decided to generate binary tag assignments based on the probabilistic output of the auto-tagger by applying a threshold to the tag prediction probabilities. Empirical analysis showed that a threshold of 0.4 produced an average tag profile of 5.2 tags which is in accordance with the average profile size of manually tagged items (5.1 tags for POIs and 5.8 for music tracks).

For a music track v whose tags are predicted using an auto-tagger, we define the tag profile X_v as:

$$X_v = \{t_i \mid p(t_i) \geq 0.4\}, \quad i = \{1, \dots, K\} \quad (14.2)$$

where K is the size of tag vocabulary (in our case $K = 24$) and $p(t_i)$ denotes the probability for a tag t_i to be relevant for the track.

Having the dataset of POIs and music tracks annotated with a common vocabulary of emotion labels and a metric that gives a similarity score for any given pair of POI and a music track (Eq. 14.1), we can evaluate our approach in a user study. In the next section we describe the evaluation and present the results.

14.3.5 Comparison to Alternative Matching Techniques

The new dataset with more diverse POIs spread out across 17 different cities made a live user study, similar to that illustrated in Sect. 14.3.3, i.e., with users visiting the POI and then listening to the system-selected music, impossible to conduct. Therefore, to evaluate the scaled-up version of our approach, we opted for a web-based study, where a text description and images of each POI were visualized, and the users were asked to listen to the suggested music tracks and evaluate if they match the displayed POI (see Fig. 14.7).

As described in the previous section, we have collected a dataset of 25 POIs and 369 music tracks belonging to 9 music genres, with a part of the music dataset tagged manually and also the full set of tracks auto-tagged by an algorithm. This setting allowed us comparing two versions of emotion-based music matching—one using only the manually annotated 123 tracks, and the other using the full set of 369 auto-tagged tracks. We call the first version *tag-based* and the other *auto-tag-based* approach. In both approaches, the POI annotations were user-generated.

For each POI in the dataset, the *tag-based* approach ranks the manually annotated 123 music tracks using the similarity score in Eq. 14.1. Likewise, the *auto-tag-based*

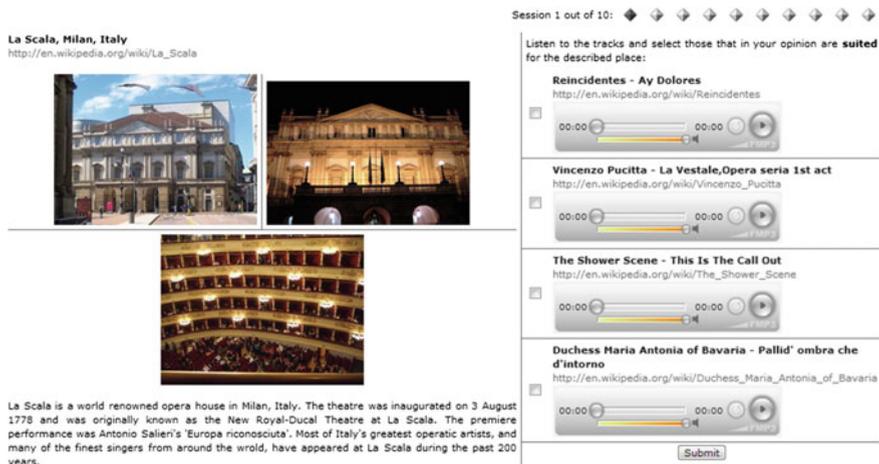


Fig. 14.7 Screenshot of the web application used to evaluate the different music matching approaches

approach ranks the 369 tracks with automatically predicted labels (Eq. 14.2). The top-ranked music track is presented to the user along with the POI.

In addition to comparing the manually and automatically generated emotion annotations, we wanted to compare the performance of emotion-based approach to alternative music matching techniques, which we describe here.

14.3.5.1 Genre-Based Approach

Traditionally, music delivery services employ personalization techniques to provide music content to users. Therefore, as a baseline approach, we employed a basic personalization technique—*genre-based* music matching which selects the music tracks based on the users' genre preferences. We aimed to compare a personalized music matching technique with the *knowledge-based* and *tag-based* approaches, which are not personalized, but rather directly match music with the POIs.

In order to obtain the users' preferences, we asked the study participants to select their preferred music genres prior to performing the evaluation. The genre taxonomy was based on the music tracks in our dataset, and included: Classical, Medieval, Opera, Folk, Electronic, Hip Hop, Jazz, Pop, and Rock. For each displayed POI, the genre-based track is randomly selected from the whole set of music tracks belonging to the user's preferred genres.

14.3.5.2 Knowledge-Based Approach

The *knowledge-based* music matching approach employs the technique presented in [23]. Given a POI, this approach ranks musicians by their relatedness to the POI. The relatedness is computed from the semantic relations between the POI and musicians extracted from the DBpedia knowledge base [3].

DBpedia—the Linked Data version of Wikipedia—contains information on more than 3.5 million entities and semantic relations between them. This information is stored and retrieved in the form of triples, which are composed of the *subject–property–object* elements, such as *<Vienna State Opera, located in, Vienna>*, *<Gustav Mahler, belongs to, Opera composers>*, where a subject and an object belong to certain classes (e.g., Building, Person, City, Music Category), and the property denotes a relation between the classes (e.g. a building *being located* in a city). Given a POI, the knowledge-based approach queries DBpedia (using the SPARQL semantic query language⁵), and builds a graph where nodes correspond to classes and edges to relations between the classes. The graph is built using a predefined set of relations (location, time, and architecture/art category relations) and contains a starting node without incoming edges that corresponds to the POI, and target nodes without outgoing edges that belong to the musician class. Then, a weight spreading algorithm is applied to the graph to compute the relatedness score for each musician

⁵<http://www.w3.org/TR/rdf-sparql-query/>.

node. Nodes that are connected to the POI through more paths in the graph receive higher scores. Finally, the highest-scored musicians are returned for the target POI. For more details on the approach, refer to [23].

As explained in Sect. 14.3.4, the knowledge-based approach was used to build the dataset of POIs and music tracks—for each of the 25 POIs, top-5 related musicians were retrieved from DBpedia. Subsequently, we have downloaded three representative music tracks for each musician. Using this approach, we assume that there are no major differences between the tracks of the same musician. Therefore, for each POI, the knowledge-based track is randomly selected from the three music tracks by the top-ranked musician.

We believe that the emotion-based and knowledge-based techniques represent two complementary ways of establishing a match between a place and a music track: a track may “feel right” for a POI, or it can be linked to a POI by factual relations (e.g., belonging to the same cultural era or composed by someone whose life is related to the POI). In previous works [5, 23], we have evaluated the two techniques independently on different datasets and against primitive baselines. It was therefore important to directly compare the performance of these techniques, and to evaluate their combination.

14.3.5.3 Combined Approach

Finally, we implemented a hybrid combination of the *knowledge-based* and *auto-tag-based* approaches, employing a rank aggregation technique [12]. Since the music-to-POI similarities produced by the two techniques have different value ranges, we used the normalized Borda count rank aggregation method to give equal importance to the two. Given a POI u , the *knowledge-based* and *auto-tag-based* approaches produce the rankings of music tracks σ_u^{kb} and σ_u^{tb} . We denote the position of a track v in these rankings as $\sigma_u^{kb}(v)$ and $\sigma_u^{tb}(v)$ respectively. Then, we compute the combined score of the track v for the POI u as:

$$\text{combined_score}(u, v) = \frac{N_{kb} - \sigma_u^{kb}(v) + 1}{N_{kb}} + \frac{N_{tb} - \sigma_u^{tb}(v) + 1}{N_{tb}} \quad (14.3)$$

where N_{kb} and N_{tb} are the total number of tracks in the corresponding rankings. For each POI, the combined approach selects the top-scored music track.

14.3.5.4 User Study

To determine which approach produces better music suggestions, we designed a web-based interface (Fig. 14.7) for collecting the users’ subjective assessments of whether a music track suits a POI. This user study was similar to the experiment conducted to determine the best-performing similarity metric for the tag-based approach (see Sect. 14.3.2, Fig. 14.3). As in the previous case, participants of the experiment were

repeatedly asked to consider a POI, and while looking at its images and description, to listen to the suggested music tracks.

While in the previous experiment the music suggestions for a POI were selected using the different metrics applied to items' tag profiles, here music tracks were selected using the proposed five approaches described above—the personalized baseline approach and the four non-personalized matching approaches. The users were asked to check all the tracks that in their opinion suit the displayed POI. The order of the music tracks was randomized, and the user was not aware of the algorithms that were used to generate the suggestions. In total, a maximum of five tracks corresponding to the top-ranked tracks given by each approach were suggested for a POI, but sometimes less tracks were shown as the tracks selected by the different approaches may overlap.

To understand which approach produces suggestions that are most appreciated by the users, we measured the precision of each matching approach M as follows:

$$\text{precision}(M) = \frac{\# \text{ of times } t^M \text{ marked as a match}}{\# \text{ of times } t^M \text{ presented to the users}} \quad (14.4)$$

where t^M is a music track suggested using the approach M .

14.3.5.5 Results

A total of 58 users participated in the evaluation study, performing 564 evaluation sessions: viewing a POI, listening to the suggested music tracks, and providing feedback. 764 music tracks were selected by the users as well-suited for POIs. Table 14.4 shows the performance of the matching approaches. All non-personalized matching techniques performed significantly better than the personalized *genre-based* track selection ($p < 0.001$ in a two-proportion z -test). This result shows that in a situation defined by a visit to a POI, it is not really appropriate to suggest a music track liked by the user, but it is more important to adapt the music to the place.

We note that both emotion-based techniques outperform the baseline approach. Furthermore, the evaluation results suggest that the *tag-based* music matching approach can be successfully scaled up using automatic tag prediction techniques. The *auto-tag-based* approach even outperformed the *tag-based* approach with marginal significance ($p = 0.078$). This can be explained by the larger variety of music

Table 14.4 Precision values for the different music matching approaches

Genre-based	Knowledge-based	Tag-based	Auto-tag-based	Combined
0.186	0.337*	0.312*	0.362*	0.456**

The values marked with * are significantly better than the *Genre-based* approach (two-proportion z -test, $p < 0.001$). The value marked with ** is significantly better than the other approaches ($p < 0.01$)

in the auto-tagged music dataset—using the *auto-tag-based* approach the tracks were selected from the full set of 369 music tracks, while the *tag-based* approach used only the subset of 123 manually annotated tracks. Scaling up the process of tag generation without harming performance is hence the vital advantage of using the auto-tagger.

Finally, the *combined* approach produced the best results, outperforming the others with statistical significance at $p < 0.01$. These results confirm our hypothesis that the users are more satisfied with music suggestions when combining the *tag-based* and *knowledge-based* techniques, which represent orthogonal types of relations between a place and a music track.

14.4 Discussion and Further Work

We believe that techniques and results described in this chapter illustrate clearly the importance of using emotions for capturing the relationship between music and places, and for developing new engaging music applications. We presented a scalable emotion-based solution for matching music to places, and demonstrated that alternative music-to-POI matching approaches may be effectively combined with that emotion-based technique. This can lead to even richer music delivery services, where emotion-based music suggestions are supported by additional relations, e.g., knowledge linking a musician to a POI.

Emotion-aware and location-aware music recommendation topics belong to a larger research area of context-aware music recommendation [21] (see also Chap. 15). This is a new and exciting research area with numerous innovation opportunities and many open challenges. Up to date, few context-aware music services are available for public use. While certain music players allow specifying the user's mood or activity as a query, most context-aware systems are research prototypes, not yet available to the public. A major challenge is understanding the relations between context factors, such as location, weather, time, mood, and music features. Some research on this topic exists in the field of music psychology [29–31], therefore, collaboration between music psychologists and music recommendation researchers is essential.

While in this work we demonstrated the effectiveness of the proposed techniques through a web-based evaluation, it is important to further evaluate the approach in real-life settings to confirm our findings. Moreover, additional evaluation would help us understand which type of associations between music and POIs—emotion-based or knowledge-based—the users prefer in different recommendation scenarios (e.g., sightseeing, choosing a holiday destination, or acquiring knowledge about a destination).

Another important future work direction is studying the semantics of emotion tags. In current work, we treated all tags equally and we did not explore their relations, while contradicting or complementary emotion tags may provide an important source of information. Moreover, our analysis of the tag vocabulary is far from complete. While we have shown that the proposed set of labels can be effectively used for

matching music to places, further user studies and may lead to a more definitive emotion vocabulary.

It is also important to address the automatic acquisition of tags for POIs (for instance, using web mining techniques and folksonomy datasets like Flickr as sources of tagging information), to analyze the importance of individual audio features when automatically tagging music tracks, and to explore more complex hybrid recommendation strategies for combining our music suggestions with personalized recommendations produced by existing music delivery services, e.g., Last.fm.

Finally, we believe that the designed music-to-POI matching solutions may be adapted to other types of content. Since music is a heavily emotion-loaded type of content, emotions may be employed to link music to any content items that may be labeled with emotions—movies, books, or paintings [6, 26, 36].

References

1. Adkins, M., Santamas, H.: Exploring the perception and identity of place through sound and image. In: Proceedings of the Invisible Places International Symposium, pp. 403–421 (2014)
2. Ankolekar, A., Sandholm, T.: Foxtrot: a soundtrack for where you are. In: Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications, pp. 26–31 (2011)
3. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: a nucleus for a web of open data. In: ISWC '08: Proceedings of the 7th International Semantic Web Conference, pp. 722–735 (2008)
4. Bachelard, G.: The Poetics of Space. Beacon Press, Boston (1958)
5. Braunhofer, M., Kaminskas, M., Ricci, F.: Location-aware music recommendation. *Int. J. Multimedia Inf. Retrieval* **2**(1), 31–44 (2013)
6. Cai, R., Zhang, C., Wang, C., Zhang, L., Ma, W.-Y.: MusicSense: contextual music recommendation using emotional allocation modeling. In: MULTIMEDIA '07: Proceedings of the 15th International Conference on Multimedia, pp. 553–556. ACM, New York, NY, USA (2007)
7. Castello, L.: Rethinking the Meaning of Place: Conceiving Place in Architecture-Urbanism. Ashgate Publishing, Ltd., London (2010)
8. Cheng, Z., Shen, J.: Just-for-me: an adaptive personalization system for location-aware social music recommendation. In: Proceedings of International Conference on Multimedia Retrieval, pp. 185–192, ACM (2014)
9. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G.: Emotion recognition in human-computer interaction. *IEEE Signal Process. Mag.* **18**(1), 32–80 (2001)
10. Debord, G.: Introduction to a critique of urban geography. *Critical Geographies A Collection of Readings* (1955)
11. Dunker, P., Nowak, S., Begau, A., Lanz, C.: Content-based mood classification for photos and music: a generic multi-modal classification framework and evaluation approach. In: MIR'08: Proceeding of the 1st ACM International Conference on Multimedia Information Retrieval, pp. 97–104. ACM, New York, NY, USA (2008)
12. Dwork, C., Kumar, R., Naor, M., Sivakumar, D.: Rank aggregation methods for the web. In: Proceedings of the 10th International Conference on World Wide Web (WWW), pp. 613–622 (2001)
13. Ellin, N.: Good Urbanism: Six Steps to Creating Prosperous Places. Island Press, Washington (2012)

14. Engler, M.: Experience at view places: an inquiry into the emotional ties between people and places. In: Proceedings of the 21st Annual Conference of the Environmental Design Research Association, pp. 222–230 (1990)
15. Gillhofer, M., Schedl, M.: Iron maiden while jogging, debussy for dinner? In: MultiMedia Modeling, pp. 380–391, Springer (2015)
16. Gretzel, U.: Intelligent systems in tourism: a social science perspective. *Ann. Tour. Res.* **38**(3), 757–779 (2011)
17. Hevner, K.: Experimental studies of the elements of expression in music. *Am. J. Psychol.* **48**(2), 246–268 (1936)
18. Huron, D.: Perceptual and cognitive applications in music information retrieval. In: Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR) (2000)
19. Iosafat, D.: On sonification of place: psychosonography and urban portrait. *Org. Sound* **14**(01), 47–55 (2009)
20. Kaminskas, M., Ricci, F.: Matching places of interest with music. In: Workshop on Exploring Musical Information Spaces, pp. 68–73 (2009)
21. Kaminskas, M., Ricci, F.: Contextual music information retrieval and recommendation: state of the art and challenges. *Comput. Sci. Rev.* **6**, 89–119 (2012)
22. Kaminskas, M., Ricci, F., Schedl, M.: Location-aware music recommendation using auto-tagging and hybrid matching. In: Proceedings of the 7th ACM Conference on Recommender Systems (RecSys 2013), Hong Kong, China (2013)
23. Kaminskas, M., Fernández-Tobías, I., Ricci, F., Cantador, I.: Knowledge-based identification of music suited for places of interest. *Inf. Technol. Tour.* **14**(1), 73–95 (2014)
24. Law, E.L.M., Von Ahn, L., Dannenberg, R.B., Crawford, M.: Tagatune: a game for music and sound annotation. In: ISMIR, vol. 3, p. 2 (2007)
25. Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based multimedia information retrieval: state of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl. (TOMCCAP)* **2**(1), 1–19 (2006)
26. Li, C.-T., Shan, M.-K.: Emotion-based impressionism slideshow with automatic music accompaniment. In: MULTIMEDIA '07: Proceedings of the 15th International Conference on Multimedia, pp. 839–842. ACM Press, New York, NY, USA, (2007)
27. Markines, B., Cattuto, C., Menczer, F., Benz, D., Hotho, A., Stumme, G.: Evaluating similarity measures for emergent semantics of social tagging. In: Proceedings of the 18th International Conference on World Wide Web, pp. 641–650, ACM (2009)
28. Noll, P., Noll, B.: List of adjectives in american english. <http://www.paulnoll.com/Books/Clear-English/English-adjectives-1.html>
29. North, A.C., Hargreaves, D.J.: Situational influences on reported musical preference. *Psychomusicology Music Mind Brain* **15**(1–2), 30–45 (1996)
30. North, A.C., Hargreaves, D.J.: *The Social and Applied Psychology of Music*. Cambridge University Press, Cambridge (2008)
31. Pettijohn, T.F., Williams, G.M., Carter, T.C.: Music for the seasons: seasonal music preferences in college students. *Curr. Psychol.* 1–18 (2010)
32. Russell, J.A.: A circumplex model of affect. *J. Pers. Soc. Psychol.* **39**(6), 1161–1178 (1980)
33. Schedl, M., Vall, A., Farrahi, K.: User geospatial context for music recommendation in microblogs. In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 987–990, ACM (2014)
34. Seyerlehner, K., Schedl, M., Knees, P., Sonnleitner, R.: A refined block-level feature set for classification, similarity and tag prediction. In: 7th Annual Music Information Retrieval Evaluation eXchange (MIREX) (2011)
35. Shi, Y., Larson, M., Hanjalic, A.: Tags as bridges between domains: improving recommendation with tag-induced cross-domain collaborative filtering. In: User Modeling, Adaption and Personalization, pp. 305–316. Springer, Heidelberg (2011)
36. Stupar, A., Michel, S.: Picasso—to sing, you must close your eyes and draw. In: 34th ACM SIGIR Conference on Research and Development in Information, pp. 715–724 (2011)

37. Thayer, R.E.: *The Biopsychology of Mood and Arousal*. Oxford University Press, Oxford (1989)
38. Turnbull, D., Barrington, L., Lanckriet, G.: Five approaches to collecting tags for music. In: *Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 225–230 (2008)
39. Urry, J.: *Consuming Places*. Psychology Press, Abingdon (1995)
40. Yang, Y.-H., Chen, H.H.: Machine recognition of music emotion: a review. *ACM Trans. Intell. Syst. Technol. (TIST)*, **3**(3), 40 (2012)
41. Zentner, M., Grandjean, D., Scherer, K.R.: Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion* **8**(4), 494–521 (2008)