Location-aware music recommendation

Matthias Braunhofer, Marius Kaminskas & Francesco Ricci

🦄 Springer

Springer

REGULAR PAPER

Location-aware music recommendation

**Matthias Braunhofer · Marius Kaminskas ·
Francesco Ricci**

Abstract The increasing amount of online music content
has opened opportunities for implementing new effective
information access services—commonly known as music
recommender systems—that support music navigation, dis-
covery, sharing, and formation of user communities. In recent
years, a new research area of contextual music recommen-
dation and retrieval has emerged. Context-aware music rec-
ommender systems are capable of suggesting music items
taking into consideration contextual conditions, such as the
user's mood or location, that may influence the user's pref-
erences at a particular moment. In this work, we consider
a particular kind of context-aware recommendation task—
selecting music content that fits a place of interest (POI). To
address this problem we have used emotional tags assigned to
both music tracks and POIs, and we have considered a set of
similarity metrics for tagged resources to establish a match
between music tracks and POIs. Following an initial web-
based evaluation of the core matching technique, we have
developed a mobile application that suggests an itinerary and
plays recommended music for each visited POI, and evalu-
ated it in a live user study. The results of the study show that
users judge the recommended music as suited for the POIs,
and that the music is rated higher when it is played in this
usage scenario.

1 Introduction

Music has always played a major role in human enter-
tainment. With the coming of digital music and Internet
technologies, a huge amount of music content has become
available to millions of users around the world. With millions
of artists and songs on the market, it is becoming increasingly
difficult for the users to search for interesting and novel music
content —there is a lot of potentially interesting music that is
difficult to discover. Furthermore, huge amounts of available
music data have opened opportunities for researchers work-
ing on music information retrieval and recommendation to
create new viable services that support music navigation, dis-
covery, sharing, and formation of user communities [7]. The
demand for such services—commonly known as music rec-
ommender systems—is high, and therefore there is a huge
economic potential for online music content.

Music recommender systems are decision support tools
that help tame the information overload by retrieving only
the items that are estimated as relevant to the user, based on
the user's profile, i.e., a representation of the user's music
preferences [20]. For example, Last.fm: [1] a popular Internet
radio and recommender system—allows a user to mark songs
or artists as favorites, tracks the user's listening habits, and
based on this information can identify and recommend music
content that is likely to be of interest to the user.

However, most of the available music recommender sys-
tems suggest music without taking into consideration the

M. Braunhofer · M. Kaminskas (✉) · F. Ricci
Faculty of Computer Science, Free University of Bozen-Bolzano,
Piazza Domenicani 3, 39100 Bolzano, Italy
e-mail: mkaminskas@unibz.it

M. Braunhofer
e-mail: mbraunhofer@unibz.it

F. Ricci
e-mail: fricci@unibz.it

---

[1] http://www.last.fm/.

user's context, e.g., her mood, or current location and activity [2]. In fact, a study on users' musical information needs [14] concluded that there is a growing need for "extra-musical information" that would "contextualize users' real-world searches" for music to provide more useful retrieval results. In response to these demands, over the last few years a new research topic of contextual music retrieval and recommendation has emerged [13]. Systems that can recommend music depending on the user's actual situation, e.g., emotional state, or any other contextual conditions that might influence the user's desire to listen to particular music, can be used in new engaging applications. For instance, location-aware systems can retrieve music content that is relevant to the user's location, e.g., by selecting music composed by artists that lived in that location [8]. Or, an in-car music player may adapt music to the landscape the car is passing [4].

In this line of research, we are considering the problem of retrieving music suited for the place of interest (POI) that the user is visiting or browsing using an information service. The intuition is that a Vivaldi concerto better suits a narrow street in Venice than a Bach organ fugue, which may better suit an old Gothic church. Being able to select music for a place can be used for creating new engaging location-aware music delivery services. In particular, in our work we are considering a scenario where a tourist is sightseeing using a mobile city guide. The guide recommends a walking itinerary and, while the user is visiting the suggested POIs, it plays music that suits the visited POIs. The goal is to enhance the user's experience, to create a more attractive travel guide tool, and to recommend music that could be better accepted and enjoyed by the user.

In order to establish a link between music and POIs, we chose to represent these items with a common set of features—a set of tags. With such a representation the matching can be performed by comparing the tag profiles of the two items. We decided to use tags that describe emotional properties of the items, since music and places can both raise emotions, and the commonality of these emotions could provide the base for establishing a degree of match between a place and a music track. Moreover, using tags to describe both music and POIs is a promising and viable approach, since there is a rapid growth in the amount of user-generated tagging data (a phenomena also known as social, or collaborative tagging) [21]. While tags and tag-based similarity have previously been used in recommender systems research, the specific tag vocabulary and its application to matching POIs and music presented herein is novel. Moreover, we believe that the technical solution proposed in this work has a great potential for the future of music recommendation services.

As a first step of our research, we have designed the music-to-POI matching approach and evaluated it in a web-based user study where the users were required to evaluate the appropriateness of the music selected by the system for a

set of POIs [12]. The main goal of this experiment was to evaluate a set of emotional tags, suggested by recent studies on music cognition [23], which we adopted for representing music tracks and POIs, and to evaluate a set of similarity metrics for tagged resources [17]. The experiment was carried out using a web application where the POI descriptions were shown and music played in the background. Since such "simulated" environment could not fully reflect the real-world settings of a visited POI (i.e., the surroundings, weather conditions, other people around), it was crucial to further evaluate our approach in real-world settings. Therefore, as a second step, we have implemented a mobile city guide for the city of Bolzano (Italy), and conducted a live user study [5]. The main goals of this study were to test the following hypotheses: (a) users agree with the music-to-POI match produced by our approach, and (b) users tend to rate the selected music tracks higher in this mobile, in-context usage scenario, compared with the rating situation where the context is not defined. Hence, we wanted to show that context, which is here represented by the location and the modality of music presentation, does have an impact on the user's evaluation for the suggested music. The results have shown that users judge the recommended music as suited for the POIs, and that the music is rated higher when played in this usage scenario. Therefore, the proposed context-aware music recommendation technology has a tangible benefit for users.

The rest of the paper is structured as follows. In Sect. 2 we review the related work. Section 3 provides details on the approach to tag-based matching of music and POIs, and describes the initial web-based evaluation of the approach. Then, the implementation of the mobile city guide is presented in Sect. 4, and the results of the live user study are presented in Sect. 5. Finally, conclusions are drawn and future work directions defined in Sect. 6.

2 Related work

2.1 Music recommender systems

Recommender systems are software tools and techniques providing suggestions of the items to be of use to a user [20]. In the music domain, recommender systems can support information search and discovery tasks by helping the user to find relevant music items, for instance, new music tracks, or discover new artists [7]. Several music recommendation techniques have been proposed in the literature, but most of the available systems use either content-, collaborative-, or social-based approaches, or even more often, a hybrid combination of these three basic approaches [7,6].

Content-based systems exploit features of the music tracks liked by the target user for recommending other, similar items that the user may like. Music features can be extracted

directly from the music content, using signal processing techniques, or they can be based on metadata of the tracks (e.g., genre, year, author). Conversely, collaborative-based systems do not use any features of items, but instead rely on user-generated evaluations of music tracks (ratings or implicit feedback). This approach is based on the assumption that similar users like similar items, and vice-versa—similar items are liked by similar users. A collaborative-based system exploits the dataset of user-generated evaluations to predict the evaluations for items unseen by the target user, and then recommends the items with the highest predicted evaluations.

A third approach, which is called social-based, is emerging in the music domain. It is based on computing similarities among the items to be recommended (music tracks or artists) through web mining techniques, or by exploiting social tagging information [7]. Social-based recommendations can be generated using the similarities of artists that in turn can be computed using the social activity of the users, for instance by analyzing the songs played by a community of users in the same listening sessions, or the tags assigned by users to songs and artists. The rationale of this approach is that items similar to those that the user liked will also probably be relevant to the user.

2.2 Establishing cross-domain item-to-item similarity

As stated in the Introduction, our goal in this research is not focused on the popular music recommendation problem, where the suggested items are those predicted to be relevant to the user, given her preferences. We aimed at finding music matching with items in another domain, to recommend music that fits the particular contextual situations defined by these items (i.e., places).

It is clearly challenging to match music to a place so that the user could recognize this adaptation or, even without explicitly recognizing it, appreciate such a selection and prefer it to other music not matching the place. The core technical issue to be solved here is related to the fact that music and POIs are different objects, and there is no obvious way to match one type of content with the other. In recommender systems literature, the similarity of two items has been established either using their feature-based descriptions, or using their ratings given by a set of users [20]. The first approach requires that the two items, whose similarity is sought, share a common set of features, while the second one requires that a large number of users co-rate the two items.

The first approach is therefore difficult to apply when the two items are not of the same type, while the second would only predict that a user who likes (dislikes) the first item will also like (dislike) the second one. However, this is not really a sign that the two items match together, and that they can be recommended together, or one in the context of the other. The problem of matching POIs with music tracks is more closely related to that found in cross-selling, e.g., recommending a type of boots that suit a kind of skis. This is a rather unpopular problem that has only been addressed by researchers working on recommending a good bundling of items, e.g., a travel package [19].

A third approach to establish cross-domain item similarity, which has not been widely used in recommender systems so far, is identifying semantic relations between items with the help of structured knowledge sources like Wikipedia. [2] For instance, Loizou [16] used Wikipedia for identifying explicit semantic relations between music artists and movies. Then, with such relations, users and items were incorporated into a graph, upon which a probabilistic recommendation model was built. As said earlier, we have opted for representing items in the two domains with a common set of tags, describing the emotional properties of music and POIs. In recommender system research, social tags have been used for cross-domain user modeling [1], but to our knowledge this work is the first to apply tags for matching items across domains.

2.3 Context-aware music recommendation

Another research area related to our work is context-aware recommendation, since a place can be considered as a context in which music listening is performed. Context-aware recommendation and retrieval in music domain is a new emerging research topic [13]. Here we review some of the works on context-aware music recommendation that exploit the location-related context information.

The user's location may have a strong impact on her perception and preferences of music. For instance, while walking down a busy city street a user might prefer listening to different music compared to when strolling in the woods. The US music duo Bluebrain is the first band to record a location-aware album. [3] In 2011, the band released two such albums—one dedicated to Washington's National Mall park, and the second dedicated to New York's Central Park. Both albums were released as iPhone apps, with music tracks pre-recorded 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. Despite the large potential of location-aware music services, up to date there has been little research exploring location-related information in music recommendations.

Reddy and Mascia [18] presented a mobile music recommender system Lifetrak that generates a playlist using the user's music library based on the current context of the user. The context information used by the authors includes location (represented by a ZIP code) as well as time, weather, and

---

activity information. The context is obtained using the sensors of the mobile device that stores the application and RSS feeds of weather and traffic information services. Similarly to our work, for Lifetrak to generate recommendations, the songs in the user's library have to be tagged by the user with tags from a controlled vocabulary. However, in contrast to our approach, in this work the tags in the vocabulary directly represent the values of the previously mentioned context parameters. So for instance, songs have to be labeled with a ZIP code of a certain area to be recommended for that location.

Gaye et al. [10] presented a system for interactive music generation in urban environments. The described system takes a wide range of input parameters (light and noise level, temperature, user's movements, etc.) for generating electronic music in real-time. The authors have implemented a wearable prototype that consists of sensors, a micro controller, and a laptop with a music programming environment. The major drawbacks of the system are hardware-related—a complex network of sensors and wires makes it difficult to use in everyday life. Furthermore, due to the specific genre of artificially generated music (electronic), this device is not suitable for a wide range of music listeners.

Although not related to music recommendation, the work of Finney and Janer [9] describes an approach to automatically generate sound effects (soundscapes) for virtual environments to enhance the users' sense of presence, i.e., "the feeling of being situated in an environment despite being physically situated in another". Their approach relies on retrieving sound clips from the Freesound Project database [4] and efficiently mixing them combining background sounds (e.g., birds, wind) with sounds that draw users' attention (e.g., church bells, narration). The approach was combined with the Google Street View application and allowed the users to browse a set of locations while listening to the automatically generated soundscapes.

More recently, Ankolekar and Sandholm [3] presented a mobile audio application, Foxtrot, that allows its users to explicitly assign audio content to a particular location. Similarly to our work, the authors also 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. However, instead of using a knowledge-driven approach, i.e., understanding which emotional characteristics link music and a location, Foxtrot relies on crowd-sourcing—the users of Foxtrot are allowed to assign audio pieces (either music tracks or sound clips) to specific locations (represented by the geographical coordinates of the user's current location), and also to specify the visibility range of the assigned audio content. The system is

then able to provide a stream of location-aware audio content to its users.

3 Matching music to POIs

As mentioned above, we address the problem of providing location-adapted music recommendations by exploiting emotional tags. These tags have been used to annotate both music and POIs, and therefore act as a common set of descriptive features for establishing a match between these two types of items. At this stage of our research, we have bootstrapped a dataset of POIs and music tracks with tags through a specially designed web application. We leave the issue of scaling-up the approach with automatic tag acquisition for future work. In this section, we first describe the data acquisition process and the similarity measures used to match music and POIs. Then we describe the web-based evaluation of the approach.

3.1 Tagging music and POIs

Figure 1 shows the interface of the web application used for tagging music tracks and POIs in our dataset. The dataset consisted of 75 music tracks (famous classical compositions and movie soundtracks), and 50 POIs in the city of Bolzano and surrounding areas (castles, churches, monuments, nature objects, etc.). The descriptions of the POIs were taken from the region's tourism website. [5] 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 tag the items using a controlled tag vocabulary consisting of adjectives from the Geneva Emotional Music Scale (GEMS) model described in [23]. The GEMS model consists of nine categories of emotions, each category containing 2–4 emotional tags (Table 1). We selected the GEMS model, instead of other popular emotion models, because it has been explicitly developed and validated for the music genre that we have used, namely classical music. Since our approach deals with both music and POIs, we could not rely solely on tags derived from a music cognition study. Therefore, in addition to the emotional tags from GEMS model, we have selected five categories of tags describing physical characteristics of items, that proved to be useful in a preliminary user study [11]. The five categories are: *Age* (Ancient, Modern), *Light and Color* (Colorful, Bright, Dark, Dull), *Space* (Open, Closed), *Weight* (Light, Heavy), and *Temperature* (Cold, Mild, Warm).

In total, during the data acquisition phase, 817 tags were collected for the POIs (16.34 tags per POI on average), and

---

Fig. 1 Screenshot of the web application used for tagging POIs and music tracks

Table 1 Emotional tags from the GEMS model

Category	Tags
Wonder	Allured, Amazed, Moved, Admiring
Transcendence	Fascinated, Overwhelmed, Thrills, Transcendence
Tenderness	Mellowed, Tender, Affectionate, In love
Nostalgia	Sentimental, Dreamy, Melancholic, Nostalgic
Peacefulness	Calm, Serene, Soothed, Meditative
Power	Triumphant, Energetic, Strong, Fiery
Joyful Activation	Joyful, Animated, Bouncy, Amused
Tension	Tense, Agitated, Irritated
Sadness	Sad, Tearful

1,025 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. Furthermore, the tagging was performed at different locations and times by a number of people with different cultural backgrounds, thus the collected tags were not uniformly biased toward certain contextual conditions.

Following the data acquisition process, we have investigated if certain tags have more potential to provide a good match between music and POIs. The distribution of tag categories in the collected dataset (Fig. 2) shows that certain types of tags have been applied to both music tracks and POIs with similar probabilities. This particularly applies to the categories *Peacefulness*, *Power*, *Light and Color*, *Weight*, and *Temperature*. To further narrow down the set of potentially important tags, we identified the tags that mostly contribute to the variability within our dataset by applying Principal Component Analysis [22] to POIs and music separately. The tags that were present in the top-ranked factors for both POIs
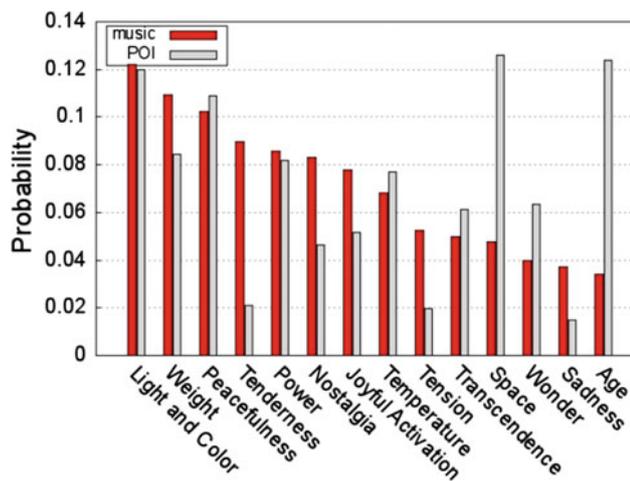
Fig. 2 Distribution of tag categories in the collected data

and music are: *Agitated*, *Animated*, *Bouncy*, *Bright*, *Calm*, *Cold*, *Dark*, *Dreamy*, *Energetic*, *Joyful*, *Mild*, *Light*, *Sad*, *Strong*, *Transcendence*. these qualities in a music track or in a place. These tags separate different POIs and music tracks, and therefore are the most likely to provide good basis for the match between the items. We have also used Feature Selection with Best First search strategy [22] to identify the tags that separate POIs from music. These tags are: *Ancient*, *Closed*, *Open*, *In love*, *Irritated*, *Tender*, and *Tense*. This is in accordance with Fig. 2, which shows that tags describing *Age* and *Space* of items were mostly applied to POIs, while the tags describing *Tenderness* and *Tension*—to music. Therefore, tags in these categories are not likely to be useful for direct matching of music and POIs, at least for our selection of POIs and tracks.

In conclusion, we believe that certain tags might contribute more to the perceived match between music and POIs. However, we leave further investigation of this issue for future work, and in the current implementation of our approach assign equal importance to all tags. These findings are important to consider when scaling-up the approach with automatic tag acquisition, but they also indicate that an effective similarity metric for this task should be robust against the differences in the overall tag distributions observed for the two types of items. We observe that in previous research that used tag-based similarity metrics this was not an issue, as the items to be matched were confined to be in a single domain [17].

We note that the granularity of the tag vocabulary is different for emotional and physical tags—while the tags in each GEMS model category are mostly synonyms, physical tags are clearly distinct from each other. Therefore, we have considered a model where the GEMS tags are replaced with their emotion categories. For instance, the tags *Allured*, *Amazed*, *Moved* and *Admiring* appearing in any item's tag profile were substituted with *Wonder*. Such merging of tags improved the

tag coverage and reduced the dimensionality of item profiles from 46 (the initial number of individual tags) to 22 (9 emotion categories + 13 physical tags). In the next section, we will present some similarity metrics that are applied either to the original tag profiles, or to the more compact tag profiles.

3.2 Similarity metrics

In order to match POIs with music, we decided to consider a well-established set of similarity metrics that are applicable to tagged resources. Markines et al. [17] evaluated the performance of different similarity metrics using classical IR evaluation measures, when computing the similarity between tagged resources. However, this study was conducted on a single folksonomy dataset [6], with the task being to predict URL-to-URL similarity. The ground truth for resources' similarity was the graph-based similarity of URLs. Since our task was to match tagged objects from different domains (music and POIs), where the ground truth similarity could only be assessed by subjective users' evaluations, we could not directly rely on the outcome of that study, and therefore needed to evaluate these metrics for our specific task.

Table 2 lists the similarity metrics that were considered in the evaluation. In these equations, u and v represent items (either a music track, or a POI), t represents a tag, X_u—the set of tags with a non-zero frequency in the tag profile of the item u, f_{ut}—the frequency of tag t in the tag profile of the item u, $p(t)$—the fraction of items (both music tracks and POIs) annotated with t, and w_{ut}—the TF-IDF weight of tag t in an item's profile X_u. In order to compute these TF-IDF weights, for each item u, all the tags assigned to the item (with repetitions) were considered as a document representing the item.

The usage of logarithms in the first four metrics is related to Shannon information theory. Intuitively, a very common tag will have a high probability and therefore a very small log probability. Thus, it will bring a small contribution to the similarity score. We observe that all six metrics range in [0, 1], making their comparison easy. Moreover, we note that they can be applied both to the original tag profiles and to the merged tag profiles introduced in Sect. 3.1. As a result, we have 12 different methods to compute the similarity between a music track and a POI.

Before selecting the most effective similarity computation methods in a user study, we had to narrow down their number, since a user cannot express many subjective relatedness judgments in a single session. Therefore, we first evaluated these metrics offline, by computing the correlation of the ranked lists produced by the similarity metrics when matching (scoring) the available music tracks to a given POI. We sorted the music tracks recommended for a given POI using

---

Table 2 The similarity metrics considered for matching POIs and music tracks

Matching	$\mathrm{MatchingSim}(u,v) = \dfrac{\sum_{t \in X_u \cap X_v} \log p(t)}{\sum_{t \in T} \log p(t)}$, where T is the set of all tags in the dataset	
Overlap	$\mathrm{OverlapSim}(u,v) = \dfrac{\sum_{t \in X_u \cap X_v} \log p(t)}{max(\sum_{t \in X_u} \log p(t), \sum_{t \in X_v} \log p(t))}$	
Jaccard	$\mathrm{JaccardSim}(u,v) = \dfrac{\sum_{t \in X_u \cap X_v} \log p(t)}{\sum_{t \in X_u \cup X_v} \log p(t)}$	
Dice	$\mathrm{DiceSim}(u,v) = \dfrac{2 \sum_{t \in X_u \cap X_v} \log p(t)}{\sum_{t \in X_u} \log p(t) + \sum_{t \in X_v} \log p(t)}$	
Cosine	$\mathrm{CosineSim}(u,v) = \dfrac{\sum_t f_{ut} f_{vt}}{\sqrt{\sum_t f_{ut}^2}\sqrt{\sum_t f_{vt}^2}}$	
Weighted Cosine	$\mathrm{WCosineSim}(u,v) = \dfrac{\sum_t w_{ut} w_{vt}}{\sqrt{\sum_t w_{ut}^2}\sqrt{\sum_t w_{vt}^2}}$, where w_{ut} is the TF-IDF weight of tag t	
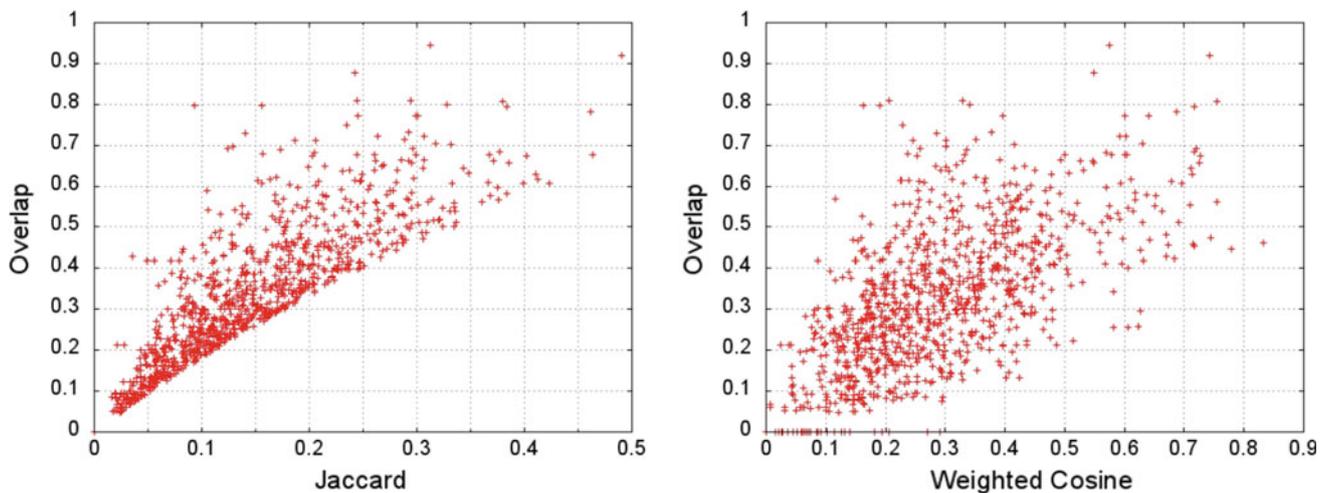
Fig. 3 Music-to-POI similarity values obtained by different similarity metrics

the different similarity metrics, and computed, pairwise, the Spearman's correlation of these ranked lists. Averaging, for each pair of metrics, the correlations of the ranked lists of music tracks computed for all the POIs in our dataset, we produced an average correlation score between every pair of metrics. The more correlated two metrics are, the more similar the music tracks recommended for a given POI will be. This initial analysis allowed us to study the general properties of the metrics, and to discard the redundant similarity computation methods.

When comparing the similarity metrics applied to the original tag profiles, i.e., without merging the GEMS tags in the same emotional category, we observed two clusters of metrics: Matching, Overlap, Jaccard, and Dice metrics all have an average correlation greater than 0.8 between each other. Likewise, Cosine and Weighted Cosine similarities

have a correlation greater than 0.7, but are less correlated with the metrics in the first cluster (e.g., correlation of Weighted Cosine with Overlap was 0.64, Cosine with Jaccard—0.68). Figure 3 shows the scatterplots of two highly correlated and two less correlated metrics.

The same clusters were observed when comparing the similarity metrics applied to the merged tag profiles. Hence, relying on these results we have selected one representative metric from each cluster—Jaccard from the first and Cosine from the second. We observe that there is a major difference between these two metrics: Cosine considers the tag frequency in items' profiles, while Jaccard only considers each co-occurring tag once. Thus, in the web-based evaluation study we used these two metrics, applied both to the original tag profiles, and to the merged tag profiles—4 similarity computation methods in total.

Fig. 4 Screenshot of the web application used for evaluating music-to-POI matching

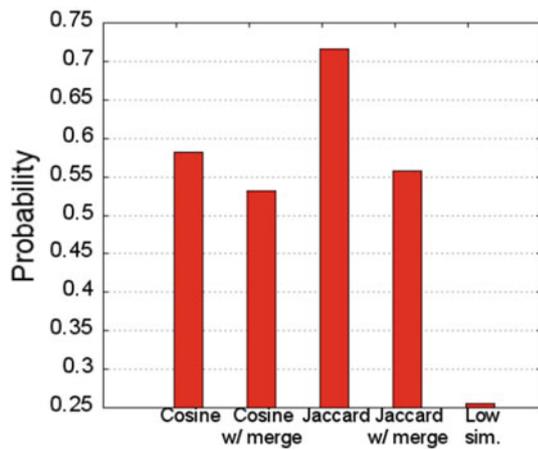## 3.3 Web-based evaluation

Having selected the four similarity computation methods for matching music and POIs, we have designed an experiment to collect the users' subjective evaluations, i.e., assessments if a music track suits a POI. We have designed a web interface (Fig. 4), where the users were repeatedly asked to consider a POI, and while looking at it, to listen to some selected music tracks. The user was asked to check all the tracks that in her opinion suit that POI.

During each evaluation step the music recommendations for a POI were selected using two out of the four considered similarity computation methods. The selected tracks included the two best matching tracks for each method (highest similarity). In addition, we introduced two tracks that were mostly different from the matching tracks, i.e., having low similarity to the given POI. Introducing the low similarity tracks allowed us to directly compare the tracks that were supposed to fit the POI with those not. In total, a maximum of six tracks were suggested for each POI, but usually less tracks were shown as the tracks selected by the similarity metrics may overlap.

The goal of this analysis was to see whether the users actually agree with the music-to-POI matching computed using our approach. We note that the outcome of this evaluation was not evident at all, since with a superficial evaluation, even the less similar tracks could be considered as suited— there are no large differences among the considered tracks (all of them are popular orchestral music).

For example, consider the evaluation step shown in Fig. 4. The POI Victory Monument was tagged as *bright*, *heavy*, *open*, *strong*, *triumphant*, *tense*, etc. In this case, the two metrics used to select the tracks are: Jaccard (suggesting tracks 1 and 5), and Jaccard applied to the merged tag profiles (suggesting tracks 3 and 5). The low similarity tracks are tracks 2 and 4. Track 1 has been tagged as *open*, *heavy*, *triumphant*, *amazed*, etc.; track 3—as *open*, *bright*, *agitated*, *bouncy*, *in love*, *triumphant*, etc.; track 5—as *open*, *heavy*, *triumphant*, *strong*, *cold*, etc. Contrastingly, tracks 2 and 4 have been tagged as *serene*, *light*, *colorful*, etc. Looking at the tag profiles, it is easy to understand why the similarity metrics suggest tracks 1, 3 and 5. However, the user is neither aware of the items' tag profiles, nor of the different ways the tracks were selected. It was therefore crucial to see if a person, just by listening to the selected music tracks, would agree with the match produced by our approach.

The online evaluation was carried out by 10 users in total performing 154 evaluation steps, that is, each user considered on average 15.4 POIs and the music suggested for these POIs. The set of study participants was disjoint from the set of users who took part in the tagging procedure (described in Sect. 3.1), with the exception of a few users. Moreover, since the two experiments took place months apart, the users could not remember how they tagged the items, and during the evaluation they could not access the tag profiles of the items they viewed.

In order to compare the effectiveness of different metrics in selecting the best tracks, we have computed the probability

Fig. 5 The selection probabilities for the five approaches evaluated in the web-based study

that a metric produces a music track that is considered suited for a POI by the users. The probability was computed as the ratio of the number of times any track produced by a metric was selected over the total number of evaluation steps where this metric was used, i.e., tracks produced by this metric were presented. Note that each time a music track, which was suggested by multiple metrics, was selected as appropriate for a POI, the probabilities for all these metrics were increased.

From the results of this experiment (Fig. 5), it is clear that all four tested similarity computation methods performed significantly better than the low similarity matching (99 % confidence level of the two-proportion z test). Among the four methods, Jaccard performed significantly better than the others (95 % confidence level of the two-proportion z test). A possible reason for the better performance of Jaccard compared to Cosine is that Jaccard metric in contrast to Cosine uses the probability that a tag can be found in a corpus; thus a frequent tag contributes to the similarity score less than a rare tag. The inferior performance of the metrics using the

merged tag profiles indicates that merging emotional tags introduces a loss of information (at least with the current dataset). A more thorough revision of the tag vocabulary could be applied in the future.

In conclusion, we can affirm that the users consider the music tracks suggested by our approach as more suited for POIs than other not matching tracks. Furthermore, Jaccard similarity metric selects the tracks that the users most frequently choose as suited for the illustrated POIs. We note that in this evaluation study the users were asked to evaluate the matching of music to POI while they were just reading a description of the POI, and not really visiting the place. In order to measure the effect of the POI-adapted music recommendations while the user is actually visiting the POI, we have implemented a mobile guide for the city of Bolzano. The next section describes the implementation of the guide, and the live user study that was conducted.

4 Mobile travel guide application

This section describes the usage scenario and the technologies used to develop PlayingGuide: an Android-based travel guide that illustrates the POI the user is close to and plays music suited for that POI. Users of PlayingGuide are tourists, possibly new to Bolzano, interested in exploring some of the city's POIs. 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. 6, 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. 6, 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.

Fig. 6 Screenshots of the mobile guide application, showing the map view, the details of a POI, and a feedback dialog
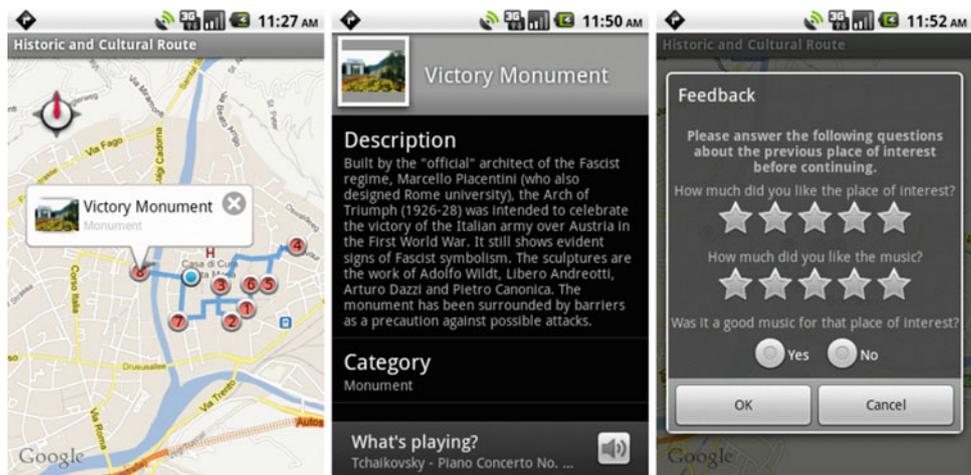
PlayingGuide has been implemented in a fat client architecture, i.e., the entire application runs locally on the mobile device and allows to synchronize local data changes with a remote server. This architecture was chosen to limit the potential problems related to the unreliable wireless data connection. The guide recommends music for a POI in the following two steps:

Step 1. Given a POI and a set of music tracks as input, the first step to generate a music recommendation is to compute the similarity between this POI and the available music tracks. Our dataset for this evaluation consisted of 75 music tracks and 32 POIs in Bolzano city center. Compared to the original dataset, presented in Sect. 3.1, the number of POIs has been reduced (from 50 to 32), since only the POIs within the walking distance of the city center were left for evaluation. Both POIs and music tracks have been tagged using a controlled tag vocabulary as described in Sect. 3.1. In order to compute the music-to-POI similarity, both POIs and music tracks are represented as vectors, with one component corresponding to each tag in our tag vocabulary, together with a weight for each component. As shown in Table 2, for the Jaccard metric we define the weight of tag t with respect to a POI (or a music track) u as $-\log p(t)$, where $p(t)$ denotes the fraction of all POIs and music tracks annotated with t. Then, the similarity between a POI and a music track is computed as the weighted Jaccard similarity of their vector representations. The Jaccard similarity metric was chosen since it performed best in the initial web-based evaluation of our approach, as described in Sect. 3.3.

Step 2. Given the music-to-POI similarity scores, the final step for delivering a music track recommendation for a POI is to sort the music tracks by decreasing similarity score, and then randomly pick out one of the top N (in our case 3) music tracks. The motivation for not always choosing the top-scoring music track for each POI is to avoid, or at least minimize, 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 visiting an itinerary.

5 Evaluation and results

In order to evaluate the proposed music-to-POI matching approach, we compared the performance of PlayingGuide 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 PlayingGuide 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). There was no overlap between these users and the participants of the tagging (Sect. 3.1) or the web-based evaluation (Sect. 3.3). We note that as in the web-based evaluation, the outcome of this comparison was not evident, 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.

The study subjects were instructed in: the purpose of the experiment; their task and the procedure of the experiment; and in the usage of the test device—a Google Nexus One mobile phone. Following this introductory phase, each subject was given a phone with earphones, and was asked to complete the "Historic and Cultural Route" to visit various POIs in Bolzano. This route required the subjects to walk for approximately 45 min in the center of Bolzano. Whenever a subject was approaching a POI, either belonging to the route or not, a notification invited the user to inspect the POI's details and to listen to the recommended music track. If the recommended 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.

Immediately after viewing the details of a POI and listening to the accompanying music, the subjects were asked, in a feedback dialog (Fig. 6, right), to answer three questions related to the POI and the recommended music, namely: (a) "How much did you like the place of interest?", (b) "How much did you like the music?", and (c) "Was it a good music for that place of interest?". The first two questions were rated on a five-star rating scale (with 1 star being the lowest score and 5 stars being the highest score), whereas the third question required a simple Yes/No answer. A total of 308 responses regarding the various visited POIs and their recommended music tracks were obtained: 157 (51 %) from subjects in the MATCH group, and 151 (49 %) from subjects in the MUSIC group.

After the "Historic and Cultural Route" had been completed, subjects were asked to fill out a paper-and-pencil questionnaire based on the Computer System Usability Questionnaire (CSUQ) [15] to assess the overall usability and effectiveness of the system. The subjects rated various statements on a 7-point Likert scale, commonly used for studies involving questionnaires and ranging from 1 (strongly disagree) to 7 (strongly agree).

Finally, 3 months later, a time period sufficient to eliminate the possible bias toward previously heard tracks, the subjects who participated in the evaluation were asked to re-rate the music tracks that were recommended for the various POIs. The new ratings were collected through a simple web interface where the music tracks were played one by one. Since the web interface presented the music tracks without

Table 3 Mean ratings for the music tracks in MATCH and MUSIC groups

Rating acquisition condition	Group of music tracks	
	MATCH	MUSIC
Rated on the mobile	3.78	3.34
Rated on the web	3.22	2.95

any reference to POIs, it enabled us to collect the subjects' ratings without any influence produced by the match between the POI and the music, or the contextual situation of the visit.

5.1 How much the users liked the places

The mean ratings for the question "How much did you like the place of interest?" were similar across both conditions, being somewhat higher in the MUSIC condition ($M = 3.93$, SD = 0.99) than in the MATCH condition ($M = 3.78$, SD = 1.08). There is, however, no significant difference between the conditions: $p = 0.21$ in a t test. This is as expected, since MATCH and MUSIC subjects visited almost the same POIs, and they were asked to rate each POI independently from the recommended music track. Hence, matching the music to a POI does not significantly alter the evaluation of the user for the POI.

5.2 How much the users liked the music

Analyzing the logging data collected during the experiment, we found that the mean listening time (in seconds) to the music recommended for each POI was slightly higher in the MATCH condition ($M = 38.43$, SD = 24.56) than in the MUSIC condition ($M = 36.91$, SD = 23.40). However, this difference is not statistically significant: $p = 0.58$ in a t test.

The mean rating for the question "How much did you like the music?" was 3.82 (SD = 1.02) in the MATCH condition, and 3.53 (SD = 1.17) in the MUSIC condition. The observed difference is statistically significant: $p = 0.023$ in a t test. This result seems to support the hypothesis that users like the suggested music more when it is matching the visited POIs. In fact, under a more careful analysis this could not be validated, since considering the ratings for the same music tracks collected via the web interface (see Table 3), i.e, where the users could rate the music tracks without any reference to POIs, we again found a significantly larger mean rating for the music tracks suggested by the MATCH variant compared to those suggested by the MUSIC variant. Table 3 shows the mean ratings acquired for the tracks in the MATCH and MUSIC conditions via the feedback dialog on the phone (i.e., mobile) and the mean ratings for the same tracks acquired via the web interface 3 months later.

Hence, the higher ratings given to music tracks by the subjects in the MATCH condition could also be determined by the fact that, in general, the users liked these tracks more. However, this data validates the hypothesis that listening to a music track on the mobile device, in this particular situation, has the effect of increasing the rating for the music track. In fact, in both the MATCH and MUSIC groups a two-tailed, paired, t test shows a significantly larger mean rating when the music tracks were rated on the mobile phone, compared to the web interface: for the MATCH group $p < 0.001$; and for the MUSIC group $p = 0.03$. In summary, we can say that the music appreciation is influenced by the device context, i.e., by the modality of music listening.

5.3 Matching music to POI

In the final, and more important evaluation of the effectiveness of the music-to-POI matching approach, we measured the proportion of "Yes" answers to the question "Was it a good music for that place of interest?". This proportion was substantially higher in the MATCH condition (0.77) than in the MUSIC condition (0.60). This difference in proportions is statistically significant, $\chi^2(1, N = 308) = 10.89, p < 0.001$. We can conclude that users evaluate the music tracks recommended 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 recommendation, 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, $\chi^2(1, N = 77) = 19.75, p < 0.001$, which proves our hypothesis that users prefer tracks for POIs that are generated by our music-to-POI matching approach.

While our hypothesis is supported by the average users' feedback for the full set of POIs, it is also interesting to look at the users' judgment of music appropriateness for the individual POIs in the itinerary (Fig. 7). We can see that in
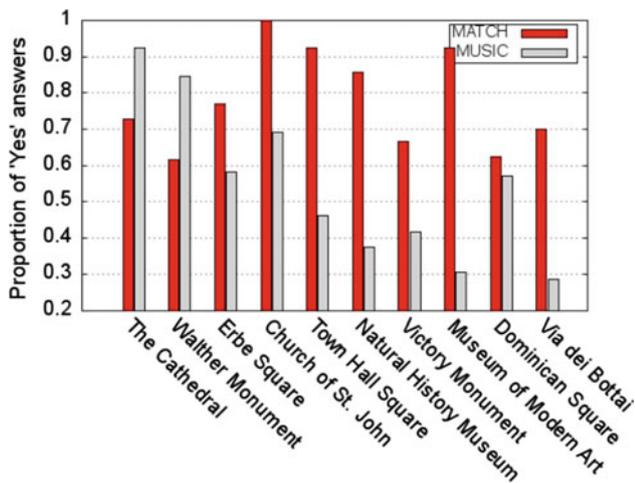
Fig. 7 Opinions about the appropriateness of music recommendations for individual POIs given by study participants in the two test groups—MATCH and MUSIC

Fig. 8 The distributions of tags from the controlled vocabulary in the profiles of two POIs

eight cases out of ten there is a larger proportion of users that evaluated the music suggested by the MATCH system as appropriate for the POI. In fact, only in two particular cases out of ten—the Cathedral of Bolzano, and the Walther Monument—the proportion of users that replied by saying that the music is good for the place is larger in the MUSIC condition.

The reason for the variation of the performance of the proposed matching method may lie in the particular nature of the POIs. In fact, when comparing the tag profiles of the Cathedral and the Museum of Modern Art, i.e., two POIs where there is a larger proportion of users that preferred the MUSIC and the MATCH music selection respectively, we can observe that the distribution of tags in the two profiles is quite different (Fig. 8). In the case of the Cathedral, more diverse tags were used, while for the Museum of Modern Art the tags were applied more consistently. This shows that for certain POIs, users may find it difficult to clearly define the emotional characteristics. Consequently, for such POIs it is difficult to recommend music in a way that the users would recognize and appreciate the match. Our intuition is that when a POI raises a few distinct emotions, it is easier to establish a meaningful match between a music track and the POI. Nevertheless, this issue deserves further investigation.

Another reason for the negative result in the two above-mentioned cases could be the influence of other contextual conditions. For instance, consider Walther Monument, which is situated in the main city square of Bolzano. During the festivities and in the evenings this is a particularly busy and lively place. However, the square's atmosphere can also be calm and relaxed when few people are around, e.g., during a working day. The users who assigned tags for Walther Monument are familiar with the more common atmosphere of this POI—the lively square. Consequently, it has been tagged
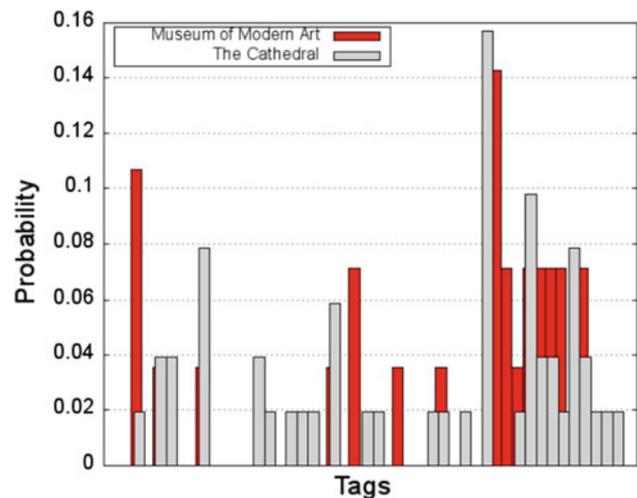
as *fiery*, *animated*, *amused*, *colorful*, etc., and the music track that scored highest for this POI is Rimsky-Korsakov's "Dance of the Bumble Bee"—a fast and animated piece. However, should the evaluation study subject enter the square on a calm midday, the atmosphere of the POI may not suit that of the music track. These observations lead us to believe that the impact of additional contextual factors (such as time, weather, or presence of other people) on the music selection process needs to be investigated.

5.4 Usability survey

Finally, Table 4 illustrates the ratings given by the subjects to each statement in the usability questionnaire. Both system variants received very positive responses. In general, the differences between the groups' mean ratings are not statistically significant. This is not surprising, since both groups tested the same system, but with different music recommendation approaches. There is, however, a marginally significant difference in statement 11 ("The music was correctly selected for each POI."), $p = 0.051$ in a t test, indicating again that MATCH subjects perceived the music recommended and played for the various POIs as more appropriate compared to MUSIC subjects.

6 Conclusions

In this paper we have analyzed a new problem in music recommendation, i.e., recommending music tracks that suit a POI. We have developed an approach that exploits user-assigned emotional tags to both music tracks and POIs. We have collected and analyzed the tagging data obtained from real users through a custom-developed interface that enables

Table 4 Usability questionnaire ratings collected after the evaluation of the mobile guide

Statement	MATCH mean (SD)	MUSIC mean (SD)
1. It was simple to use this system	6.08 (0.51)	6.46 (0.66)
2. I am able to complete my work quickly using this system	5.58 (1.00)	5.54 (1.27)
3. I feel comfortable using this system	6.00 (0.60)	5.92 (1.12)
4. It was easy to learn to use this system	6.17 (0.83)	6.54 (0.78)
5. Whenever I make a mistake using the system, I recover easily and quickly	5.60 (1.07)	5.20 (1.62)
6. The information provided with this system is clear	5.90 (1.29)	5.92 (1.04)
7. It is easy to find the information I needed	6.00 (1.18)	5.77 (0.93)
8. The organization of information on the system screens is clear	6.08 (1.24)	6.31 (1.11)
9. The interface of this system is pleasant	6.25 (0.62)	6.69 (0.63)
10. I like using the interface of this system	6.17 (0.83)	6.38 (0.65)
11. The music was correctly selected for each POI	**5.00** (0.74)	**4.08** (1.38)
12. I liked the music played for each POI	5.08 (0.67)	4.38 (1.98)
13. I would recommend it to a friend	6.00 (0.74)	5.92 (1.19)
14. Overall, I am satisfied with this system	6.00 (0.74)	6.00 (0.82)

Values in bold indicate statistically significant differences

users to uniformly tag both music tracks and POIs. Then, we have performed an experiment where the users were required to evaluate the appropriateness of the music selected by the system for POIs through a web interface. The results showed that users tend to agree with the matching produced using our proposed approach, and allowed us to select the best-performing similarity metric, the weighted Jaccard similarity, which was shown to produce the matching preferred by most of the users.

Subsequently, using results of the initial evaluation, we have developed and evaluated PlayingGuide—a novel mobile location-aware recommender system that suggests and plays music tracks while users are visiting POIs in a city. Before the evaluation of this application, we have formulated the following two experimental hypotheses: (a) users agree with the music recommendations generated by our approach, and (b) users consider the selected music tracks as more appealing if they are suggested and played on the implemented mobile application in the context of the visit to a POI. In a live user study, we were able to confirm our hypotheses.

An important next step in this research is understanding if certain tags (emotions) contribute more to the perceived match between a POI and a music track, since in the current implementation all tags had equal importance (see Sect. 3.1 for discussion). As a follow-up, we intend to revise the tag vocabulary, thus reducing the computational complexity of the approach. Another important future step is to move from manually labeled data toward automatic tag acquisition for both music and POIs. We are currently investigating public folksonomies (e.g., Flickr, Last.fm) and audio autotagging approaches as possible solutions to the scaling problem. The availability of automatically obtained tags may require us to further revise the matching approach, as in some cases a direct match of emotions can be difficult to obtain. In this line, we intend to investigate indirect relations between tagged items. For instance, a track labeled as *sad* may be considered by the users as a good match for a POI labeled with *wonder*. In addition, we want to study if personal preferences should be taken into account in this task, as currently the same match is provided for all the users. In order to perform these analyses, an evaluation with a larger set of users is required, as the current studies involved a limited number of participants. Moreover, aiming to complement the emotional relations between POIs and music tracks found by means of the tags, in a parallel direction of our research we are investigating an approach to automatically obtain semantic relations between POIs and musicians from the Linked Data repositories [8].

The topic of matching music to POIs is new, and there are many research questions that deserve further investigation. used to match Nevertheless, our results already demonstrate that recommending music for POIs is feasible, and that the proposed approach could be used to create new and appealing music recommendation services.

References

1. Abel F, Herder E, Houben GJ, Henze N, Krause D (2011) Cross-system user modeling and personalization on the social web. User Modeling and User-Adapted Interaction (UMUAI). Special Issue Pers Soc Web Syst 22(3):1–42
2. Adomavicius G, Mobasher B, Ricci F, Tuzhilin A (2011) Context-aware recommender systems. AI Mag 32(3):67–80
3. Ankolekar A, Sandholm T (2011) Foxtrot: a soundtrack for where you are. In: Interacting with sound workshop—exploring context-aware, local and social audio applications, pp 26–31

4. Baltrunas L, Kaminskas M, Ludwig B, Moling O, Ricci F, Aydin A, Lüke K-H, Schwaiger R (2011) Incarmusic: context-aware music recommendations in a car. In: Proceedings of the 12th International Conference on E-Commerce and Web Technologies (EC-Web 2011). Springer, pp 89–100

5. Braunhofer M, Kaminskas M, Ricci F (2011) Recommending music for places of interest in a mobile travel guide. In: Proceedings of the fifth ACM conference on Recommender systems. ACM, New York, pp 253–256

6. Burke R (2007) Hybrid web recommender systems. In: The adaptive web. Springer, Berlin, pp 377–408.

7. Celma Ò, Lamere P (2011) If you like radiohead, you might like this article. AI Mag 32(3):57–66

8. Fernández-Tobías I, Cantador I, Kaminskas M, Ricci F (2012) Knowledge-based music retrieval for places of interest. In: MIRUM'12—2nd International Workshop on Music Information Retrieval with User-Centered and Multimodal, Strategies

9. Finney N, Janer J (2010) Soundscape generation for virtual environments using community-provided audio databases. In: W3C Workshop—Augmented Reality on the Web. W3C, Barcelona

10. Gaye L, Mazé R, Holmquist LE (2003) Sonic city: the urban environment as a musical interface. In: NIME '03: Proceedings of the 2003 conference on New interfaces for musical expression. pp 109–115,

11. Kaminskas M, Ricci F (2009) Matching places of interest with music. In: Workshop on Exploring Musical Information Spaces. University of Alicante, pp 68–73.

12. Kaminskas M, Ricci F (2011) Location-adapted music recommendation using tags. In: 19th International Conference on User Modeling, Adaptation, and Personalization

13. Kaminskas M, Ricci F (2012) Contextual music information retrieval and recommendation: state of the art and challenges. Comput Sci Rev 6:89–119

14. Lee JH, Stephen Downie J (2004) Survey of music information needs, uses, and seeking behaviours: preliminary findings. In: ISMIR Proceedings. pp 441–446

15. Lewis JR (1995) IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum Comput Interact 7(1):57–78

16. Loizou A (2009) How to recommend music to film buffs: Enabling the provision of recommendations from multiple domains. PhD thesis, University of Southampton, Southampton

17. Markines B, Cattuto C, Menczer F, Benz D, Hotho A, Stumme G (2009) Evaluating similarity measures for emergent semantics of social tagging. In: Proceedings of the 18th international conference on World Wide Web. ACM, pp 641–650

18. Reddy S, Mascia J (2006) Lifetrak: music in tune with your life. In: Proceedings of the 1st ACM international workshop on human-centered multimedia. ACM, New York, pp 25–34

19. Ricci F, Cavada D, Mirzadeh N, Venturini A (2006) Case-based travel recommendations. In: Destination recommendation systems: behavioural foundations and applications. CABI, pp 67–93

20. Ricci F, Rokach L, Shapira B, Kantor PB (2011) Recommender systems handbook. Springer, Berlin

21. Sen S, Lam SK, Rashid AM, Cosley D, Frankowski D, Osterhouse J, Harper FM, Riedl J (2006) Tagging, communities, vocabulary, evolution. In: Proceedings of the 20th anniversary conference on Computer supported cooperative work. pp 181–190

22. Witten IH, Eibe F (2011) Data mining: practical machine learning tools and techniques, 3rd edn. Morgan Kaufmann, USA

23. Zentner M, Grandjean D, Scherer KR (August 2008) Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion 8(4):494–521