INTERACTIVE AND CONTEXT-AWARE MOBILE MUSIC EXPERIENCES

Jarno Seppänen
Nokia Research Center
Media Laboratory
Helsinki, Finland
jarno.seppanen@nokia.com

Jyri Huopaniemi
Nokia Research Center
Media Laboratory
Helsinki, Finland
jyri.huopaniemi@nokia.com

ABSTRACT

In this paper, we present compelling scenarios for future mobile music. Firstly, we provide a high-level architecture of mobile devices as contextual interactive computers. Secondly, we present research questions that arise from the combination of three domains of data: context data, collaborative listening data, and music content data. From these three domains, we propose to model the purpose of music listening and to offer purposeful music recommendations, which are both personalized and situationalized to the person and context of interest. Thirdly, we describe a realization of a mobile orchestra using existing mobile phone technology and discuss the role of personalization and preferences in future mobile music services. The paper will be concluded by future research problems and the landscape for compelling new services.

1. INTRODUCTION

Mobile devices have transformed during this decade from phones and messaging devices to ubiquitous mobile computers, with audio applications ranging from immersive communication to ubiquitous entertainment [1]. This paper concentrates on the musical use cases of current and future mobile devices. Mobile music devices have distinctive properties that differentiate them from home electronics devices or other music reproduction systems. Mobile music devices follow their user to changing situations in different environments. Mobile devices also offer rich sensing and wireless communications capabilities, making it possible to create novel music applications and services, which follow their users and intelligently react to changes in their music needs.

In this paper, we present a concept of a rich mobile music device which is capable of intelligent music consumption and performance, relying on user data, contextualization, music signal analysis, and real-time modification. We show that it is possible to realize both the use cases of intelligent music playback and performance using a single architecture.

In Figure 1, an architecture of an intelligent mobile music system is described. User model refers to the personal profile, preferences and potential artistic criteria of the user of the device. The situational context model takes sensor data (such as location, time, activity) and user data (such as user actions) as inputs, which is fed to the service interface for inference. The content model carries the musical content and the associated metadata (see, for example, [2] for further discussion). The community model refers to the input by and behavior of a single user or a group of users to the system under study. Two main categories of use cases have been defined for this type of a system - an interactive mobile orchestra where the user of the mobile device is the performer and the mobile device itself acts as the musical instrument. Another domain

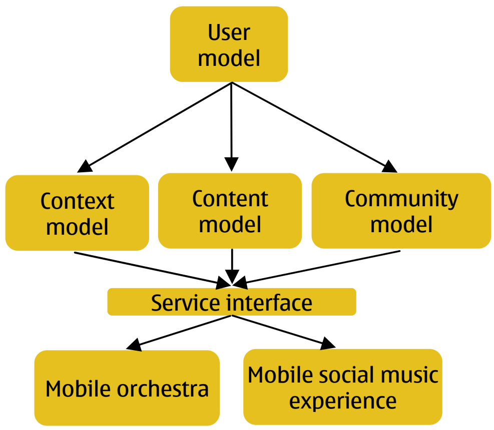

Figure 1: The architecture of intelligent mobile music system.

of applications is an interactive music listening experience, utilizing a combination of professional music content with social and personal features. In the following, the features of mobile devices pertaining to these novel applications are describe in more detail.

Off-the-shelf mobile devices afford a great range of sensor capabilities. Global Positioning System (GPS) receivers are built in to an increasing number of mobile devices, and microphones and cameras are found in practically every mobile phone built today. Such sensors can already provide an unlimited source of contextual information.

Furthermore, the devices contain integrated sensors such as ambient light brightness sensors, internal thermometers, accelerometers, and magnetometers. Mobile devices are equipped with several radio technologies that can be used for context and proximity sensing, including UMTS (universal mobile telecommunications system), WLAN (wireless local area networking), Bluetooth, and NFC (near field communications) transceivers. The phone usage patterns themselves bear contextual meaning: is the phone in general, meeting, or silent mode? Is the calendar status free or occupied? When did the user last use the phone? What applications are running in the device?

The range of available context features is broad, and in fact practical applications cannot afford to observe all features all the time. As with all mobile devices, each additional used feature draws a current from the battery, and thus consumes what is ultimately the most valuable and most critical mobile resource. Therefore, it is not generally advisable to use more sensors than required, or to run the sensors for longer time periods than necessary.

Feature selection is a common practice in machine learning, where the number and dimensionality of data vectors are reduced to optimize the classifier or predictor system performance. As explained above, feature selection for mobile context observations has an especially concrete relationship with implementation quality. In addition to optimizing system accuracy and use of computational resources, context-aware mobile systems require optimizing energy consumption as well.

Naturally, data collection must be done transparently and in full collaboration with the user. Not all music listenings or context features are collected automatically, without the user knowing, but the user is kept in full control of when and how the data collection happens. The user has full access to viewing her personal data in the service and permanently deleting any or all observations.

1.1. Related work

Personalization of Internet services is an active research topic dedicated to mass customization. A general definition of personalization is *"understanding the needs of each individual and helping satisfy a goal that efficiently and knowledgeably addresses each individual's need in a given context"* [3]. Often personalization is simply taken as automatic personalization, where user-specific customization is pursued without requiring much manual effort from the user. Personalization has really gained importance with always-connected services "in the cloud."

Context-aware applications and services use context information to provide relevant services to the user and the task at hand. Context is defined as the set of location *(where),* identity *(who),* activity *(what),* and time *(when)* observations, or more broadly, *"any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves"* [4]. Context-aware services are becoming reality on always-on and always-with mobile devices, capable of detailed context measurements. For example, simply the ambient noise can be used to determine the place or situation in a useful manner [5].

Recommender systems are intimately related to personalized services. In theory, recommender systems provide the underlying implementation of personalization – in practice, recommendation and personalization often combine into one. Recommendation is usually cast as a prediction problem: given a limited set of ratings of items by users, predict the values of missing ratings. In that scenario, a recommender system first collects user rating data, computes unknown rating predictions, and finally recommends the highest-rated items from the predicted set. Naturally, the more rating data is accumulated for a user or an item, the better the recommendations become for them. Recommenders are commonly implemented using collaborative filtering methods [6]. In pure collaborative filtering, only the ratings of items by users are used in providing recommendations. Other music recommender systems analyze the music content to provide recommendations [7]. The context-aware recommender utilizes context data as an additional input to the recommendation task, alongside information of users and items [8].

Personalized music services are a prime application of recommender systems as music rating data is often readily available. Real-world music usage behavior studies provide valuable information on the roles of music in people's everyday lives: *"our results show very clearly that people do indeed consciously and actively use music in different interpersonal and social contexts in order to produce different psychological states, that the resulting musical experiences occur on a variety of different levels of engagement, and that the value placed upon the music is dependent on these contexts"* [9]. Such results immediately raise the following question: would it be possible to model context dependency of music listening? Alternatively, is it possible to provide music recommendations which adapt to the situation of listening?

2. RESEARCH PROBLEM

Our main objective is to research the *purpose of music listening*, to the extent that can be learned from data. Here, the purpose of music refers to the subjective reasons why certain music is played in certain situations by certain persons. Our hypothesis is that music has different purposes for different persons in different situations, and to truly personalize the music offering, the music service should model the purpose of music to its listener. Therefore, we postulate that context, or situations, encode additional structure that can be utilized to improve recommendation performance. In this setting, the purpose of music listening reflects and represents the commonalities between the music, the user, and the situation of listening.

2.1. Modeling three domains

The problem is associated with three domains of data: context data, collaborative listening data, and music content data. The immediate research question arises as to how these three domains should be best combined for learning of music purpose. Figure 2 illustrates the three different domains of input features for context-aware personal music services:

1. User and community domain has information about user profiles and relationships;
2. Context domain represents the situations that link users with music; and
3. Music content domain encapsulates music metadata and music content descriptors.

The three domains are conceptually orthogonal, and as such act as independent sources of data for the music recommendation problem. From these three data sources, the music recommender aims to predict *purposeful* music selections, given the past behavior of the person in different situations, based on the user and context features. Here, we emphasize the fact that the recommendations are not only to be personalized but also to be *situationalized,* according to the learned purposes of music in similar past situations. In practice, to avoid requiring to store and process the full feature set for each prediction, the recommenders learn a context-aware and personal *music purpose model,* which is more compact than the full dataset but retains a desired prediction accuracy.

2.2. Context-aware collaborative filtering

In the existing literature, collaborative filtering algorithms often perform well in recommender systems [6]. However, no collaborative filtering method exists for taking the purpose or situation

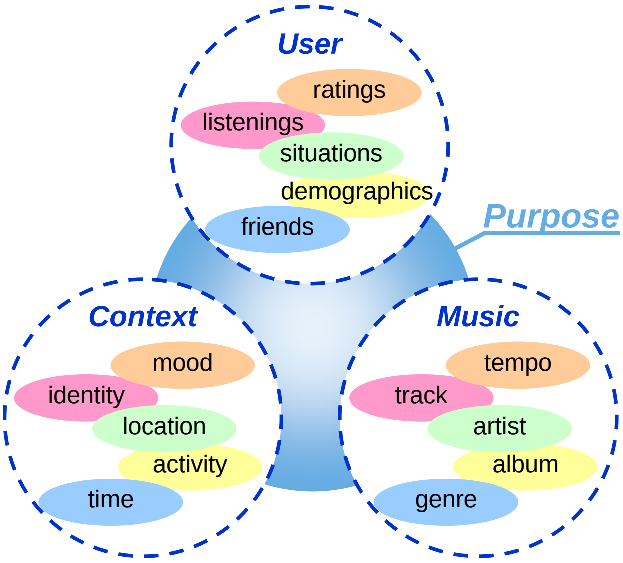

Figure 2: Context-aware music recommender feature domains.

of listening into account. More generally, we suggest research on modeling the purpose of music, and applying such models in collaborative filtering or music recommendation. Extending standard collaborative filtering methods towards context awareness is a natural approach towards purposeful recommenders.

2.3. Context features for music service

A part of the problem of modeling situations is the challenge of designing useful context features, taking into consideration the desired application, purposeful music recommendation. The research question is then, how to encode context for efficient music recommendation. How is music listening distributed in space and time, and what are their dynamics. What kind of locations and times are similar in terms of music listening; what kind of variation is found in music listening locations and times. How to partition or classify music based on its listening context.

2.4. Purposeful music recommender evaluation

As the topic of purposeful and context-aware recommenders is still in the early phases of research, we should also consider the question of evaluation methods of music recommenders which are designed to utilize the situations and predict the purposes of music listening. This question is related to finding a temporally motivated task statement beside the standard non-temporal rating prediction task, which is often used in recommender evaluations. As music listening is inherently a time-variant activity, the recommendations need also be evaluated in a temporal order.

3. USE CASES

Purposeful music recommendation. In the primal scenario, the user wishes to listen to some music. Using the context-aware service, the user automatically receives music recommendations that align well with the purpose of music that the user has in mind. When starting the daily commute, the user hears highly entertaining music to make time fly faster; when going for a jog, the user hears stimulating music to defeat the fatigue; when socializing with family or friends, the service recommends either atmospheric or activating background music, depending on the situation. Inevitably, music recommenders must evolve to take context information into account, and at that point they become purposeful recommenders. Whenever context data is available, it can be used as an independent data source to describe the subject of recommendations more precisely.

Personal music catalog. In an online music shop there can be millions and millions of tracks available for purchase. Yet the user often has something in mind already before going shopping for music – it can be a vague idea of how and where the music will be played, unless the user is shopping for a specific album. That is, the user may have an implicit purpose for the music as they browse the online shop catalog, and if the purpose is well aligned with the user's context, there is an opportunity to personalize and situationalize the music catalog view for the user. The personal music catalog filters irrelevant artists and tracks from the user's view, allowing the user to browse most suitable items to her personal preference and situation.

Mobile orchestra. The set of personalized context-aware mobile music players can also be viewed as an orchestra in itself. The players perform a composition according to their owners' situation and music preference. Rather than relying on professionally created music, the interaction between device and the user (=performer) creates the unique content.

4. IMPLEMENTATION

The context-aware personal music service comprises four parts:

1. Data collector,
2. Music purpose model,
3. Music recommender, and
4. Music performer.

Collection of data is the foundation of any personalized service. Personalization is achieved through explicitly asking user ratings and implicitly collecting purpose-related observations, where the latter approach is usually more important.

The raw data is transformed into a more manageable and useful form by music purpose model training. The purpose model is a classification or regression model which affords rating prediction for unrated artists or tracks, given users and situations.

Making music recommendations requires following the user's situation and applying the purpose model to predict suitable artists, albums, and tracks. The recommender is almost always used to construct a playlist of music, that is, using predicted ratings to generate an ordered list of tracks.

Context-aware music performance replaces automatic recommendations with a musical score to play. The score is composed in terms of user purposes and situations, so that the performed music reflects its listeners and the listening situations, as instructed by the composer.

4.1. Data collection

4.1.1. User data

As in all recommender systems, we first and foremost collect explicit user–item rating data. Users optionally rate music tracks during listening, and these ratings are directly applicable for music recommendation. If a user previously liked a track in a given situation, the same track and other similar tracks can be expected to be good recommendations when the same user and situation are encountered the next time. Respectively, if the user disliked or skipped a track, artists like it should not be recommended again.

When the user just listens to music without rating it, we collect the listening history, if the user allows. The listening and skipping data on tracks, albums, and artists characterize the purposes of music nonintrusively. Similarly, when activated by the user, the service memorizes the situations visited by the user during music listening. Also demographics can also be integrated to user data. Friendship graphs and other social networks are also part of the user profile.

4.1.2. Context data

Context data is collected to anchor the ratings and listenings to the situation. Also context can be observed explicitly, by asking the user to describe or tag the situation, and implicitly, by collecting relevant sensor readings when the user agrees. User-provided situation tags are again directly applicable for music recommendation, by associating all music listenings and ratings with the coincident tags. In practice purely manual situation labeling is not sufficient alone, because it would require large amounts of work from all users to describe all situations. A more practical context-aware system is able to learn to automatically suggest relevant tags based on location, time, activity, and other sensor values.

Considering music, one very important piece of context is the emotional state of the listener – after all, music is intrinsically communication about emotions. In practical systems, the emotions or moods of the listener cannot be directly sensed, but they can for example be asked. Also, it is important to realize that when music is listened according to its mood, we can perhaps glean information about the mental state through the listened music, which we already are collecting.

Outdoors location is precisely available with a built-in GPS receiver. Where the GPS signal is not detectable, a good enough resolution can be achieved by locating the nearest cellular network cell base station. In practice the network cell resolution ranges from hundreds of meters in dense urban areas up to tens of kilometers in sparsely populated areas. Indoors location can also be more precisely detected by WLAN and Bluetooth proximity scanning – such local-area wireless networks may be useful in detecting the floor or even the room of the user.

The accelerometer is sufficient to recognizing movement and activity to some degree. For example, standing still, walking, running, or vehicular movement can be recognized from each other by the accelerometer signals. Further, the ambient noise spectrum can tell whenever the user is in a motor vehicle [5]. Activity can also be observed from the phone usage data, starting with simple phone profile information (general, silent, meeting, etc.).

4.1.3. Music content data

Music recommender systems also harvest music metadata in addition to collecting user and context data. Such metadata contains textual titles of the genres, artists, albums, tracks, lyrics, etc., as well as acoustical features of timbre, rhythm, harmony, and melody. The music metadata is needed to be able to associate different pieces of music with each other, and to help alleviate the sparsity of the rating and listening data.

4.2. Music purpose modeling

The heart of the personalized music service is the music purpose model, which encompasses information about all music listened in the service by all users in all situations. Each user's music listenings can be combined together to form a music preference estimate. In the simplest terms, the purpose profile could be just the co-occurrences of most listened artists and most visited places in the last three months for each user. Such a "model" obviously represents a too limited view of the purpose of music listening, practically assuming that place would be the only meaningful piece of context when the purpose of music is concerned.

Varying context data such as location and time are collected together with the music listening data, and are used to build a representation of music listening situations as part of the model. Situation recognition is performed based on the raw context data, where the user's timeline is segmented into non-overlapping situation visits. Similar situations are also matched to each other. Then, given the user and situation data, each user's music listening and rating data is further decomposed into contextual sub-profiles that provide more detail on how and when music is listened.

4.3. Context-aware music recommender

The purpose of the music purpose model is to make predictions, given observations of user, music content, and context features. Predominantly the predictions are *context-aware music recommendations,* i.e., suggestions on music to be listened in the user's current situation.

The context-aware recommender needs to be able to handle situations in addition to users and items. If the situations are discrete, like users and items, then the context-aware recommender problem can be cast as an extended collaborative filtering problem. For that purpose, the three-dimensional *users* $\times$ *items* $\times$ *situations* data cube can be reduced either to a *(users* $\times$ *user-situations)* $\times$ *items* matrix, treating each user's situations as separate rows, or a *users* $\times$ *(items* $\times$ *item-situations)* matrix, enumerating each item's situations as separate columns. In either matrix form, collaborative filtering can then be applied to predict ratings in a context-aware manner.

4.4. Context-aware music performer

Context-aware composition and performance is a novel approach to how a mobile device and the surrounding context can be used simultanously to achieve an enriched experience. By contextuality we mean the ability for the composer and/or performer to take into account the sensorial input of the surroundings and/or by the user. An example of such a system created recently has been the Mobile Orchestra.[1], which is described in more detail in the next section.

5. PROTOTYPES

5.1. Mobile recommender prototype

The mobile environment sets a number of constraints for the implementation of a personalized and context-aware music service. Power and networking efficiency are among the most important aspects of any mobile application, and music services are no exception. All other mobile hardware resources such as processor,

[1] http://ccrma.stanford.edu/groups/mopho/

memory, display, sensors, or radios, have grown significantly in their capacity, but not the battery.

A natural way to implement personalized services is to create a wireless connection from the mobile clients to an online web server. The server is storing the rating data of all users and items, and therefore is able to provide music recommendations to the clients. To adapt the recommendations in real time, the clients are uploading the music rating and listening data directly from the mobile device to the server. In this architecture, the client devices are required to be online to be able to receive recommendations from the server.

Context data is also collected at the client and uploaded to the server for providing context-aware services. Observations about the location, activity, identity, and time are collected and pre-processed in the mobile device to suitable format for uploading. The context data chain must be designed primarily with energy efficiency in mind, and therefore context data should be collected sporadically and uploaded in bursts. In practice, the collector often follows a 20–80 pattern, recording context data only 20% of the time and being idle 80% of the time or more. Also, each context observation is not uploaded immediately, but the observations are buffered at the client and only uploaded when a sufficient amount of data exists.

The actual networking is implemented using HTTP/HTTPS over UMTS cellular network. As such, the communications protocol can be designed using web standards, while UMTS connectivity provides ubiquitous access to the service. While the HTTP protocol is not optimized for mobile usage, we find the benefits of using a standard protocol to be unsurpassable.

Because the recommendations are made online in the web server in the above architecture, it is not usable in offline situations such as flights. An alternate architecture is needed for such scenarios, in which a recommender data snapshot is downloaded to the clients, to enable offline recommendations. The recommender snapshot consists of artist similarity information for the user's favorite artists, for example. Using artist similarities, the mobile client is able to recommend similar artists to the ones which the user is listening to.

5.2. Mobile orchestra prototype

The use of mobile phones for music creation has become possible in recent years due to open software environments and inclusion of sound synthesizers in devices. The most traditional use of sound synthesis in mobile devices is ringing tones and alert tones, but these seldom allow for artistic creation and interaction. In early 2008, a novel use for modern mobile phones was invented through the Mobile Phone Orchestra (MoPho) [10] project at the Center for Computer Research in Music and Acoustics (CCRMA) at Stanford University. In this project, the mobile phone becomes a musical instrument and the user of the phone becomes the performer. The open Symbian and Python platform allows for development of various sound generation and performance tools, allowing real-time generation and/or modification of multichannel sound. Furthermore, in an ensemble playback situation the location context of the performer and device can result in suprising sonic landscapes. In [10], details about the project and the mobile phone compositions are given.

The software architecture of MoPho implementation is a combination of a high-level programming language, Python, and a native C++ -based programming language for Symbian-based mobile

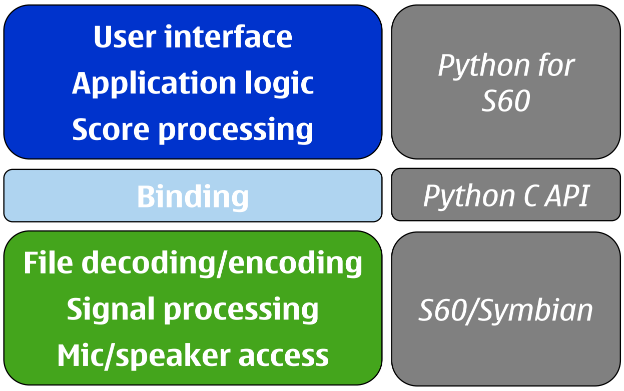

Figure 3: The implementation architecture of MoPhO mobile phone orchestra.

phones. The MoPho project uses Nokia N95 devices that feature rich multimedia and sensor capabilities, including GPS, 3-axis accelerometer (allowing gestural input and control), high-quality 5 megapixel camera, CD quality stereo audio, stereo speakers, and an open multiplatform programming environment. In Figure 3, the implementation architecture of the MoPho environment is described. The user interface and application logic is programmed using a high-level Python language, allowing for very fast and intuitive prototyping. Furthermore, the dynamic event-driven score processing is carried out in Python. The binding between native Symbian / S60 programming environment is enabled via a Python C API. The bottom part of the architecture carries out the computationally intensive real-time signal processing operations and carries out the basic file I/O and transducer (microphone, speaker) access.

In the first phase, the mobile phone orchestra has shown novel expressive performance possibilities via context-dependent and gesture -controlled sound synthesis. By taking into account full capabilities of rich mobile multimedia computers, more advanced features such as real-time networking of the devices allowing synchronous creation and playback can be envisaged. Furthermore, the form factor of the device itself could be transformed to model traditional musical instruments as shown in the recent KDDI-Yamaha mobile synthesizer project.[2]

6. DISCUSSION

We have identified great potential and interesting challenges in bringing context awareness into music experiences on mobile devices. Utilization of context promises great potential for both next-generation music recommender and performance systems, and offers a next level of personalized services, which are accurate to the needs of the user in the actual situation.

It must be mentioned, though, that we do not expect context information to improve recommendation performance unconditionally, for all users and cases. Further research and understanding is needed as to how and when can context be best utilized and integrated, and how and when should it not be acquired. The concrete applications and benefits of context awareness for personal music services have only recently appeared within reach.

[2] http://www.au.kddi.com/au_design_project/models/2008/index.html

All context features are also not equal in importance, and we intuitively expect user's mood to be one of the most important context descriptor, yet it is one which cannot directly be measured with currently available sensors. On the other hand, the recently listened music may reveal mood information indirectly, under the assumption that mood and music have a dependency. As an opposite example, the user's location is very exactly measurable using GPS receivers, yet the raw coordinates generally provide very little additional information for music recommendation. The crucial issue is the design of location features, which can be computed from latitude/longitude data, and which offer musically relevant information.

The use of mobile devices in music performance has great untapped potential, of which only the first examples have been shown recently. A new range of applications and services utilizing context-dependence, deep personalization and social aspects is expected, potentially changing the landscape of music listening and performance.

For the scientific community, the field of research presented in this paper is rich of new interesting research problems. The combination of mobile sound and music related research with human-computer interaction, context awareness, ubiquitous computing and recommender systems offers an interdisciplinary view to many fields, with the potential of very large impact into new human practices and use cases.

7. ACKNOWLEDGMENTS

We wish to thank Henri Penttinen for inspiring discussion and feedback during this work and for his efforts in realizing the MoPhO mobile phone orchestra.

8. REFERENCES

[1] Jyri Huopaniemi, "Future of personal audio: Smart applications and immersive communication," in *Proc. 30th AES International Conference on Intelligent Audio Environments*, 2007, pp. 1–7.

[2] Juha Lehikoinen, Pertti Huuskonen, Antti Aaltonen, and Ilkka Salminen, *Personal Content Experience: Managing Digital Life in the Mobile Age*, Springer, 2007.

[3] Doug Riecken, "Personalized views of personalization," *Comm. ACM*, vol. 43, no. 8, 2000.

[4] Anind K. Dey and Gregory D. Abowd, "Towards a better understanding of context and context-awareness," Tech. report GIT-GVU-99-22, Georgia Institute of Tech., Atlanta, GA, USA, 1999.

[5] Antti Eronen, Juha Tuomi, Anssi Klapuri, Seppo Fagerlund, Timo Sorsa, Gaëtan Lorho, and Jyri Huopaniemi, "Audio-based context awareness — acoustic modeling and perceptual evaluation," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP)*, 2003.

[6] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl, "GroupLens: An open architecture for collaborative filtering of netnews," in *Proc. ACM Conf. Computer Supported Cooperative Work*, Chapel Hill, NC, USA, 1994, pp. 175–186.

[7] Rainer Typke, Frans Wiering, and Remco C. Veltkamp, "A survey of music information retrieval systems," in *Proc. Int. Conf. Music Information Retrieval (ISMIR)*, 2005, pp. 153–160.

[8] Annie Chen, "Context-aware collaborative filtering system: Predicting the user's preference in the ubiquitous computing environment," in *Location- and Context Awareness*, pp. 244–253. Springer, 2005.

[9] Adrian C. North, David J. Hargreaves, and Jon J. Hargreaves, "Uses of music in everyday life," *Music Perception*, vol. 22, no. 1, pp. 41–77, 2004.

[10] Ge Wang, Georg Essl, and Henri Penttinen, "MoPhO: Do mobile phones dream of electric orchestras?," in *Proc. Int. Comp. Music Conf. (ICMC)*, 2008.

[11] Paul Resnick and Hal R. Varian, "Recommender systems," *Comm. ACM*, vol. 40, no. 3, 1997.