

Impact of User Context on Song Selection

Olufisayo Omojokun
Georgia Institute of Technology
College of Computing
Atlanta, GA 30332-0760 USA
+1 (404) 385-6348
omojokun@cc.gatech.edu

Michael Genovese
Georgia Institute of Technology
College of Computing
Atlanta, GA 30332-0760 USA
+1 (404) 385-6348
mgenovese3@mail.gatech.edu

Charles Isbell Jr.
Georgia Institute of Technology
College of Computing
Atlanta, GA 30332-0760 USA
+1 (404) 385- 6491
isbell@cc.gatech.edu

ABSTRACT

The rise of digital music has led to a parallel rise in the need to manage music collections of several thousands of songs on a single device. Manual selection of songs for a music listening experience is a cumbersome task. In this paper, we present an initial exploration of the feasibility of using song signal properties and user context information to assist in automatic song selection. Users listened to music over the course of a month while their context and song selections were tracked. Initial results suggest the use of context information can improve automated song selection when patterns are learned for each individual.

Categories and Subject Descriptors

H.5.5 [Information Interfaces and Presentation]: Sound and Music Computing – *modeling*.

General Terms

Measurement, Design, Human Factors.

Keywords

Music, Song Selection, Context, Modeling, Prediction, and Sensors

1. INTRODUCTION

The rise of digital music has led to music collections of thousands of songs on a single device. Users tend to either build playlists of songs for various occasions or allow random songs to play, skipping songs that they do not want to hear. The former is extremely time consuming and the latter can lead to jarring juxtapositions of songs.

One may take many different factors into consideration when deciding what songs to play. The person's mood, activity, and the other people around could all be expected to play a role in the decision. There is complexity in both modeling the state of the user and the content of the various songs. Ideally, a user's interaction history with a media player should provide enough

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'08, October 27–31, 2008, Vancouver, British Columbia, Canada.

Copyright 2008 ACM 978-1-60558-303-7/08/10...\$5.00.

information to automate these decisions and ensure the user a satisfying listening experience.

Preferably, an automated system would understand the user's environment and music collection without the user having to directly provide input. Sensors could be used to make inferences about the user's environment. New songs could be scanned and integrated automatically, based on acoustic information that is drawn from the songs themselves [4,7], with out-of-band metadata derived from the ID3 tags.

In the rest of this paper, we present our initial exploration of the feasibility of using context information gathered specifically from non-invasive and natively available sensors to assist in automatic song selection. After discussing related work in the next section, we motivate and describe our song selection approach in detail. We then discuss our experiments and results in Sections 4 and 5, respectively.

2. RELATED WORK

Several systems have been built to address the song selection problem. These systems include: Lifetrak, PersonalSoundtrack, DJogger, Pampalk's system, and XPod.

Lifetrak [6] is a system that incorporates sensor information into song selection. The songs are represented in terms of tags that the user assigns in order to link the songs to the appropriate contexts in which they should be played. User feedback is incorporated to make a song more or less likely to play in a given context. Context considered relevant to song selection includes location, time of operation, velocity of the user, weather, traffic and sound. User locations and velocity are determined by GPS. Location information includes tags based on zip code and whether the user is inside or outside (inferred by the presence or absence of a GPS signal). The times of the day are divided out into configurable parts of the day (morning, evening, etc). The velocity is abstracted into one of four states: static, walking, running and driving. Use of accelerometers are planned to enable indoor velocity information. If the user is driving, an RSS feed on traffic information is used to typify the state as calm, moderate or chaotic. If the user is not driving, a microphone reading is used for the same purpose. Additionally, an RSS feed provides a meteorological condition (frigid, cold, temperate, warm or hot).

PersonalSoundtrack [3] and DJogger [1] measure users' activity level and used this to select songs. Using an accelerometer, PersonalSoundtrack attempts to match a user's steps per minute to a song's beats per minute. The authors also describe using a user's location, terrain, time of day, and weather as other promising directions for describing context.

DJogger also links user pace to music pace. The difference is that the music changes pace based on adjustments needed to maintain the pace of the user's workout goals.

XPod [2] also collects physiological data and applies it to song selection. The user's physiological data is categorized as resting, active or passive. The categories are used for retrieval of a context sensitive rating. These ratings are manually applied by the user and are applied to artist and genre when selected for a song. A neural network is used to learn when a user will skip a song. The song is represented by the rating and the beats per minute.

Pampalk developed a system [5] that does not use user activity level to inform song selection. Rather, it uses song similarity and user skipping behavior to reduce skip count. When a song is skipped, similar songs are removed from the playlist. When a song is accepted similar songs are added. This approach is similar to the widely popular Pandora.com approach.

3. APPROACH

Existing systems take different factors into consideration depending upon what the system builders believe to be significant, in addition to what information is readily available. None of the systems described above take both complex signal properties of individual songs and environmental conditions into consideration.

The most common song property—BPM—does not readily discriminate between song genre, instrumentals, and other important features of a song. In our work, we extract a wide range of signal properties to yield a highly descriptive per-song feature set. These properties can be extracted for individual windows (or frames) of songs. Whole song features can subsequently be derived by calculating averages and standard deviations of each feature from these windows.

In particular, we extracted: Spectral Centroid, Spectral Rolloff Point, Spectral Flux, Compactness, Spectral Variability, Root Mean Square, Fraction of Low Energy Windows, Zero Crossings, Strongest Beat, Beat Sum, Strength Of Strongest Beat, Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Coefficients (LPC), and Method of Moments.

The MIREX (Music Information Retrieval Evaluation eXchange) competitions at past meetings of the International Conference on Music Information Retrieval provide the inspiration for the use of our particular features, as many of them have been shown to be quite useful together in song genre classification. *Compactness*, for example, describes the noisiness of a song. Further, *beat sum* represents the importance of beat regularity, while *beat histogram* measures the relative strength of different tempi (see [4] for a more exhaustive description of these features).

To gather environmental conditions, we specifically focus on non-invasive and natively available sensors for two important reasons. First, it offers low setup overhead by not requiring users to make the effort of explicitly placing special equipment on their bodies

or in their environments. Second, it has low equipment overhead by not making important assumptions about the conditions of a user's music listening environment. In fact, our sensors were actually deployed as program code on the users' music listening devices—hence they were inherently available wherever users played songs.

We focused on users' laptops as a point of data collection. Laptops are mobile, thus allowing us to flexibly capture users' listening behaviors in different kinds of environments (*e.g.* at work, home, or a coffee shop). Portable media players are quite pervasive and also mobile. Laptops, however, have the benefits of: (1) allowing us to persistently capture played songs in files and (2) easily executing third-party (sensor) code that collects environmental data, possibly using the Internet.

4. EXPERIMENTAL SETUP

A script was set up on the computers of four users who were asked to listen to music as they normally would. The script made use of sensors built into the user's systems. The users were evenly divided between male and female, students and professionals and users in their twenties and thirties. The script collected data on the songs that the user played and the context in which the song was played. Context data included:

- Time of day / day of week
- Weather - temperature, humidity, weather description in text and precipitation
- Front most process running (*i.e.* the application in which the user is currently engaged)
- Other processes running.
- Bluetooth devices (e.g. cell phone and computer) in range, as a proxy for who is around
- Noise level in the room in which the user is playing music, as a proxy for activity level
- System and application volume level

At the end of the month long data collection period, the songs that had been played were scanned for signal features using jAudio [4]. jAudio is a freely downloadable signal feature extraction application. It provides an easy to use and Java-based framework for extracting the different kinds of per-song signal features listed in the previous section.

5. RESULTS

The participants listened to 96, 126, 172 and 270 songs over the course of data collection. As Figures 1 and 2 describe, correlations were found between the signal features and the context information. The day of the week, front most process and volume level in the room were among the most highly correlated

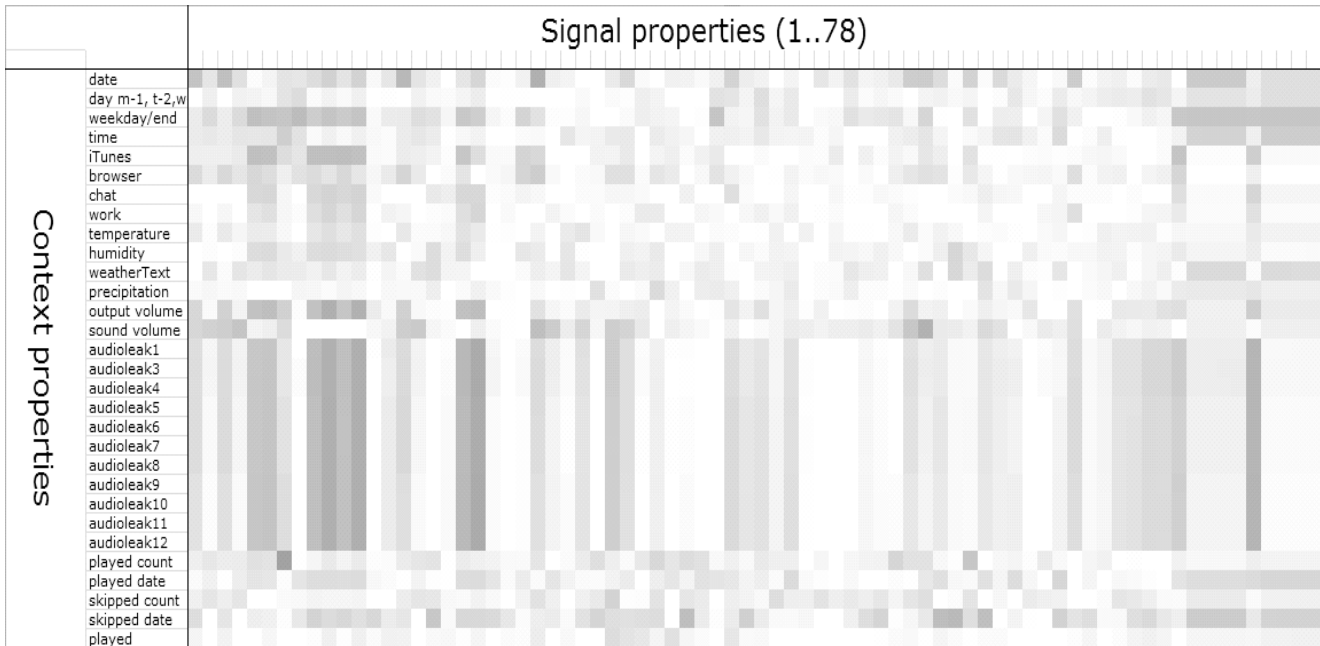


Figure 1. This image depicts the correlations between the signal properties and context variables for all of the users aggregated together. The darker the square is, the higher the absolute value of the correlation. What appears to be the strongest correlation is the result of one user with a significantly different music collection having a less sensitive microphone compared to the other users.

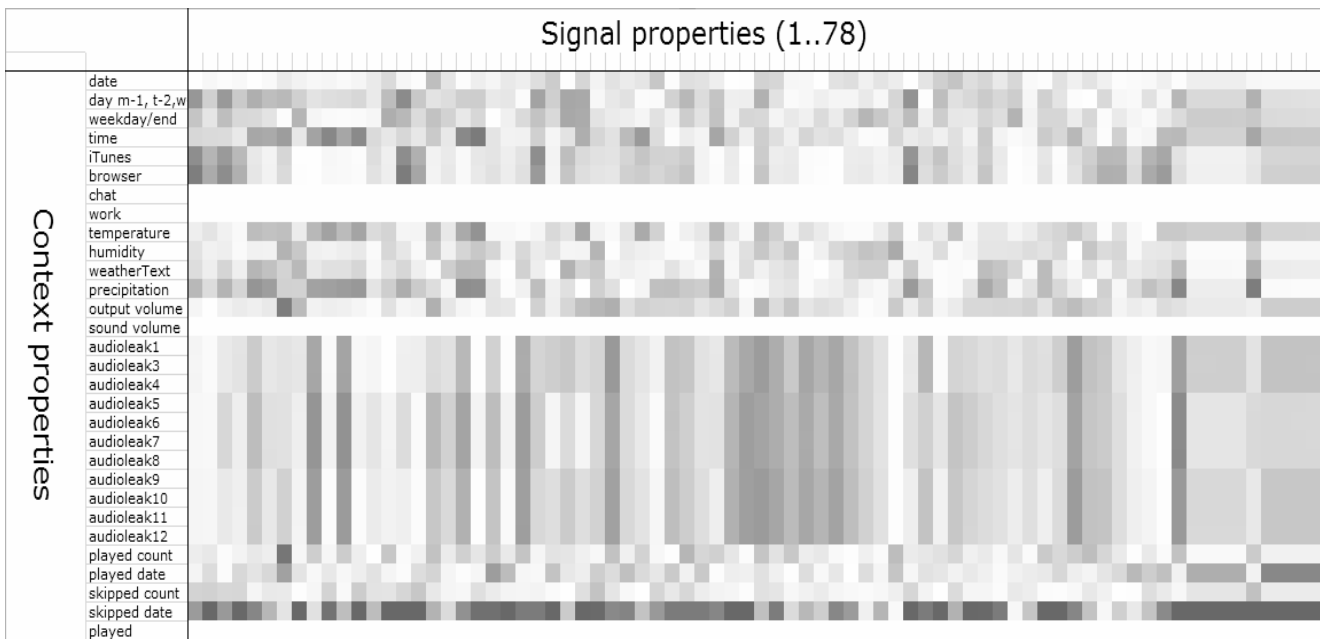


Figure 2. This image depicts the correlations between signal properties and context variables for one of the users. The correlations are stronger in this case.

context variables. The correlations found were influenced heavily by the grouping strategy employed. Patterns were dampened when the data was aggregated across users. Throughout all data collection, only one Bluetooth device was detected. We expected that evidence of different listening patterns would be found when different people were around, but our users did not turn on their Bluetooth devices. To make use of the information, the user would likely need to be prompted to turn their devices on.

Microphone quality varied significantly between users. One user had an older computer that did not have a built in microphone and one user had a microphone that showed very little variation over time. The usefulness of the data collected is limited by the quality of the microphone.

6. CONCLUSIONS AND FUTURE WORK

The use of context variables, gathered through natively available sources, is promising in assisting song selection. Using signal properties as the sole basis for song selection in an adaptive media player improves song selection over random play. As a correlation was found between the signal properties and the user context, different signal properties could be favored in different contexts. Further evidence that the context variables can be of use in song selection is found when many of the correlations between signal properties and context variables are in opposite directions for songs that were skipped and songs that were played to completion.

There did not appear to be patterns between users. As a result, we would suggest using a personalized approach in learning correlations between signal properties and context variables.

We imagine several directions for future work. Foremost, we plan to conduct a longer-term study with a larger pool of users. This will allow us to search for group patterns in listening preferences and weigh the benefits and costs of personalized recommendations. In our next study, we would like to either provide Bluetooth devices or request users to turn on their Bluetooth enabled phones in order to determine the effects of having different people around. In fact, the quick adoption of programmable phones with rich multimedia capabilities should make it easier to follow a similar process on mobile devices and facilitate obtaining other types of context information. The new

iPhone 3G and the release of the associated software developer kit provides us with a starting point.

7. REFERENCES

- [1] Biehl, J. T., Adamczyk, P. D., and Bailey, B. P. 2006. DJogger: a mobile dynamic music device. In CHI '06 Extended Abstracts on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006). CHI '06. ACM, New York, NY, 556-561.
- [2] Dornbush, S., English, J., Oates, T., Segall, Z. and Joshi, A. 2007. XPod: A human activity aware learning mobile music player. In Proceedings of the Workshop on Ambient Intelligence, 20th International Joint Conference on Artificial Intelligence, January 2007.
- [3] Elliott, G. T. and Tomlinson, B. 2006. PersonalSoundtrack: context-aware playlists that adapt to user pace. In CHI '06 Extended Abstracts on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006). CHI '06. ACM, New York, NY, 736-741.
- [4] McEnnis, D., McKay, C., Fujinaga, I., and Depalle, P. 2005. jAudio: A Feature Extraction Library. In Proceedings of ISMIR 2005.
- [5] Pampalk, E., Pohle, T. and Widmer, G. 2005. Dynamic playlist generation based on skipping behavior. In Proceedings of ISMIR 2005 Sixth International Conference on Music Information Retrieval, 2005.
- [6] Reddy, S. and Mascia, J. 2006. Lifetrak: music in tune with your life. In Proceedings of the 1st ACM international Workshop on Human-Centered Multimedia (Santa Barbara, California, USA, October 27 - 27, 2006). HCM '06. ACM, New York, NY, 25-34.
- [7] Tzanetakis, G., Essl, G. and Cook, P. 2001. Automatic Musical Genre Classification of Audio Signals. In Proc. Int. Symposium on Music Information Retrieval (ISMIR), Bloomington, Indiana, 2001.